

**CORRELATION-BASED CROSS-LAYER
COMMUNICATION IN WIRELESS SENSOR
NETWORKS**

A Thesis
Presented to
The Academic Faculty

by

Mehmet Can Vuran

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2007

CORRELATION-BASED CROSS-LAYER COMMUNICATION IN WIRELESS SENSOR NETWORKS

Approved by:

Professor Ian F. Akyildiz, Advisor
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor Ye (Geoffrey) Li
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor Raghupathy Sivakumar
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor Douglas M. Blough
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor Mostafa Ammar
Department of Computer Science
Georgia Institute of Technology

Date Approved: 7 June 2007

*To my mom and dad,
Ayla and Mehmet Vuran.*

ACKNOWLEDGEMENTS

It is a pleasure to thank the many people who made this thesis possible.

It is difficult to overstate my gratitude to my advisor, Dr. Ian F. Akyildiz. I wholeheartedly thank Dr. Akyildiz for his invaluable support during my study, for his endless belief in me during hard times, for his strong guidance, friendship, and trust throughout my Ph.D. Dr. Akyildiz has provided me the tools, the will, the power, and the strive to reach success. I am also indebted to Dr. Akyildiz for creating a family atmosphere for us "newcomers" with his lovely wife Maria Akyildiz and daughters Celine, Rengin, and Corinne during his house parties, dinners, and joyful soccer game views.

My cordial thanks also extend to Dr. Mostafa Ammar, Dr. Douglas M. Blough, Dr. Ye (Geoffrey) Li, and Dr. Raghupathy Sivakumar for being on my dissertation defense committee. I would also like to thank Dr. Douglas M. Blough, Dr. Ye (Geoffrey) Li, and Dr. Raghupathy Sivakumar for being on my dissertation proposal committee. Their invaluable comments have helped me to achieve a solid research path towards this thesis.

I would like to acknowledge the members of the Broadband and Wireless Networking Laboratory (BWN-Lab) for creating an excellent atmosphere during my study. I am especially thankful to Ozgur B. Akan for his friendship and support.

I would like to also thank my aunt and uncle Ayten and Tuncer Someren for their love and support and for making Atlanta a hometown for me. Finally, I thank my parents, Ayla and Mehmet Vuran, and my wife, Demet, for their love, support, encouragement, and sacrifices.

Allahtan doktor olalim daa...

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xiii
SUMMARY	xv
I INTRODUCTION	1
1.1 Research Objectives and Solutions	6
1.1.1 Theory of Correlation in Wireless Sensor Networks	6
1.1.2 Spatial Correlation-based Collaborative Medium Access Control in Wireless Sensor Networks	8
1.1.3 XLM: Cross-Layer Module for Efficient Communication in Wireless Sensor Networks	9
1.1.4 Cross-Layer Analysis of Error Control in Wireless Sensor Networks	10
1.1.5 Cross-layer Packet Size Optimization for Wireless Sensor Networks	11
1.2 Thesis Outline	12
II THEORY OF CORRELATION IN WIRELESS SENSOR NETWORKS	13
2.1 Motivation	14
2.2 Related Work	16
2.3 Spatio-Temporal Correlation in Wireless Sensor Networks	17
2.3.1 Architecture and Correlation Model for WSN	18
2.3.2 Spatial Correlation in WSN	20
2.3.3 Temporal Correlation in WSN	24
2.3.4 Spatio-temporal Correlation	27
2.4 Analysis and Results	32

2.4.1	Spatial Correlation	32
2.4.2	Temporal Correlation	34
2.4.3	Spatio-temporal Correlation with Point Sources	36
2.4.4	Spatio-temporal Correlation with Field Source	38
2.4.5	External Parameters	40
2.5	Exploiting Correlation in WSN	42
2.5.1	Correlation-based Medium Access Control	43
2.5.2	Correlation-based Reliable Event Transport	44
III	SPATIAL CORRELATION-BASED COLLABORATIVE MEDIUM AC- CESS CONTROL IN WIRELESS SENSOR NETWORKS	46
3.1	Motivation	46
3.2	Related Work	47
3.3	Corralaries from Spatial Correlation Theory	49
3.4	The Iterative Node Selection (INS) Algorithm	52
3.5	CC-MAC Protocol Description	57
3.5.1	Packet Structure	59
3.5.2	Event MAC (E-MAC)	60
3.5.3	Network MAC (N-MAC)	64
3.6	Simulation Results	65
3.6.1	CC-MAC Parameters	66
3.6.2	Comparative Study	70
IV	XLM: CROSS-LAYER MODULE FOR EFFICIENT COMMUNICATION IN WIRELESS SENSOR NETWORKS	77
4.1	Motivation	77
4.2	Related Work	80
4.3	XLM: Cross-Layer Module for WSN	82
4.3.1	XLM Basics and Definitions	85
4.3.2	XLM Transmission Initiation	85
4.3.3	XLM Receiver Contention	86

4.3.4	XLM Angle-based Routing	88
4.3.5	XLM Local Cross-Layer Congestion Control	90
4.3.6	XLM Duty Cycle Analysis	95
4.4	Performance Evaluation	99
4.4.1	XLM Parameters	102
4.4.2	Comparative Evaluation	104
V	CROSS-LAYER ANALYSIS OF ERROR CONTROL IN WIRELESS SEN- SOR NETWORKS	114
5.1	Motivation	114
5.2	Related Work	116
5.3	Error Control in Wireless Sensor Networks	117
5.4	Analysis Approach and System Model	120
5.5	Cross-Layer Analysis	123
5.5.1	The Expected Hop Distance	123
5.5.2	Energy Consumption Analysis	126
5.5.3	Latency Analysis	128
5.5.4	Decoding Latency and Energy	129
5.5.5	Bit and Packet Error Rate	130
5.6	Numerical Results	131
5.6.1	Hop Length Extension	132
5.6.2	Transmit Power Control	138
5.6.3	Effects of End-to-End Distance and Target PER	139
5.6.4	Hybrid Error Control	141
5.6.5	Overview of Results	143
VI	CROSS-LAYER PACKET SIZE OPTIMIZATION FOR WIRELESS SEN- SOR NETWORKS	146
6.1	Motivation	146
6.2	Related Work	148
6.3	Factors Affecting The Packet Size	149

6.3.1	Error Detection and Correction: Bit Level Energy Consumption	149
6.3.2	Collision: Packet Level Energy Consumption	151
6.4	Packet Size Optimization Framework	155
6.4.1	Optimization Metrics	157
6.4.2	Channel Model	159
6.4.3	End-to-End Energy Consumption	159
6.5	Packet Size Optimization	162
6.6	Packet Size Optimization in Wireless Underwater and Underground Sensor Networks	168
6.6.1	Underwater Channel Model	169
6.6.2	Underground Channel Model	171
6.6.3	Results	173
VII	CONCLUSIONS AND FUTURE WORK	186
7.1	Research Contributions	186
7.1.1	Theory of Correlation in Wireless Sensor Networks	186
7.1.2	Spatial Correlation-based Collaborative Medium Access Control in WSN	187
7.1.3	XLM: Cross-layer Module for Efficient Communication in WSN	188
7.1.4	Cross-layer Analysis of Error Control in WSN	189
7.1.5	Cross-layer Packet Size Optimization	189
7.2	Future Research Directions	190
7.2.1	Correlation in Multimedia Sensor Networks:	190
7.2.2	Vast Area Sensor Network Testbed	191
7.2.3	Cross-layer Communication in Cognitive Radio Networks	191
7.2.4	Integration of Sensor Networks and Wireless Internet	192
	REFERENCES	193
	VITA	201
	INDEX	202

LIST OF TABLES

1	Simulation Parameters	100
2	Parameters	132
3	Optimal l_D and ψ_{Th} values	164
4	Overview of Results	167

LIST OF FIGURES

1	Correlation model and architecture.	17
2	Observed Event Distortion for different θ_1 values according to changing number of representative nodes	33
3	Observed event distortion for varying normalized reporting frequency.	34
4	Distortion vs. sampling rate for different values of (a) M , (b) K , (c) θ_T , and (d) θ_S (Point source).	35
5	Distortion vs. sampling rate for different values of θ_T (Field source).	38
6	Number of nodes vs. sampling rate, (M, f) tuples meeting various D_{max} constraints (Field source).	39
7	Distortion vs. M for different grid sizes, $f = 13.74$ (Field source).	40
8	Distortion vs. M for different network topology and node selection procedures (Field source).	42
9	Spatial Re-usage in Sensor Networks.	51
10	Comparison between minimum distortion among 1000 random trials and the distortion found by the INS solution.	53
11	Iterative Node Selection Algorithm. RunVQ() performs VQ algorithm and finds the places of representative nodes.	54
12	16 Representative nodes chosen by VQ algorithm and the Voronoi regions representing the correlation regions.	55
13	E-MAC and N-MAC. The representative node transmits its record on behalf of the entire correlation region, while all correlation neighbors suppress their transmissions.	57
14	Structures for RTS, CTS, and DATA packets.	60
15	Two cases for E-MAC. The figure shows two cases for correlation region, r_{corr} , and transmission region, R	62
16	Energy consumption for different correlation radius values.	67
17	Distortion achieved by different correlation radius values.	68
18	Average energy consumption for various SSS durations.	69
19	Distortion values for various SSS durations.	70
20	Medium access delay introduced by different MAC protocols.	72

21	Percentage of dropped packets for different MAC protocols.	73
22	Goodput for different MAC protocols.	73
23	Average energy consumption for different MAC protocols.	74
24	Achieved distortion and the distortion found by the INS algorithm. . .	75
25	Illustration of angle-based routing.	88
26	Sample route created by angle-based routing.	90
27	Pseudocode of XLM	95
28	Avg. energy consumption vs. duty cycle for different values of D . . .	99
29	Route Failure Rate for XLM with angle-based routing and default geographical routing.	101
30	(a) Average throughput, (b) average goodput, and (c) average latency vs. duty cycle for different values of ξ_{Th}	103
31	(a) Average throughput, (b) average goodput, and (c) percentage of dropped packets due to retransmission timeout vs. duty cycle for different values of ξ_{Th}	106
32	(a) Average energy consumption per packet, (b) average hop count, and (c) average latency vs. duty cycle for layered protocol suites and XLM.	109
33	2D average packet reception rate graphs for (a) ARQ (FSK), (b) FEC (FSK), (c) ARQ (DSSS/OQPSK), (d) FEC (DSSS/OQPSK).	118
34	Reference model for the derivations.	121
35	(a) Avg. hop distance and (b) avg. energy consumption of a flow vs. ψ_{Th} (DSSS+OQPSK).	133
36	(a) End-to-end PER vs. received SNR and (b) avg. end-to-end latency vs. ψ_{Th} (DSSS+OQPSK).	134
37	Taxonomy function vs. ψ_{Th} for (a) DSSS+OQPSK, (b) FSK, and (c) FSK without E_{neigh}	136
38	(a) Avg. energy consumption and (b) avg. end-to-end latency vs. ψ_{Th} for different values of transmit power (FSK).	138
39	Average energy consumption vs. end-to-end distance for (a) Mica2 and (b) MicaZ nodes, and the average energy consumption vs. target end-to-end PER for (c) Mica2 and (d) MicaZ nodes.	140

40	Avg. energy consumption vs. ψ_{Th} for hybrid ARQ (a) Type I and (b) Type II (DSSS+OQPSK), and average energy consumption vs. target end-to-end PER for (a) Mica2 and (b) MicaZ nodes.	142
41	Minimum energy consumption vs. error correcting capability for (a) Mica2 and (b) MicaZ, and minimum latency vs. error correcting capability for (c) Mica2 and (d) MicaZ nodes with $PER_{e2e} \leq 10^{-2}$	144
42	(a) Packet error rate vs. payload length for ARQ and FEC. MAC failure rate vs. payload length for (b) different values of M and (c) different values of FEC rate.	151
43	Energy per useful bit (103) vs. payload length for (a) ARQ and (b) BCH with $t=5$. (c) Energy consumption per bit with optimum packet length and $l_D = 250$ bytes.	163
44	Boundaries for the acceptable ξ_{Th} - packet size region subject to (a) end-to-end packet error rate and (b) end-to-end delay. (c) Optimal packet length for different error control techniques.	166
45	Optimum energy consumption vs. ψ_{Th} for (a) deep water and (b) shallow water, and optimum packet size for (c) deep water and (d) shallow water.	175
46	Sensitivity of (a) energy consumption and (b) throughput to packet size for ARQ and FEC in deep water.	176
47	Optimization results for \mathbf{P}_{min}^{eng} , \mathbf{P}_{max}^{tput} , and \mathbf{P}_{Min}^{eng}	178
48	Constrained optimization results as a function of error correction capability, t , for different delay bounds in shallow water.	179
49	Optimization results for \mathbf{P}_{min}^{eng} and \mathbf{P}_{max}^{tput} in underground environment as a function of depth.	181
50	Optimization results for \mathbf{P}_{min}^{eng} and \mathbf{P}_{max}^{tput} in underground environment as a function of volumetric water content.	182
51	Constrained optimization results as a function of end-to-end delay bound in underground environment for (a-c) burial depth and (d-f) volumetric water content.	184

LIST OF ABBREVIATIONS

ARQ	Automatic Repeat reQuest.
BCH	Bose, Ray-Chaudhuri, Hocquenghem.
BLER	Block Error Rate.
CC-MAC	Correlation-based Collaborative Medium Access Control.
CSMA	Carrier Sense Multiple Access.
DCF	Distributed Coordinate Function.
DIFS	DCF Interframe Space.
DSSS	Direct Sequence Spread Spectrum.
E-MAC	Event MAC.
ESRT	Event-to-Sink Reliable Transport.
FCP	First Contention Phase.
FEC	Forward Error Correction.
FSK	Frequency Shift Keying.
HARQ	Hybrid ARQ.
INS	Iterative Node Selection.
JGRV	Jointly Gaussian Random Variable.
MAC	Medium Access Control.
MEMS	Micro Electro-Mechanical Systems.
MMSE	Minimum Mean Square Error.
N-MAC	Network MAC.
NACK	Negative Acknowledgement.
O-QPSK	Offset Quadrature Phase Shift Keying.
PCF	Point Coordinate Function.
PER	Packet Error Rate.
PIFS	PCF Interframe Space.

PRR	Packet Reception Rate.
QoS	Quality of Service.
RS	Reed-Solomon.
SNR	Signal to Noise Ratio.
SSS	Suspicious Sleep State.
UW-ASN	Underwater Acoustic Sensor Networks.
VQ	Vector Quantization.
WSN	Wireless Sensor Networks.
WSS	Wide-Sense Stationary.
WUSN	Wireless Underground Sensor Networks.
XLM	Cross-layer Module.

SUMMARY

Wireless sensor networks (WSN) are event based systems that rely on the collective effort of densely deployed sensor nodes continuously observing a physical phenomenon. The spatio-temporal correlation between the sensor observations and the cross-layer design advantages are significant and unique to the design of WSN. Due to the high density in the network topology, sensor observations are highly correlated in the space domain. Furthermore, the nature of the energy-radiating physical phenomenon constitutes the temporal correlation between each consecutive observation of a sensor node. This unique characteristic of WSN can be exploited through a cross-layer design of communication functionalities to improve energy efficiency of the network.

In this thesis, several key elements are investigated to capture and exploit the correlation in the WSN for the realization of advanced efficient communication protocols. A theoretical framework is developed to capture the spatial and temporal correlations in WSN and to enable the development of efficient communication protocols. Based on this framework, spatial Correlation-based Collaborative Medium Access Control (CC-MAC) protocol is described, which exploits the spatial correlation in the WSN in order to achieve efficient medium access. Furthermore, the cross-layer module (XLM), which melts common protocol layer functionalities into a cross-layer module for resource-constrained sensor nodes, is developed. The cross-layer analysis of error control in WSN is then presented to enable a comprehensive comparison of error control schemes for WSN. Finally, the cross-layer packet size optimization framework is described.

CHAPTER I

INTRODUCTION

With the recent advances in *Micro Electro-Mechanical Systems* (MEMS) technology, wireless communications, and digital electronics; the construction of low-cost, low-power, multifunctional sensor nodes that are small in size and communicate untethered in short distances has become feasible. The ever-increasing capabilities of these tiny sensor nodes, which consist of sensing, data processing, and communicating components, enable the realization of wireless sensor networks (WSN) based on the collaborative effort of a large number of nodes.

Wireless Sensor Networks have a wide range of applications such as environmental monitoring [71], biomedical research [65], human imaging and tracking [29], and military applications [52]. Consequently, WSN are slowly becoming an integral part of our lives. Recently, considerable amount of research efforts have enabled the actual implementation of sensor networks tailored to the unique requirements of certain sensing and monitoring applications.

In order to realize the existing and potential applications for WSN, sophisticated and extremely efficient communication protocols are necessary. WSN are composed of a large number of sensor nodes, which are densely deployed either inside a physical phenomenon or very close to it. In order to enable reliable and efficient observation and initiate right actions, physical phenomenon features should be reliably detected/estimated from the collective information provided by sensor nodes [7]. Moreover, instead of sending the raw data to the nodes responsible for the fusion, sensor nodes use their processing abilities to locally carry out simple computations and transmit only the required and partially processed data. Hence, these properties of

WSN impose unique challenges for development of communication protocols in such an architecture.

The intrinsic properties of individual sensor nodes, pose additional challenges to the communication protocols in terms of energy consumption. WSN applications and communication protocols are mainly tailored to provide high energy efficiency. Sensor nodes carry limited, generally irreplaceable power sources. Therefore, while traditional networks aim to achieve high *Quality of Service* (QoS) levels, sensor network protocols focus primarily on energy conservation. Moreover, the deployment of the WSN is another constraint that is considered in developing WSN protocols. The position of sensor nodes need not be engineered or pre-determined. This allows random deployment in inaccessible terrains or disaster relief operations. On the other hand, the random deployment constraints of WSN result in self-organizing protocols to emerge in the WSN protocol stack. In addition to the placement of nodes, the density in the network is also exploited in WSN protocols. Since generally, large number of sensor nodes are densely deployed in WSN, neighbor nodes may be very close to each other. Hence, multihop communication in sensor networks is exploited in communication between nodes since it leads to less power consumption than the traditional single hop communication.

Since energy consumption is the major constraint in WSN, most of the proposed communication protocols improve the energy efficiency to a certain extent by exploiting the collaborative nature of WSN. However, the main commonality of these protocols is that they follow the traditional layered protocol architecture. While these protocols may achieve very high performance in terms of the metrics related to each of these individual protocol layers, they are not jointly designed to maximize the overall network performance while minimizing the energy expenditure. Considering the scarce energy and processing resources of WSN, joint design of networking layers, i.e., cross-layer design, stands as the most promising alternative to inefficient traditional

layered protocol architectures.

Accordingly, an increasing number of recent work has focused on the cross-layer development of wireless sensor network protocols. In fact, recent work on WSN [75][81] reveal that cross-layer integration and design techniques result in significant improvement in terms of energy conservation. Generally, there are three main reasons behind this improvement. First, the stringent energy, storage, and processing capabilities of wireless sensor nodes necessitate such an approach. The significant overhead of layered protocols results in high inefficiency. Moreover, recent empirical studies highlight that the properties of low power radio transceivers and the wireless channel conditions need to be considered in protocol design. Finally, the event-centric approach of WSN requires application-aware communication protocols, which also mandates a cross-layer communication protocol design.

In addition to the wireless channel impact and cross-layer interactions, spatio-temporal correlation is another significant characteristic of sensor networks. The dense deployment coupled with the physical properties of the sensed phenomenon introduce correlation in the spatial and temporal domain. As a result, the spatio-temporal correlation-based protocols emerge for improved efficiency in networking wireless sensors. The correlation in WSN can be classified into two, i.e., spatial and temporal correlation. Typical WSN applications require spatially dense sensor deployment in order to achieve satisfactory coverage [20, 51]. Due to the high density in the network topology, spatially proximal sensor observations are highly correlated with the degree of correlation increasing with decreasing internode separation. Furthermore, the nature of the energy-radiating physical phenomenon constitutes the temporal correlation between each consecutive observation of a sensor node [42]. The degree of correlation between consecutive sensor measurements may vary according to the temporal variation characteristics of the phenomenon. The existence of the spatial and temporal correlations bring significant potential advantages for the development

of efficient communication protocols well-suited for the WSN paradigm.

In WSN, correlation between sensors can be exploited in terms of *aggregation*, *collaborative source coding*, or *correlation-based protocols*. Consequently, these techniques aim to reduce the redundancy in the traffic by filtering correlated data. This makes it essential for each packet to be transmitted reliably highlighting the importance of energy efficient error control. Moreover, the multi-hop features of the WSN require a unique definition of reliability other than the conventional reliability metrics that focus on point-to-point reliability. More specifically, in a WSN, when a packet is injected into the network, each node along the path to the sink consumes a certain amount of its scarce resources to relay the packet. Each packet has a different reliability notion because of the path it has already traversed. Furthermore, in WSN, the applications are interested in the collaborative information from sensors about a specific event, rather than individual readings of each sensor. Consequently, the reliability notion considered in WSN differs from the approach in traditional wireless networks, in terms of both multi-hop reliability and event-based reliability.

The main objective of the WSN is to reliably detect/estimate event features from the collective information provided by sensor nodes. While the collaborative nature of the WSN brings significant advantages over traditional sensing; the spatio-temporal correlation between the sensor observations and the cross-layer design in the WSN are another significant and unique characteristics of the WSN that needs to be exploited to drastically enhance the overall network performance. In this thesis, several key elements are investigated to capture and exploit the correlation in the WSN for the realization of cross-layer communication protocols. The performed work can be mainly classified into two topics, i.e., correlation and cross-layer communication in WSN. More specifically, the challenges addressed in this thesis can be summarized as follows:

- *Spatio-temporal Correlation*: The main goal of this research is to develop a

theoretical framework to model the spatio-temporal correlation in WSN and investigate the interdependencies between the properties of the physical phenomenon that is sensed and the network parameters. Accordingly, first spatial and temporal correlation is investigated independently. Furthermore, the theory of spatio-temporal correlation is established.

- *Spatial Correlation-based MAC*: Based on the spatio-temporal correlation framework, we propose a spatial Correlation-based Collaborative Medium Access Control (CC-MAC) protocol as a first attempt to exploit correlation in communication protocols. CC-MAC minimizes unnecessary channel access contention by filtering correlated data transmission.
- *Cross-layer Communication*: To the best of our knowledge, to date, there is no unified cross-layer communication protocol for efficient and reliable event communication which considers transport, routing, medium access functionalities with physical layer (wireless channel) effects for WSN. In this research, a unified cross-layer module (XLM) that melts common protocol layer functionalities into a cross-layer module is proposed.
- *Error Control*: The selection of error control scheme is another important design decision in WSN. Forward Error Correction (FEC) schemes improve the error resiliency compared to Automatic Repeat reQuest (ARQ). In a multi-hop network, this improvement can be exploited by reducing the transmit power or by constructing longer hops, however, at the cost of energy consumption. We propose a cross-layer analysis framework to evaluate this tradeoff and assess the performance of FEC and ARQ schemes in WSN.
- *Packet Size Optimization*: Since WSN are packet based networks, the determination of packet size also plays an important role in cross-layer design of communication protocols. Hence, we propose a cross-layer packet size optimization

framework in our research.

1.1 Research Objectives and Solutions

In order to address the challenges posed by spatio-temporal correlation characteristics of observed data in WSN and the cross-layer design requirements of communication protocols for these networks, several topics need to be investigated. In this thesis, new theoretical tools and communication protocols are developed to capture the effects of spatio-temporal correlation in cross-layer communication protocols design. The following five areas are investigated under this research and each of them is described in the following subsections:

1. Theory of Correlation in Wireless Sensor Networks
2. Spatial Correlation-based Collaborative Medium Access Control in Wireless Sensor Networks
3. XLM: Cross-Layer Module for Efficient Communication in Wireless Sensor Networks
4. Cross-Layer Analysis of Error Control in Wireless Sensor Networks
5. Cross-layer Packet Size Optimization for Wireless Sensor Networks

1.1.1 Theory of Correlation in Wireless Sensor Networks

Wireless Sensor Networks (WSN) are characterized by the dense deployment of sensor nodes that continuously observe physical phenomenon. Due to high density in the network topology, sensor observations are highly correlated in the space domain. Furthermore, the nature of the physical phenomenon constitutes the temporal correlation between each consecutive observation of a sensor node. These spatial and temporal correlations along with the collaborative nature of the WSN bring significant potential advantages for the development of efficient communication protocols well-suited

for the WSN paradigm. Based on the physical characteristics and dispersion pattern over an area, the physical phenomenon to be observed can be modeled as point source or field source. Clearly, understanding the spatio-temporal correlation characteristics of the point and field sources brings potential advantages to be exploited in the design of efficient communication protocols. Although there has been some research effort to study the correlation in WSN [31, 56, 58, 64], most of these existing studies investigate the information theoretical aspects of the correlation, and they do not provide efficient networking protocols that exploit the correlation in the WSN.

In this thesis, several key elements are investigated to capture and exploit the correlation in the WSN for the realization of advanced efficient communication protocols. A theoretical framework is developed to model the spatial and temporal correlations in sensor networks. The objective of this analysis is to capture the spatio-temporal characteristics of point and field sources in WSN. First, the spatial and temporal correlations are investigated independently. Furthermore, models for point and field sources are developed and their spatio-temporal characteristics are analytically derived along with the distortion functions. Based on the theoretical analysis, numerical simulations are performed. This analytical work provides tools for finding the feasible operating region in terms of spatial and temporal resolution for a specific distortion constraint considering spatio-temporal correlation, signal properties, and network variables in WSN. The insight obtained from this framework enables the development of efficient communication protocols that exploit these advantageous intrinsic features of the WSN paradigm. Based on this framework, possible approaches are discussed to exploit spatial and temporal correlation for efficient medium access and reliable event transport in WSN, respectively.

1.1.2 Spatial Correlation-based Collaborative Medium Access Control in Wireless Sensor Networks

The spatio-temporal framework that is developed in this thesis provides insight to developing correlation-based communication protocols. As a result, exploiting spatial correlation in the context of collaborative nature of the WSN can lead to significant performance improvement of communication protocols. Intuitively, data from spatially separated sensor nodes are more useful for the sink than the highly correlated data from closely located sensor nodes. Hence, it may not be necessary for every sensor node to transmit its data to the sink; instead, a smaller number of sensor data might be adequate to transmit certain event features to the sink within a certain distortion constraint.

In this thesis, the spatial correlation among the observations of closely location sensors is exploited at the medium access control (MAC) layer. Current studies on medium access in WSN focus mainly on the energy-latency tradeoffs and on decreasing energy consumption by modifying known medium access techniques [83], [74], [10], [59]. To the best of our knowledge, this is the first effort that exploits spatial correlation in WSN at the MAC layer. Based on the spatio-temporal framework, it is shown that a sensor node can act as a representative node for several other sensor nodes observing the correlated data. Consequently, a distributed, spatial Correlation-based Collaborative Medium Access Control (CC-MAC) protocol is designed which has two components: *Event MAC (E-MAC)* and *Network MAC (N-MAC)*. E-MAC filters out the correlation in sensor records while N-MAC prioritizes the transmission of route-thru packets. Simulation results show that CC-MAC achieves high performance in terms energy, packet drop rate, and latency.

1.1.3 XLM: Cross-Layer Module for Efficient Communication in Wireless Sensor Networks

Severe energy constraints of battery-powered sensor nodes necessitate energy-efficient communication in order to fulfill application objectives of Wireless Sensor Networks (WSN). However, the vast majority of the existing solutions are based on classical layered protocol approach. Although many of the basic functionalities of each layer are crucial for proper network operation, the layering approach leads to significant overhead. It is much more efficient to have a unified scheme which melts common protocol layer functionalities into a cross-layer module for resource-constrained sensor nodes. This view has been shared in many recent work [16], [17], [27] [50], [67], [69], [87], [86], where pair-wise cross-layer communication techniques have been proposed. However, to the best of our knowledge, to date, there is no unified cross-layer communication protocol that incorporates transport, routing, and medium access functionalities along with considering the physical layer (wireless channel) effects in a single module.

In this thesis, a unified cross-layer module (XLM), which replaces the entire traditional layered protocol architecture that has been used so far in WSN, is developed. The design principle of XLM is complete unified cross-layering such that both the information and the functionalities of traditional communication layers are melted in a single module. The objective of XLM is highly reliable communication with minimal energy consumption, adaptive communication decisions, and local congestion avoidance. To this end, the protocol operation of XLM is governed by the new cross-layer concept of *initiative determination*. Based on this concept, XLM performs receiver-based contention, initiative-based forwarding, local congestion control, and distributed duty cycle operation in order to realize efficient and reliable communication in WSN. In a cross-layer simulation platform, the state-of-the-art layered protocol configurations have been implemented along with XLM to realize a complete

performance evaluations, which show that XLM significantly improves the communication performance and outperforms the traditional layered protocol architectures in terms of both network performance and implementation complexity.

1.1.4 Cross-Layer Analysis of Error Control in Wireless Sensor Networks

Error control is of significant importance for Wireless Sensor Networks (WSN) because of their severe energy constraints and the low power communication requirements. In this thesis, a cross-layer methodology for the analysis of error control schemes in WSN is presented such that the effects of multi-hop routing and the broadcast nature of the wireless channel are investigated. More specifically, the cross-layer effects of routing, medium access and physical layers are considered. This analysis enables a comprehensive comparison of forward error correction (FEC) codes, automatic repeat request (ARQ), and hybrid ARQ schemes in WSN. So far, the performance of FEC codes have been investigated in a point-to-point fashion [46, 62, 66] in the context of WSN. To the best of our knowledge, this is the first work that considers both the broadcast wireless channel and the multi-hop structure of WSN with realistic channel models and a 2 dimensional topology.

Hybrid ARQ and FEC schemes improve the error resiliency of communication compared to ARQ. In a multi-hop network, this improvement can be exploited by reducing the transmit power (*transmit power control*) or by constructing longer hops (*hop length extension*), which can be achieved through channel-aware routing protocols. The results of our analysis reveal that for hybrid ARQ schemes and certain FEC codes, the hop length extension decreases both the energy consumption and the end-to-end latency subject to a target PER compared to ARQ. This decrease in end-to-end latency is crucial for delay sensitive, real-time applications, where both hybrid ARQ and FEC codes can be strong candidates. On the other hand, transmit power control results in significant savings in energy consumption at the cost of increased

latency for certain FEC codes. The results of our analysis indicate the cases where ARQ outperforms FEC codes for various end-to-end distance and target PER values.

1.1.5 Cross-layer Packet Size Optimization for Wireless Sensor Networks

The unique characteristics of WSN necessitate rethinking of classical wireless networking in all aspects [7]. Among these, the determination of the optimal packet size tailored for these unique characteristics constitutes a fundamental and still unexplored problem in WSN. Traditionally, packet size optimization is performed considering a point-to-point link, where the goal is to ensure a successful and efficient transmission based on an efficiency metric [44], [55], [62]. However, in this traditional approach, the influence of multi-hop and broadcast nature of wireless communication in WSN cannot be captured.

In this thesis, a cross-layer solution for packet size optimization in WSN is introduced such that the effects of multi-hop routing, the broadcast nature of the physical wireless channel, and the effects of error control techniques are captured. A key result of this research is that contrary to the conventional wireless networks, longer packets reduce the collision probability. On the other hand, packets cannot be too long because of the wireless channel errors, which lead to retransmissions and increased energy consumption. Our cross-layer analysis clearly reveals this tradeoff. An optimization solution is formalized by using three different objective functions, i.e., throughput, energy consumption, and resource utilization. Each objective function highlights a different aspect of communication in WSN and can be selected according to the requirements of the application in use. Furthermore, the effects of end-to-end latency and reliability constraints are investigated that may be required by a particular application. As a result, a flexible, cross-layer optimization framework is developed to determine the optimal packet size in WSN. From this framework, the optimal packet sizes under various network parameters are determined. Furthermore,

the optimal packet size optimization framework is extended for wireless underwater and underground sensor networks and the results are presented.

1.2 Thesis Outline

This thesis is organized as follows: Chapter 2 presents the spatio-temporal framework, which is developed to capture the spatial and temporal correlations in WSN. The objective of this framework is to enable the development of efficient communication protocols which exploit the correlation. Based on this framework, Chapter 3 introduces the spatial Correlation-based Collaborative Medium Access Control (CC-MAC) protocol, which exploits the spatial correlation in the WSN in order to achieve efficient medium access. Chapter 4 describes the cross-layer module (XLM) that melds common protocol layer functionalities into a cross-layer module for resource-constrained sensor nodes. Chapter 5 presents the cross-layer analysis of error control in WSN. This framework enables a comprehensive cross-layer comparison of ARQ, FEC, and hybrid ARQ schemes in WSN. Chapter 6 introduces the cross-layer packet size optimization framework. Finally, Chapter 7 summarizes the research results and suggests a number of problems for future investigation.

CHAPTER II

THEORY OF CORRELATION IN WIRELESS SENSOR NETWORKS

In this chapter, we develop a theoretical framework for the spatio-temporal correlation in wireless sensor networks. First, the architecture and the correlation model are presented. Based on this model, we first investigate spatial and temporal correlation in WSN separately. More specifically, the distortion functions for spatial and temporal correlation are derived. Then, the spatio-temporal correlation in WSN is investigated based on two source models, i.e., point and field sources. The results of this analysis are then further explored in the context of correlation-based cross-layer module design in the following sections. This framework was first introduced in [8] and then significantly enhanced in [78] and [77]. The remainder of this chapter is organized as follows. The motivation for this work is provided in Section 2.1. In Section 2.2, the related work in this area is summarized. The theoretical framework is developed to model the spatial and temporal correlations in wireless sensor networks in Section 2.3. Based on the theoretical analysis, in Section 2.4, the numerical simulations for spatio-temporal correlation characteristics of WSN data sources are performed. The results of this study yield a significant theoretical background to be used in the development of efficient communication protocols. In Section 2.5, applications of the results of this framework are explored for energy-efficient communication in WSN.

2.1 Motivation

The main objective of the WSN is to reliably detect/estimate event features from the collective information provided by sensor nodes. Therefore, the energy and hence processing constraints of small wireless sensor nodes are overcome by this collective sensing notion which is realized via their networked deployment. While the collaborative nature of the WSN brings significant advantages over traditional sensing including greater accuracy, larger coverage area, and extraction of localized features; the spatio-temporal correlation among the sensor observations is another significant and unique characteristic of the WSN which can be exploited to drastically enhance the overall network performance. The characteristics of the correlation in the WSN can be summarized as follows:

- *Spatial Correlation:* Typical WSN applications require spatially dense sensor deployment in order to achieve satisfactory coverage [20, 51]. As a result, multiple sensors record information about a single event in the sensor field. Due to the high density in the network topology, spatially proximal sensor observations are highly correlated with the degree of correlation increasing with decreasing internode separation.
- *Temporal Correlation:* Some of the WSN applications such as event tracking may require sensor nodes to periodically perform observation and transmission of the sensed event features. The nature of the energy-radiating physical phenomenon constitutes the temporal correlation between each consecutive observation of a sensor node [42]. The degree of correlation between consecutive sensor measurements may vary according to the temporal variation characteristics of the phenomenon.

The spatial and temporal correlation characteristics of a WSN application depends also on the type of the source that generates the physical phenomenon. Based on the

application, the physical phenomenon to be observed can be modeled as *point source*, e.g., target detection/tracking, or *field source*, e.g., monitoring of magnetic field and seismic activities. In general, events generating signal that originates from a single point in the field can be modeled as a point source. The cases, where the physical phenomenon is dispersed over the field, can be modeled as a field source. Clearly, it is of great importance to capture the spatio-temporal characteristics of point and field sources to be able design energy-efficient communication protocols which can exploit the potential advantages of correlation in WSN.

In addition to the collaborative nature of the WSN, the existence of above mentioned spatial and temporal correlations bring significant potential advantages for the development of efficient communication protocols well-suited for the WSN paradigm. For example, intuitively, due to the spatial correlation, data from spatially separated sensors is more useful to the sink than highly correlated data from nodes in proximity. Therefore, it may not be necessary for every sensor node to transmit its data to the sink. Instead, a smaller number of sensor measurements might be adequate to communicate the event features to the sink within a certain reliability/fidelity level. Similarly, for a certain event tracking application, the measurement reporting frequency, at which the sensor nodes transmit their observations, can be adjusted such that temporal-correlated phenomenon signal is captured at the sink within a certain distortion level and with minimum energy-expenditure.

Consequently, despite the existing proposed works for medium access control and transport layer solutions in sensor networks discussed above, these protocols do not exploit the correlation in the sensed phenomenon. In this section, several key elements are investigated to capture and exploit the correlation in the WSN for the realization of advanced efficient communication protocols. First, a theoretical framework is developed to model the spatial and temporal correlations in sensor networks. This framework is further extended by considering the joint effects of spatio-temporal

correlation for both point and field sources. To this end, first, the model for point and field sources are introduced and their spatio-temporal characteristics are derived along with the distortion functions. The objective of the correlation framework is to enable the development of efficient communication protocols which exploit these advantageous intrinsic features of the WSN paradigm. Based on this framework, possible approaches are discussed to exploit spatial and temporal correlation for efficient medium access and reliable event transport in WSN, respectively.

2.2 Related Work

There has been some research efforts to study the correlation in WSN [11], [24], [25], [26], [30], [31], [32], [56], [58], [64], [32]. However, most of these existing studies investigate the information theoretical aspects of the correlation, and they do not provide efficient networking protocols which exploit the correlation in the WSN. For example, in [31], a joint source-channel coding paradigm is proposed for optimal performance in a class of sensor networks. On the other hand, there exist some proposals that attempt to exploit spatial correlation in WSN [58, 56]. However, these schemes aim to find the optimum rate to compress redundant information in the sensor observations and they also do not propose to exploit correlation for developing efficient communication protocols for the WSN. The joint routing and source coding is introduced in [64] to reduce the amount of traffic generated in dense sensor networks with spatially correlated records. While joint routing and source coding reduces the number of transmitted bits; from the network point of view, the number of transmitted packets remains unchanged, which can be further minimized by regulating the network access based on the spatial correlation between the sensor nodes.

In [25], the problem of correlated data gathering with the objective of energy minimization is studied using different coding models. Similar analysis is performed in [24] for lossy network and high-resolution coding under distortion constraints. The

optimal network density under the total distortion constraint for delay-sensitive real-time data gathering of correlated physical phenomenon in WSN is investigated in [26]. Similarly, in [30], the joint optimization of sensor placement and transmission structure for data gathering is proposed. However, these studies consider only the spatial correlation in the WSN and do not incorporate the temporal correlation in their analysis. In [32], a lower bound on the best achievable end-to-end distortion for different coding schemes is derived as a function of the network parameters. Moreover, in [11], the relation between spatial and temporal sampling rate on the overall network delay and energy consumption is studied. However, these work neither consider nor capture the spatio-temporal correlation characteristics of the physical phenomenon.

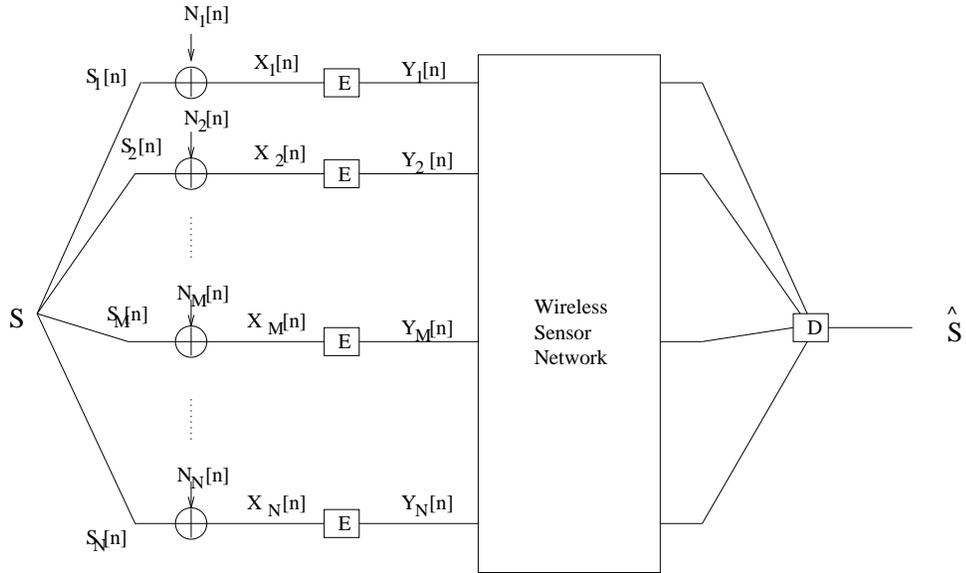


Figure 1: Correlation model and architecture.

2.3 Spatio-Temporal Correlation in Wireless Sensor Networks

In this section, we develop the theoretical framework for the spatio-temporal correlation in wireless sensor networks. The results of this analysis are then further explored in the context of correlation-based medium access control and reliable event transport approaches for WSN.

2.3.1 Architecture and Correlation Model for WSN

In a sensor field, each sensor observes the noisy version of a physical phenomenon. The sink is interested in observing the physical phenomenon using the observations from sensor nodes with the highest accuracy. The physical phenomenon in interest can be modeled as a spatio-temporal process $s(t, x, y)$ as a function of time t and spatial coordinates (x, y) .

Depending on the specific sensor application, the physical phenomenon may be a spatio-temporal process generated by a point source in case of applications such as object tracking. In this case, the sink is interested in reconstructing the source signal at a specific location (x_0, y_0) based on sensor observations. In other applications, the spatio-temporal process may be a combination of multiple point sources where the sink is interested in reconstructing the signal in multiple locations or over an event area. Although the reconstruction is application specific, the properties of the observations can be modeled based on the spatio-temporal process $s(t, x, y)$.

The model for the information gathered by N sensors in an event area is illustrated in Fig. 1. The sink is interested in estimating the event source, S , according to the observations of the sensor nodes, n_i , in the event area. Each sensor node n_i observes $X_i[n]$, the noisy version of the event information $S_i[n]$, which is spatially correlated to the event source, S . In order to communicate this observation to the sink through the WSN, each node has to encode its observation. The encoded information, $Y_i[n]$, is then sent to the sink through the WSN. The sink, at the other end, decodes this information to get the estimate, \hat{S} , of the event source S . The encoders and the decoders are labelled as E and D in Fig. 1, respectively. Using this model, we will exploit various aspects of correlation among sensor readings both in terms of time and space.

Each observed sample, $X_i[n]$, of sensor n_i at time n is represented as

$$X_i[n] = S_i[n] + N_i[n] , \quad (1)$$

where the subscript i denotes the spatial location of node n_i , i.e. (x_i, y_i) , $S_i[n]$ is the realization of the space-time process $s(t, x, y)$ at time $t = t_n$ ¹ and $(x, y) = (x_i, y_i)$, and $N_i[n]$ is the observation noise. $\{N_i[n]\}_n$ is a sequence of i.i.d Gaussian random variables of zero mean and variance σ_N^2 . We further assume that the noise each sensor node encounters is independent of each other, i.e., $N_i[n]$ and $N_j[n]$ are independent for $i \neq j$ and $\forall n$.

As it is shown in Fig. 1, each observation $X_i[n]$ is then encoded into $Y_i[n]$ by the source-coding at the sensor node as

$$Y_i[n] = f_i(X_i[n]) \quad (2)$$

and then sent through the network to the sink. The sink decodes the received data to reconstruct an estimation \hat{S} of the source S

$$\hat{S} = g(Y_1[n_1], \dots, Y_1[n_\tau]; \dots; Y_N[n_1], \dots, Y_N[n_\tau]) \quad (3)$$

based on the data received from N nodes in the event area over a time period $\tau = t_{n_\tau} - t_{n_1}$. The sink is interested in reconstructing the source S according to a distortion constraint

$$D = E \left[d(S, \hat{S}) \right] . \quad (4)$$

In the next subsections, the general distortion function in (4) will be used to independently obtain the distortion functions for spatial and temporal correlation in the WSN, which can then be explored in discussing the correlation-based medium access control and reliable event transport approaches for WSN in Sections 2.5.1 and 2.5.2, respectively. Furthermore, we extend the independent analysis of spatial and temporal correlation by considering the joint effects of spatio-temporal correlation.

¹Note that, we use a discrete-time model since each node is assumed to sample the physical phenomenon synchronously after the initial wake-up.

2.3.2 Spatial Correlation in WSN

In this section, we model the spatial correlation between observations of each sensor node. The information gathered by N sensors in an event area can be modeled as shown in Fig. 1. The sink is assumed to be interested in a point source S . Since we only consider the spatial correlation between nodes, in this analysis, we assume that the samples are temporally independent. Hence, by dropping the time index n , (1) can be restated as

$$X_i = S_i + N_i, \quad i = 1, \dots, N. \quad (5)$$

The sink is interested in reconstructing the source S according to observations of nodes n_i which observe the spatially correlated version of S at (x_i, y_i) , i.e., S_i . The physical phenomenon is modeled as jointly gaussian random variables (JGRVs) at each observation point as

$$\begin{aligned} E\{S_i\} &= 0, \quad i = 1, \dots, N, \\ \text{var}\{S_i\} &= \sigma_S^2, \quad i = 1, \dots, N, \\ \text{cov}\{S_i, S_j\} &= \sigma_S^2 \text{corr}\{S_i, S_j\}, \\ \text{corr}\{S_i, S_j\} &= \rho_{i,j} = K_\vartheta(d_{i,j}) = \frac{E[S_i S_j]}{\sigma_S^2}, \end{aligned}$$

where $d_{i,j} = \|\mathbf{s}_i - \mathbf{s}_j\|$ denotes the distance between nodes n_i and n_j located at coordinates \mathbf{s}_i and \mathbf{s}_j , respectively and $K_\vartheta(\cdot)$ is the covariance function. The covariance function is assumed to be non-negative and decrease monotonically with the distance $d = \|\mathbf{s}_i - \mathbf{s}_j\|$, with limiting values of 1 at $d = 0$ and of 0 at $d = \infty$. Generally, covariance models can be classified into four groups [13]:

- *Spherical* :

$$K_\vartheta^S(d) = \begin{cases} 1 - \frac{3}{2} \frac{d}{\theta_1} + \frac{1}{2} \left(\frac{d}{\theta_2}\right)^3 & \text{if } 0 \leq d \leq \theta_1 \\ 0 & \text{if } d > \theta_1 \end{cases} ; \theta_1 > 0.$$

In this model, two observations taken more than θ_1 distance apart are uncorrelated

- *Power Exponential* :

$$K_{\vartheta}^{PE}(d) = e^{(-d/\theta_1)^{\theta_2}}; \theta_1 > 0, \theta_2 \in (0, 2].$$

For $\theta_2 = 1$, the model becomes exponential, while for $\theta_2 = 2$ squared exponential.

- *Rational Quadratic* :

$$K_{\vartheta}^{RQ}(d) = \left(1 + \left(\frac{d}{\theta_1}\right)^2\right)^{-\theta_2}; \theta_1 > 0, \theta_2 > 0.$$

- *Matérn* :

$$K_{\vartheta}^M(d) = \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)} \left(\frac{d}{\theta_1}\right)^{\theta_2} \mathcal{K}_{\theta_2}\left(\frac{d}{\theta_1}\right); \theta_1 > 0, \theta_2 > 0,$$

where $\mathcal{K}_{\theta_2}(\cdot)$ is the modified Bessel function of second kind and order θ_2 .

The correlation model can be chosen according to the properties of the physical phenomenon the sink is interested in. Since we are interested in S , which is also a JGRV, we use a special notation with

$$\begin{aligned} \text{var}\{S\} &= \sigma_S^2, \\ \text{corr}\{S, S_i\} &= \rho_{s,i} = K_{\vartheta}(d_{s,i}) = \frac{E[SS_i]}{\sigma_S^2}, \end{aligned}$$

where $d_{s,i}$ denotes the distance between the source S and the node n_i . The observation noise N_i of each node n_i is modeled as i.i.d. Gaussian random variable with zero mean and variance σ_N^2 , i.e., $N_i \sim \mathcal{N}(0, \sigma_N^2)$.

As each sensor node n_i observes an event information X_i , this information is encoded and then sent to the sink through the WSN. In traditional point-to-point communication, the optimum performance is obtained by compressing the information

according to the source statistics and then adding redundant information to accommodate the errors introduced in the wireless channel. This technique is known as the *separation principle*. In WSN, where multiple nodes try to send information about the same event, however, it is known that joint source-channel coding outperforms separate coding [31, 57]. In addition, for Gaussian sources, if the source is Gaussian and the cost on the channel is the encoding power, then uncoded transmission is optimal for point to point transmission [33]. Furthermore, for sensor networks with finite number of nodes, uncoded transmission outperforms any approach based on the separation paradigm leading to the optimal solution for infinite number of nodes [31]. Hence, we adopt uncoded transmission for the sensor observations in this work. Each node n_i sends to the sink, a scaled version, Y_i , of the observed sample X_i according to encoding power constraint P_E .

$$Y_i = \sqrt{\frac{P_E}{\sigma_S^2 + \sigma_N^2}} X_i, \quad i = 1, \dots, N, \quad (6)$$

where σ_S^2 and σ_N^2 are the variances of the event information S_i and the observation noise N_i , respectively.

The sink needs to calculate the estimation of each event information, S_i , in order to estimate the event source S . Since uncoded transmission is used, it is well known that minimum mean square error (MMSE) estimation is the optimum decoding technique [54]. Hence, the estimation, Z_i , of the event information S_i is simply the MMSE estimation of Y_i , which is given by

$$Z_i = \frac{E[S_i Y_i]}{E[Y_i^2]} Y_i. \quad (7)$$

Note that the estimated values of Z_i 's are spatially correlated since the actual event information S_i 's are spatially correlated. This spatial correlation results in redundancy in each event information sent to the sink. Although the sink is interested in estimating the event source, S , with a distortion constraint, intuitively, this constraint can still be met by using a smaller number of sensor nodes rather than all the

nodes in the event area. In order to investigate the distortion achieved when smaller number of nodes sending information, we assume that only M out of N packets are received by the sink, where N is the total number of sensor nodes in the event area. Since the sink decodes each Y_i using the MMSE estimator, the event source can simply be computed by taking the average of all the event information received at the sink. Then, \hat{S} , the estimate of S , is given as,

$$\hat{S}(M) = \frac{1}{M} \sum_{i=1}^M Z_i . \quad (8)$$

The distortion achieved by using M packets to estimate the event S is given as

$$D(M) = E[(S - \hat{S}(M))^2] , \quad (9)$$

where we use the mean-squared error as the distortion metric. Using (5) and (6) in (7), the estimate Z_i of each event information S_i can be written as

$$Z_i = \frac{E[S_i Y_i]}{E[Y_i^2]} \sqrt{\frac{P}{\sigma_S^2 + \sigma_N^2}} (S_i + N_i) . \quad (10)$$

Denoting $\alpha = \sqrt{\frac{P}{\sigma_S^2 + \sigma_N^2}}$,

$$\begin{aligned} E[S_i Y_i] &= \alpha \sigma_S^2 , \\ E[Y_i^2] &= \alpha^2 \left(\sigma_S^2 + \sigma_N^2 \right) , \end{aligned}$$

then, (10) is restated as

$$Z_i = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_N^2} (S_i + N_i) . \quad (11)$$

Using (11) and (8) in (9), the distortion function $D(M)$ is found to be

$$D(M) = \sigma_S^2 - \frac{\sigma_S^4}{M(\sigma_S^2 + \sigma_N^2)} \left(2 \sum_{i=1}^M \rho_{(s,i)} - 1 \right) + \frac{\sigma_S^6}{M^2(\sigma_S^2 + \sigma_N^2)^2} \sum_{i=1}^M \sum_{j \neq i}^M \rho_{(i,j)} . \quad (12)$$

$D(M)$ shows the distortion achieved at the sink as a function of number of nodes M that send information to the sink and correlation coefficients $\rho_{(i,j)}$ and $\rho_{(s,i)}$ between nodes n_i and n_j , and the event source S and node n_i , respectively. Based

on the distortion function, we discuss possible approaches that can be used in the Medium Access Control (MAC) protocols for WSN in Section 2.5.1. Consequently, the correlation-based collaborative MAC (CC-MAC) protocol is described in Chapter 3.

2.3.3 Temporal Correlation in WSN

The energy-radiating physical phenomenon constitutes the temporal correlation between each consecutive observation of a sensor node [42]. For the periodic sensing applications such as event tracking, each consecutively taken sensor observations are temporally correlated to a certain degree. In this section, we establish the theoretical analysis for this temporal correlation, which will be further elaborated in the context of correlation-based reliable event transport approach discussed in Section 2.5.2.

Here, we consider the temporal correlation between the sensor observations and hence we omit the spatial variation in this analysis. We are interested estimating the signal $s(t)$ in a decision interval of τ . In our theoretical analysis, we model an event-to-sink distortion metric, where all the information coming from the sensor nodes in the event area is considered as if it is generated by a single source node during the decision interval τ .

Assume that the sensed information from the sensors are sent to the sink using a reporting frequency of f . In this case, we seek to control the reporting frequency f such that a desired distortion level is not exceeded in the estimation of the event features at the sink. The event signal $s(t)$ is assumed to be a Gaussian random process with $\mathcal{N}(0, \sigma_s^2)$. The sink is interested in finding the expectation of the signal $s(t)$ over the decision interval τ , i.e., $S(\tau)$. Assuming the observed signal $s(t)$ is wide-sense stationary (WSS), the expectation of the signal over the decision interval τ can be calculated by the time average of the observed signal [35], i.e.,

$$S(\tau) = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} s(t)dt , \quad (13)$$

where t_0 is the time the sensor node wakes up for the sampling of the signal. With a change of variables, $S(\tau)$ can be shown as

$$S(\tau) = \frac{1}{\tau} \int_0^\tau s(t_0 + \Gamma) d\Gamma . \quad (14)$$

We define the value of the signal at each sampling interval as

$$S[n] = s\left(t_0 + \frac{n}{f}\right) , \quad (15)$$

where f is the sampling frequency and $S[n]$ are JGRV with $\mathcal{N}(0, \sigma_s^2)$ ². For the derivation of the distortion function, the following definitions are needed:

$$\begin{aligned} E\{S[n]\} &= 0 , \\ E\{(S[n])^2\} &= \sigma_s^2 , \\ E\{S[n]S[m]\} &= \sigma_s^2 \hat{\rho}_S(n, m) , \\ E\{s(t)s(t + \delta)\} &= \sigma_s^2 \rho_S(\delta) , \end{aligned}$$

where $\hat{\rho}_S(n, m) = \rho_S(|m - n|/f)$ is the covariance function that depends on the time difference between signal samples. Although our results about the distortion function apply to all the covariance models introduced in Section 2.3.2, we use the power exponential model in the derivation since the physical event information such as electromagnetic waves is modeled to have an exponential autocorrelation function [72]. Hence, the covariance function becomes

$$\rho_S(\delta) = e^{-|\delta|/\theta_1} . \quad (16)$$

Each sensor node observes the noisy version of the signal given as

$$X[n] = S[n] + N[n] \quad (17)$$

and the transmitted signal is expressed by

$$Y[n] = \sqrt{\frac{P_E}{\sigma_S^2 + \sigma_N^2}} X[n] \quad (18)$$

²Note that, the samples of a Gaussian random process are jointly Gaussian [35]

based on the discussion in Section 2.3.2. Using the MMSE estimator at the sink, each sample is estimated as

$$Z[n] = \frac{E[S[n]Y[n]]}{E[Y^2[n]]} Y[n]. \quad (19)$$

Hence, each estimated sample from the sensor nodes can be represented as

$$Z[n] = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_N^2} (S[n] + N[n]). \quad (20)$$

After collecting all the samples of the signal in the decision interval τ , the sink estimates the expectation of the signal over the last decision interval by

$$\hat{S}(\tau) = \frac{1}{\tau f} \sum_{k=1}^{\tau f} Z[k], \quad (21)$$

where τf is the total number of sensor samples taken within a decision interval with duration of τ . As a result, the distortion achieved by using τf samples to estimate the event is given as

$$D = E \left[\left(S(\tau) - \hat{S}(\tau) \right)^2 \right]. \quad (22)$$

Using the definitions above and substituting (14), (20), and (21) into (22); the distortion function can easily shown to be

$$\begin{aligned} D(f) &= \sigma_S^2 + \frac{\sigma_S^4}{\tau f (\sigma_S^2 + \sigma_N^2)} + \\ &+ \frac{\sigma_S^6}{\tau^2 f^2 (\sigma_S^2 + \sigma_N^2)^2} \sum_{k=1}^{\tau f} \sum_{l \neq k} e^{-\frac{|k-l|}{f}/\theta_1} \\ &- \frac{2\sigma_S^4 \theta_1}{\tau^2 f (\sigma_S^2 + \sigma_N^2)} \sum_{k=1}^{\tau f} \left(2 - e^{-\frac{k}{f\theta_1}} - e^{-\frac{\tau-k}{f}/\theta_1} \right). \end{aligned} \quad (23)$$

It is observed from (23) that the distortion in the estimation decreases with increasing f . Note that the distortion level, D , for the estimation of event features from the sensor observations corresponds to the reliability level of the event-to-sink communication in the WSN. In Section 2.5.2, this distortion function will be further explored in the context of reliable event transport in WSN.

2.3.4 Spatio-temporal Correlation

In this section, we extend the independent analysis of spatial and temporal correlation by considering the joint effects of spatio-temporal correlation. The objective of this analysis is to capture the spatio-temporal characteristics of physical phenomenon modeled by point and field sources in WSN. In the following, first, the model for point and field sources are introduced and their spatio-temporal characteristics are derived along with the distortion functions.

2.3.4.1 Point Source

In many WSN applications such as target detection and fire detection, the goal is to estimate the properties of an event generated by a single point source through collective observations of sensor nodes. In this section, we first introduce our model for the point source and formulate its spatio-temporal characteristics. Then, we derive the distortion function for the estimation of the point source.

The point source is assumed to generate a continuous signal which is modeled as a random process $f_S(s, t)$, where s denotes the outcome and t denotes time. For ease of illustration, we use $f_S(t)$ in the remaining. We model the point source, $f_S(t)$, as a gaussian random process such that $f_S(t)$ is first-order stationary, i.e., $\mu_S(t) = \mu_S$ and has a variance σ_S^2 . Without loss of generality, we assume $\mu_S = 0$.

For ease of illustration, we assume the coordinate axis is centered at the point source. As a result, the received signal, $f(x, y, t)$, at time t at a location (x, y) can be modeled as

$$f(x, y, t) = f_S\left(t - \frac{\sqrt{x^2 + y^2}}{v}\right) e^{-\frac{\sqrt{x^2 + y^2}}{\theta_s}}, \quad (24)$$

which is the delayed and attenuated version of the signal $f_S(t)$. In this model, we assume that the event signal travels with the speed, v , and is attenuated based on an exponential law, where θ_s is the attenuation constant. Note that, the function $f(x, y, t)$ is also a Gaussian random process and the samples taken by the sensors are

jointly Gaussian random variables (JGRVs). Since, $\mu_S = 0$, the mean of the received signal, $\mu_E = 0^3$. The variance of the received signal is also given as follows:

$$\sigma_E^2(x, y) = E[f^2(x, y, t)] = \left(\sigma_S e^{-\sqrt{x^2+y^2}/\theta_s}\right)^2. \quad (25)$$

An interesting result from (25) is that, the variance of the signal observed at location (x, y) depends on the distance between the observation location and the point source. The received signal at time t_k by a sensor n_i at location (x_i, y_i) is given by

$$S_i[k] = f(x_i, y_i, t_k) . \quad (26)$$

Assuming wide-sense stationarity, the *spatio-temporal correlation function* for two samples of a point source taken at locations (x_i, y_i) and (x_j, y_j) , and at times t_k and t_l , respectively, is given by

$$\begin{aligned} \rho_p(i, j, k, l) &= \frac{E[S_i[k] S_j[l]]}{\sigma_E(x_i, y_i) \sigma_E(x_j, y_j)} , \\ &= \rho_S(\Delta_t) , \end{aligned} \quad (27)$$

where $\Delta_t = |t_k - t_l - (d_i - d_j)/v|$, $d_i = \sqrt{x_i^2 + y_i^2}$ is the distance of the sensor n_i to the point source, and $\rho_S(\Delta_t) = E[f_S(t) f_S(t + \Delta_t)]/\sigma_S^2$ is the correlation function of the point source which is given by $\rho_S(\Delta_t) = e^{-\Delta_t/\theta_t}$, where θ_t is a constant governing the degree of correlation. Note that the spatio-temporal correlation between two samples, $\rho_p(i, j, k, l)$, depends mainly on the difference between sample times t_k and t_l since generally $v \gg (d_i - d_j)$.

In WSN, we are interested in estimating the signal generated by the point source using the samples collected by the sensor nodes. The expectation of the generated signal, $f_S(t)$, over an interval τ is given by

$$S(\tau) = \frac{1}{\tau} \int_0^\tau f_S(t) dt . \quad (28)$$

³The subscripts S and E , which are used throughout this section, represent the *source* and *event*, respectively.

Each sensor node, n_i , receives the attenuated and delayed version of the generated signal $f_S(t)$, i.e., $S_i[k]$. Due to the impurities in the sensor circuitries, the sampled signal is the noisy version of this received signal which is given by

$$X_i[k] = S_i[k] + N_i[k] , \quad (29)$$

where the subscript i denotes the location of the node n_i , i.e., (x_i, y_i) , k denotes the sample index which corresponds to time $t = t_k$ ⁴, $X_i[k]$ is the noisy version of the actual sample $S_i[k]$, and $N_i[k]$ is the observation noise, i.e., $N_i[k] \sim \mathcal{N}(0, \sigma_N^2)$. $S_i[k]$ is given by (24) and (26).

The observed information, $X_i[k]$, is then encoded and sent to the sink through the WSN. It has been shown that joint source-channel coding outperforms separate coding. Moreover, for WSN with finite number of nodes, uncoded transmission outperforms any approach based on the separation paradigm leading to the optimal solution for infinite number of nodes [78]. Under the light of these results, we assume that uncoded transmission is deployed in each node. Hence, the transmitted observation, $Y_i[k]$, is given by

$$Y_i[k] = \sqrt{\frac{P_E}{\sigma_S^2 + \sigma_N^2}} X_i[k], \quad i = 1, \dots, N , \quad (30)$$

where σ_S^2 and σ_N^2 are the variances of the event information $S_i[k]$ and the observation noise $N_i[k]$, respectively.

The transmitted information is decoded at the sink. Since uncoded transmission is used, it is well known that minimum mean square error (MMSE) estimation is the optimum decoding technique [78]. Hence, the estimation, $Z_i[k]$, of the event information $S_i[k]$ is simply the MMSE estimation of $Y_i[k]$, which is given by

$$Z_i[k] = \frac{\sigma_E^2(x_i, y_i)}{\sigma_E^2(x_i, y_i) + \sigma_N^2} (S_i[k] + N_i[k]) . \quad (31)$$

⁴Note that we use a discrete-time model since each node is assumed to sample the physical phenomenon synchronously after the initial wake-up.

The sink is interested in estimating the expected value of the event during a decision interval τ which is given by (28). Assuming each sensor node sends information at a rate of f samples/sec, this estimation can simply be found by

$$\hat{S}(\tau, f, M) = \frac{1}{\tau f M} \sum_{i=1}^M \sum_{k=1}^{\tau f} Z_i[k], \quad (32)$$

where M is the number of sensor nodes that send samples of the observed point source. M nodes are chosen among the nodes in the network to represent the point source, and hence, are referred to as *representative nodes*. Consequently, the distortion achieved by this estimation is given by

$$D_p(\tau, f, M) = E \left[(S(\tau) - \hat{S}(\tau, f, M))^2 \right], \quad (33)$$

where the subscript p denotes the point source. Using (24), (25), (28), (31), and (32), (33) can be expressed as

$$\begin{aligned} D_p(\tau, f, M) &= \sigma_S^2 - \frac{2}{\tau^2 f M} \sum_{i=1}^M \sum_{k=1}^{\tau f} \frac{\sigma_S^4 e^{-3d_i/\theta_s}}{\sigma_S^2 e^{-2d_i/\theta_s} + \sigma_N^2} \theta_t \\ &\quad \times \left[2 - e^{-(t_k+d_i/c)} - e^{-(\tau-t_k-d_i/c)/\theta_t} \right] \\ &\quad + \frac{\sigma_N^2}{\tau f M^2} \sum_{i=1}^M \frac{\sigma_S^4 e^{-2d_i/\theta_s}}{(\sigma_S^2 e^{-d_i/\theta_s} + \sigma_N^2)^2} \\ &\quad + \frac{1}{\tau^2 f^2 M^2} \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^{\tau f} \sum_{l=1}^{\tau f} \alpha \rho(i, j, k, l), \end{aligned} \quad (34)$$

where

$$\alpha = \frac{\sigma_S^8 e^{-2(d_i+d_j)/\theta_s}}{(\sigma_S^2 e^{-d_i/\theta_s} + \sigma_N^2)(\sigma_S^2 e^{-d_j/\theta_s} + \sigma_N^2)},$$

$d_i = \sqrt{(x_i + y_i)}$, and $\rho(i, j, k, l)$ is the spatio-temporal correlation function given in (27).

2.3.4.2 Field Source

There exist applications such as temperature monitoring and seismic monitoring, where the physical phenomenon is dispersed over the sensor field, and, hence, can be

modeled as a field source. Thus, here, we explore the spatio-temporal characteristics of observing such a phenomenon in WSN.

The event signal $f(x, y, t)$ is assumed to be a Gaussian random process with $\mathcal{N}(0, \sigma_s^2)$. The sink is interested in estimating the signal $f(x_0, y_0, t)$ over the decision interval τ at location (x_0, y_0) . Assuming the observed signal $f(x, y, t)$ is wide-sense stationary (WSS), the expectation of the signal over the decision interval τ , i.e., $S(\tau)$ can be calculated by the time average of the observed signal:

$$S(\tau) = \frac{1}{\tau} \int_0^\tau f(x_0, y_0, t) dt , \quad (36)$$

where (x_0, y_0) is the event location. The signal, $S_i[k]$ received at time t_k by a sensor node at location (x_i, y_i) is defined as in (26) and $S_i[k]$'s are JGRV with $\mathcal{N}(0, \sigma_s^2)$. The covariance of two samples, $S_i[k]$ and $S_j[l]$, is given by:

$$\text{cov}\{S_i[k], S_j[l]\} = \sigma_s^2 \rho_s(i, j) \rho_t(\delta) ,$$

where

$$\rho_s(i, j) = e^{-d_{i,j}/\theta_s} , \quad \text{and} \quad \rho_t(\delta) = e^{-|\delta|/\theta_t} \quad (37)$$

are spatial and temporal correlation functions, respectively, $\delta = (k - l)/f$, f is the sampling rate, $d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ is the distance between two nodes n_i and n_j , and θ_s and θ_t are spatial and temporal correlation coefficients, respectively.

Following the discussion and derivations in Section 2.3.4.1, the noisy version of the signal, $X_i[k]$, and the transmitted signal, $Y_i[k]$ are given by (29) and (30), respectively. The estimation $Z_i[k]$ can be found as follows:

$$Z_i[k] = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_N^2} \left(S_i[k] + N_i[k] \right) . \quad (38)$$

After collecting the samples of the signal in the decision interval τ from M nodes, the sink estimates the expectation of the signal over the last decision interval as given in (32). As a result, the distortion achieved by this estimation is given as in

(33). Using the definitions above and substituting (36), (38), and (32) into (33); the distortion function can be derived as

$$\begin{aligned}
D_f(\tau, f, M) = & \sigma_S^2 - \frac{2\sigma_S^2}{\tau^2 f M (\sigma_S^2 + \sigma_N^2)} \sum_{i=1}^M \rho_s(i, s) \sum_{k=1}^{\tau f} \theta_t \left[2 - e^{-k/(f\theta_t)} - e^{-\left(t-\frac{k}{f}\right)/\theta_t} \right] \\
& + \frac{\sigma_S^4 \sigma_N^2}{\tau f M (\sigma_S^2 + \sigma_N^2)^2} \\
& + \frac{\sigma_S^6}{(\tau f M (\sigma_S^2 + \sigma_N^2))^2} \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^{\tau f} \sum_{l=1}^{\tau f} \rho_s(i, j) \rho_t(|k-l|/f). \quad (39)
\end{aligned}$$

2.4 Analysis and Results

In order to gain more insight to our intuitions, we performed case studies using the distortion functions derived in Section 2.3 for spatial, temporal, and spatio-temporal for point and field sources, respectively. The results of this analysis is provided in the following.

2.4.1 Spatial Correlation

In Section 2.3.2, the theoretical framework of spatial correlation in WSN is developed and the distortion resulting from sending a subset M of total N nodes in the event area is derived (12). The relations between the positions of the sensor nodes in the event area and the event estimation reliability is also important for exploiting spatial correlation.

In order to gain more insight to our intuitions, we performed a case study using the distortion function (12). In a 500 by 500 grid, we deployed 50 sensor nodes randomly. We use the *Power Exponential* model with $\theta_2 = 1$ and $\theta_1 = \{10, 50, 100, 500, 1000, 5000, 1000\}$ as the covariance model for the covariance function, $K_\theta(\cdot)$ in (1). The parameter, θ_1 , controls the relation between the distance of the nodes and the correlation coefficient. For each value of θ_1 , we calculate the distortion function (12) by varying the number of sensor nodes sending information. Starting from 50 nodes, we decrease the number of nodes that send event information to the sink. We

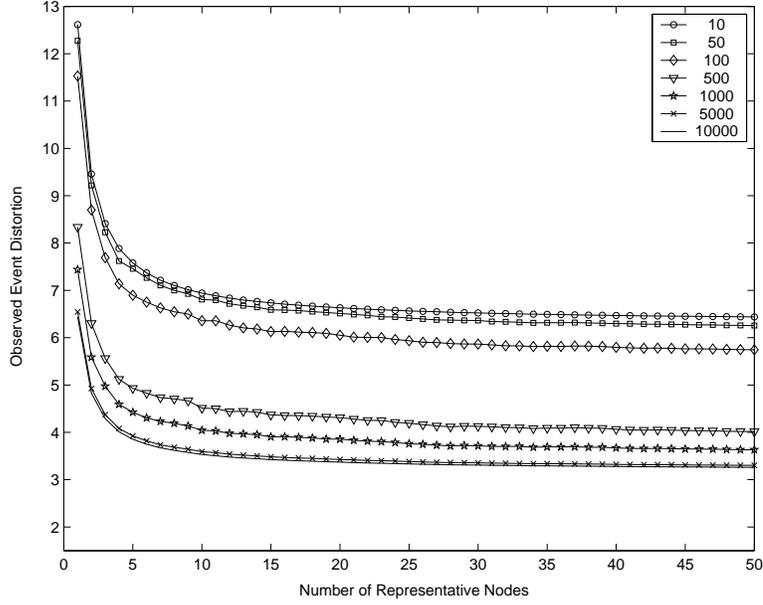


Figure 2: Observed Event Distortion for different θ_1 values according to changing number of representative nodes

refer to these nodes as the *representative nodes*.

The simulations are performed for a fixed topology with 1000 trials for each number of representative nodes. Representative nodes are selected randomly among the 50 nodes for each trial and the distortion function is calculated according to the locations of these nodes. The average distortion calculated from these simulations and the distribution of the distortion for each number of representative nodes is shown in Fig. 2.

As shown in Fig. 2, the achieved distortion stays relatively constant when the number of representative nodes is decreased from 50 to 15. This behavior is due to the highly redundant data sent by the sensor nodes that are close to each other. In addition, with increasing θ_1 , the observed event distortion decreases since close nodes become less correlated with increasing θ_1 .

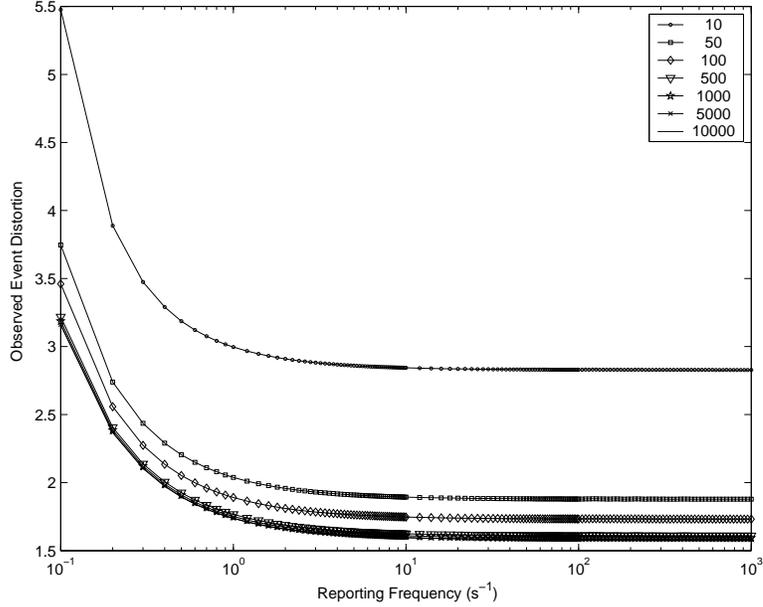


Figure 3: Observed event distortion for varying normalized reporting frequency.

2.4.2 Temporal Correlation

As derived in Section 2.3.3, the distortion $D(f)$ observed in the estimation of the signal S being tracked depends on the reporting frequency f used by the sensor nodes sending their readings to the sink in the decision interval τ . A case study with the same network configuration and parameters in Section 2.5.1 is also performed to observe the variation of the observed event distortion at the sink for varying reporting frequency f , i.e., distortion function $D(f)$ in (23). It is observed from (23) and Fig. 3 that the observed event distortion at the sink decreases with increasing f . This is because the number of samples received in a decision interval i increases with increasing f conveying more information to the sink from the event area. Note that however, above a certain reporting frequency, f , the observed event distortion cannot be further reduced. Therefore, a significant energy saving can be achieved by selecting small enough f that achieves desired event distortion D^* and does not lead to an overutilization of the scarce sensor resources.

On the other hand, any f chosen arbitrarily small to achieve a certain distortion

bound D^* using (23) may not necessarily achieve the desired distortion level and hence assure the event transport reliability. This is mainly because all of the sensor samples generated with this chosen reporting frequency may not be received because of packet losses in the sensor network due to link errors and network disconnectivity. Similarly, as very high values of f do not bring any additional gain in terms of observed event distortion as shown in Fig. 3; on the contrary, it may endanger the event transport reliability by leading to congestion in the sensor network. Let f_{max} be the maximum reporting frequency which the network capacity can accommodate. Thus, $f > f_{max}$ leads to congestion and hence packet losses resulting in an increase in the observed event distortion.

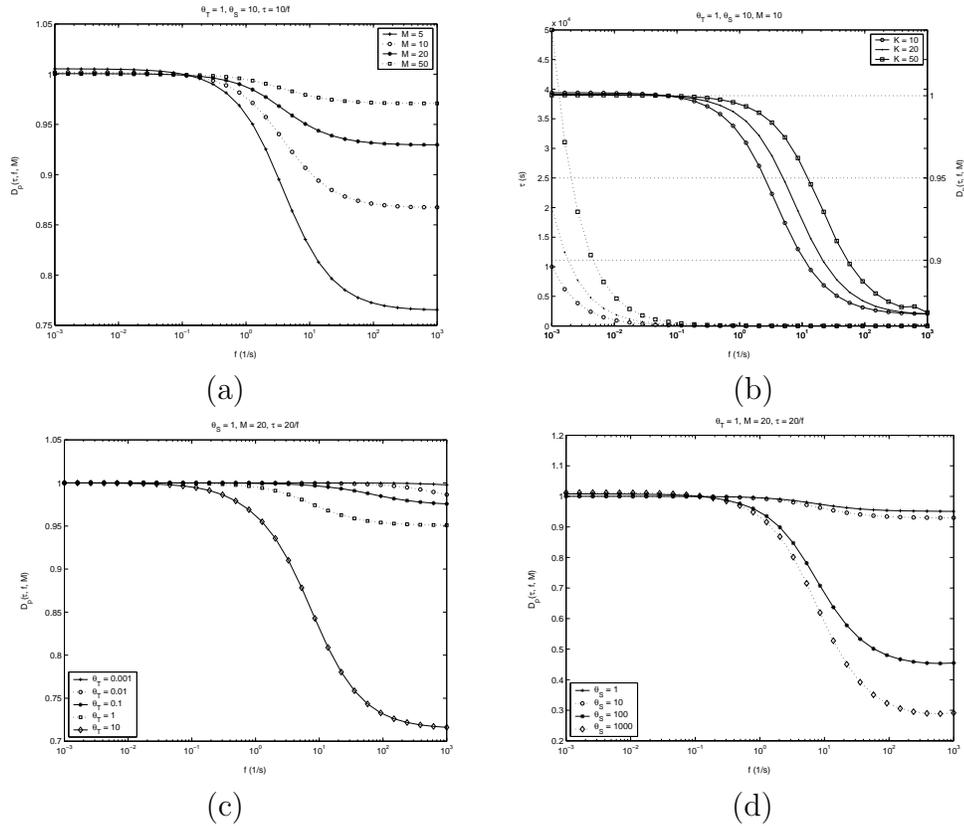


Figure 4: Distortion vs. sampling rate for different values of (a) M , (b) K , (c) θ_T , and (d) θ_S (Point source).

2.4.3 Spatio-temporal Correlation with Point Sources

In this section, we provide numerical simulation results for spatio-temporal correlation characteristics of point sources using the distortion functions given by (34). A sensor network of a grid topology of $50\text{m} \times 50\text{m}$ with 120 nodes is used for the evaluations. For the point source evaluations, the location of the point source is the center of the grid.

For each evaluation, the closest M nodes to the center are chosen to send information. This selection method, which we refer to as *ordered selection*, enables us to observe the spatial correlation effects as the M value is increased, since higher value of M corresponds to nodes farther from the center being chosen. Another important parameter is the number of packets sent by a single sensor node during the decision interval τ , which is denoted as $K = \tau f$, where f is the sampling rate of the sensor nodes. In the evaluations, the parameter K is fixed and τ is determined as the f changes, which enables us to investigate the effect of a large range of sampling rate values.

In this section, the behavior of the distortion function in (34) is investigated for various values for sampling rate, f , number of representative nodes, M , number of samples, K , and spatial and temporal correlation coefficients, θ_S and θ_T , respectively. The influence of sampling rate, f , on distortion is shown in Fig. 4(a). It is clearly seen that, as the sampling rate increases, distortion decreases, which show the effect of temporal resolution on event estimation. The rate of decrease is significantly large for a specific range of f values, e.g., $0.1 < f < 100$ for $M = 5$. Moreover, above this range of f values, the distortion remains relatively constant. This observation reveals that there is an optimal value, f_{opt} , for temporal resolution such that further increase in sampling rate, f , does not influence the distortion.

The effect of number of representative nodes, M , on distortion is also shown in Fig. 4(a). It is clear that, increasing M degrades the distortion function for high values

of sampling rate, f . As the sampling rate increases, consecutive samples become sufficient to extract the characteristics of the source. However, as M is increased, distortion increases due to decrease in spatial correlation. On the other hand, for lower sampling rates, e.g., $f < 0.1s^{-1}$, an increase in M improves the distortion since the temporal resolution is not sufficient in this case. As a result of increased M , the spatial correlation helps build a more accurate estimation of the signal. However, increasing M above a specific value, e.g., $M = 10$, has no impact on distortion. This result reveals that, there is an optimal value, M_{opt} , for efficient estimation, which we will investigate in detail later.

In Fig. 4(b), the effect of number of samples, K , is shown. The solid lines represent the distortion values for each K , and the dotted lines show the corresponding decision interval values, i.e., τ . Note that, when the sampling rate is low, τ increases significantly so that required number of samples can be collected (note the logarithmic scale on x -axis). However, this increase in τ results in temporally uncorrelated samples to be collected, which leads to higher distortion. However, for higher values of f , τ decreases and the temporal resolution becomes sufficient. Moreover, for the transition region, where the distortion function decreases significantly, the number of samples, K , has an important influence on distortion. In this region, lower K also improves the estimation since more closely sampled instances are taken into account. However, when the sampling rate, f , is further increased, the temporal resolution becomes so fine that, any number of samples does not effect the distortion, where the lines for different K values intersect.

In Fig. 4(c) and 4(d), the influence of the temporal and spatial correlation coefficients, θ_T and θ_S , on distortion are shown, respectively. A higher value of either of these parameters corresponds to a higher correlation in either temporal or spatial domain. As shown in Fig. 4(c), as θ_T is increased, i.e., the temporal correlation between nodes are increased, an increase in sampling rate results in a much more

decrease in the distortion. This is also observed for θ_S as shown in Fig. 4(d). An interesting result is that although the number of nodes are fixed in Fig. 4(d), higher spatial correlation improves the effect of temporal resolution since each selected node is highly correlated with the point source.

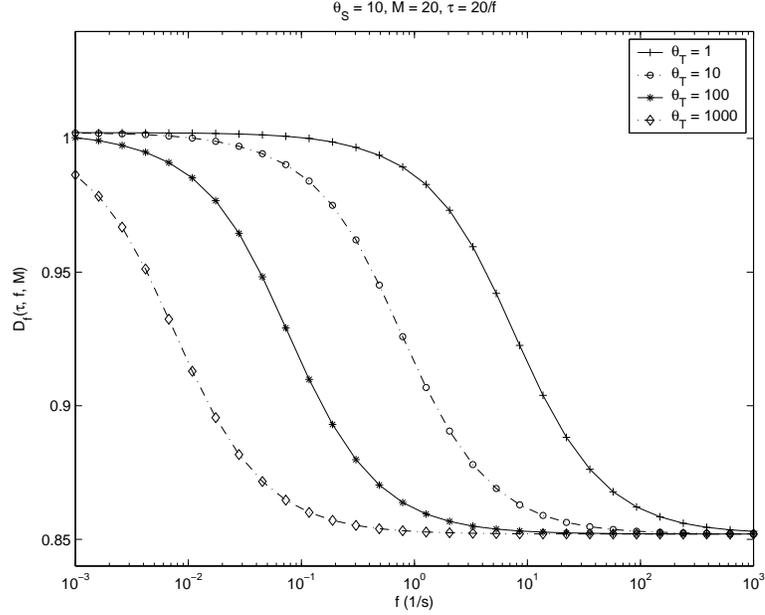


Figure 5: Distortion vs. sampling rate for different values of θ_T (Field source).

2.4.4 Spatio-temporal Correlation with Field Source

In this section, the behavior of the distortion function in (39) for field sources is investigated for the same set of parameters used in Section 2.4.3. For the field source evaluations, the network aims to estimate the signal value at the center of the grid using the samples of the sensors located on the grid. The effect of the number of representative nodes, M , and the number of samples, K are found to be similar to the point source case and hence, are not reproduced here. Similar observations can be made as in Section 2.4.3.

In Fig. 5, the effect of temporal correlation coefficient on the distortion is shown. Contrary to our results for point source, the decrease in distortion does not depend on

θ_T . However, the distortion plot shifts to the left when θ_T is increased. This affects the optimal sampling rate value, f_{opt} , i.e., higher θ_T results in smaller f_{opt} value. On the other hand, the relation between spatial correlation coefficient θ_S with distortion is similar to the case in point source shown in Fig. 4 (d), and hence, is not reproduced here. Similarly, the decrease in distortion depends on θ_S , which shows that spatial correlation is more important in decreasing the distortion for field sources.

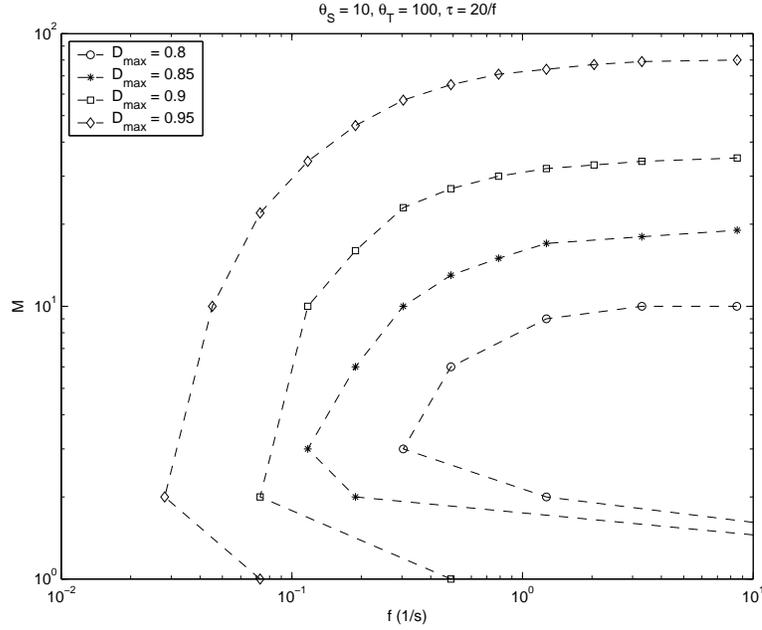


Figure 6: Number of nodes vs. sampling rate, (M, f) tuples meeting various D_{max} constraints (Field source).

In Fig. 6, the tradeoff between spatial and temporal resolution is shown. Each point represents the boundary of the feasible region for (M, f) values that meet a certain distortion constraint, D_{max} . The figure can be read as follows: For each allowed distortion D_{max} , the tuples represent the boundary of the feasible region inside which the distortion constraint is guaranteed. An important result is that, for each D_{max} value, there is an optimum operating point, where minimum number of nodes can be used with low sampling rate. Increasing M above this value also requires increase in temporal resolution. Moreover, a decrease in maximum allowed distortion value,

D_{max} , results in a smaller feasible region, as expected. This also results in a smaller range for feasible values of M . Consequently, aggressively collecting information from each sensor node in the field, does not necessarily correspond to more accurate estimation. This figure serves as an important guideline to the design of communication protocols, network topology, and deployment for a particular distortion requirement.

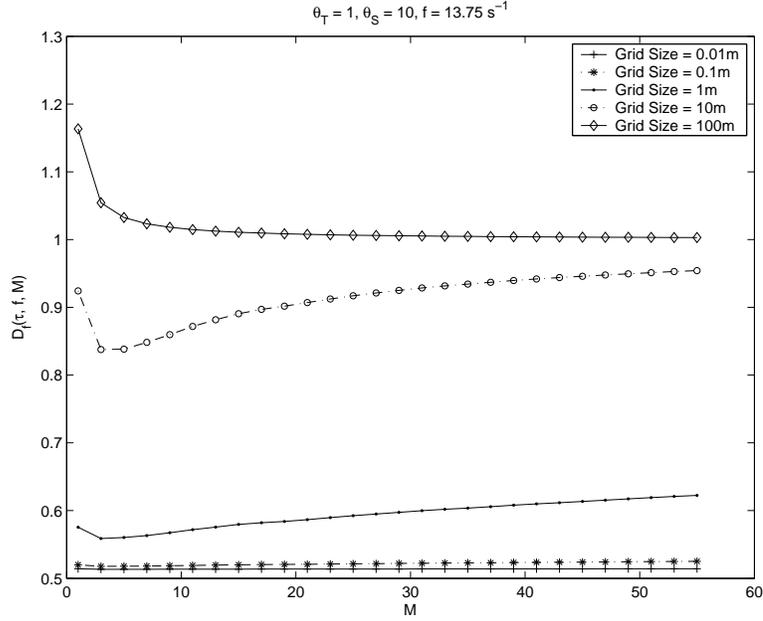


Figure 7: Distortion vs. M for different grid sizes, $f = 13.74$ (Field source).

2.4.5 External Parameters

In this section, we investigate the effects of the network topology properties, such as grid size and the effects of using a random topology for field sources. The results also apply to point sources.

We present the effects of network topology on event distortion in Fig. 7, where distortion is shown versus M for different grid sizes at sampling rate $f = 13.74 \text{ s}^{-1}$. *Grid size* is defined as the minimum distance between two neighbor nodes in the grid topology. Smaller grid sizes correspond to higher granularity at the cost of

higher density and larger number of sensors to cover a certain area, which affect the deployment cost and energy consumption of the WSN. It is shown in Fig. 7 that there is an optimal point for M for small grid sizes (< 10 m). Moreover, increasing grid size also increases the distortion. As explained before, when M is increased, nodes far from the field center are chosen. Hence, as the grid size increases, these nodes become spatially uncorrelated. In this case, increasing M helps decrease distortion. On the other hand, when the grid size is small, an increase in M does not affect the distortion. This shows that a suitable internode distance needs to be chosen for efficient coverage of the physical phenomenon as well as an optimum M value.

So far, a grid topology has been considered for evaluations. However, a grid topology may not be practical for some WSN and the effect of using a random topology needs to be considered. Moreover, in our evaluations, we use a specific method for representative node selection, such that the closest M nodes to the source are chosen. This selection assumes that the location of the event source is known. However, in some applications, the location of the source may not be available to the sink. This is important especially in tracking applications.

The effect of randomness in the network topology and the node selection method is investigated using three different scenarios. The first scenario corresponds to a random topology of 120 nodes in a $50\text{m} \times 50\text{m}$ topology, where M representative nodes are randomly selected. For Scenario 2, again random topology is used with ordered selection method. Scenario 3 corresponds to grid topology with ordered selection. The achieved distortion from these scenarios are shown in Fig. 8. For scenarios 1 and 2, the average of 1000 random topologies are shown. The results show that when ordered selection is used, the randomness of the network topology improves the achieved distortion. It can also be observed that, when the nodes are randomly chosen, i.e., Scenario 1, the achieved distortion is significantly high because the representative nodes can be located anywhere in the network. However, a higher

value of M , in this case, helps reduce the distortion. Hence, in practical applications where the source location is not known, at first, higher number of representative nodes can be used to locate the signal source, and once the location of the signal source is found, the number of nodes can be decreased to the optimum M value to improve the accuracy of estimation.

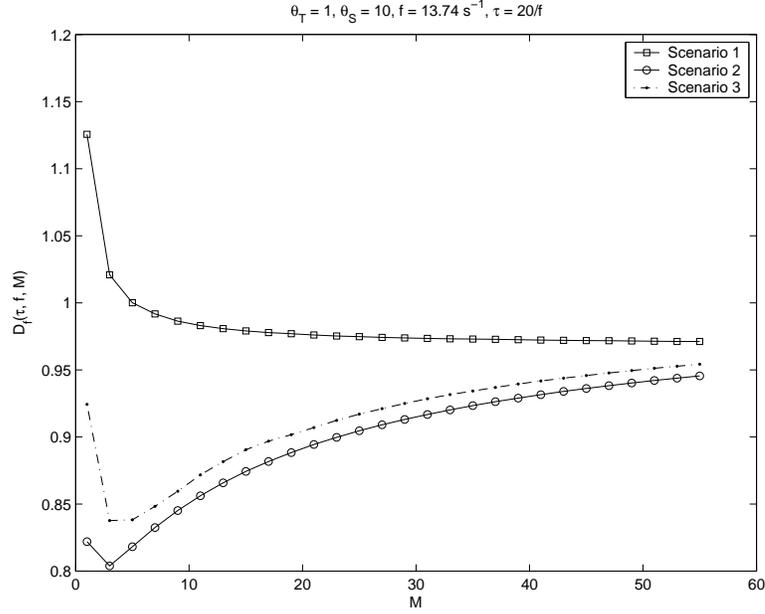


Figure 8: Distortion vs. M for different network topology and node selection procedures (Field source).

2.5 Exploiting Correlation in WSN

Spatial and temporal correlation, in addition to the collaborative nature of the WSN, bring significant potential advantages for the development of efficient communication protocols well-suited for the WSN paradigm. In this section, we discuss possible approaches exploiting spatial and temporal correlation to achieve energy-efficient medium access and reliable event transport in WSN, respectively.

2.5.1 Correlation-based Medium Access Control

The shared wireless channel between sensor nodes and energy considerations of the WSN make the Medium Access Control (MAC) a crucial part of the wireless sensor phenomenon. The distributed architecture and the application-oriented traffic properties of the WSN pose interesting challenges for the development of MAC protocols. Furthermore, the scarce energy sources of sensor nodes necessitate energy aware MAC protocols. Hence, MAC protocols for WSN should be developed tailored to the physical properties of the sensed phenomenon and the specific network properties so that the access to the channel is coordinated with minimum collisions without effecting the connectivity throughout the network.

In WSN, many individual nodes deployed in large areas sense events and send corresponding information about these events to the sink. When an event occurs in the sensor field, all the nodes in an event area collect information about the event taking place and try to send this information to the sink. Due to the physical properties of the event, this information may be highly correlated in nature according to the spatial correlation between sensor nodes. Intuitively, data from spatially separated sensors is more useful to the sink than highly correlated data from closely located sensors. Hence, it may not be necessary for every sensor node to transmit its data to the sink; instead, a smaller number of sensor measurements might be adequate to communicate the event features to the sink within a certain distortion constraint. As a result, the MAC protocol can *reduce the energy consumption of the network by exploiting spatial correlation in the WSN without compromising on the access latency as well as the distortion achieved.*

Consequently, due to the spatial correlation between sensor observations, significant energy saving can be achieved by choosing representative nodes among the nodes in the event area without degrading the achieved distortion at the sink. It is clear that reduced number of nodes transmitting information decreases contention in the

wireless medium resulting in decreased energy consumption. In Chapter 3, the lessons learned from this analysis will be exploited to develop a MAC protocol that exploits the spatial correlation between closely located sensor nodes that regulates medium access and prevents redundant transmissions from closely located sensors.

2.5.2 Correlation-based Reliable Event Transport

In order to realize the potential gains of the WSN, it is imperative that desired event features are reliably communicated to the sink. To accomplish this, a reliable transport mechanism is required in addition to an efficient medium access scheme as discussed in Section 2.5.1. The main objective of the transport layer mechanism in WSN is to achieve reliable collective transport of event features from the sensors within the coverage of the phenomenon, i.e., event area, to the sink. In order to provide reliable event detection at the sink, possible congestion in the forward path should also be addressed by the transport layer. Once the event is sensed by a number of sensor nodes within the event area, significant amount of traffic is triggered by these sensor nodes, which may easily lead to congestion in the forward path. Furthermore, the error and congestion control objectives must be achieved with minimum possible energy expenditure. Energy efficiency must be also considered in transport mechanism design by shifting the burden to the high-powered sink in the WSN in order to conserve limited sensor resources.

Unlike traditional communication networks, the sensor network paradigm necessitates that the event features are estimated within a certain distortion bound, i.e., required reliability level, at the sink as discussed in Section 2.3. Reliable event detection at the sink is based on collective information provided by source nodes and not on any individual report. Hence, conventional end-to-end reliability definitions and solutions are inapplicable in the WSN regime and would only lead to over-utilization

of scarce sensor resources. On the other hand, the absence of reliable transport altogether can seriously impair event detection which is the main objective of WSN deployment. Hence, the WSN paradigm necessitates a collective *event-to-sink* reliability notion rather than the traditional end-to-end notion [4]. The main rationale behind such *event-to-sink reliability* notion is that the data generated by the sensors are temporally correlated which tolerates individual packets to be lost to the extent where the desired event distortion D^* is not exceeded. Consequently, the event-to-sink reliable transport (ESRT) protocol has been developed that exploits the results obtained from this research [4].

CHAPTER III

SPATIAL CORRELATION-BASED COLLABORATIVE MEDIUM ACCESS CONTROL IN WIRELESS SENSOR NETWORKS

In this chapter, the *spatial correlation-based collaborative medium access control (CC-MAC)* protocol is presented. CC-MAC exploits spatial correlation, which is inherent in Wireless Sensor Networks (WSN), at the MAC layer. CC-MAC was first presented in [80]. The motivation for this work is provided in Section 3.1. The recent work on spatial correlation in WSN in addition to existing MAC protocols are summarized in Section 3.2. Based on the theoretical framework that was developed in Chapter 2, the corollaries are discussed in Section 3.3 and a node selection algorithm is developed for correlation filtering in Section 3.4. Details of the distributed CC-MAC protocol that regulates medium access to reduce the number of nodes transmitting information are presented in Section 3.5. CC-MAC performance analysis and simulation results are presented in Section 3.6.

3.1 Motivation

Wireless sensor networks (WSN) are event-based systems that rely on the collective effort of several microsensor nodes observing a physical phenomenon. Typical WSN applications require spatially dense sensor deployment to achieve satisfactory coverage [51]. As a result, several sensor nodes record information about a single event in a sensor field. Because of the high density in the network topology, the sensor records may be spatially correlated subject to an event. The degree of spatial correlation increases with decreasing internode separation.

Exploiting spatial correlation in the context of collaborative nature of the WSN can lead to significant performance improvement of communication protocols. Possible approaches to utilize spatial and temporal correlation in WSN are already investigated in [77] and [78]. Intuitively, data from spatially separated sensor nodes is more useful for the sink than the highly correlated data from closely located sensor nodes. Hence, it may not be necessary for every sensor node to transmit its data to the sink; instead, a smaller number of sensor data might be adequate to transmit certain event features to the sink within a certain distortion constraint.

In this chapter, we show how this spatial correlation can be exploited at the medium access control (MAC) layer. The shared wireless channel between sensor nodes and the energy considerations make the MAC layer a crucial part in WSN. The existing MAC protocols cannot be applied here because of the event-based traffic properties and the energy constraints in WSN. Also, the channel access must be coordinated in such a way that possible collisions are minimized or eliminated. These requirements can be satisfied by intelligent management of transmission attempts among nodes by exploiting the spatially correlated nature of the event information. Based on the theoretical framework for spatial correlation in WSN developed in Chapter 2, we develop a distributed, spatial correlation-based collaborative MAC (CC-MAC) protocol that regulates medium access and prevents redundant transmissions from closely located sensors.

3.2 Related Work

There exists some research to study the correlation in WSN [31, 56, 64] in recent years. In these work, the information theoretical aspects of the correlation are explored. In other words, these studies aim to find the optimum rate to compress redundant information in the sensor observations. More recently, the relation between distortion, spatio-temporal bandwidth, and power for large sensor networks is investigated [32].

However, correlation (spatial or temporal) between sensor observations is not considered in this work. Moreover, none of the above solutions develop communication network protocols.

In [84], spatial and temporal correlations are exploited to eliminate the acknowledgements in the communication. While the number of acknowledgements is considerably reduced, the number of redundant packets is still large in the network. The joint routing and source coding is introduced in [64] to reduce the amount of traffic generated in dense sensor networks with spatially correlated records. While joint routing and source coding reduces the number of transmitted bits, the number of transmitted packets remains unchanged from the network point of view. In our opinion, the number of transmitted packets can be further minimized by regulating the network access based on the spatial correlation between the sensor nodes. Moreover, the relation between spatial and temporal sampling rate on the overall network delay and energy consumption is studied in [11]. However, the spatial and temporal correlations between sensor observations are not investigated.

Current studies on medium access control in WSN focus mainly on the energy-latency tradeoffs. S-MAC [83] aims to decrease the energy consumption by using sleep schedules with virtual clustering. T-MAC [74], a variant of S-MAC, incorporates variable sleep schedules to further decrease the energy consumption. However, in both protocols, sensor nodes keep sending redundant data with increased latency because of periodic sleep durations. In [10], an energy-aware TDMA-based MAC protocol is presented where the sensor network is assumed to be composed of clusters and gateways. Each gateway acts as a cluster-based centralized network manager and assigns slots in a TDMA frame. The protocol assumes a cluster-based topology that results in significant additional processing complexity and overhead in the overall sensor network. An energy efficient collision-free MAC protocol that is based on a time-slotted structure is presented in [59]. Each node determines its own time slot

using a distributed election scheme based on traffic requirements of its every two-hop neighbor. Although the protocol achieves high delivery ratio with tolerable delay, the performance of the protocol depends on the two-hop neighborhood information in each node. Since this information is collected through signaling, in the case of high density sensor networks, the signaling cost increases significantly resulting in either incomplete neighbor information because of collisions or high energy consumption.

So far, the existing MAC solutions focus on decreasing energy consumption by modifying known medium access techniques. In our opinion, event-based MAC protocols that exploits topology and traffic properties of WSN need to be developed. In Chapter 2 and [78], we introduced the characteristics of spatial and temporal correlations in WSN. Based on this framework, we develop the spatial correlation-based collaborative MAC (CC-MAC) protocol that aims to *reduce the energy consumption of the network by exploiting spatial correlation in WSN without compromising the channel access latency and the achieved distortion.*

3.3 Corralaries from Spatial Correlation Theory

The theoretical framework developed in Chapter 2 reveals that significant energy savings can be achieved by exploiting the spatial correlation between sensor observations. More specifically, the event distortion, $D_E(M)$, that is achieved at the sink as a function of the number of sensor nodes M was found to be

$$D_E(M) = \sigma_S^2 - \frac{\sigma_S^4}{M(\sigma_S^2 + \sigma_N^2)} \left(2 \sum_{i=1}^M \rho_{(s,i)} - 1 \right) + \frac{\sigma_S^6}{M^2(\sigma_S^2 + \sigma_N^2)^2} \sum_{i=1}^M \sum_{j \neq i}^M \rho_{(i,j)}, \quad (40)$$

where σ_S^2 and σ_N^2 are the variance of the received signal and the observation noise, respectively, $\rho_{(i,j)}$ and $\rho_{(s,i)}$ are the correlation coefficients between nodes n_i and n_j , and the event S location and the sensor node n_i , respectively. It follows from our previous discussions in Chapter 2 that the following can be concluded about the achieved distortion at the sink:

Remark 1: The minimum distortion is achieved when all the sensor nodes in the event area send information to the sink. However, the achieved distortion at the sink can be preserved even though the number of the representative nodes is decreased. As a result, significant energy saving is possible by allowing less number of sensor nodes to send information.

Remark 2: Based on (40), there are two factors affecting the distortion other than the number of representative nodes.

1. The correlation coefficient, $\rho_{(s,i)}$, between a sensor node n_i sending the information and the event source S affects the distortion function negatively. The distortion increases as the distance between the event source S and the node n_i increases. Intuitively, if a representative sensor node is chosen apart from the source, it observes inaccurate data resulting in higher distortion at the sink.
2. The correlation coefficient, $\rho_{(i,j)}$, between each representative node n_i and n_j affects the distortion positively. As the distance between sensor nodes increases, the distortion decreases. Since further apart sensor nodes observe less correlated data, the distortion is decreased if these nodes are chosen as the representative nodes.

Consequently, because of the spatial correlation between sensor observations, significant energy savings can be achieved by choosing representative nodes among the sensor nodes in the event area without degrading the achieved distortion at the sink. It is clear that the reduced number of sensor nodes transmitting information decreases contention in the wireless medium resulting in decreased energy consumption.

As a result, our aim is to find the minimum number of representative nodes that achieve the distortion constraint given by the sensor application. This minimum

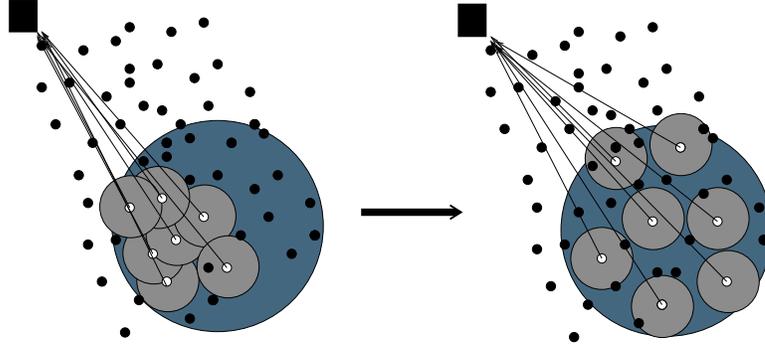


Figure 9: Spatial Re-usage in Sensor Networks.

number is given as

$$M^* = \underset{M}{\operatorname{argmin}} \{D_E(M) < D_{max}\}$$

where D_{max} is the maximum distortion allowed by the sensor application.

It is important to note that the minimum number of representative nodes, M^* , depends on the locations of the representative nodes. It follows from our previous discussions that for a fixed number of representative nodes, the minimum distortion can be achieved by choosing these nodes such that (i) they are located as close to the event source S as possible and (ii) they are located as far apart from each other as possible. As an example, as illustrated in Fig. 9, choosing representative nodes such that they are spread over the event area results in a decrease in distortion because of less redundant data sent by these nodes. Note that such a formation also improves the medium access success during the transmission of the information. Since the representative nodes are not located close to each other, the probability of collision in the wireless medium decreases. As a result, *exploiting spatial correlation not only improves the distortion but also utilizes the wireless channel* because of the spatial reuse property of the wireless medium.

To apply our observations about the distortion function, $D_E(M)$, a node selection technique is required to select the representative nodes resulting in minimum

energy expenditure. For this purpose, we introduce the *iterative node selection (INS)* algorithm in Section 3.4. The INS algorithm is a control agent running at the sink that determines the minimum number of representative nodes based on the distortion constraint, D_{max} . Accordingly, the average distance between the representative nodes is determined and each sensor node in the WSN is informed about this average distance value. Each node then performs the spatial correlation-based collaborative MAC (CC-MAC) protocol, explained in Section 3.5, that exploits spatial correlation distributively.

3.4 The Iterative Node Selection (INS) Algorithm

According to the observations in Section 3.3, we introduce the *iterative node selection (INS)* algorithm to find the number of the representative nodes in WSN. INS tries to find the ideal locations of representative sensor nodes such that the required distortion can be maintained at the sink. Based on the INS algorithm results, the CC-MAC protocol is performed distributively by each sensor node to achieve the required performance. The INS algorithm requires the statistical properties of the node distribution as input and provides a correlation radius value for distributed operation as output. As pointed out in Section 3.3, the locations of the representative sensor nodes should be chosen such that the redundancy between event information is decreased. The selection of locations of correlated points based on a distortion constraint has been investigated with the well-established Vector Quantization (VQ) methods in coding theory [49]. Hence, these methods can be exploited by suitable mapping to our problem. We first give an overview about the VQ design problem and then introduce our INS algorithm.

The VQ design can be stated as follows: Given a vector source with its statistical properties known, given a distortion constraint, and given the number of codevectors, the VQ algorithm tries to find a codebook and a partition that result in the smallest

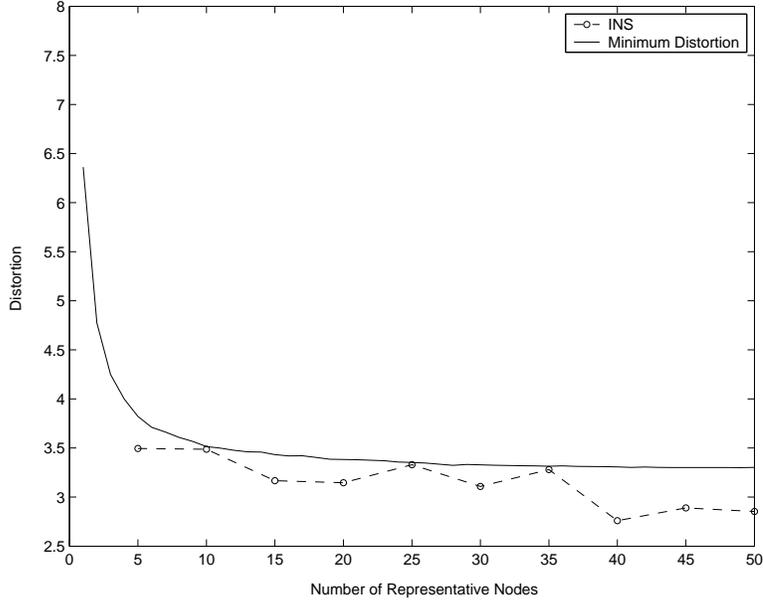


Figure 10: Comparison between minimum distortion among 1000 random trials and the distortion found by the INS solution.

average distortion. More specifically, the VQ algorithm aims to represent all possible codewords in a code space by a subset of codewords, i.e., the codebook, within the distortion constraint. Hence, the VQ solutions fit well with our problem, where we want to represent all the sensor nodes in an event area with smaller number of representative nodes. If we choose two dimensional codevectors, the code-space in the VQ approach can be mapped to the network topology with the node places as the codeword spaces. Once the codebook is determined, the VQ algorithm uses Voronoi regions to determine the partition of a code such that any information in this partition is represented by the codevector. The Voronoi regions determine the areas closest to the points representing the area. By applying the VQ algorithms, e.g., [49], to our representative node selection problem, the codebook and the partitions can be found. The codebook represents the locations of the representative nodes, while the partitions represent the areas of which the representative nodes are responsible for.

Since the VQ algorithms require only the statistical properties of the code space, for the selection of representative nodes, only the statistical properties of the topology

are required at the sink. These properties constitute the density of the network and the type of node distribution, e.g., uniform, Gaussian, or Poisson. Therefore, the INS algorithm does not require the exact locations of the nodes to be collected at the sink. It is assumed that the statistical properties the WSN topology is known by the INS algorithm¹. Based on the statistical properties, the INS algorithm first forms a sample topology. Then, as shown in Fig. 11, the algorithm starts with selecting all the nodes in the event region as representative nodes. Then, the number of representative nodes, M , is iteratively decreased. For each value of M , the positions of the nodes are found such that the distortion, $D_E(M)$, is minimized.

The INS algorithm decreases the number of representative nodes until the distortion constraint, D_{max} , is met. The VQ solution is used to find the positions of the representative nodes for each iteration using the sample topology created from the statistical properties of the network. The distortion, $D_E(M)$, is then calculated using (40). Once the maximum allowed distortion is met, the algorithm terminates.

Using a VQ toolbox [3], we simulated the INS algorithm using the same topology used in Chapter 2. The locations of representative nodes are determined for 5 to 50 representative nodes. The distortion found from these selections is shown in Fig. 10.

¹Note that the density and node distribution can be evident from the initial deployment of the WSN.

```

INS()
   $M = N$ ;
  Calculate  $D_E(M)$ 
  while ( $D_E(M) \leq D_{max}$ )
     $M = M - k$ 
    RunVQ()
    Calculate  $\rho_{(s,i)}$  and  $\rho_{(i,j)} \quad \forall i, j$ 
    Calculate  $D_E(M)$ 
  end
end;

```

Figure 11: Iterative Node Selection Algorithm. RunVQ() performs VQ algorithm and finds the places of representative nodes.

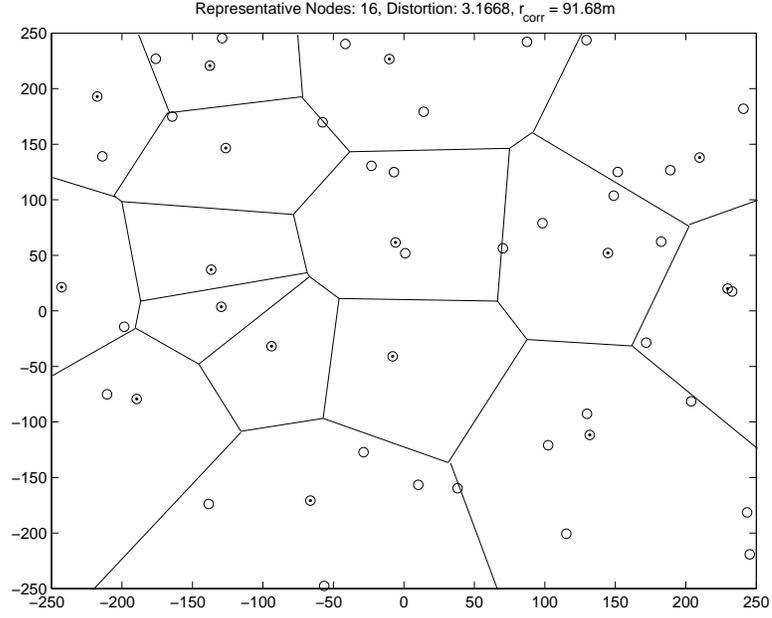


Figure 12: 16 Representative nodes chosen by VQ algorithm and the Voronoi regions representing the correlation regions.

Moreover, we show the minimum distortion found among the 1000 random trials in Chapter 2 for $\theta_1 = 5000$. Note that by choosing the representative node locations based on the VQ algorithm rather than the random selection, significant improvement in the achieved event distortion is possible as shown in the Fig. 10. In Fig. 12, we show the locations of the selected nodes for 16 representative nodes. Each representative node is shown by a circle and a dot. The representative nodes are the only nodes that transmit their event information to the sink, while the other nodes are suppressed from sending information. The Voronoi regions determine the areas where only one node is allowed to transmit its event information to the sink.

As explained before, the INS algorithm creates a sample topology for the sensor network to exploit spatial correlation and filter correlation between the nodes. Fig. 12 shows that if a node transmits data, then the nodes in its proximity are not required to send data. We call this area, specified by the INS algorithm, the *correlation region* of the representative node. Based on these observations, we introduce the following definitions:

- *Correlation Radius (r_{corr}):* The radius of the correlation region is called the *correlation radius*, r_{corr} . The INS algorithm determines this value from the average radius of Voronoi regions. Nodes with internode distance, $d_{(i,j)}$, smaller than the correlation radius, r_{corr} , are assumed to contain highly correlated data. This distance is assumed to be known a-priori from the exchange of local information during network initialization or to be estimated from the received control signal strength as discussed in [63].
- *Correlation Neighbor:* A node n_j is said to be the *correlation neighbor* of node n_i if its distance, $d_{(i,j)}$, to the node n_i is smaller than the correlation radius, r_{corr} . The correlation neighbors are shown as circles in Fig. 12.

When the INS algorithm converges, the average radius of the Voronoi regions, i.e., the correlation radius, r_{corr} , is calculated and the distributed CC-MAC protocol, as explained in Section 3.5, is performed. To exploit the spatial correlation between sensor nodes and to improve the performance of the WSN, our MAC protocol tries to create the correlation regions distributively. Note that the INS algorithm determines the representative nodes that would achieve the minimum distortion given the number of representative nodes. However, since this centralized information is not suitable for distributed control, the correlation radius, r_{corr} , is informed to the individual nodes so that they try to form the correlation regions in a distributed manner and choose the representative nodes, accordingly. Since the INS algorithm resides at the sink and requires no location information, no additional energy consumption is introduced at the sensor nodes, which perform only the CC-MAC protocol.

During the lifetime of the network, the network topology can change because of node failure or battery drain of sensor nodes. However, since the distortion depends on the physical phenomenon, such a change should not affect the distortion achieved at the sink unless the number of sensors decreases significantly. In such a case, the event information cannot be captured at the desired distortion level even if all the

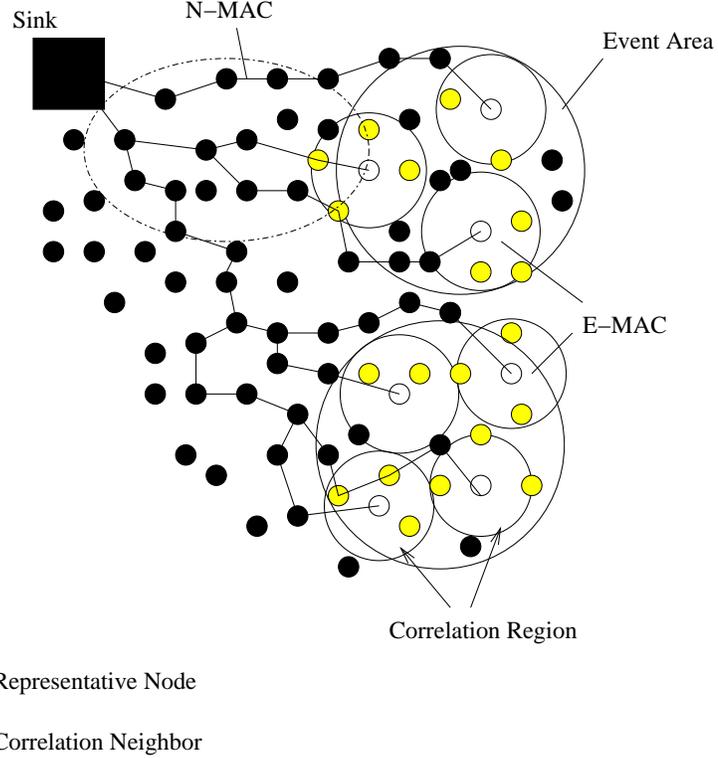


Figure 13: E-MAC and N-MAC. The representative node transmits its record on behalf of the entire correlation region, while all correlation neighbors suppress their transmissions.

nodes send information, i.e., $r_{corr} = 0$. Hence, new nodes can be deployed. If new nodes are deployed in the sensor field to increase the spatial resolution or to improve the connectivity, the CC-MAC operation is not affected since the desired distortion is achieved at the sink.

Overall, the goals of the CC-MAC protocol are to determine representative nodes without any explicit internode communication, to create the correlation regions, and to prevent the correlation neighbors from transmitting their event information.

3.5 *CC-MAC Protocol Description*

Based on the results presented in Section 3.3 and Section 3.4, we propose the spatial correlation-based collaborative MAC (CC-MAC) protocol that aims to collaboratively regulate sensor node transmissions. It follows from our earlier discussion in Section

3.3 that the distortion constraint can be achieved even though the number of nodes sending information about an event is decreased. Furthermore, by intelligently selecting the locations of the representative nodes, the distortion, $D_E(M)$, can be further reduced. To achieve these goals, the INS algorithm, which resides at the sink, determines the correlation radius, r_{corr} , for a distortion constraint, D_{max} , as explained in Section 3.4. This information is then broadcast to each sensor node during the network setup. The CC-MAC protocol, which is implemented at each sensor node, then performs MAC distributively. CC-MAC exploits spatial correlation in the MAC layer by using the correlation radius, r_{corr} , to suppress the redundant information.

We now present the principles of CC-MAC protocol in detail. When a specific source node, n_i , transmits its event record to the sink, all of its correlation neighbors have redundant information with respect to the distortion constraint, D_{max} . This redundant information, if sent, increases the overall latency and contention within the correlation region, as well as wasting scarce WSN energy resources. Our proposed CC-MAC protocol aims to *prevent the transmission of such redundant information and prioritize the forwarding of filtered data to the sink*.

In WSN, the sensor nodes have the dual functionality of being both data originators and data routers. Hence, the medium access is performed for two reasons:

- *Source Function:* Source nodes with event information perform medium access to transmit their packets to the sink.
- *Router Function:* Sensor nodes perform medium access in order to forward the packets received from other nodes to the next destination in the multi-hop path to the sink.

According to the spatial correlation between observations in WSN, the medium access attempts related to the source function of the sensor nodes should be coordinated such that the transmission of the redundant information to the sink is collaboratively

prevented. However, once a packet is injected into the network, it has to be reliably transmitted to the sink. Hence, the route-thru packet is more valuable at an intermediate node than its own generated data packet.

To address these two different contention attempts in WSN, our CC-MAC protocol contains two components corresponding to the source and router functionalities. *Event MAC (E-MAC)* filters out the correlated records and *Network MAC (N-MAC)* ensures prioritization of route-thru packets. More specifically, a node performs E-MAC when it wants to transmit its sensor reading to the sink, while N-MAC is performed when a node receives a packet and tries to forward it to the next hop. A typical WSN with the E-MAC and N-MAC application areas are shown in Fig. 13.

Since centralized medium access is not preferred in WSN, we use a distributed protocol to determine the representative nodes. Both E-MAC and N-MAC use a CSMA/CA-based medium access control with appropriate modifications and enhancements. The information about correlation formation is embedded inside the RTS/CTS/DATA/ACK packets. Each node is informed about the correlation information about a node using these packets. As a result, additional signaling is not required for our CC-MAC protocol. We explain the packet structure and the principles of both E-MAC and N-MAC in the following sub-sections.

3.5.1 Packet Structure

To address the unique characteristics of WSN, i.e., spatially correlated information and higher priority route-thru packets, a bit in the reserved space of RTS, CTS, and DATA packet structures is used as a new field called *first hop (FH) field* as shown in Fig. 14. The *FH field helps the sensor nodes to differentiate the type of packet*, i.e., newly generated packet (source functionality) or a route-thru packet (router functionality), and perform E-MAC or N-MAC accordingly.

When a sensor node records an event, it sets the FH field of the RTS and DATA

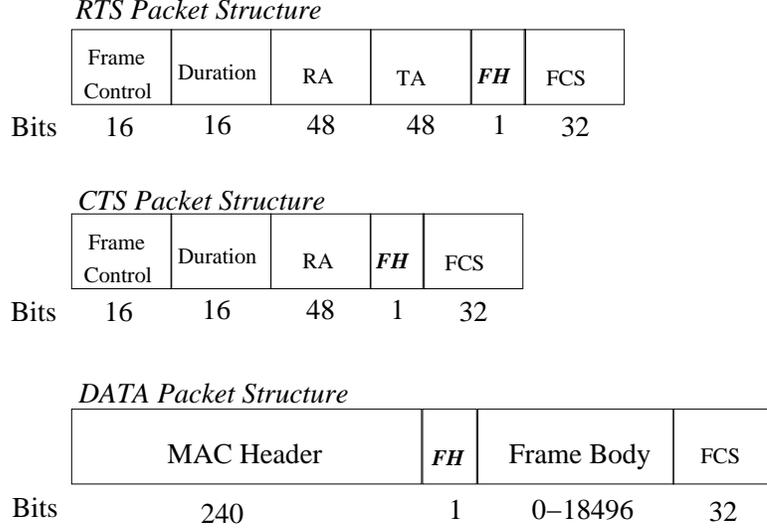


Figure 14: Structures for RTS, CTS, and DATA packets.

packets related to the transmission of the sensor record. All nodes overhearing the RTS with FH field set, determine that the transmission is related to a source functionality and perform E-MAC as will be explained in Section 3.5.2. The recipient of this RTS packet sets the FH field of the CTS packet that is sent back to the source node. As a result, each neighbor of the sender and the receiver is informed about the type of the packet being transferred. Once a node receives the DATA packet, it clears the FH field, indicating that the packet is a route-thru packet. The node, then, simply forwards the packet to the next hop. Nodes accessing the medium for router functionality do not set the FH field in their RTS packets and perform N-MAC as will be explained in Section 3.5.3.

3.5.2 Event MAC (E-MAC)

The Event MAC (E-MAC) protocol aims to filter out correlated event information by forming correlation regions based on the correlation radius, r_{corr} , obtained from the INS algorithm as shown in Section 3.4.

In each correlation region, a single representative sensor node transmits data for a specific duration, while all other nodes stop their transmission attempts.

After each transmission duration, a new representative node is selected as a result of the contention protocol. All sensor nodes contend for the medium for the first time so that the representative nodes are selected by the help of the spatial-reuse property of the wireless channel. This initial phase is called as the *first contention phase* and is explained as follows.

- *First Contention Phase (FCP)*: In the first contention phase, all nodes with event information contend for the medium for the first time using RTS / CTS / DATA / ACK structure [36]. Each of these nodes sets the FH field of the RTS packet and tries to capture the medium for transmission. At the end of this phase, some of the sensor nodes access the channel while others have to backoff. Note that more than one node can access the channel in this first phase because of the spatial reuse as shown in Fig. 13.

When a node n_i captures the channel after the FCP, it becomes the representative node of the area determined by the correlation radius r_{corr} . The node, n_i , continues to send information to the sink as a sole representative of its correlation region. Using the information in the ongoing transmission, other nodes hearing the transmission can determine whether they are correlation neighbors of node n_i . Every node n_j that listens to the RTS packet of the node n_i looks at the FH field and determines that the transmission is related to a source functionality. In addition, each node n_j determines $d_{(i,j)}$, its distance to node n_i . If $d_{(i,j)}$ is found to be less than the correlation radius, r_{corr} , then the node n_j determines that it is a correlation neighbor of the node n_i and stops its transmission. If the node is outside the correlation region of node n_i , then it contends for the medium if it has a packet to send.

The protocol procedure for the correlation neighbors depends on the relation between the transmission range, R , of the sensor nodes and the correlation radius, r_{corr} . For the case when $r_{corr} \leq R$, the transmission area of a node n_i contains the correlation area of the node as shown in Fig. 15. Hence, all the correlation neighbors

of the node n_i can hear the transmission of node n_i . As a result, the redundancy because of correlation can be totally removed by the already ongoing transmission from the representative node. However, in the case when $r_{corr} > R$, some of the correlation neighbors of node n_i cannot hear the transmission of node n_i . Hence, the redundancy cannot be filtered out completely with respect to the total distortion constraint. Based on these observations, we first explain the procedure for the correlation neighbors for the first case and then point out the modifications for the second case.

Case 1: $r_{corr} \leq R$

In this case, when a node n_j determines that it is a correlation neighbor of a node n_i it suppresses its data transfer to the sink for a specific amount of time and performs necessary procedures for forwarding the packets. In addition, in order to conserve energy during the transmission of node n_i , each correlation neighbor enters *suspicious sleep state (SSS)* of duration T_{SSS} . As a result, during the *SSS* period, the representative node n_i continues sending information to the sink as a sole representative of its correlation region. Furthermore, the correlation neighbors defer contending for the medium for T_{SSS} .

In *SSS*, the correlation neighbors switch to sleep state for the duration of the transmission, i.e., T_{tx} , that is extracted from the duration field of the RTS, CTS or

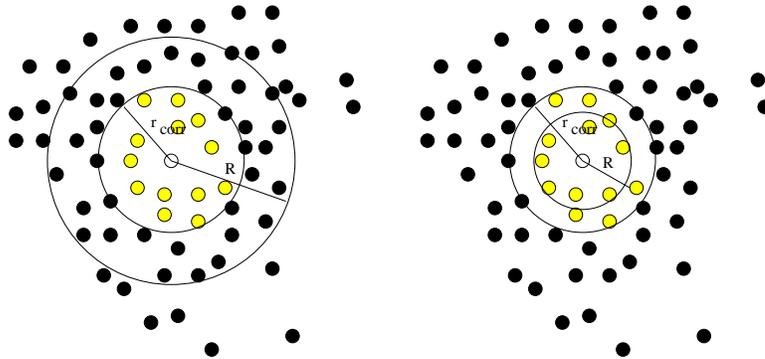


Figure 15: Two cases for E-MAC. The figure shows two cases for correlation region, r_{corr} , and transmission region, R .

DATA packets. However, in order to be able to forward the route-thru packets and to maintain network connectivity, correlation neighbors start to listen to the channel after a random sleep interval of t_{rs} , such that $T_{tx} < t_{rs} < T_{nextTx}$, where T_{nextTx} is the time when node n_i will begin the next transmission of the sensor record. T_{nextTx} is assumed to be set by higher layer protocols and is related to the transmission rate of the application. As a result, a correlation neighbor stays in sleep state during t_{rs} . The correlation neighbors then listen to the channel for EIFS sec and if there is a transmission destined to the neighbor, the node performs N-MAC. Otherwise, it switches to sleep state again.

After T_{SSS} duration, node n_i releases its representative role by switching into SSS , leaving the medium to other nodes. The remaining nodes then go through the first contention phase again. As a result, an equal load-sharing among sensor nodes is provided.

Case 2: $r_{corr} > R$

In this case, some nodes correlated to the representative node n_i cannot be informed about their correlated data. This results in unnecessary contention for the medium from some of the correlation neighbors outside the communication range. However, trying to inform these nodes about node n_i 's transmission requires additional transmission and contention for channel access, which increases the overhead of the protocol. Hence, there is a trade-off between correlation filtering and protocol overhead in this case.

Another problem that needs to be addressed in this case is the routing of node n_i 's packets out of the correlation region. Since all communication neighbors are also correlation neighbors, the packets generated by n_i should be routed through these nodes. However, if all correlation neighbors go into SSS as explained in Case 1, node n_i would not be able to send its packets outside the correlation region. To overcome

this problem, we introduce the *directional sleeping technique*. The directional sleeping technique works as follows. After the first contention phase, all the correlation neighbors drop their in-queue packets. They do not try to send their packets for T_{SSS} seconds but continue to listen to the channel to route packets from node n_i . If, for a certain number κ of transmissions, a correlation neighbor does not receive an RTS packet destined for itself, it determines that the path from node n_i to the sink does not include itself and enters the *SSS*.

Other nodes that are in the route from node n_i to the sink continue to listen to the channel. This principle results in directional sleeping where nodes that are not in the path from node n_i to the sink can switch to sleep state. In addition, *SSS* helps those nodes not in the path to route packets coming from other sources to the sink as explained in Case 1. After the *SSS*, the remaining nodes enter the FCP and the whole process is repeated.

3.5.3 Network MAC (N-MAC)

As a node records an event and transmits its packets using E-MAC, these packets are forwarded through the network by intermediate nodes that perform the router functionality. In addition, node deployment over large sensor fields may have to deal with multiple concurrent events. Hence, when a packet is routed to the sink, it may traverse through nodes corresponding to other concurrent events. However, since the correlation has already been filtered out using E-MAC, the route-thru packet must be given priority over the packets generated by another concurrent event. This is the reason why network MAC (N-MAC) component is required.

When an intermediate node receives a DATA packet, it performs N-MAC to further forward that packet to the next hop. The route-thru packet is given precedence in two phases. When a correlation neighbor receives an RTS regarding a route-thru packet during the random listening period of the *SSS*, it switches from *SSS* to receive

state and receives the packet. During the transmission, the representative node defers its transmission and the route-thru packet is received by the correlation neighbor.

To further exploit the higher priority of the route-thru packet, we use a priority scheme similar to the IEEE 802.11 point coordinate function (PCF)[36]. A node in a correlation region with a route-thru packet listens to the channel for *PCF inter frame space* (PIFS) time units, which is smaller than the *distributed coordination function (DCF) inter frame space* (DIFS) used by the nodes performing E-MAC. The router node then sets its backoff window size to a random number between $[0, CW'_{max} - 1]$, where CW'_{max} is a value smaller than the actual CW_{max} used by the representative node. Such a principle increases the probability that the router node captures the channel since the router node begins backoff before the representative node of the correlation region. As a result, the route-thru packet is given precedence. Since backoff procedure is still used, the collision between multiple route-thru packets that may be in the same correlation region is prevented. If, on the other hand, the representative node receives the route-thru packet, it simply gives precedence to the route-thru packet and forwards the route-thru packet.

The two components of CC-MAC, i.e., E-MAC and N-MAC, provide a complete solution for exploiting spatial correlation at the MAC layer. E-MAC filters the highly correlated information by defining a correlation region for transmitting nodes and preventing correlation neighbors in that region from transmitting their data. N-MAC provides precedence to the filtered out data while it traverses through the WSN. In Section 3.6, we investigate the performance of the CC-MAC protocol as well as the effects of protocol parameters.

3.6 Simulation Results

We use $ns - 2$ [1] for our simulations. To gain more insight into the protocol operation, we first evaluate the effects of CC-MAC protocol parameters on the overall

performance such as energy consumption and distortion achieved at the sink. We then present a comparative study of various MAC protocols. Along with our CC-MAC protocol, we simulated T-MAC [74], TRAMA [59], and S-MAC [83], which are energy aware MAC protocols designed specifically for wireless sensor networks, in addition to IEEE 802.11[36] and a simple carrier sense multiple access (CSMA) protocol.

We present simulation results for sensor topology of 50 nodes randomly deployed in a $500 \times 500 m^2$ sensor field. We assume that one of the sensor nodes is a sink and all other nodes send their event information to that sink. The sensor nodes are modeled according to the $ns - 2$ wireless node module and energy model. The transmission range of each node is $100m$ with average energy consumption of $24.75mW$, $13.5mW$, and $15\mu W$ during transmitting, receiving, and sleeping, respectively. We assume that nodes consume the same energy for idle listening as receiving. The parameters CW_{max} and CW'_{max} are chosen as 1024 and 512. In each simulation, an event occurs with an event source located at the center of the sensor field. Each sensor node reports their event information to the sink. To investigate the effect of the traffic load, the simulation results are presented by varying the reporting period of the sensor nodes. The reporting period determines the period each node creates packets about the event information. Each simulation is performed for $600s$.

3.6.1 CC-MAC Parameters

Two parameters that are required for the CC-MAC operation are the *correlation radius*, r_{corr} , defined in Section 3.4 and the *suspicious sleep state interval*, T_{SSS} , mentioned in Section 3.5. We present the effects of these parameters on the CC-MAC performance according to the following performance metrics:

- *Average Energy Consumption (J)*: is the average energy a node consumes during the simulation. The average energy consumed is the most important metric

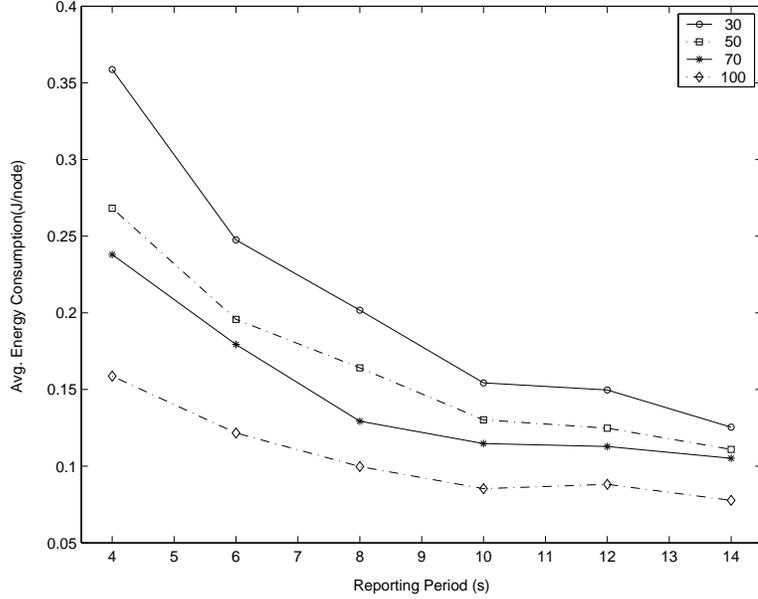


Figure 16: Energy consumption for different correlation radius values.

since the WSN rely on energy awareness.

- *Distortion:* is the average distortion achieved at the sink according to the received sensor information. For each reporting interval, the distortion is calculated using (40), according to the information received at that interval. We use the *Power Exponential* model with $\theta_2 = 1$ and $\theta_1 = 5000$ as the covariance model. The distortion metric is presented to evaluate the reliability of the CC-MAC protocol in terms of distortion.

The first set of experiments are performed for the evaluation of the effect of correlation radius, r_{corr} , defined in Section 3.4. The correlation radius, r_{corr} , is required by the CC-MAC protocol to form the correlation regions for achieving the distortion constraint, D_{max} , as explained in Section 3.5. To observe the effect of the correlation radius, r_{corr} , on the performance of CC-MAC, we performed simulations by varying the correlation radius, r_{corr} from $30m$ to $100m$. We use SSS duration of $T_{SSS} = 50s$ for this set of experiments. This value is chosen such that the representative nodes could send multiple packets even for low reporting rates, such that nodes do not enter

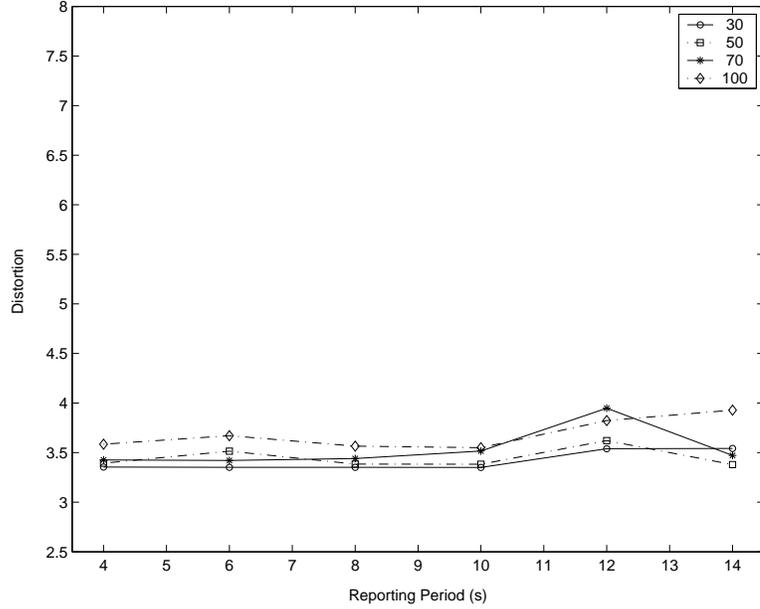


Figure 17: Distortion achieved by different correlation radius values.

the *first contention phase* (FCP), which is explained in Section 3.5, for every packet. Note that this selection is made so that our CC-MAC protocol does not behave as a pure CSMA/CA with sleep periods. The results are shown in Fig. 16 and 17.

Fig. 16 shows the energy consumption for 4 different correlation radius values, i.e., $r_{corr} = \{30m, 50m, 70m, 100m\}$. The average energy consumption decreases as the reporting period is increased, as expected, since less packets are generated during the simulation. It is clearly seen that significant energy conservation is possible by increasing the correlation radius, r_{corr} . Since less representative nodes transmit information to the sink when correlation radius, r_{corr} is increased, less contention attempts are made in the overall network. The gain achieved by increasing the correlation radius is even more significant for high traffic load. As an example, increasing correlation radius from $r_{corr} = 30m$ to $r_{corr} = 100$ results in energy conservation of 56% for reporting period of 4s, while the energy conservation of 38% is achieved for reporting period of 10s. When the traffic load is increased, i.e., the reporting period is decreased, the collision probability during medium access also increases because of

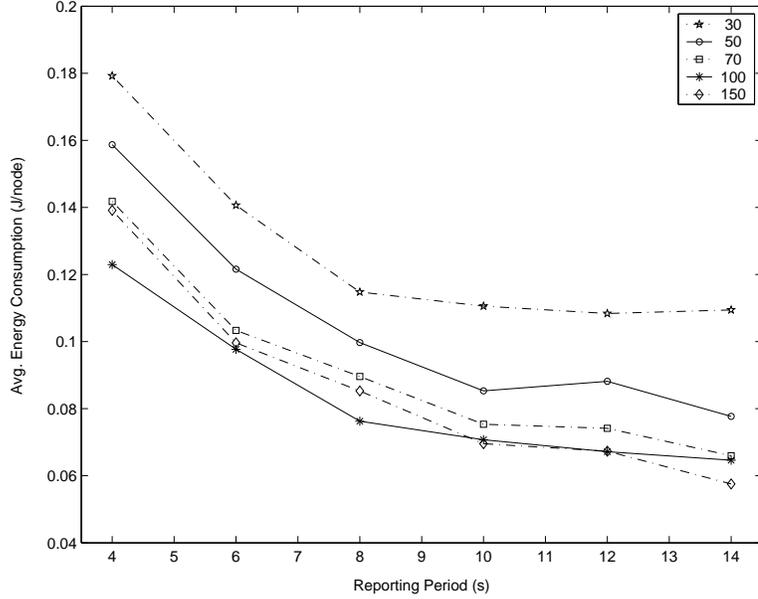


Figure 18: Average energy consumption for various SSS durations.

increased packet flow in the WSN. Hence, the effect of filtering correlated information has an increased impact on the energy conservation for high traffic load.

We show the effect of varying correlation radius, r_{corr} on the achieved distortion in Fig. 17. Fig. 17 shows that the achieved distortion is insensitive to the reporting interval. In addition, the distortion achieved stays relatively constant when the correlation radius, r_{corr} , is varied. These results are consistent with our theoretical results in Chapter 2. The theory concludes that the achieved distortion stays relatively constant when the number of representative nodes is higher than 15. Since the correlation radius, r_{corr} , in effect, determines the number of representative nodes in CC-MAC, the same distortion behavior is also achieved in Fig. 17.

The second set of experiments considers the effect of the SSS duration, T_{SSS} , on the performance metrics. The SSS duration determines how long a node will stay as a representative node after capturing the channel during the FCP as explained in Section 3.5. Small T_{SSS} results in equal sharing of energy consumption by increasing the probability that every node becomes a representative node. Moreover, small T_{SSS} increases the probability that a node will be in the awake state for a route-thru packet.

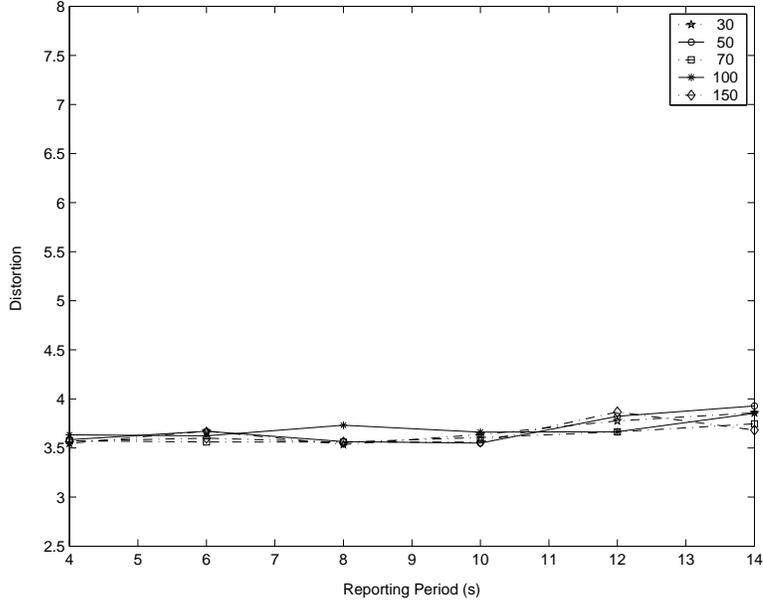


Figure 19: Distortion values for various *SSS* durations.

We performed simulations by changing the *SSS* duration, T_{SSS} , from 30s to 150s. We used $r_{corr} = 100m$ for this set of experiments.

The variation of energy consumption for various T_{SSS} is shown in Fig. 18. A decrease in the *SSS* duration from 150s to 30s, leads up to 60% increase in the energy consumption. As the *SSS* duration decreases, the selection of the representative nodes, i.e., FCP, is performed more frequently. Hence, nodes consume more energy due to the increase in number of contentions.

The distortion performance is shown in Fig. 19. The results show that the distortion is not affected by the choice of T_{SSS} . This follows from the fact that the same number of representative nodes sends information, regardless of the *SSS* duration. Hence, the choice of *SSS* duration, T_{SSS} , can be determined without any consideration about the distortion achieved at the sink.

3.6.2 Comparative Study

In this section we compare the performance results of our CC-MAC with TRAMA [59], S-MAC [83], T-MAC [74], IEEE 802.11 [36], and the basic CSMA protocol. Using

the same sensor network setup in Section 3.6.1, we evaluate the following performance metrics of then above protocols:

- *Medium Access Delay (s)*: is the average time spent between the time a packet is handed to the MAC layer and the time it is received at the next hop. This delay accounts for the contention delay in the case of contention-based protocols [74], [83] and scheduling delay in schedule-based protocols [10], [59].
- *Packet Drop Rate*: is the fraction of packets that is dropped during the medium access. It is calculated as the percentage of dropped packets to the total packets sent from the MAC layer throughout the simulation. This metric shows the performance of the MAC protocol in terms of medium access overhead introduced in terms of wasted number of packets.
- *Goodput*: is the ratio between the total number of packets received at the sink and the total number of packets generated by all sensor nodes. As a result, the efficiency of the MAC protocol is investigated.
- *Average Energy Consumption (J)*: is the average energy a sensor node consumes during the simulation.

TRAMA [59] is a schedule-based MAC protocol, designed specifically for WSN. Each node performs traffic-based scheduling using two-hop neighbor information in the network based on a schedule interval set according to the traffic rate. We use a schedule interval of 100 transmission slots of duration $6.82ms$ for the TRAMA protocol. S-MAC [83] and T-MAC [74], on the other hand, are contention-based MAC protocols that incorporate sleep cycles to conserve energy. Since T-MAC has been proposed to improve the energy efficiency of S-MAC by further trading off throughput and latency, we include T-MAC in our energy consumption analysis. Since the throughput and latency performance of T-MAC is designed to be inferior

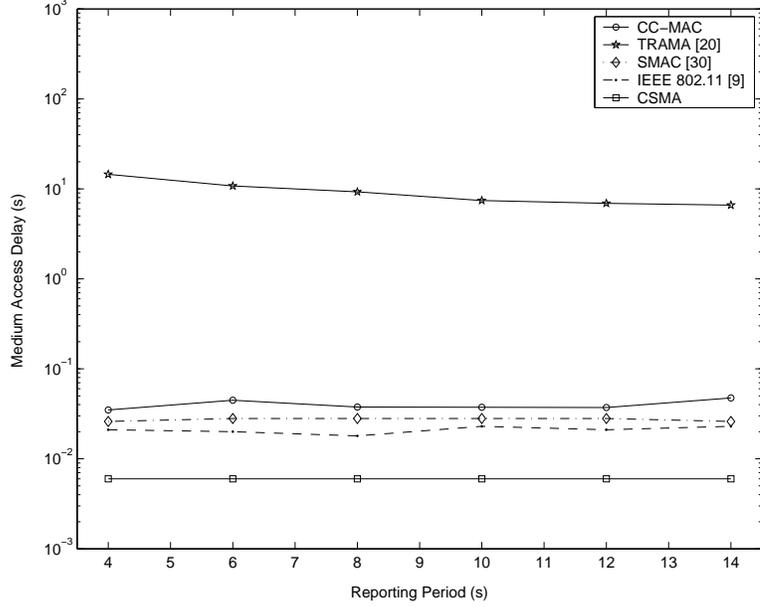


Figure 20: Medium access delay introduced by different MAC protocols.

to S-MAC [74], we consider S-MAC for the other performance metrics described above. We simulated the S-MAC protocol with frequency of schedule update of 10s and 10% duty cycle. The T-MAC protocol is simulated with minimum idle listening period TA of 15ms. The IEEE 802.11 is performed according to the $ns - 2$ module [1]. According to the INS protocol, we set the correlation radius $r_{corr} = 91.68m$ to achieve a distortion constraint of $D_{max} = 3.16$ that accounts for 16 representative nodes in the INS solution. The SSS duration is set as $T_{SSS} = 100s$.

Fig. 20 shows the medium access delay achieved by each MAC protocol. CSMA has the lowest medium access delay of $6ms$, which is because of the lack of collision avoidance mechanism and energy-awareness. CC-MAC performs close to the IEEE 802.11 and S-MAC performance with medium access delay below $50ms$. Note that the delay performance of three protocols is relatively constant for variable traffic load. TRAMA has a medium access delay of $10s$ that is significantly higher than the medium access delay introduced by the contention-based protocols. This difference is because of the schedule-based medium access scheme.

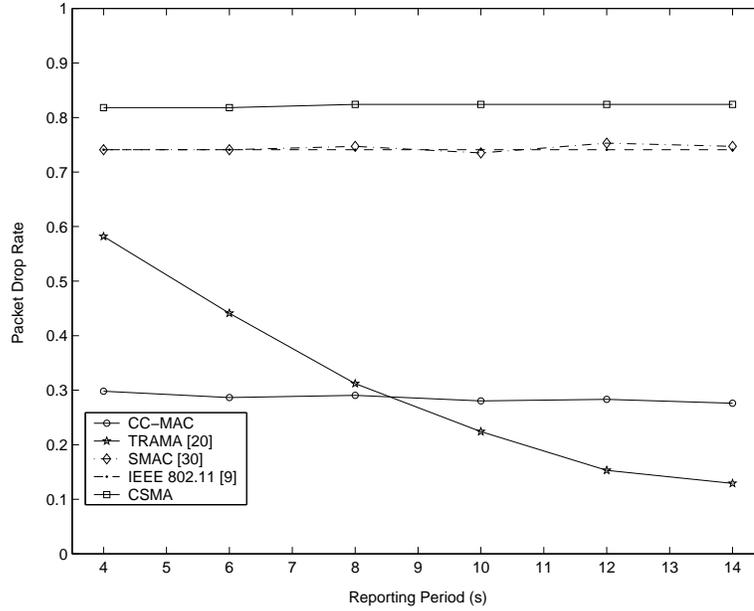


Figure 21: Percentage of dropped packets for different MAC protocols.

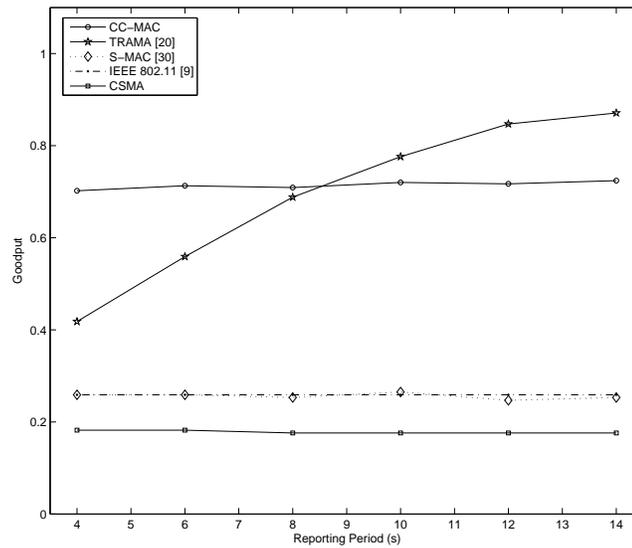


Figure 22: Goodput for different MAC protocols.

The packet drop rate is shown for each of the MAC protocols in Fig. 21. S-MAC, IEEE 802.11, and CSMA achieve similar packet drop rates, which are in the order of 80%, while TRAMA has a packet drop rate between 60% and 20% depending on the traffic load. CC-MAC protocol outperforms all the contention-based protocols, as

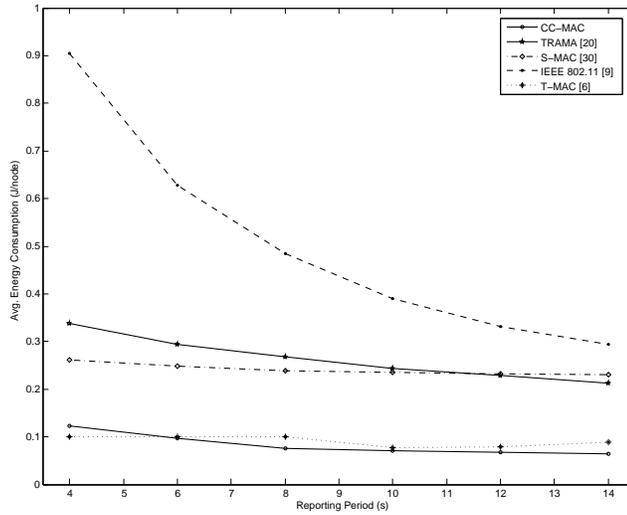


Figure 23: Average energy consumption for different MAC protocols.

well as TRAMA for high traffic load. Note that the packet drop rate is insensitive to traffic load in the case of S-MAC, IEEE 802.11, CSMA, and CC-MAC. As a result, the packet drop rate depends on the number of nodes contending for the medium. Since CC-MAC prevents correlation neighbors from contending for the medium, the packet drop rate is significantly lowered. The packet drop rate of the TRAMA protocol varies according to the traffic load. This is related to the scheduling approach of the protocol. Although collisions are prevented by scheduling transmissions, as the load increases, packet drop rate also increases since the packets cannot be accommodated in the transmission slots. For reporting period higher than 10s, TRAMA has lower packet drop rate than CC-MAC.

The goodput of CC-MAC is shown in Fig. 22 along with CSMA, IEEE 802.11, S-MAC, and TRAMA protocols. Note that we do not compare the throughput since the efficiency of the protocols is more important than the total number of packets received at the sink. It is clear that CC-MAC would achieve lower throughput compared to other protocols since it filters the redundant data injected into the network without compromising the distortion limits. However, sink is interested in uncorrelated data

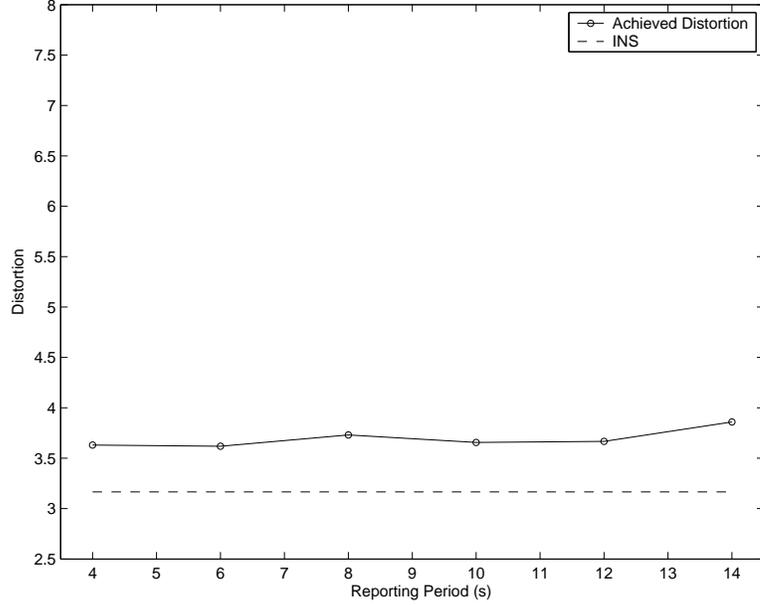


Figure 24: Achieved distortion and the distortion found by the INS algorithm.

to construct the event features rather than the highly redundant data from each sensor node. Since the redundant contention attempts are eliminated, the goodput of CC-MAC is significantly higher than of the other contention-based protocols achieving up to 180% increase compared to S-MAC [83] goodput. Moreover, CC-MAC outperforms TRAMA [59], which is a reservation-based protocol, when the nodes are transmitting at a high frequency. As the reporting period is increased, the frame structure of TRAMA can accommodate more nodes and goodput increases compared to CC-MAC. However, note that the number of nodes sending information to the sink is still much lower in CC-MAC than in TRAMA because of suppressed correlation neighbors. Hence, less packets are introduced into the network with less nodes transmitting data. As a result, energy consumption is decreased as shown in Fig. 23.

The energy consumption performance of CC-MAC with three energy-aware protocols, TRAMA, S-MAC, and T-MAC is shown in Fig. 23. We also compare these four protocols IEEE 802.11 for completeness. As shown in Fig. 23, CC-MAC has

significant energy conservation compared to other energy-aware protocols. While S-MAC outperforms TRAMA for high reporting frequency, TRAMA achieves better energy consumption for increased reporting period. Since, T-MAC provides variable sleep schedules based on the traffic, the energy consumption is significantly lower than S-MAC and TRAMA. However, with the help of spatial correlation-based approach of the CC-MAC protocol, the WSN consumes 25% less energy compared to T-MAC, 70% less energy compared to S-MAC and TRAMA, and 85% less energy compared to IEEE 802.11. As a result, CC-MAC protocol provides significant energy savings without compromising latency and throughput. This significant gain in energy consumption increases the lifetime of the network.

CC-MAC also achieves distortion within 10% of the distortion constraint found theoretically by the INS algorithm as shown in Fig. 24. The increase in the distortion is because of the lost packets during the transmission of event information to the sink. As shown in Fig. 21, 30% of the packets are dropped during transmission. The dropped packets result in an increase in the achieved distortion since less information about the event is received at the sink than anticipated by the INS algorithm. Hence, a reliability guarantee is required from higher layers for the exact achievement of required distortion. Moreover, the distributed nature of the CC-MAC protocol brings additional offset to the achieved distortion. Since the INS algorithm determines the correlation radius based on the statistical properties of the network, the realization of the network may be different because of changes in the statistical properties and different realizations. However, it is clear from our results in Section 3.6.1 that the achieved distortion can further be decreased by decreasing the correlation radius. Hence, by using less number of sensor nodes and filtering out the correlation between sensor observations, CC-MAC protocol achieves significant gains in terms of energy consumption in the MAC layer without compromising from latency, goodput, and distortion.

CHAPTER IV

XLM: CROSS-LAYER MODULE FOR EFFICIENT COMMUNICATION IN WIRELESS SENSOR NETWORKS

In this chapter, a unified cross-layer module (XLM) is presented. XLM achieves efficient and reliable event communication in wireless sensor networks (WSN) with minimum energy expenditure. XLM was first presented in [9]. The motivation for this work is provided in Section 4.1. In Section 4.2, we first provide a review of existing work on cross-layer design in WSN. The XLM protocol basics, overview, and the protocol description are introduced in Section 4.3. In Section 4.4, we provide performance evaluations of the XLM solution and provide a comparative analysis with five layered suites.

4.1 Motivation

Wireless sensor networks (WSN) are event-based systems that exploit the collective effort of densely deployed sensor nodes which continuously observe certain physical phenomenon. In general, the main objective of any WSN application is to reliably detect/estimate event features from the collective information provided by sensor nodes. Nevertheless, the main challenge for achieving this objective is mainly posed by the severe energy and processing constraints of low-end wireless sensor nodes.

Clearly, the collaborative sensing notion of the WSN achieved by the networked deployment of sensor nodes help to overcome the characteristic challenge of WSN, i.e., resource constraints. To this end, there has been significant amount of research effort that aims to develop collaborative networking protocols in order to achieve

communication with maximum energy efficiency.

In addition to the collaborative sensing and networking in WSN, spatio-temporal correlation is another significant characteristic of sensor networks. Dense deployment of sensor nodes makes the sensor observations highly correlated in the space domain with the degree of correlation increasing with internode proximity. Similarly, some of WSN applications such as event tracking require sensor nodes to periodically sample and communicate the sensed event features. The nature of the energy-radiating physical phenomenon yields temporal correlation between each consecutive observation of a sensor node. It has been shown in Chapter 2 and Chapter 3 that exploiting the spatial and temporal correlation further improves energy-efficiency of communication in WSN.

Most of the proposed communication protocols that exploit the collaborative nature of WSN and their correlation characteristics improve energy efficiency to a certain extent. However, the main commonality of these protocols is that they follow the traditional layered protocol architecture. More specifically, the majority of these communication protocols are individually developed and optimized for different networking layers, i.e., transport, network, medium access control (MAC), and physical layers. While these protocols may achieve very high performance in terms of the metrics related to each of these individual layers, they are not jointly designed in order to maximize the overall network performance while minimizing the energy expenditure. Considering the scarce energy and processing resources of WSN, joint design of networking layers, i.e., cross-layer design, stands as the most promising alternative to inefficient traditional layered protocol architectures.

In fact, recent work on WSN [61], [75], [81] reveal that cross-layer integration and design techniques result in significant improvement in terms of energy conservation in WSN. There exists some research on the cross-layer interaction and design in developing new communication protocols [16], [17], [27] [50], [67], [69], [86], [87].

However, as discussed in Section 4.2 in detail, these studies either provide analytical results without any communication protocol design, or perform cross-layer design within a limited scope, e.g., only routing and MAC, which do not consider all of the protocol layers involved in the communication in WSN.

Clearly, there is still much to be gained by rethinking the protocol functions of protocol layers in a unified way so as to provide a single communication module for efficient communication in WSN. To the best of our knowledge, to date, there is no unified cross-layer communication protocol for efficient and reliable event communication that incorporates transport, routing, medium access functionalities along with considering the physical layer (wireless channel) effects for WSN.

In this chapter, a unified cross-layer module (XLM) is developed and presented, which achieves efficient and reliable event communication in WSN with minimum energy expenditure. XLM melts common protocol layer functionalities into a cross-layer module for resource-constrained sensor nodes. The operation of the XLM is devised based on the new cross-layer notion of *initiative determination*, which constitutes the core of the XLM and implicitly incorporates the intrinsic communication functionalities required for successful communication in WSN. Based on the initiative concept, XLM performs receiver-based contention, initiative-based forwarding, local congestion control, and distributed duty cycle operation in order to realize efficient and reliable communication in WSN. In a cross-layer simulation platform, the state-of-the-art layered protocol configurations have been implemented along with XLM to provide a complete performance evaluation. Analytical performance evaluation and simulation experiment results show that XLM significantly improves the communication performance and outperforms the traditional layered protocol architectures in terms of both network performance and implementation complexity.

4.2 *Related Work*

A cross-layer integrated MAC/routing protocol has been independently proposed by [69] and [86]. In [86] and [87], the performance analysis of a cross-layer geographical random forwarding (GeRaF) algorithm is presented. This algorithm introduces receiver-based routing with MAC and routing layer cross-layer interaction. However, the GeRaF algorithm requires a sensor node with two radios for signaling, which may not be feasible in some scenarios. In [85], the MAC protocol is modified for a single radio node. However, for all studies [85], [86], [87], the given analysis considers a perfect channel model and is based purely on geographical relations. As a result, the behavior of the proposed cross-layer protocol presented in [85], [87], [86] may not be realistic.

In [69], a receiver-based routing is again proposed, where the performance of the protocol is analyzed based on a simple channel model and lossless links. Moreover, the latency performance of the protocol is presented based on different delay functions and collision rates. Also, the effect of the physical layer are not considered in the protocol operation.

A new integrated MAC/routing solution is proposed in [60] for geographical routing in WSN. The proposed solution considers a realistic channel model including fading channel statistics. However, this work considers only the interaction between MAC and routing layers and omits the transport layer and physical layer issues. Furthermore, the integrated MAC/routing scheme proposed in this work does not explicitly address the energy-efficiency requirement of energy-constrained sensor networks.

In [16] and [17], a cross-layer optimization solution for power control at the physical layer and congestion control at the transport layer is considered. Moreover, a cross-layer analysis of the impact of physical layer constraints on link-level and network-level performance of CDMA sensor networks is presented in [28]. This work

underlines important tradeoffs between topology control and receiver design principles. However, both these studies apply only to CDMA-based wireless multihop networks which may not apply to WSN where CDMA technology may not be the most efficient scheme.

In [67], new forwarding strategies for geographic routing are proposed based on an analytical work in [88]. The authors provide expressions for the optimal forwarding distance for networks with and without ARQ. However, the forwarding protocol does not consider the impact of medium access and uses a classical sense of forwarding that results in a very high overhead. Moreover, the analysis for the distribution of optimal hop distance is based on a linear network structure, which may not be realizable in WSN using geographical routing, where a 2-dimensional terrain exists.

In [27], joint routing, MAC, and link layer optimization is proposed. Although the optimization problems presented in this work are insightful, no communication protocol for practical implementation is proposed. Moreover, the transport layer issues such as congestion and flow control are not considered. Furthermore, in [47] and [48], a thorough investigation of optimization techniques for cross-layer design in wireless networks is performed. It has been stated that scheduling constitutes the bottleneck in optimization due to the nonconvex nature of the scheduling problem. Consequently, in [47], a distributed cross-layer congestion control and scheduling algorithm is developed. However, this solution considers a pair-wise cross-layer design including only transport and link layers.

Consequently, these studies either provide analytical results without any communication protocol design, or perform cross-layer design within limited scope, e.g., only routing and MAC, which do not consider all of the networking layers involving in the communication in WSN.

4.3 XLM: Cross-Layer Module for WSN

XLM is a cross layer communication module for WSN, which replaces the entire traditional layered protocol architecture that has been used so far in WSN. The design principle of XLM is a complete unified cross-layering such that both the information and the functionalities of traditional communication layers are melted in a single module. *Consequently, XLM incorporates the required functionalities and aims to address the corresponding responsibilities of transport, network and MAC layers of classical layered networking approach by taking the physical layer and channel effects into account.*¹ The objective of XLM is highly reliable communication with minimal energy consumption, adaptive communication decisions and local congestion avoidance. Since XLM replaces all of the traditional communication layers, it has all of the required functionalities. To this end, the cross-layer module incorporates new cross-layer notion of initiative determination, receiver-based contention, initiative-based forwarding, local congestion control, and distributed duty cycle operation. The details of these functionalities are explained in the following sections. Before explaining the specifics of the XLM operation, in this section, we first provide an overview of the cross-layer operation of XLM.

The basis of communication in XLM is build on the so-called *initiative* concept, which provides freedom for each node to decide on participating in communication. Consequently, a completely distributed and adaptive network operation is realized. In WSN, the major goal of a communication suite is to successfully transport event information by constructing (possibly) multi-hop paths to the sink. During XLM operation, the next-hop in each communication is not determined in advance. Instead,

¹Note that the sensor nodes equipped with XLM will still have RF transceiver which has all necessary physical layer functionalities, e.g., modulation/demodulation, channel coding, RF power control, specified according to the specific deployment and application requirements.

an *initiative determination* procedure is used for each node to decide on participating in communication, i.e., routing and forwarding of the event information. As will be explained next, initiative determination is performed by each node based on its current capabilities related to all communication layers, e.g., link quality and relative location deduced from the channel information, current incoming forwarding request load, buffer level, and remaining energy level; melting these information into a unified decision incentive driving its level of willingness of participating in the communication in the sensor network. Therefore, the cross-layer initiative determination concept constitutes the core of the XLM and implicitly incorporates the intrinsic communication functionalities required for successful communication in WSN.

A node initiates transmission by broadcasting an RTS packet to indicate its neighbors that it has a packet to send. Upon receiving an RTS packet, each neighbor of node i decides to participate in the communication or not. This decision is given through *initiative determination* considering the current status of the node's communication capabilities. The initiative determination is a binary operation where a node decides to participate in communication if its initiative is 1. Denoting the initiative as \mathcal{I} , it is determined as follows:

$$\mathcal{I} = \begin{cases} 1, & \text{if } \begin{cases} \xi_{RTS} \geq \xi_{Th} \\ \lambda_{relay} \leq \lambda_{relay}^{Th} \\ \beta \leq \beta^{max} \\ E_{rem} \geq E_{rem}^{min} \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (41)$$

The initiative is set to 1 if all four conditions in (41) are satisfied. Each condition in (41) constitutes a certain communication functionality in XLM. The first condition, i.e., $\xi_{RTS} \geq \xi_{Th}$, ensures reliable links be constructed for communication based on the current channel conditions. For this purpose, it requires that the received signal

to noise ratio (SNR) of an RTS packet, ξ_{RTS} , is above some threshold ξ_{Th} for a node to participate in communication. The effect of this threshold on routing and energy consumption performance will be analyzed and the most efficient value of this threshold will be chosen in Section 4.3 . The second, i.e., $\lambda_{relay} \leq \lambda_{relay}^{Th}$, and the third, i.e., $\beta \leq \beta^{max}$, conditions are used for local congestion control in XLM. As explained in Section 4.3.5, the second condition in this component prevents congestion by limiting the traffic a node can relay. More specifically, a node participates in the communication if its relay input rate, λ_{relay} , is below some threshold λ_{relay}^{Th} . The third condition ensures that the node does not experience any buffer overflow and hence, also prevents congestion. The last condition, i.e., $E_{rem} \geq E_{rem}^{min}$, ensures that the remaining energy of a node E_{rem} stays above a minimum value, E_{rem}^{min} . This constraint helps preserve uniform distribution of energy consumption throughout the network.

The cross-layer functionalities of XLM lie in these constraints that define the *initiative* of a node to participate in communication. Using the initiative concept, XLM performs receiver-based contention, initiative-based forwarding, local congestion control, hop-by-hop reliability, and distributed operation. The details of XLM operation are explained next. More specifically, we first define the basic parameters for the operation of XLM in Section 4.3.1. For a successful communication, a node first initiates transmission as explained in Section 4.3.2. Then, the nodes that hear this initiation perform initiative determination. The nodes that decide to participate in communication perform receiver-based contention and angle-based routing as described in Section 4.3.3 and Section 4.3.4. Moreover, the local congestion control component ensures energy efficient as well as reliable communication by two-step congestion control as explained in Section 4.3.5. Finally, based on this protocol operation description, the most efficient operation point for XLM is analytically investigated in Section 4.3.6.

4.3.1 XLM Basics and Definitions

We assume the following network model for the operation of XLM. Each node performs a distributed duty cycle operation. The value of the duty cycle is denoted by δ and defines the ratio of the time a node is active. Each node is implemented with a *sleep frame* with length T_S sec. As a result, a node is active for $\delta \times T_S$ sec and sleeps for $(1 - \delta) \times T_S$ sec. Note that the start and end times of each node's sleep cycle are by no means synchronized. As a result, a distributed duty cycle is employed.

In the network, nodes send information to the sink if an event occurs in their vicinity. The area that an event occurs is denoted as the *event area*. Based on this model and the XLM mechanism, each node in XLM, contributes to the transmission of event information to the sink based on its duty in the network and the current network conditions. The duty of a node in WSN can be classified in two classes:

- *Source Duty*: Source nodes with event information need to transmit their packets to the sink. Hence, these nodes perform transmission rate selection based on the congestion in the network.
- *Router Duty*: Sensor nodes also forward the packets received from other nodes to the next destination in the multi-hop path to the sink. These nodes indicate their initiative on accepting new flows through their path to the destination.

Based on these duties, each node will determine its initiative to participate in the transmission of an event as explained above. The protocol operation details explained in the following sections incorporate the aforementioned definitions.

4.3.2 XLM Transmission Initiation

Since a node may be spatially correlated with its neighbors, when it has a packet to send, it also checks if its information is correlated with the transmitting source nodes, abandoning the transmission if a correlated node exists [80]. This is accomplished

by when a node has a packet to transmit, it first listens to the channel for a specific period of time, T_{SSS} , as we have previously investigated in [80]. If the channel is occupied, the node performs backoff based on its contention window size CW_{RTS} . When the channel is idle, the node broadcasts an RTS packet, which contains the location of the sensor node i and the location of the sink. This packet serves as a link quality indicator and also helps the potential destinations to perform receiver-contention which is explained in Section 4.3.3. When a node receives an RTS packet, it first checks the source and destination locations. It is clear that, in order to route a packet to the destination, the next hop should be closer to the sink than node i . We refer to this region where the neighbors of a node that are closer to the sink reside as the *feasible region*. Similarly, the region where the neighbors of a node that are farther to the sink is referred to as the *infeasible region*. Hence, a node receiving a packet first checks if it is inside the feasible region of the transmitting node i . In order to save energy, the nodes inside the infeasible region of node i switch to sleep. The nodes inside the feasible region perform initiative determination as explained in Section 4.3. If a node decides to participate in communication, it performs *receiver contention* as explained in Section 4.3.3.

4.3.3 XLM Receiver Contention

The receiver contention operation of XLM is based on the receiver-based routing approach [69], [87]. After an RTS packet is received, if a node has an initiative, i.e., $\mathcal{I} = 1$, to participate in the communication, it performs receiver contention to forward the packet. The receiver contention is based on the routing level of each node which is determined by its location. The routing level of a node is decided based on the progress a packet would make if the node forwards the packet. The feasible region is divided into N_p priority regions corresponding to an increasing progress, i.e., A_i , $i = 1, \dots, N_p$. The nodes with the longer progress have higher priority over other

nodes. This prioritization is performed by the contention mechanism for medium access.

Each priority region, A_i , corresponds to a backoff window size, CW_i . Based on the location, a node determines its region and backs off for $\sum_{j=1}^{i-1} CW_j + cw_i$, where $cw_i \in [0, CW_i]$. This backoff scheme helps differentiate nodes of different progress into different prioritization groups. Only nodes inside the same group contend with each other. The winner of the contention sends a CTS packet to node i indicating that it will forward the packet. On the other hand, if during backoff, a potential receiver node hears a CTS packet, it determines that a node with a longer progress has accepted to forward the packet and switches to sleep.

When node i receives a CTS packet from a potential receiver, it determines that receiver contention has ended and sends a DATA packet indicating the position of the winner node in the header. The CTS and DATA packets both indicate the other contending nodes the transmitter-receiver pair. Hence, other nodes stop contending and switch to sleep. In the case of two nodes sending CTS packets without hearing each other, the DATA packet sent by the node can resolve the contention. Note that node i may not receive a CTS packet because of three cases:

- CTS packets collide,
- There exists no potential neighbors with $\mathcal{I} = 1$, or
- There exists no nodes in the feasible region.

However, node i cannot differentiate these three cases by the lack of a CTS packet. Hence, the neighbors of node i send a keep alive packet after $\sum_{j=1}^{\mathcal{N}_{pr}} CW_j$ if no communication is overheard. The existence of a keep alive packet notifies the sender that there are nodes closer to the sink, but the initiative shown in (41) is not met. With the reception of this packet, node continues retransmission. However, if a keep alive packet is not received, the node continues retransmission in case there is a CTS packet

collision. If no response is received after some retries, node i determines that a local minimum is reached and switches to angle-based routing mode as explained next.

4.3.4 XLM Angle-based Routing

XLM constructs routes through initiative-based forwarding in a hop-by-hop fashion. This technique generally results in reliable and efficient end-to-end paths as will be shown in Section 4.4. Since the routing decisions depend, in part, on the locations of the receivers, however, there may be cases where the packets reach local minima. In other words, a node cannot find any feasible nodes that are closer to the sink than itself. This problem is a well known phenomenon in geographical routing-based approaches and is generally resolved through face routing techniques [39, 41]. Face routing, however, necessitates a node to communicate with its neighbors to establish a planarized graph and construct routes to traverse around the "void" generated by the local minimum. Since this communication will increase the protocol overhead and violate the receiver-based principle of XLM, we introduce an angle-based routing technique.

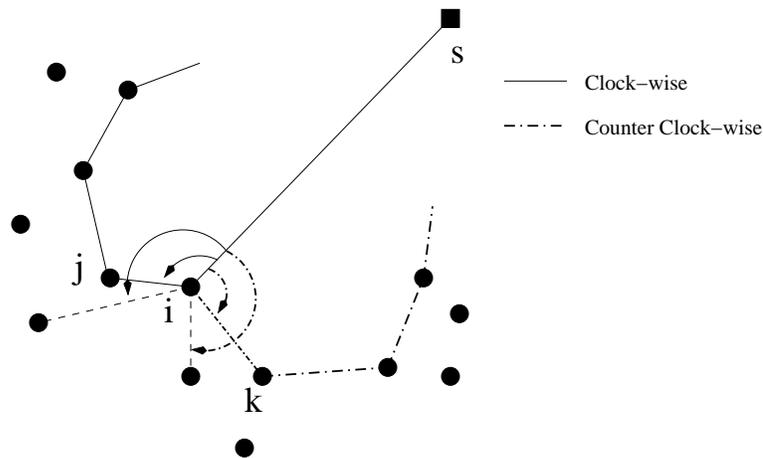


Figure 25: Illustration of angle-based routing.

The main principle of the angle-based routing can be seen in Fig. 25. When a packet reaches node i , which is a local minimum towards the sink, the packet has

to be routed either in clock-wise direction (through node j) or in counter clock-wise direction (through node k). Assume that lines are drawn between the node i and the sink, s , as well as between node i and its neighbors. If we compare the angles between the line $i - s$ and the other lines, line $i - j$ (line $i - k$) has the smallest angle in the clock-wise (counter clock-wise) direction. Using this geometric property, routes can be constructed. Once a measuring direction is set (clock-wise or counter clock-wise), the packet can traverse around the void by using the same direction. Hence, for angle-based routing we introduce the term *traversal direction* to indicate this direction. Next, we explain how the angle-based routing works.

When a node switches to angle-based routing mode as explained in Section 4.3.3, it also sets the traverse direction to clock-wise and sends an RTS packet, which indicates both the routing mode and the traverse direction. The nodes that receive this packet calculate their angle relative to the source-sink direction. Denoting the angle by θ_{ij} , node j sets its contention window to $c \theta_{ij} + cw_i$, where cw_i is a random number, and c is a constant.² The node with the smallest contention window (hence, the smallest angle) sends a CTS packet and the data communication takes place. This procedure is repeated until the packet reaches a local minimum. In this case, the traverse direction is set to counter-clockwise and the procedure is repeated. Angle-based routing is terminated and default routing is performed when the packet reaches a node that is closer to the sink than the node that started angle-based routing. A sample route found by this algorithm is shown in Fig. 26. XLM switches to angle-based routing mode in clock-wise direction at node a . At nodes b and f the traverse direction is changed while at node d angle-based routing mode is terminated and at node e , it is used again.

²The constant can be selected according to the latency requirements and the density of the network.

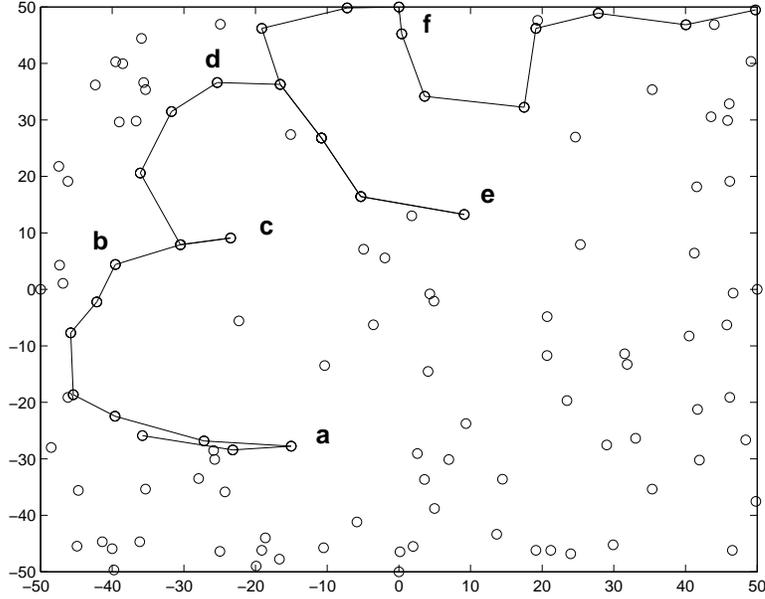


Figure 26: Sample route created by angle-based routing.

4.3.5 XLM Local Cross-Layer Congestion Control

XLM incorporates a new hop-by-hop local cross-layer congestion control component which is devised based on the buffer occupancy analysis presented here. The objective of this component is to perform hop-by-hop and local congestion control by exploiting the local information in the receiver-contention in order to avoid the need for an end-to-end congestion control. It also exploits the local reliability measures taken by the channel access functionality hence does not necessitate traditional end-to-end reliability mechanisms.³

As discussed in Section 4.3.1, a sensor node has two duties in WSN, i.e., source duty and router duty. Accordingly, here, we consider two sources of traffic as an input to the buffer of each node:

³The sink is only interested in reliable detection of event features from the collective information provided by numerous sensor nodes and not in their individual reports. As the correlated data flows originated from the event area are loss tolerant to the extent that event features are reliably communicated to the sink, the need for end-to-end reliability may not exist due to the sheer amount of correlated data flows [4]. Hence, the local reliability measures of XLM suffice for an event in the sensor field to be tracked with a certain accuracy at the sink. This is also observed in the performance evaluation of XLM in Section 4.4.

- *Generated packets:* The first source is the application layer, i.e., the sensing unit of a node, which senses the event and generates the data packets to be transmitted by the sensor node during its source duty as discussed in Section 4.3.1. We refer to these packets as the *generated packets*. For a node i , the rate of the generated packets is denoted by λ_{ii} .
- *Relay packets:* In addition to generated packets, as a part of its router duty, a node also receives packets from its neighbors to be forwarded to the sink due to multi-hop nature of sensor networks. These packets are referred as the *relay packets*. The rate at which node i receives relay packets from node j is denoted as λ_{ji} .

The input rate to the buffer of node i is hence the combination of the input rates of these two types of packets. Since the sensor nodes utilize a duty cycle, the buffer occupancy of the nodes build up while they sleep due to the generated packets.

Hence, based on the above definitions, the local cross-layer congestion control component of XLM has two main congestion control measures. The main idea of XLM cross-layer congestion control is to regulate the congestion:

- *in router duty*, by providing the sensor nodes with the freedom of deciding whether or not participating in the forwarding of the relay packets based on the current load on the node due to its relaying functionality, and
- *in source duty*, by explicitly controlling the rate of the generated data packets.

Here, we first aim to analyze the upper bound for the total relay packet rate a sensor node can accommodate in order to obtain a decision measure for local congestion control in router duty. This bound, denoted by λ_{relay}^{Th} , is used in the XLM initiative determination as presented in (41) in Section 4.3.

Denoting the generated packet rate of a node i by λ_{ii} , the input packet rate at the node i 's buffer, λ_i , can be represented as

$$\lambda_i = \lambda_{ii} + \lambda_{i,relay} = \lambda_{ii} + \sum_{j \in \mathcal{N}_i^{in}} \lambda_{ji} \quad (42)$$

where \mathcal{N}_i^{in} is the set of nodes which have node i as the next hop and λ_{ji} is the packet rate from node j ($\in \mathcal{N}_i^{in}$) to node i . Moreover, the output rate of node i can be given by

$$\mu_i = (1 + e_i)(\lambda_{ii} + \lambda_{i,relay}) \quad (43)$$

where e_i is the packet error rate. A node is active on the average δ fraction of time. Hence, the average time the node i spends in transmitting, receiving and listening during the active period is given by

$$\begin{aligned} T_{rx} &= \lambda_{i,relay} T_{PKT} , \\ T_{tx} &= (1 + e_i)(\lambda_{ii} + \lambda_{i,relay}) T_{PKT} , \\ T_{listen} &= \delta - [(1 + e_i)\lambda_{ii} + (2 + e_i)\lambda_{i,relay}] T_{PKT} , \end{aligned}$$

respectively, where T_{PKT} is the average duration required to successfully transmit a packet to another node, λ_{ii} is the generated packet rate, and $\lambda_{i,relay}$ is the total input relay packet rate of node i .

In order for a node to prevent buffer overflow and maintain its duty cycle, $T_{listen} \geq 0$. Consequently, the input relay packet rate, $\lambda_{i,relay}$ is bounded by

$$\lambda_{i,relay} \leq \lambda_{i,relay}^{Th} , \quad (44)$$

where the relay rate threshold, $\lambda_{i,relay}^{Th}$, is given by

$$\lambda_{i,relay}^{Th} = \frac{\delta}{(2 + e_i) T_{PKT}} - \frac{1 + e_i}{2 + e_i} \lambda_{ii} . \quad (45)$$

As a result, XLM incorporates a hop-by-hop congestion control which is devised based on this buffer occupancy analysis. Nodes participate in routing packets as long

as (44) is satisfied. According to (45), the relay rate threshold is directly proportional to the duty cycle value, δ . This suggests that the capacity of the network will decrease as δ is reduced. However, since lower δ results in less energy consumption, this tradeoff needs to be analyzed carefully. Moreover, note that the input relay rate of source nodes, i.e., $\lambda_{ii} > 0$, should be lower than that of the nodes that are only relays, i.e., $\lambda_{ii} = 0$. This provides homogenous distribution of traffic load to sensor nodes, where source nodes relay less traffic.

In addition to congestion control based on regulating the relaying functionality as discussed above, the XLM local congestion control component also takes an active control measure in case of network congestion, by directly regulating the amount of traffic generated and injected into the network.

During the receiver-contention mechanism described in Section 4.3.3, node i may not receive any CTS packets but receive keep alive packets. In this case, node i decides that there is a congestion in the network. Then, it reduces its transmission rate by decreasing the amount of traffic generated by itself. In other words, since the traffic injected by any node due to its router duty is controlled based on (44), the active congestion control is performed by controlling the rate of generated packets λ_{ii} at the node i itself.

Therefore, in case of congestion, XLM node reduces the rate of generated packets λ_{ii} multiplicatively, i.e., $\lambda_{ii} = \lambda_{ii} \cdot 1/\mu$, where μ is defined to be the transmission rate throttle factor. If there is no congestion detected, then the packet generation rate can be increased conservatively in order not to lead to oscillation in the local traffic load. Therefore, XLM node increases its generated packet rate linearly for each ACK packet received, i.e., $\lambda_{ii} = \lambda_{ii} + \alpha$. Here, we select $\mu = 2$, i.e., the rate of generated packets is halved in case of congestion, and $\alpha = \lambda_{ii_0}/10$, where λ_{ii_0} is the initial value of the generated packet rate set by the sensing application. Here, note also that XLM adopts a rather conservative rate control approach. This is mainly because it

has two functionalities to control the congestion for both source and router duties of a sensor node. As the node decides to take part in the forwarding based on its buffer occupancy level, it already performs congestion control as part of the XLM's forwarding mechanism. Hence, XLM node does not apply its active congestion control measures, i.e., linear increase and multiplicative decrease, to the overall transmission rate. Instead, it only updates the generated packet rate, λ_{ii} .

Note also that since the local congestion control is specific to certain regions and may not apply to the entire event area, nodes inside a congested region may reduce their transmission rates and the overall event reliability may still be met at the sink from the data from other nodes due to the sheer amount of correlated data flows as in [4]. Thus, instead of an inefficient end-to-end reliability mechanism, this local cross-layer congestion control exploits the local congestion control and reliability in order to maintain high network utilization and reliability in a local and distributed manner. In fact, this is also clearly observed in the performance evaluation experiment results as presented in Section 4.4.

The overall XLM protocol operation and its pseudo-algorithm is also presented in Fig. 27. The lines 1-13 show the algorithm for a source node i , which has a packet to send. Note that node i performs either retransmission or switches to angle-based routing based on the reception of a CTS packet, a Keep Alive packet, or the lack of them (lines 3 - 10). The lines 14-29 show the XLM operation for a neighbor node j , which receives a packet. In the case of an RTS packet, the initiative is calculated and the backoff timer is set based on whether default or angle-based routing is used as explained in Section 4.3.3 and Section 4.3.4 (lines 16-19). The expiration of the timers result in a node to transmit a CTS packet. Similarly, if node j receives a CTS, DATA, or ACK packet, these timers are reset. Finally, the values for λ_{ii} and $\lambda_{i,relay}$ are updated for each successful or unsuccessful communication according to (45) (lines 5, 8, and 24).

```

1 if has packet to send then
    // Source node i
2   Perform carrier sense and transmit RTS
3   if CTS received then
4     Transmit DATA
5     if ACK received then Increase  $\lambda_{ii}$ , update  $\lambda_{i,relay}$ 
6     else Retransmit RTS
7   else if Keep alive packet received then
8     Decrease  $\lambda_{ii}$ , update  $\lambda_{i,relay}$ 
9     Retransmit RTS
10  else
11    Switch to angle-based routing
12    Retransmit RTS
13  end
14 else if packet received then
    // Neighbor node j
15  switch packet type do
16    case RTS
17      Calculate  $\mathcal{I}$ 
18      if  $\mathcal{I}=1$  then Set backoff timer
19      else Set backoff for keep alive
20    case CTS
21      Reset backoff timers, switch to sleep
22    case DATA
23      if Destined for itself then
24        Transmit ACK, update  $\lambda_{j,relay}$ 
25      else Reset backoff timers, switch to sleep
26    case ACK
27      Reset backoff timers
28  end
29 end

```

Figure 27: Pseudocode of XLM

4.3.6 XLM Duty Cycle Analysis

XLM employs a distributed duty cycle operation as described in Section 4.3.1. Hence, the choice of the duty cycle value, δ , is important for the performance of XLM. Based on the XLM operation specifics described in the previous sections, here, we investigate the effect of duty cycle on the network performance using an energy consumption

analysis. The objective of this analysis is to find the optimum operating point for XLM in terms of duty cycle, δ . In this respect, the energy consumed by the network for a packet sent to the sink as a function of the distance of its source to the sink is investigated.

The total energy consumed as a result of a single flow from a source node at distance D from the sink can be found as

$$E_{flow}(D) = E_{per-hop} E[n_{hops}(D)] \quad (46)$$

where $E_{per-hop}$ is the average energy consumed in one hop for transmitting a packet, and $E[n_{hops}(D)]$ is the expected hop count from a source at distance D to the sink. An accurate approximation for the expected hop count is given in [87] as

$$E[n_{hops}(D)] \simeq \frac{D - R_{inf}}{E[d_{next-hop}]} + 1 \quad (47)$$

where $E[d_{next-hop}]$ is the expected hop distance, which we will derive in Chapter 5 [79] and R_{inf} is the approximated transmission range.

The energy consumed in one hop has three components as given by

$$E_{per-hop} = E_{TX} + E_{RX} + E_{neigh} \quad (48)$$

where E_{TX} is the energy consumed by the node transmitting the packet, E_{RX} is the energy consumed by the node receiving the packet, and E_{neigh} is the energy consumed by the neighbor nodes. Note that similar energy consumption analysis has also been performed in the literature in a node-centric manner which required models/assumptions for the generated traffic, e.g., [69], [87]. On the other hand, the effect of neighbor nodes has not been considered [34]. In our analysis, we investigate the energy consumption to transmit a single packet to the sink with the effect of neighbor nodes, which provides a clearer insight on the energy consumption.

In order to successfully transmit the packet, a node needs to complete the four-way handshaking. Assume that the distance between the nodes transmitting and

receiving packet is $d_h = E[d_{next-hop}]$ and the probability that a data and a control packet is successfully received at this distance are $p_s^D(d_h)$ and $p_s^C(d_h)$, respectively⁴. When a transmitter node sends an RTS packet, it is received by the receiver node with probability $p_s^D(d_h)$ and the node replies with a CTS packet. If the CTS packet is received (with probability $p_s^C(d_h)$), the transmitter node sends a DATA packet, and the communication is concluded with an ACK packet. In every failure event, the node begins retransmission. Therefore, the expected energy consumed by the transmitting node, E_{TX} , is

$$E_{TX} = \frac{K}{(p_s^C)^3 p_s^D} , \quad (49)$$

where

$$\begin{aligned} K = & E_{sense} + (p_s^C)^2 [E_{tx}^R + E_{wait}^C + E_{rx}^C] \\ & + \left(1 - (p_s^C)^2\right) E_{t/o}^C + (p_s^C)^3 p_s^D [E_{tx}^D + E_{rx}^A] \\ & + (p_s^C)^2 (1 - p_s^C p_s^D) E_{i/o}^A \end{aligned}$$

where E_{sense} is the energy consumption spent for sensing the region, E_{tx}^x and E_{rx}^x are the packet transmission and reception energies spent for packets, where the superscripts R , C , D , A refer to RTS, CTS, DATA and ACK packets, respectively. E_{wait}^{CTS} is the expected energy consumption for waiting for a receiver CTS, and $E_{t/o}$ is the energy consumed before the transmitter node times out, deciding that a suitable router does not exist.

The two terms in (49), E_{wait}^C and $E_{t/o}^C$ are the only system dependent terms. The expected waiting time for the next hop E_{wait}^C is calculated next.

According to the discussion in Section 4.3.3, on the average, each node in priority region, A_i , waits for $CW_i/2$ in its priority slot as well as the previous priority slots. Denoting the probability that the next hop for node i , \mathcal{N}_i , exists in A_k by $P_k =$

⁴We reasonably assume that the length of RTS, CTS and ACK packets are the same.

$P\{N_i = j, \text{ s.t. } j \in A_k\}$, the average waiting time for the next hop is given by

$$E_{wait}^C = e_{rx} \left\{ \sum_{i=1}^{N_P} \left[\sum_{k=1}^{i-1} CW_k + \frac{CW_i}{2} \right] P_i \right\}, \quad (50)$$

where

$$P_k = \left(1 - p_{[A(\gamma_{k-1}), \xi_{Th}]} \right) p_{[A(\gamma_k), \xi_{Th}]}, \quad (51)$$

$p_{[A(\gamma_k), \xi_{Th}]} = 1 - p_k$, p_k is given in Chapter 5 [79], e_{rx} is the energy consumption for receiving and γ_k is maximum distance from the sink for nodes in A_k .

Using the same approach, the energy consumption of the receiver node can be found as:

$$E_{RX} = \frac{1}{(p_s^C)^3 p_s^D} \{ E_{rx}^R + E_{wait}^C + E_{tx}^C + E_{rx}^D + E_{rx}^A \} \quad (52)$$

The last term in (48), E_{neigh} , is the energy consumed by the neighbors of the transmitter and the receiver nodes. First, all the nodes inside the transmission region of the transmitter consume energy for RTS packet reception if they receive the packet. Moreover, there exist nodes that listen only the CTS message sent by the receiver node. Considering these principles, the energy consumption of the neighbor nodes, E_{neigh} is expressed by:

$$E_{Neigh} = \frac{1}{(p_s^C)^2 p_s^D} \left\{ \rho \delta (\pi R_{inf}^2 - 2) p_s^C E_{rx}^R + \left(\rho \delta A(D, R_{inf}, D) - 2 \right) \times \left(E_{wait}^C + E_{rx}^C + \frac{E_{rx}^D}{2} \right) \right\}. \quad (53)$$

Finally, the probability that a packet is received is given by [88]

$$p_s = \left(1 - \frac{1}{2} e^{-\frac{\xi}{1.28}} \right)^{16l} \quad (54)$$

where ξ is the received SNR, and l is the packet length, which is l_C and l_D for p_s^C and p_s^D , respectively.

Using (49), (52), (53), and (54) in (48), the overall energy consumption of a flow can be found. Using numerical integration methods, the effect of distance, D , on the

energy consumption of a flow is found and shown in Fig. 28. Considering Fig. 28, energy consumption of a flow is minimal for $\delta \sim 0.002$. However, in relatively small sized networks of < 1000 nodes, this operating point may not provide connectivity in the network. On the other hand, note also that the energy consumption has a local minima around $\delta = 0.2$. We will also show by comprehensive performance evaluations in Section 4.4 that, this value is a suitable operating point for XLM in terms of energy efficiency.

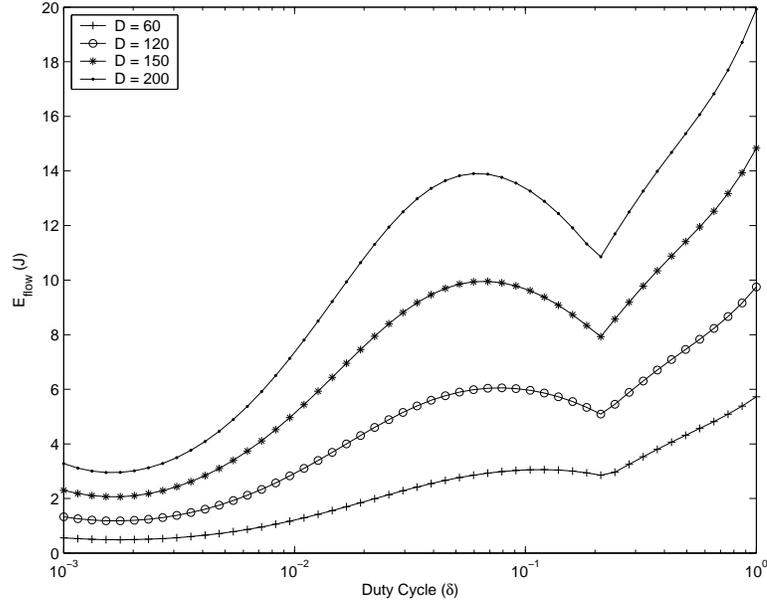


Figure 28: Avg. energy consumption vs. duty cycle for different values of D .

4.4 Performance Evaluation

In order to gain more insight into the protocol operation, we first investigate the effects of XLM protocol parameters on the overall network performance. We, then, present a comparative study between XLM and five different layered protocol suites consisting of state-of-the-art protocols. Finally, we discuss the overall communication complexity for both XLM and the layered protocol suites.

The existing sensor network simulation platforms are not suitable for cross-layer communication suite design due to their layered architecture. For this reason, we

Table 1: Simulation Parameters

Parameter	Value	Parameter	Value
Re-tx. Limit	7	P_t	5 dBm
μ	2	PL(d_0)	55 dB
α	$\lambda_{ii_0}/10$	P_n	-105 dBm
Buffer Length	30	n	3
$l_{control}$	20 bytes	σ	3.8
l_{data}	100 bytes	$T_{coherence}$	16 ms
Frame Length	5s	E_{rx}	13.5 mW
Energy Threshold	100 μ J	E_{tx}	24.75 mW
ξ_{Th}	10 dB	E_{sleep}	15 μ W
T_{sss}	5s		

evaluate XLM and various layered protocol suites in cross-layer simulator (XLS) developed at our laboratory in C++. XLS consists of a realistic channel model and an event-driven simulation engine. We present simulation results for a sensor topology of 300 nodes randomly deployed in a $100 \times 100 m^2$ sensor field. The sink is located at coordinates (80,80). The simulation parameters for both sensor nodes and the communication suites are given in Table 1. In each simulation, an event occurs in an *event area* located at coordinates (20,20) with an event radius of 20m. Each source node reports its event information to the sink. To investigate the effect of duty cycle, each simulation is performed for duty cycle values of $\delta \in [0.1, 1]$. Each simulation lasts for 60s and the results are the average of ten trials for each of ten different random topologies.

In the evaluations, we investigate the following performance metrics:

- *Throughput* is the average number of bits per second received at the sink during the simulations. In calculating this metric, only unique packets are considered since multiple copies of a packet can be received at the sink due to either broadcast nature of some protocols or retransmissions.
- *Goodput* is the ratio between the total number of unique packets received at the sink and the total number of packets generated by all the source nodes. As

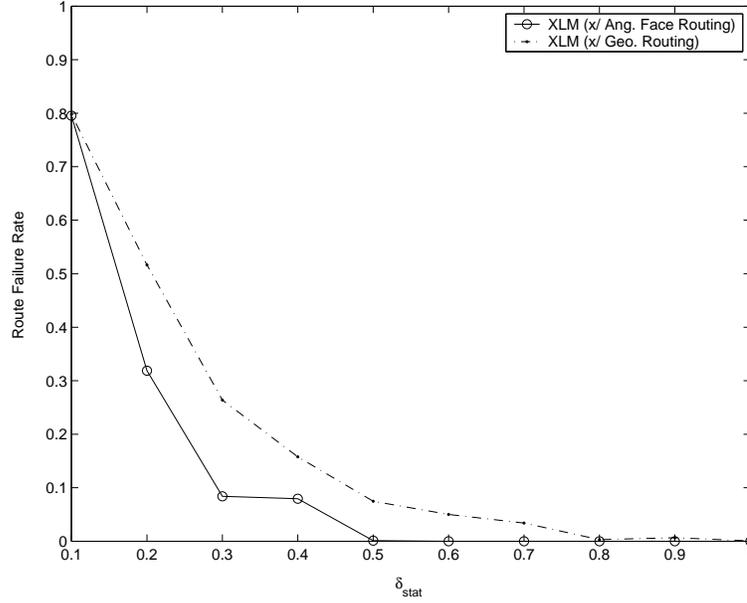


Figure 29: Route Failure Rate for XLM with angle-based routing and default geographical routing.

a result, the overall communication reliability of the suites is investigated.

- *Energy Efficiency* is the most important metric in WSN. In our simulations, we consider the average energy consumption per hop and per unique packet that is received at the sink, which can be considered as the inverse of energy efficiency. Hence, a lower value refers to a more efficient communication.
- *Number of Hops* is the average number of hops each received packet traverses to reach the sink. This metric is used to evaluate the routing performance of each suite.
- *Latency* is the average time it passes between the time a packet is generated at a source node and the time it is received at the sink. This delay accounts for the queuing delay and the contention delay at the nodes as well as specific protocol operation.

4.4.1 XLM Parameters

The parameters that affect the XLM operation are the angle-based routing, *SNR threshold*, ξ_{Th} , and *duty cycle*, δ . We present the effects of these parameters on the XLM performance in this section.

The effect of angle-based routing is shown in Fig. 29, where the route failure rate vs. duty cycle parameter δ_{stat} is shown for XLM with angle-based routing and with only default geographical routing. In these experiments, only a snapshot of the network is considered and the routes are found considering a static topology. The routing failure is the ratio of the number of unsuccessful routes between each nodes in the network and the number of all possible routes. The results show that route failure rate increases as the static duty cycle δ_{stat} is decreased. However, angle-based routing limits the route failure rate to less than 10% for $\delta_{stat} > 0.2$. This leads to up to 70% decrease in failure rate. Note that failure rate of XLM with angle-based routing also increases as δ_{stat} is further decreased since the network becomes partitioned.

In Fig. 30 (a), the total throughput received at the sink is shown. The x -axis shows the duty cycle, δ , and the throughput is shown for different SNR threshold, ξ_{Th} , values. The network throughput increases as the duty cycle, δ , is increased. An increase in the duty cycle increases the number of nodes that are active at a given time. Consequently, the capacity of the network increases. This fact is also evident from our buffer occupancy analysis in Section 4.3.5. The effect of the SNR threshold, ξ_{th} is also shown in Fig. 30 (a). The first curve on the figure, i.e., No ξ_{Th} , is the case where the first condition in (41) is not used by the nodes. In other words, nodes contend for participating in routing irrespective of the received SNR value. It can be observed that, increasing the SNR threshold, ξ_{Th} improves the network throughput upto a certain ξ_{Th} . Above this value, the network throughput degrades. This shows that a very conservative operation of XLM leads to performance degradation.

In Fig. 30 (b), the goodput performance is shown. It can be clearly seen that

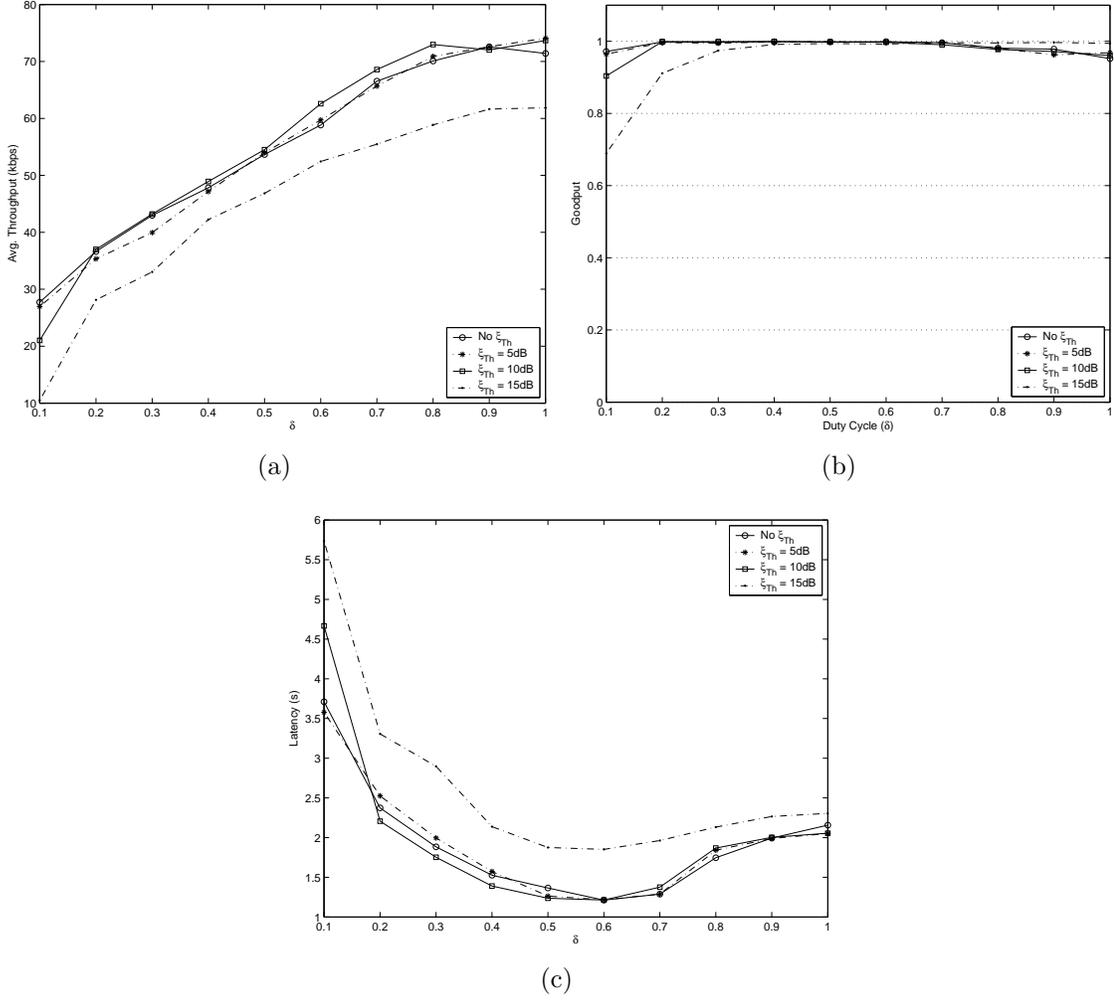


Figure 30: (a) Average throughput, (b) average goodput, and (c) average latency vs. duty cycle for different values of ξ_{Th} .

XLM provides reliability above 95% for $\delta > 0.1$ and $\xi_{Th} \leq 10dB$. The decrease in goodput at $\delta = 0.1$, is due to the fact that the connectivity of the network cannot be maintained since, on the average, only 10% of the nodes in the network are active at a given time. Moreover, for $\xi_{Th} = 15dB$, the goodput decreases up to 0.7 as the duty cycle is decreased. This is due to the fact that potential receivers with the desired channel quality cannot be found and the reliability of the XLM degrades.

In Fig. 30 (c), the end-to-end latency is shown, which reveals that increasing SNR threshold, ξ_{Th} , improves the end-to-end latency performance up to a certain ξ_{Th} value. $\xi_{Th} = 10dB$ results in the lowest latency. Since for all above performance

metrics, $\xi_{Th} = 10dB$ results in the most efficient performance, we use this value in the following evaluations.

4.4.2 Comparative Evaluation

In this section, we compare the performance of the XLM with five different layered protocol architectures. We first identify the protocol configurations implemented in our study along with the implementation issues. Then, we present the results of our comparative evaluation for networks with and without duty cycle operation. The complexity issues are also discussed.

4.4.2.1 Protocol Configurations

The protocol configurations implemented for the comparative evaluation are as follows:

Flooding: This configuration serves as the baseline for the other configurations. In this case, each node broadcasts its packet and the nodes that are closer to the sink rebroadcast this packet until it reaches sink. At the MAC layer, a simple CSMA type broadcast mechanism is used such that each node senses the channel and performs exponential backoff in case the channel is busy. No retransmission mechanism is used. At the transport layer, packets are injected at constant rate and no rate control is used. The results shown include the unique packets received at the sink excluding the duplicate packets.

[GEO]: Geographical Routing + CC-MAC + ESRT: This protocol configuration is composed of ESRT [4], geographical routing [67], and CC-MAC [80] at transport, routing, and MAC layers, respectively. The CC-MAC protocol is implemented using $r_{corr} = 7m$, and $T_{SSS} = 5s$. In the routing protocol [67], *distance-based blacklisting* is used such that, the nodes in the farthest 20% of the radio range are blacklisted and the next closest node to the sink is selected as the next hop.

[PRR]: PRR-based Geographical Routing + CC-MAC + ESRT: This protocol configuration is similar to *GEO* with the exception of the routing algorithm. In this configuration, the routing decisions are based on the channel quality of each node with its neighbors. The channel quality is measured in terms of packet reception rate (PRR) as discussed in [67]. The node that maximizes the *PRR*×*geographical advancement* product is selected as the next hop.

[PRR-SMAC]: PRR-based Geographical Routing + SMAC + ESRT: This protocol configuration is similar to *PRR* with the MAC layer replaced by the SMAC protocol [83]. In this configuration, the duty cycle operation proposed in [83] is implemented instead of the distributed duty cycle operation.

[DD-RMST]: Directed Diffusion + RMST: This case consists of RMST [70], directed diffusion [38] and a simple CSMA scheme. The RMST protocol is implemented with hop-by-hop recovery and caching, and no link-layer ARQ is used at the link layer as presented in [70]. *DD-RMST* is used in the comparative evaluations for operation without duty cycle, i.e., $\delta = 1$.

XLM: Our proposed cross layer module (XLM) is implemented according to the protocol description in Section 4.3 and pseudocode in Fig. 27 with SNR threshold $\xi_{Th} = 10dB$.

It is important to note that the existing protocols that we have implemented in the layered protocol suites are usually proposed considering only their related layers with reasonable assumptions about the other layers. As an example, in the geographical routing protocols [67], each node is assumed to know the locations of their neighbors. However, actual implementation and operation of such an information exchange procedure is important especially when comparing these solutions to the proposed XLM solution. Since the receiver-based approach employed in the XLM does not require such an explicit information exchange, this constitutes a major overhead for the layered protocol suites using such an approach. Moreover, since duty cycle is deployed

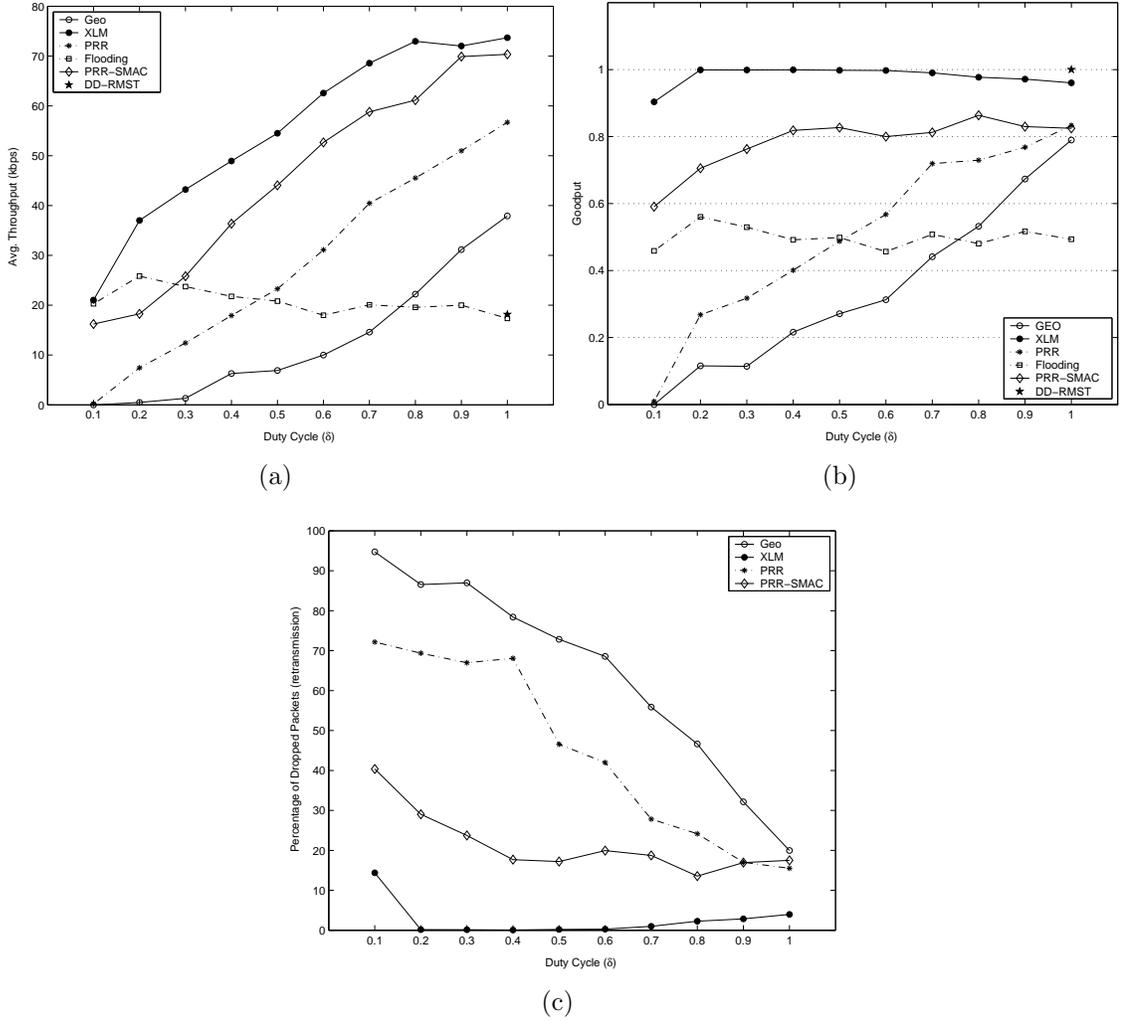


Figure 31: (a) Average throughput, (b) average goodput, and (c) percentage of dropped packets due to retransmission timeout vs. duty cycle for different values of ξ_{Th} .

in our solution, each neighbor of a node may not always be active. Hence, in order for each protocol to work together in the protocol suites, we have made some implementation modifications.

Accordingly, in *GEO*, *PRR*, and *PRR-SMAC*, each node broadcasts a beacon to indicate its position and the remaining time to sleep. This beacon is sent at the beginning of each sleep frame when a node wakes up. Each neighbor that receives this beacon determines that the specific node will be active for the duration specified in the beacon. In the case of *PRR* and *PRR-SMAC*, this beacon also serves as a

channel quality indicator. In order to optimize the network performance, in *GEO* and *PRR*, the beacons are piggybacked if there is a packet in the queue. In *PRR-SMAC*, a pairwise cross-layering is used and the routing beacons are sent with the SYNC packets. Similarly, SYNC packets are piggybacked if there is a packet in the queue.

We have indicated that *DD-RMST* is used only for operation without duty cycle, i.e., $\delta = 1$. This decision is due to the fact that neither directed diffusion nor RMST considers duty cycle operation [38], [70]. Therefore, the *DD-RMST* protocol configuration is evaluated only for $\delta = 1$ for fairness and completeness of the evaluations.

We next present the results for operation with duty cycle, by changing the duty cycle δ from 0.1 to 1 in Section 4.4.2.2. Since *DD-RMST* is only considered for operation without duty cycle, the performance metrics corresponding to this configuration are shown as a single point at $\delta = 1$ in the figures.

4.4.2.2 Results

In Fig. 31 (a), the throughput comparison for XLM and the layered protocol suites is presented. The throughput achieved by XLM is significantly higher than that of the other five layered protocol suites (more than 80% improvement for $\delta = 0.3$). This shows the clear advantage of using a cross-layer approach. In the layered protocol suites, the cross-layer information is not efficiently exploited for each functionality. For example, in *PRR* and *PRR-SMAC*, route selection is only performed based on location information and link quality, whereas the congestion level at a specific node is not considered. Another important result emerges in the comparison between XLM and *PRR-SMAC*. *GEO* and *PRR* use CC-MAC at the MAC layer. CC-MAC results in smaller number of nodes sending information in an event area as a representative of all the nodes in that area [80]. XLM also exploits this spatial correlation for medium access. However, *SMAC* [83] does not exploit this property and all the

nodes inside an event area sends information to the sink. This results in almost 3-times increase in the number of source nodes. The higher throughput value of *PRR-SMAC* compared to *GEO* and *PRR* can be explained with this fact. However, XLM still outperforms *PRR-SMAC* in terms of total throughput although less number of nodes send information, which shows that the network capacity is exploited in a more efficient manner. This also results in higher temporal resolution at the sink since the representative nodes send information at a higher rate.

Note that the total throughput achieved by the *DD-RMST* is significantly lower than XLM, *PRR*, and *GEO* and comparable to *Flooding*. This is due to two main reasons. The first reason is the additional traffic created for recovering lost packets. This additional traffic both increases the contention in the wireless channel and decreases the capacity of the network. The second reason for throughput degradation is the control packets of directed diffusion. Especially, the interest and exploratory packets constitute a significant amount of traffic due to their broadcast nature. Consequently, these two types of additional traffic significantly decrease the throughput of *DD-RMST*.

The goodput of the communication suites are shown in Fig. 31 (b). Irrespective of the duty cycle value, δ , XLM provides very high reliability. The cross-layer communication paradigm of the XLM that is adaptive to the network topology enables such high performance even when the network operates at low duty cycle. Coupled with the high throughput of XLM as shown in Fig. 31 (a), our cross-layer approach enables highly efficient communication. Moreover, *DD-RMST* provides 100% reliability while XLM results in a reliability of 96% for operation without duty cycle, i.e., $\delta = 1$. Note that RMST protocol uses hop-by-hop recovery with negative acknowledgments to request missing packets. On the other hand, XLM aims to first prevent link losses by constructing non-congested, high quality paths and then ensures high reliability by hop-by-hop ARQ technique. This approach results in reliability comparable to

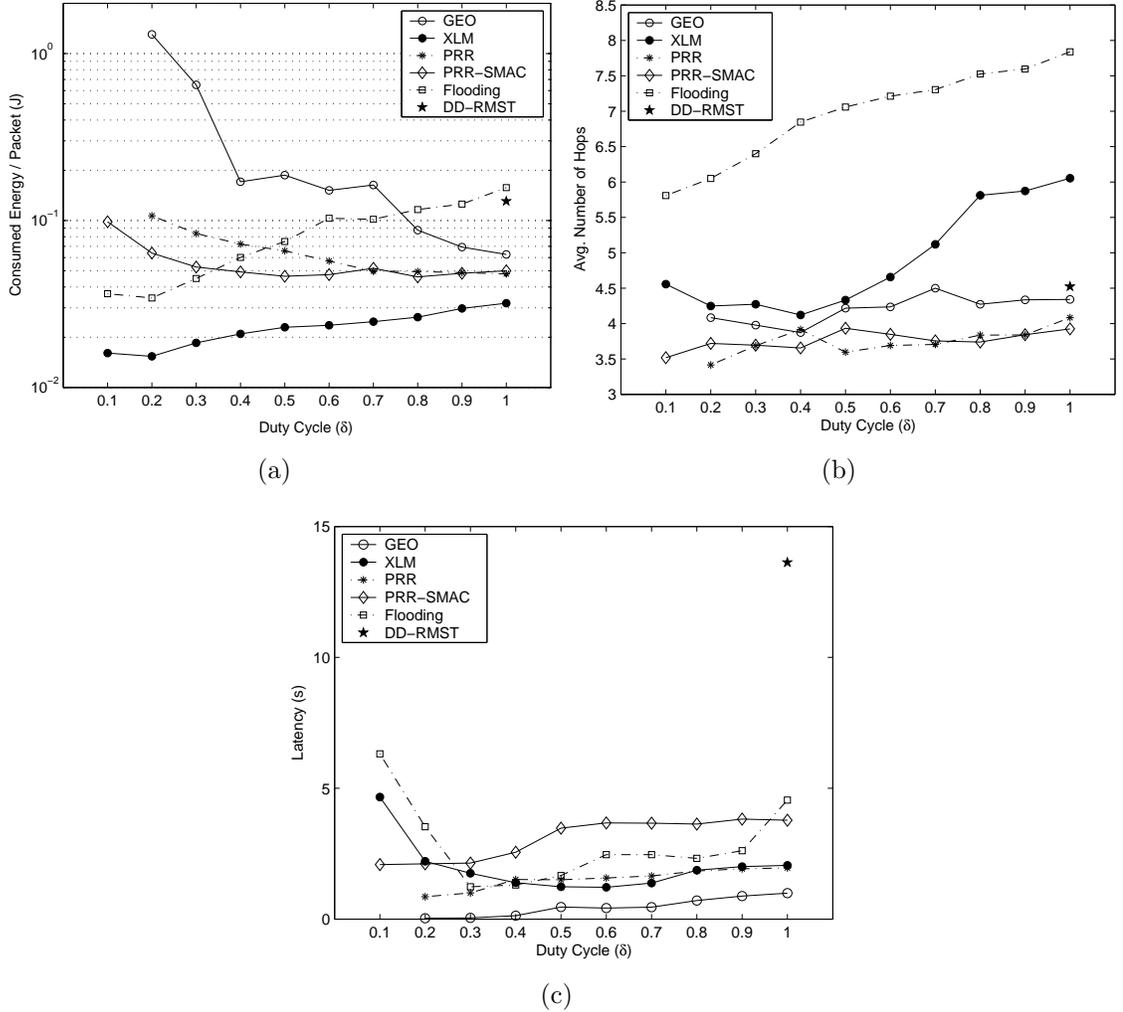


Figure 32: (a) Average energy consumption per packet, (b) average hop count, and (c) average latency vs. duty cycle for layered protocol suites and XLM.

RMST at a significantly lower cost as we will discuss next.

The decrease in reliability for the other layered protocol suites is mainly because of the significant number of packet drops due to retransmission timeouts as shown in Fig. 31 (c)⁵. This suggests that nodes cannot find their intended next hops due to either low channel quality or because the nodes switch to sleep state before receiving any packets. This is exacerbated especially in the case of low duty cycle. As a result, the reliability of the network is hampered significantly.

⁵Since *Flooding* and *DD-RMST* does not deploy ARQ, they are not included in Fig. 31 (c).

In Fig. 32 (a), the energy consumption per packet is shown. In Fig. 32 (a), the values for *GEO* and *PRR* at $\delta = 0.1$ are not shown since no packets are received by the sink. It can be seen that XLM consumes significantly less energy per packet and hence is highly energy efficient when compared to the other layered protocol suites. This difference is mainly because of the periodic broadcast of beacon packets in *GEO* and *PRR*, and SYNC packets in *PRR-SMAC*. Furthermore, the significant percentage of retransmission timeouts as shown in Fig. 31 (c) indicate significant energy wastage due to packets that cannot be transmitted to the sink. Since the network and MAC layers operate independently, the nodes chosen by the routing layer cannot be reached and significant energy consumption occurs.

An interesting result is the significantly low energy efficiency of *DD-RMST*. Although this configuration provides 100% reliability as shown in Fig. 31 (b), the layered structure of the routing, transport and MAC functionalities results in a high penalty. As explained before, the routing layer, i.e., directed diffusion, incurs significant amount of overhead in order to maintain end-to-end paths between sources and the destination. On the contrary, XLM employs an adaptive routing technique that provides an energy efficient path in terms of both link quality and energy consumption distribution. Another important observation from Fig. 32 (a) is that the energy consumption per packet for XLM has a minimum at $\delta = 0.2$. This is consistent with the mathematical analysis provided in Section 4.3.6. Hence, we observe that the duty cycle value of $\delta = 0.2$ provides the most energy efficient performance for the operation of XLM.

On the other hand, the advantages of using a separate routing layer in the layered protocol suites can be seen from Fig. 32 (b), where the average hop count is shown. *GEO*, *PRR*, *PRR-SMAC*, and *DD-RMST* result in less number of hops for the packets that reach the sink than XLM. This is due to the fact that the routing algorithms in these layered protocol suites aim to find the smallest number of hops. While

this result may be incorporated as a disadvantage when only routing layer is taken into account, the overall performance of XLM reveals that, routing layer performance alone does not provide efficient communication in WSN. In other words, while smaller number of hops might seem optimal in terms of routing efficiency, other effects such as link quality, contention level, congestion level, and overall energy consumption, necessitate a cross-layer approach in route selection for overall efficiency.

Furthermore, as shown in Fig. 32 (c), XLM incurs an end-to-end latency comparable to *PRR*. *GEO* results in smaller end-to-end delay since the routing is performed based only on geographical location. On the other hand, *PRR-SMAC* results in higher end-to-end latency due to the clustered scheduling of nodes. Fig. 32 (c) also clearly shows the tradeoff of *DD-RMST* in achieving high reliability. This configuration results in significantly high latency values when compared to the other configurations.

The end-to-end latency for *Flooding* is significantly higher for the limiting cases, i.e., $\delta = 1$ and $\delta \leq 0.2$. When all the nodes are active, flooding causes significant amount of contention and congestion leading to higher buffer occupancy time for each packet at each hop leading to higher latency. On the other hand, when the duty cycle is low, each time a node receives a packet, it has to go through one duty cycle before it can re-broadcast the packet. This, in turn, increases the end-to-end latency. Similarly, the end-to-end latency of XLM increases for low δ . The reason for this increase is evident from Fig. 31 (c). Note that for $\delta = 0.1$, 14% of the transmitted packets are dropped due to retransmission timeout. This is due to the fact that, sender nodes cannot find any neighbors that satisfy the constraints in (41) discussed in Section 4.3. As a result, the end-to-end latency increases due to retransmissions.

4.4.2.3 Implementation Complexity

In addition to the performance of our XLM module in terms of network metrics, the complexity and implementation issues of cross-layer design are also important. In

this section, we provide a qualitative comparison of cross-layer design and layered protocol architectures implemented in our simulation environment.

One of the major advantages of cross-layer design for communication protocols is the implementation efficiency. In a traditional layered protocol architecture, each layer has clear boundaries. This layered structure leads to computation delays due to the sequential handling of a packet. For example, in TinyOS [2], each layer has to wait for the lower layers to process the packet since a single buffer is used for a packet for all layers. XLM, however, melts the functionalities of traditional transport, routing, and MAC layers into a unified cross-layer communication module by considering physical layer and channel effects as shown in Fig. 27. Hence, these functionalities are performed as a whole and overall protocol efficiency can be improved using this module.

In addition to the simulation performance, the implementation issues are also important for a complete comparison. As explained in Section 4.3, XLM does not require any tables or extra buffer space for routing and transport layer functionalities. The routing is performed based on receiver initiatives which eliminates the need for a routing table at each node. As shown in Fig. 27, the implementation of XLM is both simple and compact. On the other hand, in *PRR-SMAC*, SMAC protocol maintains schedule table for each of one-hop neighbors to provide synchronized sleeping cycles. Similarly, in *DD-RMST*, at the routing layer, each node has to implement reinforcement table for each source indicating the next hop in the reinforced path. In case a node is a source node, it also has to keep track of multiple neighbors which has a path to the sink for exploratory messages. At the transport layer, RMST [70] requires a separate queue to cache data locally to support loss recovery at all hops. These requirements, due to either layered operation of the protocol stack or the internal protocol structure at each layer, places burden in memory space for communication in

sensor nodes. This extra space required by the communication stack limits the available space to develop new applications for sensor networks. On the other hand, the careful use of code space and cross-layer implementation of communication functionalities in XLM provides a much efficient operation in WSN. When coupled with the noticeably better communication performance as discussed in Section 4.4.2.2, XLM becomes a successful candidate for communication protocols in WSN.

CHAPTER V

CROSS-LAYER ANALYSIS OF ERROR CONTROL IN WIRELESS SENSOR NETWORKS

The cross layer module (XLM) framework, which is explained in Chapter 4, realizes energy efficient communication through cross-layer design of transport, routing and MAC functionalities. Another important factor in wireless communications is the error control technique for reliable communication. In this chapter, we develop a cross-layer analysis of error control schemes based on the cross-layer framework developed in Chapter 4. More specifically, the effects of multi-hop routing and the broadcast nature of the wireless communication are investigated to derive the equations governing the energy consumption, latency and packet error rate (PER) performance of error control schemes. This analysis enables a comprehensive comparison of forward error correction (FEC) and automatic repeat request (ARQ) schemes in WSN. The cross-layer analysis framework was first introduced in [79]. The rest of this chapter is organized as follows: The motivation for this chapter is provided in Section 5.1. In Section 5.2, an overview of previous analysis on error control schemes in WSN is provided. Our approach and the system model for cross-layer analysis are explained in Section 5.4. In Section 5.5, the cross-layer analysis of ARQ, FEC, and hybrid ARQ schemes is presented. The numerical evaluations are explained in Section 5.6 along with their implications on the tradeoffs of error control schemes.

5.1 Motivation

Wireless Sensor Networks (WSN) are characterized by collaborative information transmission from multiple sensor nodes observing a physical phenomenon [7]. Severe

energy constraints of battery-powered sensor nodes necessitate energy-efficient communication protocols in order to fulfill application objectives. Moreover, the low power communication constraints of sensor nodes exacerbate the effects of the wireless channel leading to error-prone links. In WSN, where correlation between sensors can be exploited in terms of aggregation, collaborative source coding, or correlation-based protocols, energy efficient error control is of extreme importance. Since these techniques aim to reduce the redundancy in the traffic, it is essential for each packet to be transmitted reliably. Moreover, the strict energy consumption requirements, the multi-hop structure of the WSN, and the broadcast nature of the wireless channel necessitate a cross-layer investigation of the effects of error control schemes.

In this chapter, a cross-layer analysis of error control schemes is presented. More specifically, the effects of multi-hop routing and the broadcast nature of the wireless communication are investigated to derive the equations governing the energy consumption, latency and packet error rate (PER) performance of error control schemes. As a result, a cross layer analysis considering routing, medium access and physical layers is devised. This analysis enables a comprehensive comparison of forward error correction (FEC), automatic repeat request (ARQ), as well as hybrid ARQ schemes in WSN. So far, the performance of FEC codes have been investigated in a point-to-point fashion [46, 62, 66]. To the best of our knowledge, this is the first work that considers both the broadcast wireless channel and the multi-hop structure of WSN with realistic channel models and a 2 dimensional topology. Moreover, a practical comparison of these schemes is provided by considering two major hardware architectures for WSN, i.e., Mica2 [22] and MicaZ [23] nodes. It should be emphasized that in this work, we do not propose a new FEC code for WSN. Rather, we devise a framework to assess the performance of FEC, ARQ, and hybrid ARQ schemes. Furthermore, our goal is to indicate the situations where one of the error control schemes should be favored.

5.2 *Related Work*

Although there have been several studies on error control techniques in wireless networks and especially in cellular networks, none of them are directly applicable to the WSN. Especially the limited energy consumption requirements of WSN and the low complexity in the sensor hardware necessitate energy efficient error control and prevent high complexity codes to be deployed. Recently, there have been some work that considers the energy consumption analysis of error control techniques in WSN.

In [68], the energy consumption profile of convolutional codes has been presented based on a specific sensor node architecture, i.e., μ AMPS node. Although FEC codes have been shown to provide flexible error control capabilities over high variety of ranges between nodes, such an advantage is limited in scenarios where limited error probabilities are acceptable. More specifically, for convolutional codes, no coding provides better energy efficiency for probability of bit error, $P_b > 10^{-5}$ [68]. This is due to the fact that at high P_b , the transceiver energy dominates the overall energy consumption. Since the packet length is increased due to coding, overall energy consumption increases. Similarly, in [62], the energy efficiency of convolutional codes is compared to the energy efficiency of BCH codes in a framework to optimize the packet size in WSN. The results of this work reveal that BCH codes outperform the most energy efficient convolutional code by almost %15. This is due to the high energy consumption of Viterbi decoding that is required for decoding of convolutional codes. Consequently, we do not consider convolutional codes in our work due to their energy inefficiency.

In [66], an analysis of different modulation schemes and two BCH codes is presented based on their energy consumption efficiency. More specifically, the effects of signal to noise ratio (SNR), 2-ary and M-ary modulation schemes, and two BCH codes on the energy consumption of a sensor node are investigated. However, in this analysis, the energy consumption for transmitting redundant bits is considered as

the only overhead of error control coding without considering the decoding energy. Furthermore, a single-hop WSN is considered, which decouples multi-hop routing and the effects of error control codes.

In [40], the effect of error control coding on the energy consumption of multihop WSN is studied. However, this analysis considers a linear topology, where the distances between each hop are fixed and equal. Moreover, for each link, the probability of error is assumed to be the same. Consequently, the fading effects of the wireless channel and the random route construction cannot be captured with the presented framework in [40]. Furthermore, the end-to-end latency has never been considered in the context of FEC codes in WSN before.

5.3 Error Control in Wireless Sensor Networks

In WSN, where correlation between sensors can be exploited in terms of aggregation, collaborative source coding, or correlation-based protocols, error control is of extreme importance. Since the abovementioned techniques aim to reduce the redundancy in the traffic by filtering correlated data, it is essential for each packet to be transmitted reliably. In general, the error control mechanisms in communication networks can be categorized into three main approaches; automatic repeat request (ARQ), forward error correction (FEC), and hybrid ARQ.

- *Automatic Repeat Request (ARQ)*: ARQ-based error control mainly depends on the retransmission for the recovery of the lost data packets/frames. ARQ protocols enable transmissions of failed packets by sending explicit acknowledges upon reception and detection of missing acknowledgements. It is clear that ARQ error control mechanisms incur significant additional retransmission cost and overhead in case of errors. On the other hand, in the case of good channel quality, overhead of the ARQ protocols is low since a packet is sent with minimum overhead compared to FEC schemes. The efficiency of ARQ in sensor

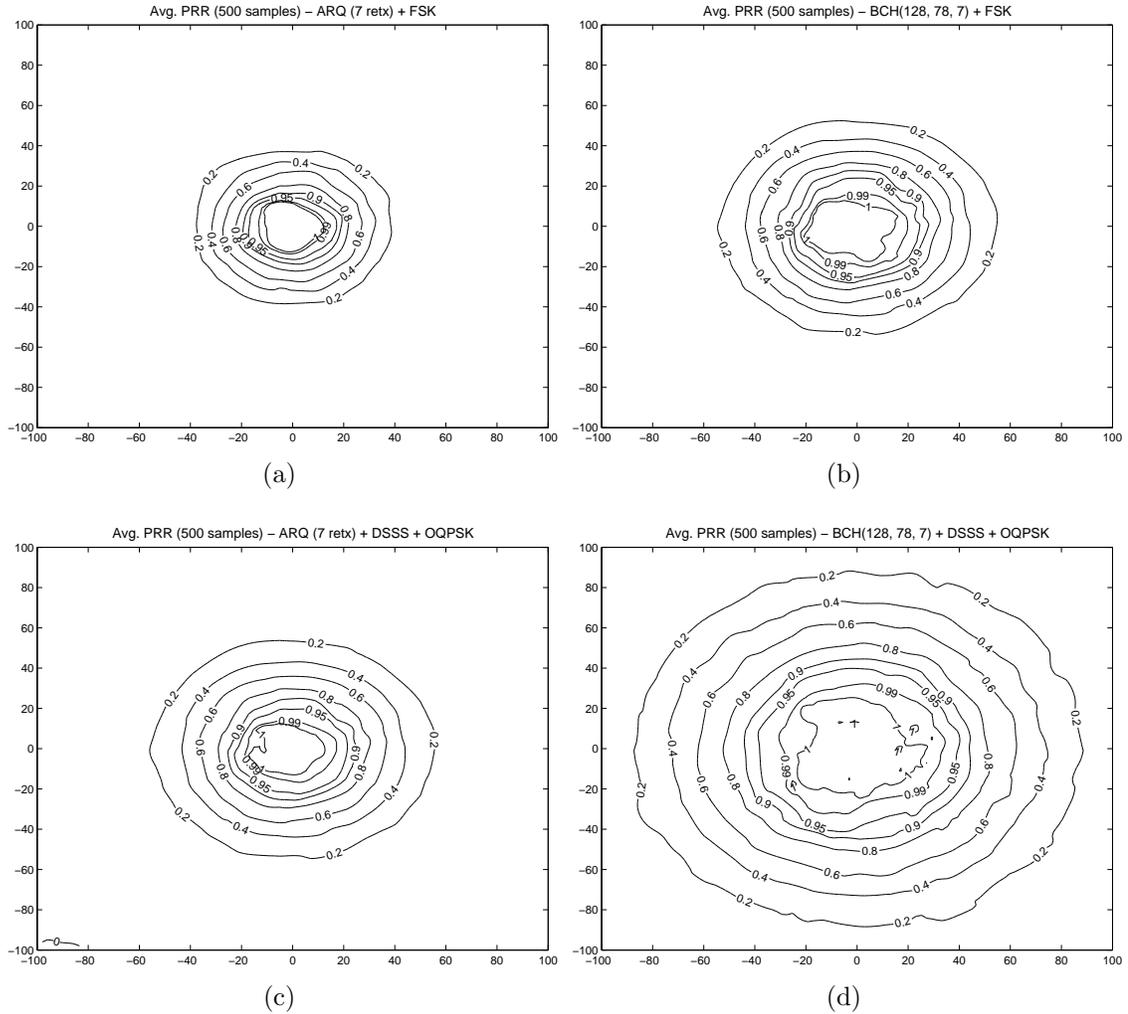


Figure 33: 2D average packet reception rate graphs for (a) ARQ (FSK), (b) FEC (FSK), (c) ARQ (DSSS/OQPSK), (d) FEC (DSSS/OQPSK).

network applications is limited due to the scarcity of the energy and processing resources of the sensor nodes.

- *Forward Error Correction (FEC)*: FEC adds redundancy to the transmitted packet such that it can be received at the receiver error-free even if the limited number of bits are received in error. There exist various FEC codes that are optimized for specific packet sizes, channel condition, and reliability such as linear block codes (BCH and Reed-Solomon (RS) codes) as well as convolutional

codes. On the other hand, for the design of efficient FEC schemes, it is important to have good knowledge of the channel characteristics and implementation techniques.

- *Hybrid ARQ (HARQ)*: Hybrid ARQ schemes aim to exploit the advantages of both FEC and ARQ schemes by incrementally increasing the error resiliency of a packet through retransmissions. Mainly, two types of HARQ schemes exist: Type I and Type II. With HARQ-I techniques, first an uncoded packet or a packet coded with a lower error correcting capability is sent. If this packet is received in errors, the receiver sends a negative acknowledgement (NACK) to the sender, which re-sends the packet coded with a more powerful FEC code. The difference in Type II is that for retransmissions, only the redundant bits are sent. While Type II decreases the bandwidth usage of the protocol, Type I does not require the previously sent packets be stored.

Forward error control (FEC) coding and hybrid ARQ schemes improve the error resiliency compared to ARQ schemes by sending redundant bits through the wireless channel. Therefore, lower signal to noise ratio (SNR) values can be supported to achieve the same error rate as an uncoded transmission. This advantage can be exploited in two ways in wireless networks:

- *Transmit Power Control*: The improved error resiliency provided through FEC codes has generally been exploited by reducing the transmit power in cellular networks. This technique, which we refer to as *transmit power control*, improves the capacity of cellular networks by reducing the interference to other users.
- *Hop Length Extension*: In multi-hop networks, the advantage of FEC coding can also be exploited by improving the transmission range of a node. This is illustrated in Figs. 33, where the packet error rates of ARQ and FEC codes for a packet of 38 bytes around a transmitter node is shown. In Fig. 33(a) and

Fig. 33(b), the packet error rates are shown for frequency shift keying (FSK) modulation, which is used in Mica2 nodes, while Fig. 33(c) and Fig. 33(d) show direct sequence spread spectrum offset quadrature phase shift keying (DSSS/O-QPSK), which is used in MicaZ nodes. It is clear that, FEC codes increase the transmission range of a node, which can be exploited to construct longer hops in a multi-hop network. We refer to this technique as *hop length extension*, which can be achieved through channel-aware cross-layer routing protocols.

In the following, we investigate the tradeoffs between ARQ, FEC, and hybrid ARQ schemes in terms of energy consumption, latency and end-to-end PER considering the transmit power control and hop length extension techniques to exploit the error resiliency of FEC techniques.

5.4 Analysis Approach and System Model

In our analysis, we consider a network composed of sensor nodes that are distributed according to a 2-D Poisson distribution with density ρ . Duty cycle operation is deployed such that each node is active for δ fraction of the time and is in sleep mode otherwise [9]. Moreover, we consider a monitoring application such that the reporting rates of sensors are low but the messages should be transmitted reliably.

In order to realize hop length extension, we consider a channel-aware routing algorithm. In this algorithm, the next hop is determined according to the received signal to noise ratio (SNR) of a packet sent from a specific node i at a distance D from the sink. Among the neighbors of i , the neighbor, j , that is closest to the sink and with SNR value, $\psi_j > \psi_{Th}$ is selected as the next hop, where ψ_{Th} is the received SNR threshold. Note that this approach can be implemented using a cross-layer approach as in [9] or through signaling [67]. The medium access is performed through RTS-CTS-DATA exchange in addition to ACK and retransmissions for ARQ and NACK and retransmissions for hybrid ARQ.

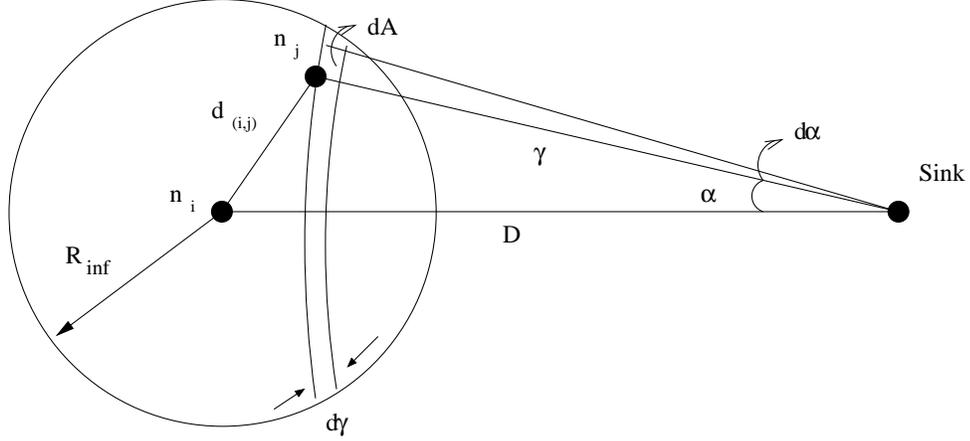


Figure 34: Reference model for the derivations.

Accordingly, first, the expected hop distance is derived as a function of the network parameters. Then, the end-to-end energy consumption, latency, and packet error rate (PER) of a single flow is derived. We use the model shown in Fig. 34 and the log-normal shadow fading channel model [88] for our derivations. Note that in such a model, the transmission range of a node is essentially infinite due to the shadow fading component. In our analysis, we approximate the transmission range of a node to R_{inf} , which is the distance at which the probability that a packet can be successfully received is negligible. Moreover, the hop distance at each hop is considered independent since duty cycle operation is performed. As a result, the state of the network will change at each hop since different nodes will be awake at different time instants. In [76], duty cycle operation is not considered and hence, a different approach has been taken. Note that similar energy consumption analysis has been performed in a node-centric manner for routing algorithms in [69], [87]. However, the effect of neighbor nodes and the effect of routing decisions has not been investigated in this context yet. In our analysis, we investigate the energy consumption to transmit a single packet to the sink with the effect of neighbor nodes and wireless channel effects, which provides a clearer insight into the energy consumption.

Our cross-layer analysis framework enables comprehensive comparison of ARQ,

FEC, and hybrid ARQ schemes. To illustrate specific results, we consider block codes because of their energy efficiency and lower complexity compared to convolutional codes [62, 68]. We consider a block code, which is represented by (n, k, t) , where n is the block length, k is the payload length, and t is the error correcting capability in bits. In our analysis, we use $(128, 50, 13)$, $(128, 78, 7)$, and $(128, 106, 3)$ extended BCH codes, which enable the evaluation of the effect of error detection capability, t , by fixing the block length, n . Furthermore, we use the $(7, 3, 2)$, $(15, 9, 3)$, and $(31, 19, 6)$ RS codes. Other FEC schemes can also be used in our framework. For the hybrid ARQ schemes, we consider three different configurations. For type-I hybrid ARQ (HARQ-I), we consider the case where first an uncoded packet is sent followed by a packet encoded by BCH codes. Furthermore, a combination of two BCH codes is also considered for the HARQ-I. The type-II hybrid ARQ schemes necessitate incremental error control coding since only the difference is sent through the retransmissions. This is generally accomplished through punctured codes [82]. Since this procedure is energy consuming, we consider the case where the packet is encoded by a BCH code and then the payload is sent first. The receiver uses only the CRC bits in the payload for decoding. In case of errors, the receiver sends a NACK packet and the transmitter sends the redundant bits for BCH decoding. We consider only one retransmission is performed for the hybrid ARQ schemes. Throughout the chapter, the hybrid ARQ schemes are indicated by a tuple (t_1, t_2) , where the first parameter, t_1 indicates the error correction capability in the first packet and t_2 is that in the second packet. As an example, HARQ-II $(0, 3)$ refers to type-II hybrid ARQ, where first an uncoded packet is sent, i.e., $t_1 = 0$, and in case retransmission is requested, the redundant bits of the BCH $(128, 106, 3)$ encoded packet is sent.

5.5 Cross-Layer Analysis

In this section, we first derive the expected hop distance by considering the effects of broadcast channel and the routing algorithm discussed in Section 5.4. The results of this analysis are then used to analyze the end-to-end energy consumption, latency, and packet error rate (PER) of FEC, ARQ, and hybrid ARQ schemes. The total energy consumed as a result of a single flow from a source node at distance D from the sink can be found as

$$E_{flow}(D) = E[E_h] E[n_h(D)] , \quad (55)$$

where $E[E_h]$ is the expected energy consumption per hop and $E[n_h(D)]$ is the expected hop count from a source at distance D to the sink. Similarly, the end-to-end latency of a flow is given by

$$T_{flow}(D) = E[T_h] E[n_h(D)] , \quad (56)$$

where $E[T_h]$ is the expected delay per hop.

A good approximation for the expected hop count is given in [87] as

$$E[n_h(D)] \simeq \frac{D - R_{inf}}{E[d_h]} + 1 , \quad (57)$$

where $E[d_h]$ is the expected hop distance. In Sections 5.5.1, 5.5.2, and 5.5.3, we derive the expressions for the expected hop length, $E[d_h]$, the expected energy consumption per hop, $E[E_h]$, and the expected latency per hop, $E[T_h]$, respectively.

5.5.1 The Expected Hop Distance

Consider a node j at coordinates (γ, α) with respect to the sink as shown in Fig. 34. The distance from node j to node i is, hence, given by

$$d_{(i,j)} = d(D, \gamma, \alpha) = \sqrt{\gamma^2 + D^2 - 2\gamma D \cos \alpha} . \quad (58)$$

The expected hop distance, $E[d_h]$, can be found as

$$E[d_h] = \int_{\gamma_{min}}^D \int_{-\alpha_\gamma}^{\alpha_\gamma} d_{(i,j)} dP\{\mathcal{N}_i = j\} , \quad (59)$$

where $\gamma_{min} = D - R_{inf}$, $d_{(i,j)}$ is the distance between nodes i and j as given by (58), $dP\{\mathcal{N}_i = j\}$ is the probability that node j is selected as the next hop, and $\alpha_\gamma = \text{acos}[(\gamma^2 + D^2 - R_{inf}^2)/(2\gamma D)]$.

In order for node j to be selected as the next hop, first, the received SNR, ψ_k , at each node, k , that is closer to the sink than node j should satisfy $\psi_k < \psi_{Th}$. Moreover, the received SNR of node j should satisfy, $\psi_j > \psi_{Th}$. The probability that node j is selected as the next hop is, hence, given by

$$dP\{\mathcal{N}_i = j\} = P\{N_{A(d\gamma)} = 1\}P\{\psi_j > \psi_{Th}\}P\{d_{(j,s)} \geq \gamma\}, \quad (60)$$

where $N_{A(d\gamma)}$ is the number of nodes in the area, dA , at distance γ from the sink, $P\{\psi_j > \psi_{Th}\}$ is the probability that the received SNR of a node j is above ψ_{Th} , and $P\{d_{(j,s)} \geq \gamma\}$ is the probability that the next hop is at least at a distance γ from the sink, s . $P\{N_{A(d\gamma)} = 1\}$ can be approximated by

$$\begin{aligned} P\{N_{A(d\gamma)} = 1\} &\simeq 1 - e^{-\rho\delta\gamma d\gamma d\alpha} \text{ as } d\gamma \rightarrow 0 \\ &\simeq \rho\delta\gamma d\gamma d\alpha, \end{aligned} \quad (61)$$

where we use the approximation $e^{-x} \simeq 1 - x$ for the last step since $(\rho\delta\gamma d\gamma d\alpha) \rightarrow 0$ as $d\gamma \rightarrow 0, d\alpha \rightarrow 0$.

For the calculation of $P\{\psi_j > \psi_{Th}\}$ and $P\{d_{(j,s)} \geq \gamma\}$, we first introduce the log-normal channel model [88], where the received power at a receiver at distance d from a transmitter is given by

$$P_r(d) = P_t - PL(d_0) - 10\eta \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma, \quad (62)$$

where P_t is the transmit power in dBm, $PL(d_0)$ is the path loss at a reference distance d_0 in dB, η is the path loss exponent, and X_σ is the shadow fading component, with $X_\sigma \sim \mathcal{N}(0, \sigma)$. Moreover, the signal to noise ratio (SNR) at the receiver is given by $\psi(d) = P_r(d) - P_n$ in dB, where P_n is the noise power in dBm.

Considering the shadow fading component, X_σ , the probability that ψ_j , is above some threshold, ψ_{Th} , is

$$\begin{aligned} \text{P}\{\psi_j > \psi_{Th}\} &= \text{P}\{X_\sigma > \beta(d_{(i,j)}, \psi_{Th})\}, \\ &= Q\left(\frac{\beta(d_{(i,j)}, \psi_{Th})}{\sigma}\right), \end{aligned} \quad (63)$$

where

$$\beta(d, \psi_{Th}) = \psi_{Th} + P_n - P_t + PL(d_0) + 10\eta \log_{10}\left(\frac{d}{d_0}\right) \quad (64)$$

and $Q(x) = 1/\sqrt{2\pi}(\int_x^\infty e^{-t^2/2})dt$.

According to the channel model above, by denoting the area that consists of nodes that are closer to the sink than node j as $A(\gamma)$, $\text{P}\{d_{(j,s)} \geq \gamma\}$ can be found as

$$\begin{aligned} \text{P}\{d_{(j,s)} \geq \gamma\} &= \sum_{i=0}^{\infty} \text{P}\{N_{A(\gamma)} = i\} p_k^i, \\ &= \sum_{i=0}^{\infty} \frac{e^{-M} M^i}{i!} p_k^i \\ &= e^{-M(1-p_k)}, \end{aligned} \quad (65)$$

where $N_{A(\gamma)}$ is the number of nodes in $A(\gamma)$, $M = \rho\delta A(\gamma)$, and $A(\gamma)$ is the area of intersection of two circles with centers separated by D and with radii R_{inf} and γ , respectively. Moreover, $p_k = \text{P}\{\psi_k \leq \psi_{Th}, k \in A(\gamma)\}$ is the probability that for a node k in $A(\gamma)$ the received SNR $\psi_k \leq \psi_{Th}$, which is given by

$$p_k = \int_{\gamma_{min}}^{\gamma} \int_{-\alpha\gamma}^{\alpha\gamma} \left[1 - Q\left(\frac{\beta}{\sigma}\right)\right] \frac{1}{A(\gamma)} d\alpha d\gamma. \quad (66)$$

Using (60), (61), (63), (65), and (66) in (59), the expected hop distance can be calculated as follows:

$$\text{E}[d_h] = \rho\delta \int_{\gamma_{min}}^D \int_{-\alpha\gamma}^{\alpha\gamma} \gamma d_{(i,j)} Q\left(\frac{\beta}{\sigma}\right) e^{-M(1-p_k)} d\alpha d\gamma, \quad (67)$$

which will be used for energy consumption and latency analysis of FEC, ARQ, and hybrid ARQ schemes according to (55), (56), and (57).

5.5.2 Energy Consumption Analysis

The expected energy consumption and latency per hop are also calculated by considering a node j as shown in Fig. 34. We first derive the expected energy consumption per hop and present the expected latency per hop accordingly. Denoting the expected energy consumption of node j by $E[E_j]$ and using (67), the expected energy consumption per hop can be calculated as

$$E[E_h] = \rho\delta \int_{\gamma_{min}}^D \int_{-\alpha\gamma}^{\alpha\gamma} \gamma E[E_j] Q\left(\frac{\beta}{\sigma}\right) e^{-M(1-p_k)} d\alpha d\gamma . \quad (68)$$

Since a node can become a next hop if its received SNR value is above a certain threshold, the expected energy consumption, $E[E_j]$, can be found as

$$E[E_j] = \int_{\psi_{Th}}^{\infty} E_{comm}(\psi, d_{(i,j)}) f_{\Psi}(\psi, d_{(i,j)}) d\psi , \quad (69)$$

where $E_{comm}(\psi, d_{(i,j)})$ is the energy consumption for communication between nodes i and j given that they are at a distance $d_{(i,j)}$ and the SNR value at node j is ψ . Moreover, $f_{\Psi}(\cdot)$ is the pdf of the SNR. Since, $P(\Psi \leq \psi) = P(X_{\sigma} \leq \beta(\psi, d_{(i,j)}))$, $f_{\Psi}(\cdot)$ is found as:

$$f_{\Psi}(\psi, d_{(i,j)}) = f_{X_{\sigma}}(\beta(\psi, d_{(i,j)})) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\beta^2}{2\sigma^2}} . \quad (70)$$

The first component, $E_{comm}(\psi, d_{(i,j)})$, in (69) is the energy consumption to transmit a packet between two nodes at a distance $d_{(i,j)}$ with received SNR ψ . $E_{comm}(\psi, d_{(i,j)})$ has three components as given by¹

$$E_{comm} = E_{TX} + E_{RX} + E_{neigh} , \quad (71)$$

where E_{TX} is the energy consumed by the node transmitting the packet (node i), E_{RX} is the energy consumed by the node receiving the packet (node j), and E_{neigh} is the energy consumed by the neighbor nodes.

¹We drop the indices ψ and $d_{(i,j)}$ for ease of illustration.

In order to successfully transmit the packet, a node needs to complete the four-way RTS-CTS-DATA-ACK handshake for ARQ, three-way RTS-CTS-DATA handshake for FEC codes, and RTS-CTS-DATA-NACK exchange for hybrid ARQ. We denote the probability that a data and a control packet is successfully received at distance $d_{(i,j)}$ by p_s^D and p_s^C , respectively². Due to the low traffic load, we assume that collisions are avoided through control packets and the probability of RTS collisions is negligible. Accordingly, E_{TX} is given as follows:

$$E_{TX}^{ARQ} = \frac{1}{(p_s^C)^3 p_s^D} \left\{ E_{sense} + (p_s^C)^2 [E_{tx}^R + E_{rx}^C] + \left(1 - (p_s^C)^2\right) E_{t/o}^C + (p_s^C)^3 p_s^D (E_{tx}^D + E_{rx}^A) + (p_s^C)^3 (1 - p_s^C p_s^D) E_{t/o}^A \right\}, \quad (72)$$

$$E_{TX}^{FEC} = E_{sense} + E_{tx}^R + E_{rx}^C + E_{dec}^C + E_{tx}^D, \quad (73)$$

$$E_{TX}^{HARQ} = E_{sense} + E_{tx}^R + E_{rx}^C + E_{dec}^C + E_{tx}^{D1} + (p_s^C)^2 (1 - p_s^D) (E_{rx}^N + E_{dec}^C + E_{tx}^{D2}), \quad (74)$$

for ARQ, FEC and hybrid ARQ, respectively, where E_{sense} is the energy consumption for sensing the region, E_{tx}^x , E_{rx}^x , and E_{dec}^x are the packet transmission, receiving, and decoding energies for packets, where the superscripts R , C , D , A , and N refer to RTS, CTS, DATA, ACK, and NACK packets, respectively, and $E_{t/o}$ is the energy consumed before timeout. The subscripts $D1$ and $D2$ in (74) refer to the transmitted packets for the first and second transmission in hybrid ARQ. In our calculations, we assume that RTS and CTS packets are also encoded in order to fully exploit the advantages of FEC codes. Using the same approach, the energy consumption of the

²We consider the length of RTS, CTS, ACK, and NACK packets the same.

receiver node is given as follows:

$$E_{RX}^{ARQ} = \frac{1}{(p_s^C)^3 p_s^D} \{ E_{rx}^R + E_{tx}^C + E_{rx}^D + E_{tx}^A \} , \quad (75)$$

$$E_{RX}^{FEC} = E_{rx}^R + E_{dec}^R + E_{tx}^C + E_{rx}^D + E_{dec}^D , \quad (76)$$

$$E_{RX}^{HARQ} = E_{rx}^R + E_{dec}^R + E_{tx}^C + E_{rx}^{D1} + E_{dec}^{D1} \\ + (p_s^C)^2 (1 - p_s^D) (E_{tx}^N + E_{rx}^{D2} + E_{dec}^{D2}) . \quad (77)$$

The last term in (71), E_{neigh} , is the energy consumed by the neighbors of the transmitter and the receiver nodes, which is given as:

$$E_{neigh}^{ARQ} = E_{neigh}^{HARQ} = \frac{1}{(p_s^C)^3 p_s^D} \left\{ (\rho\delta\pi R_{inf}^2 - 2) E_{rx}^R \right. \\ \left. + [\rho\delta (\pi R_{inf}^2 - A(D, R_{inf}, D)) - 2] E_{rx}^C \right\} , \quad (78)$$

$$E_{neigh}^{FEC} = (\rho\delta\pi R_{inf}^2 - 2) E_{rx}^R + [\rho\delta (\pi R_{inf}^2 - A(D, R_{inf}, D)) - 2] E_{rx}^C , \quad (79)$$

for ARQ, hybrid ARQ, and FEC codes³. Using these derivations in (55), the end-to-end energy consumption can be calculated.

5.5.3 Latency Analysis

The expression for end-to-end latency of a flow is found using the similar approach above. The delay per hop is given by

$$E[T_h] = \rho\delta \int_{\gamma_{min}}^D \int_{-\alpha\gamma}^{\alpha\gamma} \gamma E[T_j] Q \left(\frac{\beta}{\sigma} \right) e^{-M(1-p_k)} d\alpha d\gamma , \quad (80)$$

where

$$E[T_j] = \int_{\psi_{Th}}^{\infty} T_{comm}(\psi, d_{(i,j)}) f_{\Psi}(\psi, d_{(i,j)}) d\psi , \quad (81)$$

³We assume the header information is sufficient for backoff.

and T_{comm} is given as

$$T_{comm}^{ARQ} = \frac{1}{(p_s^C)^3 p_s^D} \left\{ T_{sense} + 2 (p_s^C)^2 T^{Ctrl} + \left(1 - (p_s^C)^2 \right) T_{t/o}^C + (p_s^C)^3 p_s^D (T^D + T^{Ctrl}) + (p_s^C)^3 (1 - p_s^C p_s^D) T_{t/o}^A \right\}, \quad (82)$$

$$T_{comm}^{FEC} = T_{sense} + 2T^{Ctrl} + 2T_{dec}^C + T^D + T_{dec}^D, \quad (83)$$

$$T_{comm}^{HARQ} = T_{sense} + 2T^{Ctrl} + 2T_{dec}^C + T^{D1} + T_{dec}^{D1} + (p_s^C)^2 (1 - p_s^D) (T^{Ctrl} + T_{dec}^C + T^{D2} + T_{dec}^{D2}). \quad (84)$$

for ARQ, FEC, and hybrid ARQ, respectively, where T_{sense} is the time spent for sensing, T^{Ctrl} and T^D are the control and data packet transmission time, respectively, $T_{t/o}$ is the timeout value, and T_{dec}^{Ctrl} and T_{dec}^D are the decoding latency for control and data packets, respectively.

5.5.4 Decoding Latency and Energy

The major overhead of FEC codes is the energy consumption for encoding and decoding packets and the delay associated with it. It is well known that the encoding energy for block codes is negligible [46]. Hence, we only consider the decoding energy and latency in our calculations in Sections 5.5.2 and 5.5.3. The Mica2 and MicaZ nodes that we consider for our analysis do not provide hardware support for FEC coding [22], [23]. Hence, we assume that FEC coding is implemented in software. According to [46], the latency of decoding for a block code (n, k, t) is given as

$$T_{dec}^{BL} = (2nt + 2t^2) (T_{add} + T_{mult}), \quad (85)$$

where T_{add} and T_{mult} are the energy consumption for addition and multiplication, respectively, of field elements in $\text{GF}(2^m)$, $m = \lceil \log_2 n + 1 \rceil$ [62]. Both Mica2 and MicaZ nodes are implemented with 8-bit microcontrollers [21], which can perform addition and multiplication of 8 bits in 1 and 2 cycles, respectively. As a result

$$T_{add} + T_{mult} = 3 \left\lceil \frac{m}{8} \right\rceil t_{cycle}, \quad (86)$$

where t_{cycle} is one cycle duration, which is 250 ns [21]. Consequently, the decoding energy consumption is $E_{dec}^{BL} = I_{proc}VT_{dec}^{BL}$, where I_{proc} is the current for processor, V is the supply voltage, and T_{dec}^{BL} is given in (85).

5.5.5 Bit and Packet Error Rate

In this section, we derive the expressions for bit and packet error rate for Mica2 and MicaZ nodes. Since the modulation schemes used in these nodes are significantly different, it is necessary to investigate the effects of FEC and hybrid ARQ on these nodes separately. Mica2 nodes are implemented with non-coherent FSK modulation scheme. The bit error rate of this scheme is given by [46]

$$p_b^{FSK} = \frac{1}{2}e^{-\frac{Eb/No}{2}}, \quad Eb/No = \psi \frac{B_N}{R}, \quad (87)$$

where ψ is the received SNR, B_N is the noise bandwidth, and R is the data rate. The modulation scheme used in MicaZ nodes is offset quadrature phase shift keying (O-QPSK) with direct sequence spread spectrum (DSSS). The bit error rate of this scheme is given by [43]

$$p_b^{OQPSK} = Q(\sqrt{(Eb/No)_{DS}}), \quad (88)$$

where

$$(Eb/No)_{DS} = \frac{2N \times Eb/No}{N + 4Eb/No(K - 1)/3},$$

where N is the number of chips per bit, and K is the number of simultaneously transmitting users.

Based on the bit error rate p_b , the PER for the error control schemes can be calculated as follows. For ARQ, the CRC-16 error detection mechanism is deployed in both Mica nodes. Assuming all possible errors in a packet can be detected, the PER of a single transmission of a packet with payload l bits is given by

$$PER^{CRC}(l) = 1 - (1 - p_b)^l. \quad (89)$$

For the BCH codes, assuming perfect interleaving at the transceiver, the block error rate (BLER) is given by

$$BLER(n, k, t) = \sum_{i=t+1}^n \binom{n}{i} p_b^i (1 - p_b)^{n-i}. \quad (90)$$

The block error rate for RS codes are found through simulations. More specifically, Berlekamp-Massey algorithm is implemented and simulated to find the relationship between block error rate and bit error rate.

Since a packet can be larger than the block length n , especially where small block lengths are used, the PER for FEC is given by

$$PER^{FEC}(l, n, k, t) = 1 - (1 - BLER(n, k, t))^{\lceil \frac{l}{k} \rceil}, \quad (91)$$

where $\lceil \frac{l}{k} \rceil$ is the number of blocks required to send l bits and $\lceil \cdot \rceil$ is the ceiling function. Using (89), (90), and (91), the PER for the hybrid ARQ schemes are also found.

5.6 Numerical Results

In this section, we investigate the effects of FEC and hybrid ARQ schemes in terms of PER, energy consumption and end-to-end latency in a multi-hop network via numerical evaluations in MATLAB. The cases where FEC and hybrid ARQ can be more favorable than ARQ are discussed. Moreover, an energy and latency-based taxonomy is devised to qualitatively compare FEC schemes with ARQ. For this comparison, two sensor node architectures are considered, i.e., Mica2 [22] and MicaZ [23]. We consider a multi-hop network, where a random access scheme and a channel-aware routing protocol is deployed as discussed in Section 5.4. Unless otherwise noted, the parameters in Table 2 are used for the numerical results. For Mica2 nodes, the experimental values in [73] are used while for MicaZ, the values on the datasheet are used [23].

The expected hop distance d_{hop} , which is found in (59), is shown in Fig. 35 (a) as a function of the received SNR threshold, ψ_{Th} , for different transmit power values, P_t . It

Table 2: Parameters

D	300 m	l_C	8 bytes
P_t	0, -5, -15 dBm	l_D	38 bytes
PL_{d0}	55 dB	t_{cycle}	250 ns
P_n	-105 dBm	I_{proc}	8 mA
η	3	V	3 V
σ	3.8		
	Mica2	MicaZ	
e_{rx}	21 mJ		59.1 mJ
$e_{tx} (P_t=0)$	24 mJ		52.2 mJ
$e_{tx} (P_t=-5)$	21.3 mJ		42 mJ
$e_{tx} (P_t=-15)$	16.2 mJ		29.7 mJ
$t_{bit} = 1/R$	62.4 μ s		4 μ s
N	N/A		16 chips
K	N/A		2

can be observed that for small values of the received SNR threshold, ψ_{Th} , the average hop distance increases. Since lower ψ_{Th} allows nodes with lower channel quality to be chosen as the next hop, further nodes may become the next hop. Therefore, the number of hops from a node to a sink decreases for smaller ψ_{Th} values. Moreover, when the transmit power of a node is decreased, the expected hop distance decreases as expected.

In the following, we present the effects of two techniques to exploit FEC codes in WSN, i.e., hop length extension and transmit power control. First, two FEC schemes, BCH and RS codes, are compared with the ARQ scheme. Then, the effects of end-to-end distance and the end-to-end target PER on the choice of error control scheme are also discussed. Furthermore, the results for hybrid ARQ are presented.

5.6.1 Hop Length Extension

In Fig. 35 (b), the end-to-end energy consumption per useful bit is shown as a function of the SNR threshold, ψ_{Th} . The energy consumption is shown for ARQ with 7 retransmissions, three BCH schemes, and three RS schemes as discussed in Section 5.4. As shown in Fig. 35 (b), the energy consumption of a flow decreases for smaller

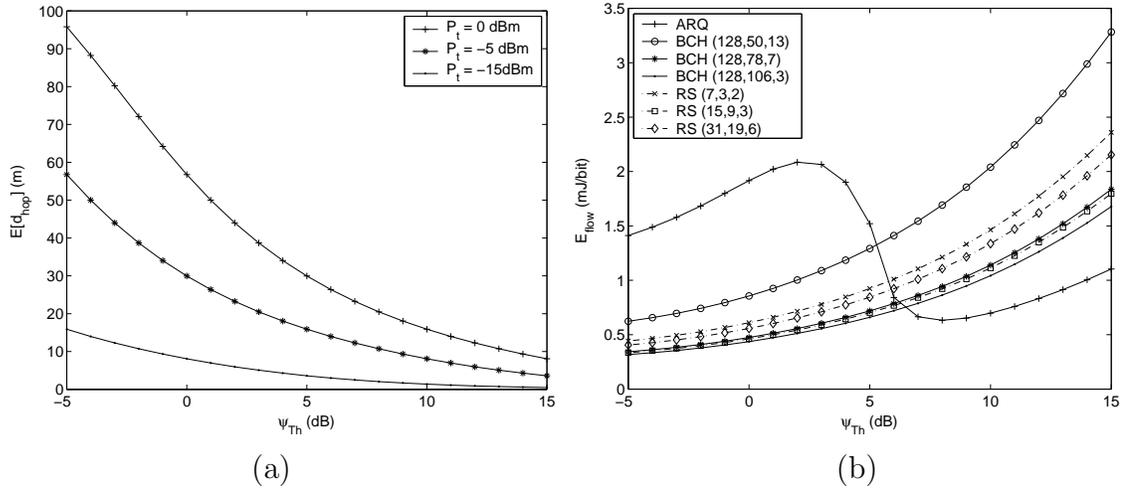


Figure 35: (a) Avg. hop distance and (b) avg. energy consumption of a flow vs. ψ_{Th} (DSSS+OQPSK).

ψ_{Th} value. This is mainly because of the increase in expected hop distance as shown in Fig. 35 (a). However, the energy consumption for ARQ significantly increases as ψ_{Th} is decreased below a specific value, e.g., 7 dB. A lower ψ_{Th} results in nodes with lower channel quality to be selected as the next hop. As a result, retransmissions occur, which increase the energy consumption per hop. Although the expected number of hops decreases, the increase in energy consumption per hop dominates the total energy consumption for ARQ. Note that for ARQ, the energy consumption curve reaches a peak point and decreases as ψ_{Th} is decreased. This point corresponds to the case that the maximum number of retransmissions is not sufficient for reliable communication.

When the FEC codes are considered, the energy consumption is proportional to the error correcting capability, t , of the code. Since the code rate is decreased for higher t , the energy consumption per useful bit increases. When ARQ and FEC codes are compared, for high ψ_{Th} values, ARQ slightly outperforms the FEC codes. However, BCH (128,106,3), BCH (128,78,7), and RS (15,9,3) codes are more energy efficient for higher ψ_{Th} . Although this figure clearly shows the energy consumption of the two schemes as a function of ψ_{Th} , the operating points of ψ_{Th} for ARQ and

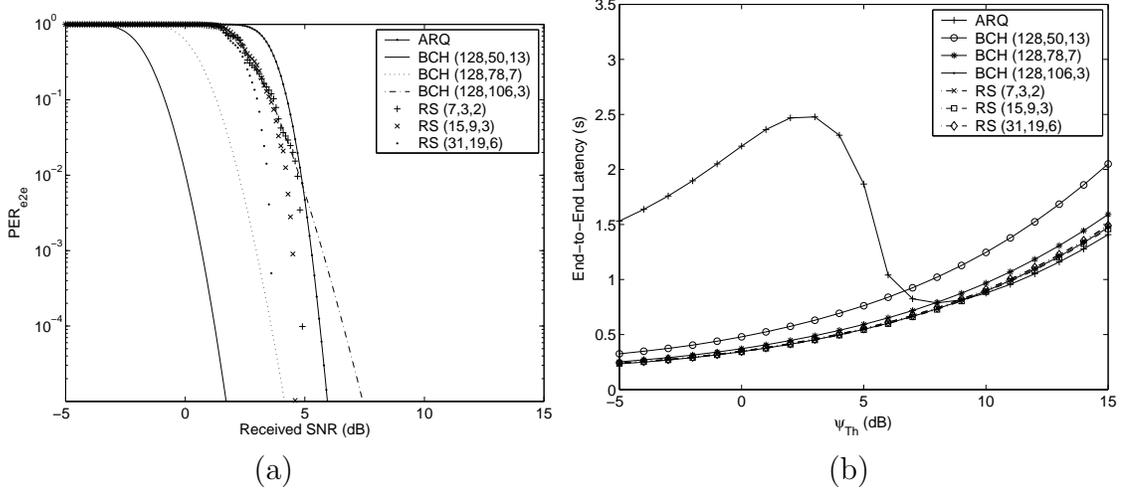


Figure 36: (a) End-to-end PER vs. received SNR and (b) avg. end-to-end latency vs. ψ_{Th} (DSSS+OQPSK).

FEC has to be determined. Hence, next, we investigate the end-to-end packet error rate (PER) performance.

The PER for CRC and FEC codes are given in (89) and (91), respectively. Since these equations show the PER for a single hop, here we extend these equations for the multi-hop case. Note that WSN applications are interested in the achievable end-to-end PER bound rather than the single hop PER. Hence, the relation between ψ_{Th} and the end-to-end PER bound can be used to determine the optimal point for ψ_{Th} . Denoting the PER of a hop i by PER_i , there exists a π such that

$$PER_i \leq \pi, \text{ for } \psi_i \geq \psi_{Th},$$

where ψ_i is the received SNR of the hop and $\pi = f(\psi_{Th})$, which can be calculated using (87) - (91) depending on the modulation scheme, the error control scheme, the channel characteristics, and the packet length. Since the end-to-end PER is

$$PER_{e2e} = 1 - \prod_{i=1}^{n_h} (1 - PER_i),$$

where n_h is the number of hops, PER_{e2e} is bounded by

$$PER_{e2e} \leq 1 - (1 - \pi)^{n_h}, \text{ for } \psi_i \geq \psi_{Th}, \forall i. \quad (92)$$

Now assume that the end-to-end PER needs to be bounded by a certain threshold, PER_{e2e}^* , according to the application requirements. Accordingly, the route selection needs to be performed such that

$$\psi_{Th} = f^{-1} \left(1 - [1 - PER_{e2e}^*]^{1/n_h} \right) . \quad (93)$$

The relationship between the end-to-end PER, PER_{e2e} , and ψ_{Th} is shown in Fig. 36 (a) for ARQ, BCH, and RS codes for MicaZ nodes. Note that for RS codes, the results of the simulations are used whereas for ARQ and BCH, (89) and (91) are used, respectively. According to Fig. 36 (a), the operating point for ψ_{Th} corresponding to a target end to end PER can be found. As an example, if the target PER of an application is 10^{-2} , the minimum value for ψ_{Th} corresponds to ~ 7 dB for ARQ, ~ 3 dB for BCH(128,78,7), and 6.8 dB for RS(15,9,3). As a result, it can be observed from Fig. 35 (b) that BCH (128,78,7) is slightly more energy efficient than ARQ. On the other hand, the RS (15,9,3) code results in higher energy consumption compared to the ARQ scheme. It is clear that more energy is consumed per hop for FEC codes due to both transmission of redundant bits and decoding. However, since the effective error rate of the channel is decreased with FEC codes, lower SNR values can be supported. By deploying a routing protocol that exploits this property, longer hop distances can be achieved leading to lower end-to-end energy consumption. However, note that the energy consumption of the BCH (128,50,13) is still above the minimum value achieved by ARQ as shown in Fig. 35 (b), which suggests that an optimal value for t should be chosen.

Exploiting FEC schemes with channel-aware routing not only improves energy consumption performance but the end-to-end latency can also be decreased significantly as shown in Fig. 36 (b). It is clear that all the six FEC schemes outperform ARQ since their optimal ψ_{Th} value is lower than ARQ. This is due to both longer hops for FEC codes and the additional retransmissions of ARQ. Since the decoding delay of the FEC codes is lower than the time consumed for retransmission of a packet,

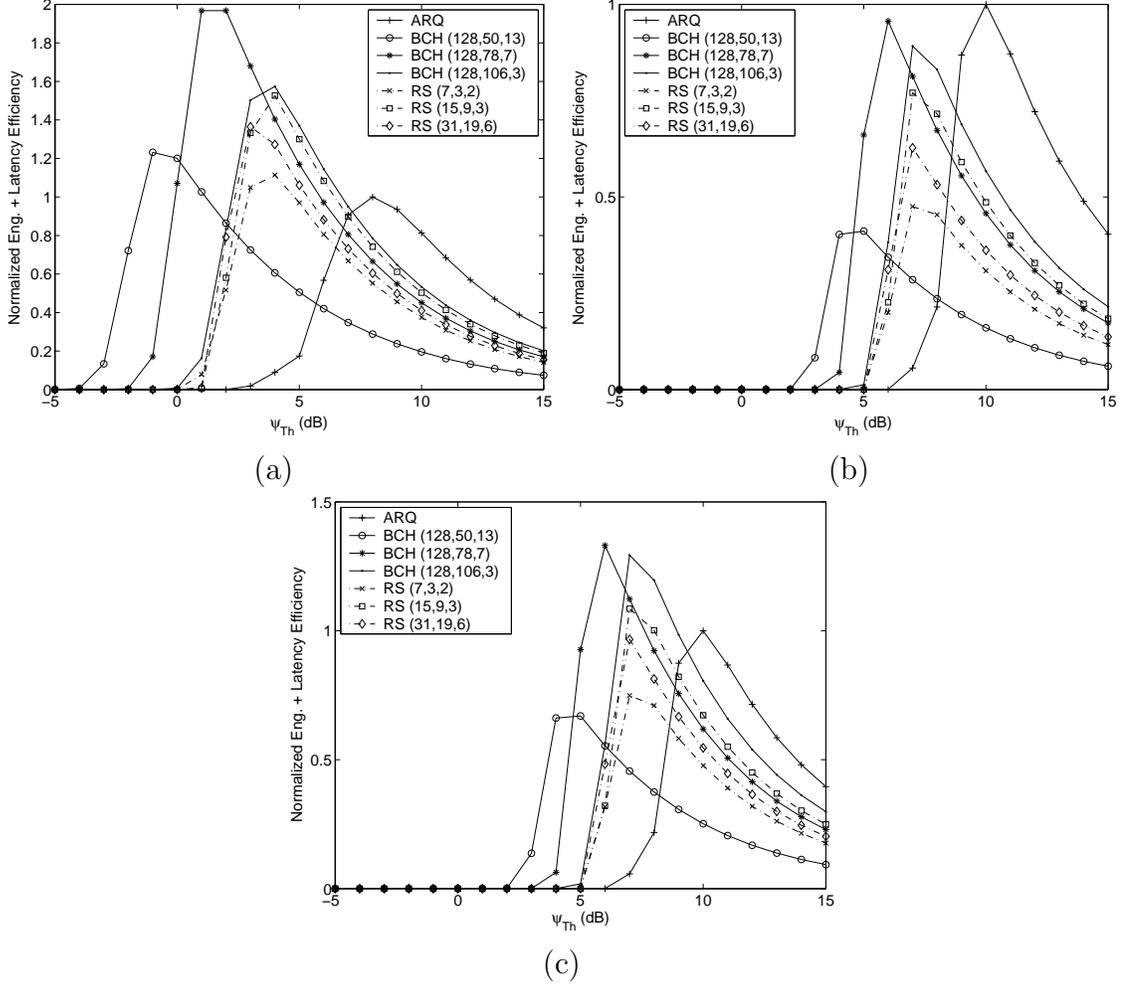


Figure 37: Taxonomy function vs. ψ_{Th} for (a) DSSS+OQPSK, (b) FSK, and (c) FSK without E_{neigh} .

FEC schemes improve the latency performance of the WSN. Furthermore, RS codes provide lower end-to-end latency when compared to the BCH codes. This is related to the better error correction capability of RS codes when same number of redundant bits are sent. Consequently, the end-to-end latency is slightly decreased.

In order to capture the efficiency of an error control scheme, we propose a taxonomy function that consists of the energy consumption, latency and PER performance. This function is given as follows:

$$\mathcal{T} = \frac{l_D}{E_{flow} T_{flow}} (1 - PER_{e2e}), \quad (94)$$

where l_D is the payload length, E_{flow} , T_{flow} , and PER_{e2e} are the end-to-end energy

consumption, latency and PER, respectively. Note that a similar efficiency function has been proposed in [62], where the energy efficiency and the reliability of a single hop has been considered. In our approach, we also consider the cost for latency and propose a multi-hop taxonomy function.

In Fig. 37 (a) and (b), the taxonomy function is evaluated for MicaZ and Mica2 nodes, respectively. The taxonomy function is normalized for the maximum value of ARQ. It is clear from (94) that a higher value of \mathcal{T} corresponds to higher efficiency. It can be observed from Fig. 37 (a) that for the MicaZ nodes, the FEC codes outperform ARQ. Moreover, RS codes are less efficient compared to the BCH codes due to their higher energy consumption. Moreover, an optimal error correction capability, t , can be found for both BCH and RS codes that leverages the PER with energy consumption and latency. On the other hand, for Mica2 nodes, ARQ is more efficient than the FEC codes. This interesting result advocates that there is no clear winner for error control techniques in WSN and their performance directly depends on the node hardware.

The reason behind the difference between MicaZ and Mica2 nodes can be explained as follows: In Fig. 37 (c), the taxonomy function is re-evaluated for Mica2 without considering the energy consumption of neighbor nodes, E_{neigh} , given in (71). In this case, BCH (128,106,3), BCH (128,78,7), and RS (15,9,3) are more efficient than ARQ. The major differences between Mica2 and MicaZ nodes are the data rate of the transceivers and the modulation schemes. As shown in Table 2, the time consumed for transmitting a bit is 15 times higher for Mica2 than MicaZ. This corresponds to significant energy consumption for communication. Since the peak of the ARQ curve corresponds to no retransmissions, it is clear that the FEC codes consume more energy primarily due to the transmission of redundant bits. When the energy consumption of the neighbors are considered, the energy consumption significantly increases. Moreover, it can be observed from Figs. 33 that FEC codes lead to smaller increase in expected hop length for Mica2 nodes than MicaZ nodes. As a result,

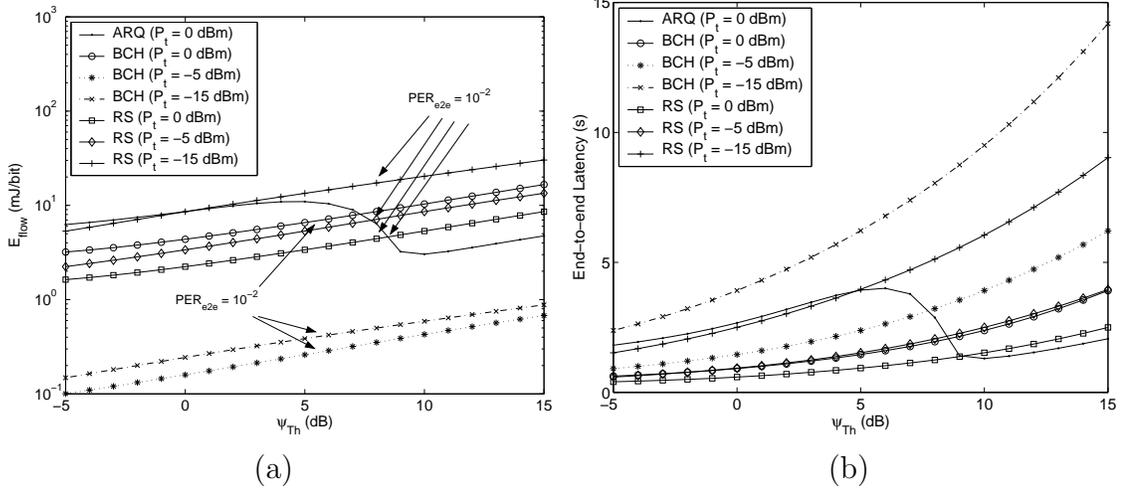


Figure 38: (a) Avg. energy consumption and (b) avg. end-to-end latency vs. ψ_{Th} for different values of transmit power (FSK).

for Mica2 nodes, the advantage of larger hop length provided by the FEC codes is outweighed by the increase in energy consumption of neighbor nodes. This favors ARQ for Mica2 nodes. It is also important to note that consideration of E_{neigh} is important to accurately assess the performance of ARQ and FEC.

5.6.2 Transmit Power Control

Another technique to exploit FEC codes is to match the average hop distance of FEC codes with ARQ. This can be achieved by decreasing the transmit power, P_t . In order to investigate the effect of transmit power, P_t , we consider three power levels, i.e., 0, -5, and -15 dBm supported by both Mica2 and MicaZ. Intuitively, decreasing transmit power can improve the energy efficiency of the FEC schemes, since less power is consumed for transmission of longer encoded packets. Although the receive power is fixed, since the interference range of a node decreases, the number of neighbors that consume idle energy also decreases. On the other hand, decreasing transmit power increases the number of hops. In Fig. 38 (a), the energy consumption of BCH (128,50,13) and RS (15,9,3) are shown for three different transmit power levels and ARQ at $P_t = 0$ dBm. Furthermore, the operating points of these schemes

corresponding to $PER_{e2e} \leq 10^{-2}$ are indicated. Note that the decrease in transmit power decreases the end-to-end energy consumption for BCH (128,50,13), which outperforms the ARQ scheme. On the other hand, RS (15,9,3) leads to higher energy consumption when the transmit power is decreased. This difference is mainly due to the relationship between decreased energy consumption per hop and the increased number of hops. For the BCH (128,50,13), decreasing transmit power from 0 dBm to -5 dBm and -15 dBm leads to a decrease of 97% and 99% in energy consumption per hop, respectively. On the other hand, for RS (15,9,3), these values are 15% and 60% for -5 dBm and -15 dBm, respectively. Decreasing transmit power leads to a decrease of 47% and 88% in hop length as shown in Fig. 36 (a), which increases energy consumption since the number of hops increase. Accordingly, for BCH (128,50,13), the decrease in energy consumption per hop dominates, which leads to overall decrease in the end-to-end energy consumption. However, RS (15,9,3) results in higher end-to-end energy consumption when the transmit power is decreased.

While transmit power control provides energy efficiency for particular FEC codes, its drawback is shown in Fig. 38 (b), where the end-to-end latency is shown. Contrary to the hop distance extension, since controlling transmit power has no effect in the time required for transmitting a packet, the end-to-end latency depends on the number of hops. Since transmit power control increases the number of hops, this technique introduces a significant increase in latency, which is a tradeoff for BCH (128,50,13) code.

5.6.3 Effects of End-to-End Distance and Target PER

In this section, we investigate the effects of end-to-end distance, D , and the target PER, PER_{e2e}^* , on the performance of ARQ and FEC schemes. The end-to-end energy consumption per useful bit is shown in Fig. 39 (a) and (b) as a function of the end-to-end distance, D , for Mica2 and MicaZ nodes, respectively. In these figures, the

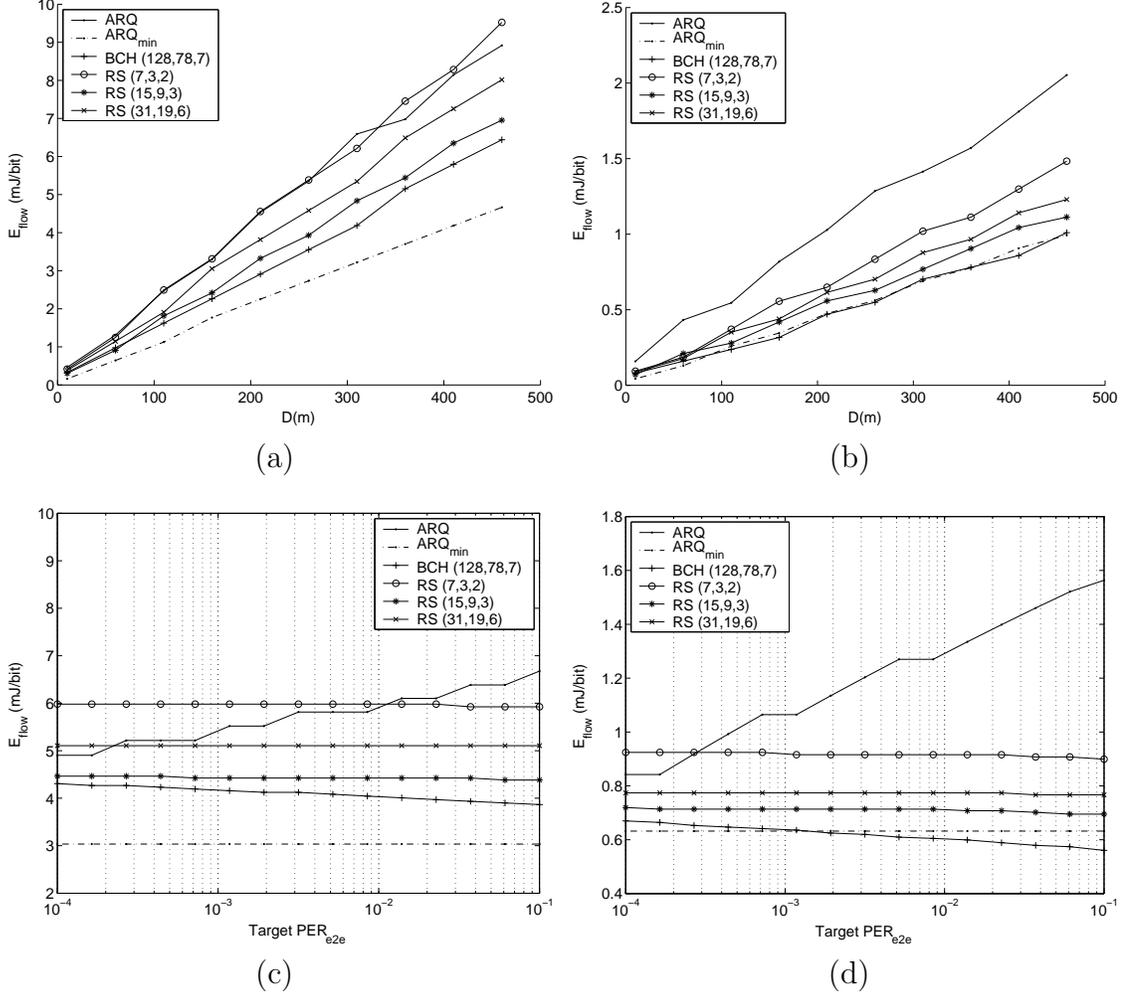


Figure 39: Average energy consumption vs. end-to-end distance for (a) Mica2 and (b) MicaZ nodes, and the average energy consumption vs. target end-to-end PER for (c) Mica2 and (d) MicaZ nodes.

minimum ψ_{Th} is selected for each BCH and RS code such that $PER_{e2e}^* \leq 10^{-2}$ is satisfied⁴. For ARQ, two curves are shown. The solid line corresponds to the minimum ψ_{Th} that satisfies the PER constraint. However, note from Fig. 38 (a) that the minimum energy consumption of ARQ corresponds to a higher ψ_{Th} value than the one found here. Since the energy consumption of the ARQ is minimum without retransmissions, a higher link quality is required to achieve the optimum performance. Hence, in Figs. 39 (a)-(d), the line denoted by ARQ_{min} corresponds

⁴For clarity, only BCH (128,78,7) code is shown with the three RS codes. Results for the other BCH codes can be found in [79].

to ψ_{Th} that achieves the minimum energy consumption for ARQ. Accordingly, for Mica2 nodes, ARQ is more energy efficient than both of the FEC codes irrespective of the end-to-end distance. On the other hand, for MicaZ nodes, BCH code with $t = 7$ results in energy consumption comparable to ARQ. In particular, ARQ is more energy efficient for end-to-end distances up to 80 m, which corresponds to ~ 5 hops for ARQ and ~ 3 hops for BCH (128,78,7) code. For hop counts higher than these values, the BCH code consumes slightly less energy compared to ARQ. Furthermore, RS codes result in energy consumption that is higher than the BCH (128,78,7) code.

As explained before, the operating point of ψ_{Th} is determined according to the target PER of the WSN application. The effect of target PER is investigated in Figs. 39 (c)-(d) for Mica2 and MicaZ nodes, respectively. Similar to our observations above, ARQ is more energy efficient for Mica2 nodes irrespective of the target PER. However, when MicaZ architecture is considered, BCH code outperforms ARQ for target $PER_{e2e} > 0.002$. When the target PER is increased, the optimal value of ψ_{Th} is decreased, which improves the energy efficiency of FEC codes. As a result, the energy consumption of BCH (128,78,7) code is more favorable than ARQ. On the contrary, irrespective of the end-to-end PER requirement, RS codes always result in higher energy consumption than BCH (128,78,7).

5.6.4 Hybrid Error Control

Hybrid ARQ schemes exploit the advantages of both ARQ and FEC techniques. In this section, we compare the end-to-end energy consumption and the latency characteristics of these schemes with the ARQ scheme and the BCH (128,78,7), which is found to be the most energy efficient FEC scheme in the previous sections.

The energy consumption of the type I and type II hybrid ARQ schemes are shown in Fig. 40 (a) and (b), respectively, for MicaZ. An important result is that type II hybrid ARQ schemes are more energy efficient than both ARQ and FEC schemes.

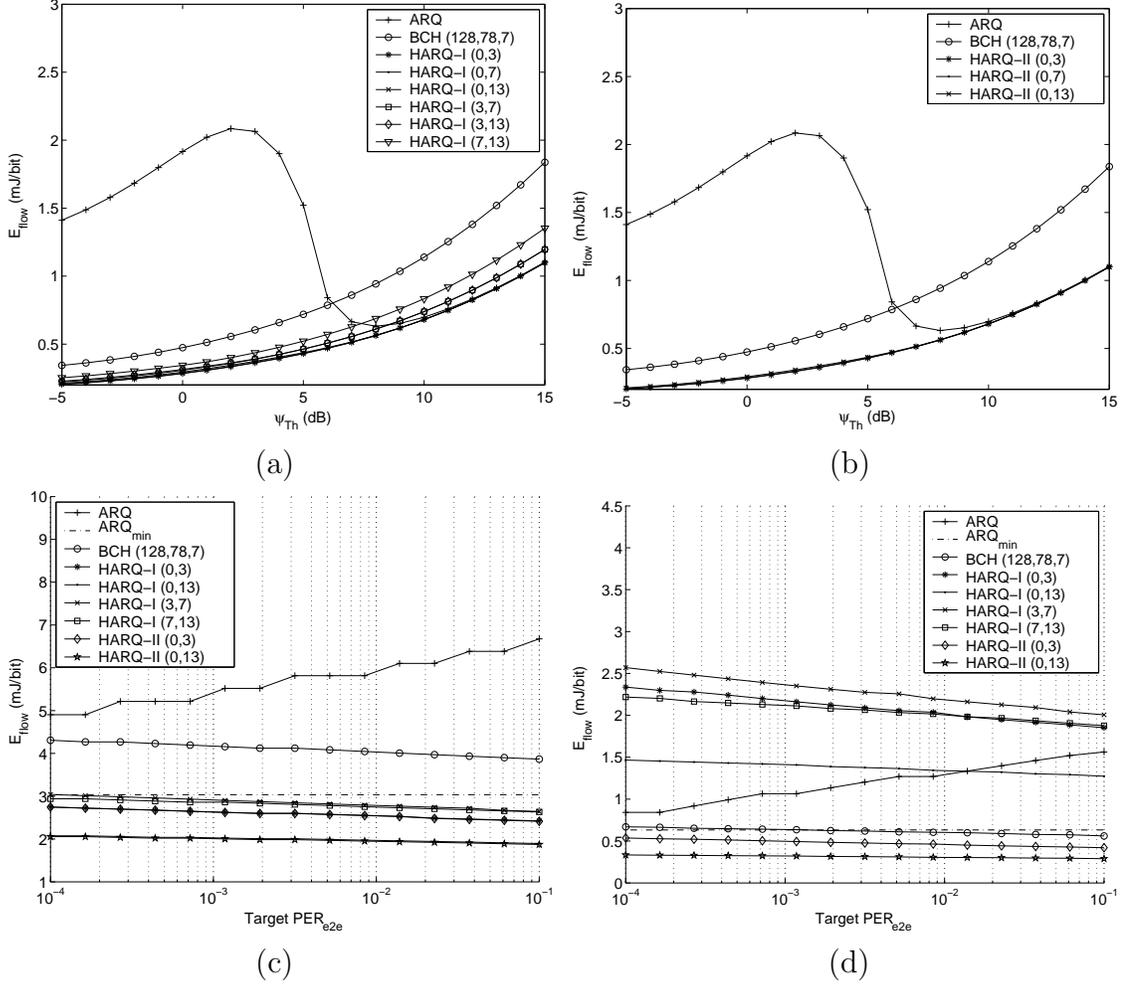


Figure 40: Avg. energy consumption vs. ψ_{Th} for hybrid ARQ (a) Type I and (b) Type II (DSSS+OQPSK), and average energy consumption vs. target end-to-end PER for (a) Mica2 and (b) MicaZ nodes.

As shown in Fig. 40 (b), the energy consumption of different HARQ-II schemes are similar for a given ψ_{Th} value. However, since the error resiliency of these protocols depend on the BCH code used, the operating point of these schemes differ based on the target PER. It is shown in Fig. 40 (a) that the energy consumption of the HARQ-I scheme is dependent on the error correction code used in the first transmission. Consequently, HARQ-I (7,13) scheme results in the highest energy consumption for a given ψ_{Th} value. Our results also show that the end-to-end latency of HARQ-I and HARQ-II schemes indicate that the hybrid ARQ schemes provide similar latency

compared to the FEC scheme. This is particularly important for real-time traffic in WSN.

The energy consumption graphs as a function of the SNR threshold, ψ_{Th} , illustrate the effect of this threshold on the performance of these error control schemes. However, for a given end-to-end reliability requirement, the operating point can be different for each scheme because of their different error resiliency. In Figs. 40 (c) and (d), the end-to-end energy consumption is shown as a function of the target PER for Mica2 and MicaZ, respectively. In both figures, it is shown that HARQ-II schemes outperform ARQ, FEC, as well as HARQ-I schemes. This is particularly appealing since the HARQ-II scheme is implemented through only a single BCH code. Hence, the implementation cost of the HARQ-II scheme in consideration is also low compared to the HARQ-I case, where two different encoding schemes can be implemented. Furthermore, the energy efficiency of HARQ-II schemes improve when more powerful FEC schemes are used. An important difference between the Mica2 and the MicaZ hardware is that for MicaZ nodes, HARQ-I results in higher energy consumption than both FEC and ARQ schemes. This is due to the fact that HARQ-I schemes send the whole packet for retransmissions. Moreover, the error resiliency of the HARQ-I is not high enough to compensate this overhead through hop length extension. On the other hand, HARQ-II schemes introduce minimal overhead by sending only the redundant bits for retransmission. Consequently, the error resiliency is improved by minimum overhead.

5.6.5 Overview of Results

An overview of the energy and latency performance of the error control schemes that are considered in this chapter is shown in Figs. 41. In the figures, the minimum end-to-end energy consumption and latency of ARQ, BCH, RS, HARQ-I, and HARQ-II schemes subject to an end-to-end PER target of 10^{-2} are shown as a function of

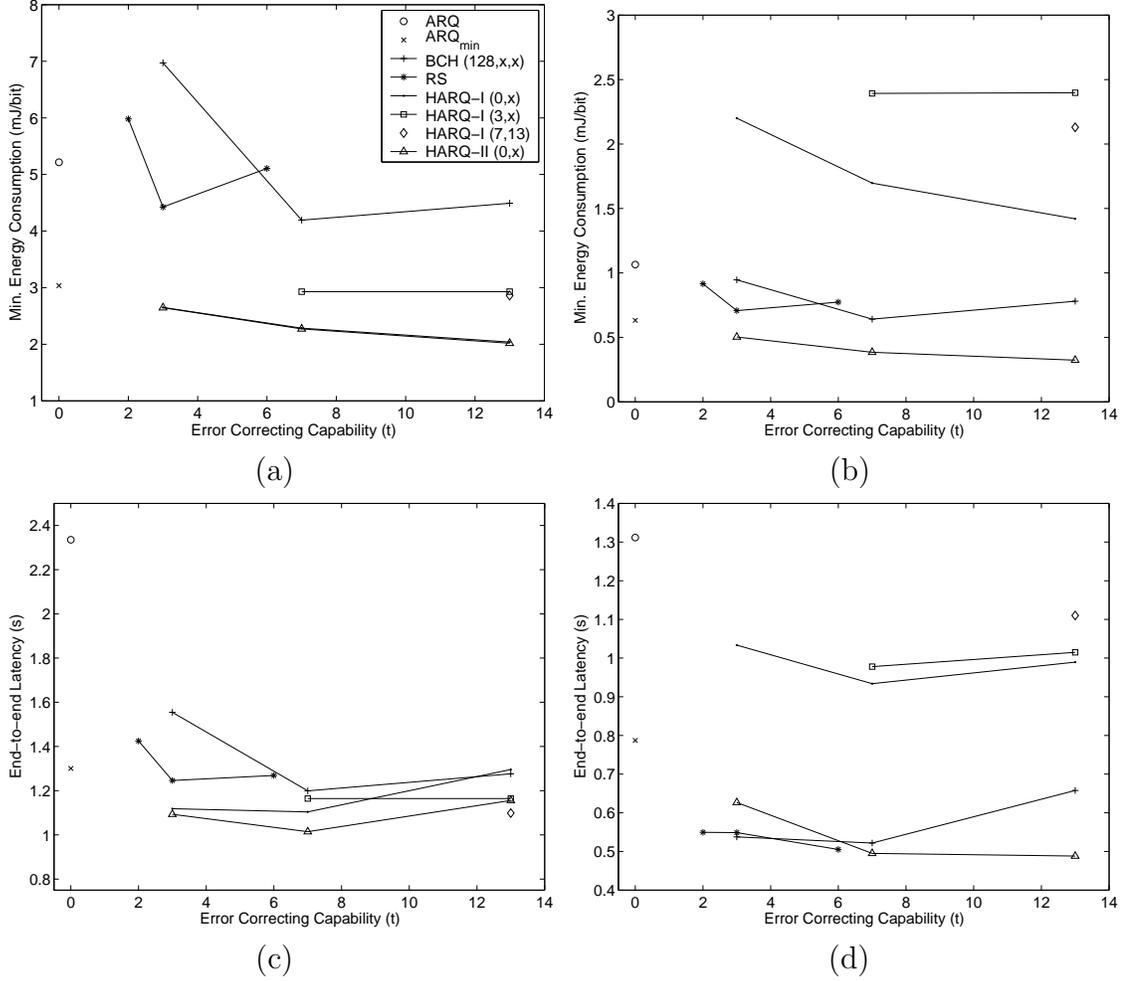


Figure 41: Minimum energy consumption vs. error correcting capability for (a) Mica2 and (b) MicaZ, and minimum latency vs. error correcting capability for (c) Mica2 and (d) MicaZ nodes with $PER_{e2e} \leq 10^{-2}$.

their error correction capability. In particular, in Fig. 41 (a) and (c), the minimum energy consumption and latency of these schemes are shown for Mica2 nodes, respectively. Consistent with our previous observations, both type-I and type-II hybrid ARQ schemes outperform other error control schemes. Moreover, it can be observed that for both BCH and RS error codes, an optimum error correction capability, t , value can be found to minimize energy consumption and latency. The results for MicaZ as shown in Fig. 41 (b) and (d) reveal that type-I hybrid ARQ codes are inefficient in terms of both energy consumption and latency. Type-II hybrid ARQ

scheme is more energy efficient compared to ARQ, BCH, and RS codes. Moreover, the energy consumption of these three schemes, ARQ, BCH, and RS, are similar. Furthermore, it can be observed from Fig. 41 (d) that RS (31,19,6) code performs very close to the HARQ-II scheme in terms of end-to-end latency, which make both of these schemes a suitable candidate for real-time traffic.

CHAPTER VI

CROSS-LAYER PACKET SIZE OPTIMIZATION FOR WIRELESS SENSOR NETWORKS

The final component of cross-layer communication in WSN is the determination of the optimal packet size. In this chapter, a cross-layer methodology for packet size optimization in WSN is proposed considering our cross-layer communication framework. We first discuss the factors affecting the packet size optimizations and then, propose the packet size optimization framework. The remainder of the chapter is organized as follows: In Section 6.1, the motivation for this work is provided. In Section 6.2, related work on packet size optimization in wireless networks is summarized. A breakdown of various cross-layer effects of packet size on network performance is presented in Section 6.3. More specifically, the effects of packet size on reliability and collisions are investigated. Based on this investigation, in Section 6.4, the cross-layer solution for packet size optimization is presented. The results of the optimization framework are presented in Section 6.5. We further extend our work for underwater and underground environment in Section 6.6.

6.1 Motivation

The unique characteristics of Wireless Sensor Networks (WSN) necessitate rethinking of classical wireless networking in all aspects [7]. Among these, the determination of the optimal packet size tailored for these unique characteristics constitutes a fundamental and still an unexplored problem in WSN. Especially, the cross-layer effects of communication due to low power communication constraints and intrinsic properties of low cost sensor nodes call for a cross-layer assessment of optimal packet size for

communication in WSN.

Traditionally, packet size optimization is performed considering a point-to-point link, where the goal is to ensure a successful and efficient transmission based on an efficiency metric [44], [55], [62]. However, in this traditional approach, the influence of multi-hop and broadcast nature of wireless communication in WSN cannot be captured. It is well-known that the packet size directly affects the reliability of the communication since longer packet sizes are susceptible to wireless channel errors given a certain level of link quality. However, in multi-hop WSN, the quality of the communication links depend on the routes established in the network. Moreover, the existence of neighbor nodes that contend for the shared wireless medium affect the communication performance significantly leading to degradation in communication success. Furthermore, the communication success also depends on both the characteristics of the wireless channel and the error control technique deployed. Considering these various factors that originate from different layers of the communication stack, a cross-layer assessment of packet size optimization for WSN is crucial.

Another aspect in packet size optimization is the nature of the WSN. WSN has found wide application areas since the development of highly sophisticated wireless sensor nodes [7]. These networks are mainly being deployed in terrestrial areas such as forests, factories, buildings, etc. In addition, recently, WSN research has developed in to underwater and underground environments. Underwater acoustic wireless sensor networks (UW-ASN) are characterized by very low bandwidth and high error rate underwater channels [5]. Similarly, recently, the applications and requirements of wireless underground sensor networks (WUSN) have been investigated [6], [45]. These networks impose additional challenges in terms of channel characteristics. As a result, an optimal packet size for these environments should also be determined.

In this chapter, a cross-layer solution for packet size optimization in wireless terrestrial, underwater, and underground sensor networks is presented such that the cross-layer effects of multi-hop routing, the broadcast nature of the wireless channel, and the effects of error control techniques are captured. For terrestrial sensor networks, the effect of packet length on the collision probability is investigated. Moreover, the relationship between routing decisions and the packet size is highlighted. Furthermore, the effects of packet size on different performance metrics such as throughput, energy consumption, latency, and success rate are investigated considering these cross-layer effects. Finally, requirements of various types of applications in WSN are considered to develop a comprehensive framework for packet size optimization. The results of this framework provide optimal packet size values tailored for different application types. The optimization framework is also extended to determine the optimal packet size in two challenged environments, where wireless sensor networks find application areas. More specifically, underwater and underground environments are investigated.

6.2 Related Work

Packet size optimization is a highly investigated topic in cellular networks [55]. However, the existing work mainly considers a single hop communication and performs optimization accordingly. Therefore, these results cannot be directly applied to WSN because of the multiple hop paradigm.

In [44], adaptive frame sizes are determined based on the varying properties of the wireless channel, i.e., for bad channel conditions, shorter packets with powerful error correction is used while longer packet sizes are selected for good channel conditions by which a high reliability can be achieved. However, variable packet sizes are not preferred in WSN due to strict hardware and computation constraints of wireless sensor nodes. As a result, we advocate to use fixed packet sizes.

The most relevant work in this topic is [62], where an optimal packet size framework is proposed. In this work, the effects of error correction on energy efficiency are investigated to determine the optimal packet size based on an energy channel model. However, the energy channel model is based only on one hop behavior and does not capture the effects of multi-hop routing or MAC collisions are not addressed in this work.

6.3 Factors Affecting The Packet Size

We investigate the factors that affect the performance of the network both on bit and packet levels.

6.3.1 Error Detection and Correction: Bit Level Energy Consumption

One of the main factors influencing the performance of communication protocols on the bit level is the error detection and correction mechanism. Here we explore the packet size effects of Automatic Repeat ReQuest (ARQ) and Forward Error Correction (FEC).

The energy consumption in WSN mainly depends on the energy required to transmit a packet and the reliability of the network. Usually small packet sizes lead to increased reliability due the decreased chance of bit errors over the wireless channel. On the other hand, small packet sizes lead to inefficient transmission due to the overhead caused by network protocols and error correcting codes if applicable. In our analysis, we consider a packet of length c with a header of α , payload of l_D , and trailer of τ such that $c = \alpha + l_D + \tau$. The header length α depends on the control information at each layer and for IEEE 802.15.4, has been determined as 15 bytes including the CRC bits [37]. The trailer length τ depends on the error correcting capability of an FEC code and is 0 for the ARQ scheme.

For the ARQ scheme, denoting the raw bit error rate of the channel as p and the number of maximum retransmissions as T , the probability that a packet is successfully

transmitted through a single hop is shown as

$$1 - PER = (1 - p)^{l+\alpha} \left(\sum_{i=0}^{T-1} \left(1 - (1 - p)^{l+\alpha} \right)^i \right), \quad (95)$$

where PER denotes the packet error rate. However, since we are interested in the probability that the packet is transmitted to the sink successfully, a network reliability notion is required. Assuming the packet belongs to a node with minimum and maximum hop count of n_{min} and n_{max} , the reliability μ can be written as

$$\mu = \sum_{n=n_{min}}^{n_{max}} p(n) \prod_{k=1}^n \left(1 - PER_k \right) \quad (96)$$

where PER_k is the packet error rate at each link and $p(n)$ is the probability that the packet routes through n hops to reach the sink. This value depends on the routing protocol and the location of the node sending the packet, which will be investigated in more detail in Section 6.4.

FEC codes pose tradeoff between the energy consumption due to increased packet size and energy gains due to error correcting capabilities of the codes. If a packet contains errors that cannot be corrected by the FEC code, the whole energy consumption to encode, transmit, and decode a packet is wasted. Based on the same notion in (96), the reliability of FEC, μ_{FEC} , can also be shown as

$$\mu_{FEC} = \sum_{n=n_{min}}^{n_{max}} p(n) \prod_{k=1}^n \sum_{j=0}^t \binom{c}{j} p_k^j (1 - p_k)^{c-j}, \quad (97)$$

where t is the error correction capability of the FEC code.

Figure 42(a) illustrates the effect of packet length on reliability (1-PER) for a single hop. The values are found using (96) and (97) with $p = 10^{-3}$ for ARQ with 4 retransmissions and 5 different BCH codes. It can be observed that when the error correcting capability of the FEC code is low, i.e., $t = 2$, ARQ outperforms FEC for small packet sizes. On the other hand, when the packet size is increased, FEC codes provide higher resiliency for higher error correcting capability, t . This advantage, however, is provided at the cost of increased energy consumption and

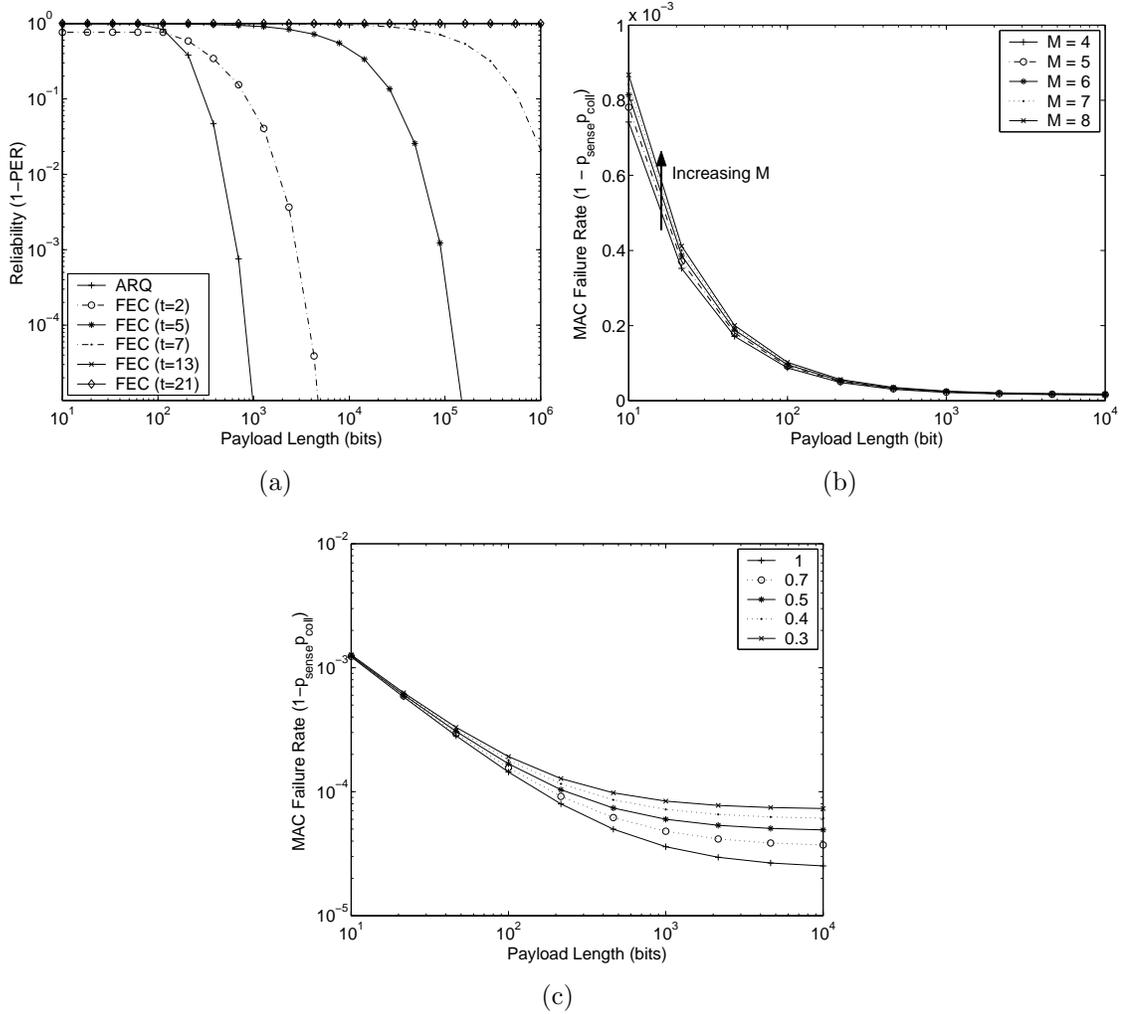


Figure 42: (a) Packet error rate vs. payload length for ARQ and FEC. MAC failure rate vs. payload length for (b) different values of M and (c) different values of FEC rate.

latency due to encoding/decoding and transmission of longer packets. This tradeoff will be investigated in Section 6.4.

6.3.2 Collision: Packet Level Energy Consumption

One of the most important source of energy consumption is collisions that occur in the high density environment such as WSN. While the MAC scheme in use affects the collision rate of the network, here, we show that the packet size has also an important effect on collisions in WSN.

Generally, the packet size is assumed to have negative effect on collisions. In cellular networks and, recently, ad hoc networks, it is accepted that longer packet sizes increase the collision rate of the network. This is due to the fact that a node occupies the channel for a longer time during and the probability that another terminal transmits a packet during this transmission increases. This is particularly accurate considering the traffic properties of cellular and ad hoc networks. These networks are characterized by independent traffic properties of each user. As a result, a MAC layer frame size is calculated assuming a fixed traffic load.

In WSN, however, the generated traffic is directly related to the physical phenomenon that is being sensed and the properties of the onboard sensors¹. This dependence results in close correlation of generated traffic in sensors that are closely located to each other. Consequently, since the collisions occur due to transmissions of these closely located sensor nodes, the dependence of collision rate on packet size can easily be analyzed.

For medium access, we assume that a node performs carrier sense mechanism to assess the availability of the channel and transmits a packet thereafter². Many work focus on the investigation of medium access performance of carrier sense mechanisms [12, 14, 19, 66]. Without loss of generality, here, we refer to the work in [66]. Note that our contribution is not to produce yet another analysis of carrier sense mechanism. Rather, we aim to illustrate the impact of packet size on collision rate in WSN, which, to the best of our knowledge, has not been performed before.

We first illustrate the formulation in [66] and then present our results. Accordingly, a successful allocation of the channel depends on both successful carrier sense and the fact that the transmission encounters no collisions. Denoting the probability of

¹Although control packets that are generated by various protocols exist, we consider the majority of the traffic is constituted by the sensed information.

²Note that if a reservation-based protocol is used, collisions may not occur. In those cases, the remainder of our framework still applies.

successful carrier sense by p_{cs} , it can be denoted as follows [66]:

$$p_{cs} = 1 - (1 - p_{cf})^{K+1} \quad (98)$$

where K is the number of re-sensings allowed for one transmission and p_{cf} is the probability of sensing the channel free, which is given by:

$$p_{cf} = e^{-\lambda_{net}(\tau_{cs}+T_{comm})} \quad (99)$$

where τ_{cs} is the carrier sense period and T_{comm} is the duration of a packet transmission. After a successful carrier sense, a collision can only occur if another node transmits during the vulnerable period of τ_{cs} . As a result, the probability of no collisions, p_{coll} , is given by

$$p_{coll} = e^{-\lambda_{net}\tau_{cs}} \quad (100)$$

The term λ_{net} that appears both in (99) and (100) refers to the overall traffic that is generated by the nodes inside the transmission range of a node, which is given by

$$\lambda_{net} = \lambda \frac{p_{cs}}{p_{comm}} \left(1 - (1 - p_{comm})^{L+1} \right), \quad (101)$$

where λ is the total generated traffic in the transmission range of a node and p_{comm} is the probability of successful transmission. Accordingly, the probability that a node can successfully acquire the channel is given by $p_{cs}p_{coll}$, which can be found by solving equations (98), (100), and (101). The effect of packet size on the MAC failure rate, i.e., $1 - p_{cs}p_{coll}$, can be observed clearly through (99) and (100). First of all, the probability of sensing the channel free, p_{cf} , depends on the duration of a packet transmission, T_{comm} . Longer packet sizes decrease this probability since a node will acquire the channel for a longer time. Second, the packet size, l_D , impacts the generated traffic rate λ , which affects both the probability of no collisions, p_{coll} , and the probability of successful carrier sense, p_{cs} .

The total generated packet rate, λ , depends on both the generated traffic rate and the size of the packet. The generated traffic in WSN can be characterized in two

main classes, *periodic* and *event based* [7]. Periodic traffic is generated especially in monitoring applications, where a certain characteristic of the physical phenomenon, e.g., temperature, humidity, etc, is constantly monitored. For event-based traffic, data is only generated if the sensing results match a certain criterion, i.e., an event occurs. Since collision is higher for higher data rate, in this analysis, we consider a worst case scenario where each node monitors the physical phenomenon and generates packets accordingly.

Let us assume that the sensor node has an average sampling rate of b bits/s. Denoting the length of the packet payload as l_D , on the average, the packet generation rate of a node i is $\lambda_{ii} = b/l_D$ pkts/s. Since a node will also relay packets from other nodes to the sink, the packet transmission rate of a node is higher than this value. If a routing scheme that equally shares the network load among nodes is considered, on the average, the packet transmission rate of a node is $\lambda_i = c_i \lambda_{ii}$, where $c_i > 1$. Consequently, λ in (101) is given by $\lambda = \sum_{i=1}^M \lambda_i$, where the number of nodes that are in the transmission range of a node is given by $M - 1$. It can be clearly seen that an increase in packet size directly affects the rate of channel access attempts and hence the traffic on the channel. The traffic rate affects both the probability of successful carrier sense, p_{cs} , and the probability of no collisions, p_{coll} . Although the packet transmission duration, T_{comm} , is increased with increasing packet length, the overall probability of successful channel access will improve.

The MAC failure rate, $1 - p_{cs}p_{coll}$, which is found using (98), (99), (100), and (101) as a function of payload length for different values of M and $b = 5$ bits/s for a WSN is given in Fig. 42(b). The above observation, where an increase in payload length decreases the MAC failure rate, can also be seen in Fig. 42(b). It is a well known fact that longer packet sizes result in higher efficiency, when the packet overhead is concerned. This important observation further motivates longer packet sizes when effects of collisions in WSN are considered.

One of the main shortcomings of using longer packet sizes is the increased susceptibility to channel errors. In this respect, FEC codes can be exploited to mitigate the effects of wireless errors on longer packets. Note that compared to ARQ protocols, when FEC codes are used, the generated traffic rate λ is not affected. On the other hand, the duration of packet transmission increases due to redundant bits in the transmitted packet. Although the error performance of the communication is improved, the collision rate may increase when FEC codes are used. In order to investigate the effect of FEC codes, the MAC failure rate, $1 - p_{cs}p_{coll}$, is also found for different code rates as shown in Fig. 42(c). The curve for 1 corresponds to no FEC coding, while the code rate is decreased corresponding to powerful codes. It is clearly seen that, higher packet sizes still decrease the MAC failure rate. However, this effect is small for higher rate FEC codes since the packet transmission duration, T_{comm} , is increased.

The results of this analysis reveal that longer packet sizes are favorable in WSN when collisions are considered alone. This is motivated by the cross-layer interdependency of generated traffic and the packet size in WSN. Moreover, the susceptibility of longer packets to wireless errors can be alleviated through the use of FEC. Although these results are significant, an overall assessment of the packet size on the network performance is required. Especially the energy consumption of the overall communication needs to be investigated to provide a complete solution for packet size optimization in WSN. In the following sections, we provide a comprehensive analysis of packet size based on the observations in this section.

6.4 Packet Size Optimization Framework

As explained in Section 6.2, most of the work on packet size determination in wireless networks focus on point-to-point links. Generally, the optimal packet size was determined based on an efficiency metric [62], where the overall energy consumption

to transmit a packet in one hop is compared with the ideal energy consumption to transmit the payload of the packet.

In multi-hop WSN, however, where the broadcast nature of the wireless channel plays an important role, it is not trivial to present an *ideal* way to transmit a packet from a sensor node to the sink. First, the multi-hop nature of the WSN necessitates an end-to-end analysis since an optimization framework considering only a single link may not reveal the intrinsic properties of the multi-hop communication. More specifically, it is not clear which route is the *ideal* route between a particular node and a sink. Second, the broadcast nature of the wireless channel results in neighbor nodes to overhear a communication, and hence to *waste* energy. On the other hand, this overhearing may be beneficial to the network with respect to routing, clustering or channel quality measurements and thus, needs to be considered in the optimization framework. In light of these cross-layer effects of communication in WSN, a definition of *efficiency metric* in WSN may not be feasible since an *ideal* way of communication may not be defined.

Consequently, we formalize our optimization framework based on end-to-end metrics. More specifically, the communication performance metrics such as energy consumption, throughput, latency, and reliability are derived for a flow that is generated at a sensor node until it reaches the sink. We formalize our optimization solution by using three different objective functions. Each objective function highlights a different aspect of communication in WSN and can be selected according to the requirements of the application in use. Furthermore, we investigate the effects of end-to-end latency and reliability constraints that may be required by a particular application. As a result, a flexible optimization framework is developed to determine the optimal packet size in WSN. Before we present the optimal packet size results, in the next sections, we first define these objective functions.

6.4.1 Optimization Metrics

We define three optimization metrics as packet throughput, energy consumption, and resource utilization.

6.4.1.1 Packet Throughput

The *packet throughput* function considers the end-to-end packet success rate and the end-to-end delay to transmit a packet of payload l_D . Consequently, the packet throughput objective function, \mathcal{U}_{tput} , is defined as follows:

Definition 1 Packet throughput:

$$\mathcal{U}_{tput} = \frac{l_D (1 - PER_{e2e})}{T_{flow}}, \quad (102)$$

where l_D is the payload length, PER_{e2e} is the end-to-end packet error rate, which considers the entire packet including header and trailer, and T_{flow} is the end-to-end latency, which is the time spent between a packet is generated at a sensor and received at the sink through the multi-hop route.

6.4.1.2 Energy Consumption

The energy consumption for useful bit between a particular node and the sink. More specifically the utility function, \mathcal{U}_{eng} , is the energy per useful bit, which is defined as the total energy consumed for each bit successfully received at the sink as defined below:

Definition 2 Energy per useful bit:

$$\mathcal{U}_{eng} = \frac{E_{flow}}{l_D (1 - PER_{e2e})}, \quad (103)$$

where E_{flow} is the end-to-end energy consumption to transport a packet from a source to a destination. Minimizing \mathcal{U}_{eng} results in optimal packet size values that achieve high energy efficiency.

6.4.1.3 Resource Utilization

Minimizing energy consumption is the main goal in WSN. However, \mathcal{U}_{eng} in (103) does not consider the delay associated with the optimal value of the packet size that minimizes this function. Since end-to-end delay is also important in WSN, the third objective function considers the energy-delay product consumed for useful bit. We refer to this function as resource utilization:

Definition 3 Resource utilization:

$$\mathcal{U}_{res} = \frac{E_{flow}T_{flow}}{l_D(1 - PER_{e2e})} \quad (104)$$

Minimizing \mathcal{U}_{res} leverages the tradeoff between energy consumption and latency. We believe this utility function serves best for practical implementation of WSN, especially for delay sensitive traffic.

In the remainder of this section, we present the derivation of each term of the utility functions, \mathcal{U}_{tput} in (102), \mathcal{U}_{eng} in (103), and \mathcal{U}_{res} in (104). For the derivation, we consider a channel-aware routing algorithm, where the next hop is determined according to the received signal to noise ratio (SNR) of a packet sent from a specific node i at a distance D from the sink. Among the neighbors of i , the neighbor, j , that is closest to the sink and with SNR value, $\psi_j > \psi_{Th}$ is selected as the next hop, where ψ_{Th} is the received SNR threshold. Note that this approach can be implemented using a cross-layer approach as in [9] or through signaling [67]. The medium access is performed through RTS-CTS-DATA exchange in addition to ACK and retransmissions for ARQ. To illustrate specific results for FEC codes, we consider block codes due to their energy efficiency and lower complexity compared to convolutional codes [68, 62]. We consider a block code, which is represented by (n, k, t) , where n is the block length, k is the payload length, and t is the error correcting capability in bits. In our analysis, we use extended BCH codes, which enable the evaluation of the effect of error correction capability, t , by fixing the block length, $n = 128$. However, other

FEC schemes can also be used in our framework.

First, we explain the channel model used for the analysis. Then, the general analysis model and the derivation of each component is given.

6.4.2 Channel Model

For our derivations, we use the log-normal channel model, which has been experimentally shown to model the low power communication in WSN accurately [88]. In this model, the received power at a receiver at distance d from a transmitter is given by

$$P_r(d) = P_t - PL(d_0) - 10\eta \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma, \quad (105)$$

where P_t is the transmit power in dBm, $PL(d_0)$ is the path loss at a reference distance d_0 in dB, η is the path loss exponent, and X_σ is the shadow fading component, with $X_\sigma \sim \mathcal{N}(0, \sigma)$. Moreover, the SNR at the receiver is given by $\psi(d) = P_r(d) - P_n$ in dB, where P_n is the noise power in dBm.

Considering the shadow fading component, X_σ , the probability that the received SNR, ψ_j , of the channel between two nodes n_i and n_j that are at a distance $d_{(i,j)}$ is above some threshold, ψ_{Th} , is

$$\begin{aligned} \text{P}\{\psi_j > \psi_{Th}\} &= \text{P}\{X_\sigma > \beta(d_{(i,j)}, \psi_{Th})\} \\ &= Q\left(\frac{\beta(d_{(i,j)}, \psi_{Th})}{\sigma}\right), \end{aligned} \quad (106)$$

where

$$\beta(d, \psi_{Th}) = \psi_{Th} + P_n - P_t + PL(d_0) + 10\eta \log_{10}\left(\frac{d}{d_0}\right) \quad (107)$$

and $Q(x) = 1/\sqrt{2\pi}(\int_x^\infty e^{-(t^2/2)}dt)$. Based on this channel model, the end-to-end energy consumption for a packet, E_{flow} , will be derived as will be shown in the following section.

6.4.3 End-to-End Energy Consumption

The end-to-end energy consumption, E_{flow} , of a packet sent from a node at a distance D from the sink is derived based on the model in Chapter 5 and [79], which is extended

here to accommodate medium access collisions according to the discussion in Section 6.3.2. We first summarize the end-to-end energy consumption framework and the extensions are discussed then. The end-to-end energy consumption for a packet, E_{flow} , is given as follows:

$$E_{flow} = E[E_h] \left(\frac{D - R_{inf}}{E[d_h]} + 1 \right), \quad (108)$$

where $E[E_h]$ is the expected energy consumption per hop, D is the distance between the source node and the sink, R_{inf} is the approximated transmission range of a node, and $E[d_h]$ is the expected hop distance. Accordingly, the expected hop distance is given by

$$E[d_h] = \rho \delta \int_{\gamma_{min}}^D \int_{-\alpha\gamma}^{\alpha\gamma} \gamma d_{(i,j)} Q \left(\frac{\beta}{\sigma} \right) e^{-M(1-p_k)} d\alpha d\gamma, \quad (109)$$

where ρ is the node density, δ is the duty cycle value, γ is the distance between the next hop and the sink, $d_{(i,j)}$ is the distance between the source node and the next hop, β is as given in (107), and $e^{-M(1-p_k)}$ is the probability that the next hop is at least at a distance γ from the sink.

Similarly, the expected energy consumption per hop, $E[E_h]$, in (108) can be found as

$$E[E_h] = \rho \delta \int_{\gamma_{min}}^D \int_{-\alpha\gamma}^{\alpha\gamma} \gamma E[E_j] Q \left(\frac{\beta}{\sigma} \right) e^{-M(1-p_k)} d\alpha d\gamma, \quad (110)$$

where $E[E_j]$ is the expected energy consumption, which is given as

$$E[E_j] = \int_{\psi_{Th}}^{\infty} E_{comm}(\psi, d_{(i,j)}) f_{\Psi}(\psi, d_{(i,j)}) d\psi. \quad (111)$$

In (111), $f_{\Psi}(\psi, d_{(i,j)}) = 1/(\sigma\sqrt{2\pi})e^{-\frac{\beta^2}{2\sigma^2}}$ is the pdf of the SNR. Moreover $E_{comm}(\psi, d_{(i,j)})$ is the energy consumption for communication between nodes i and j given that they are at a distance $d_{(i,j)}$ with a SNR value of ψ at node j , which has three components as given by³

$$E_{comm} = E_{TX} + E_{RX} + E_{neigh}, \quad (112)$$

³We drop the indices ψ and $d_{(i,j)}$ for ease of illustration.

where E_{TX} is the energy consumed by the node transmitting the packet (node i), E_{RX} is the energy consumed by the node receiving the packet (node j), and E_{neigh} is the energy consumed by the neighbor nodes [79].

In Chapter 5, E_{TX} , E_{RX} , and E_{neigh} are found considering the four-way RTS-CTS-DATA-ACK handshake for ARQ or the three-way RTS-CTS-DATA handshake for FEC codes. However, the affect of collision rate on the success of the transmission of each packet were not considered. The energy consumption for transmitter node, E_{TX} , for ARQ and FEC are found to be:

$$E_{TX}^{ARQ} = \frac{p_{cs}}{1 - p_{cs} + p_{cs}p_{coll}(p_s^C)^3p_s^D} \left\{ E_{sense} + E_{tx}^C + p_{coll}p_s^C E_{rx}^C + (1 - p_{coll}p_s^C) E_{t/o}^C \right. \\ \left. + p_{coll}(p_s^C)^2 E_{tx}^D + p_{coll}(p_s^C)^2 p_s^D E_{rx}^C + p_{coll}(p_s^C)^2 (1 - p_s^D) E_{t/o}^D \right\} \quad (113)$$

$$E_{TX}^{FEC} = \frac{p_{cs}}{1 - p_{cs} + p_{cs}p_{coll}(p_s^C)^2} \left\{ E_{sense} + E_{tx}^C + p_{coll}p_s^C (E_{rx}^C + E_{dec}^C) \right. \\ \left. + (1 - p_{coll}p_s^C) E_{t/o}^C \right\} + p_{cs}p_{coll}(p_s^C)^2 E_{tx}^D \quad (114)$$

respectively, where the effects of collisions are also reflected. The first term in each equation is the retransmission rate, where p_{cs} is the probability of successful carrier sense given in (98), p_{coll} is the probability of no collisions given in (100), and p_s^C and p_s^D are the probability of success for control and data packets, respectively, which are found using (96) and (97). The terms in the parentheses in (113) and (114) denote the average energy consumption resulting from transmission and reception of control and data packets. Similarly, the values for E_{RX} and E_{neigh} can also be found, which are not shown here for space limitations. Once the E_{comm} in (112) is found, the end-to-end energy consumption, E_{flow} , is calculated using (109-111) in (108). Moreover, the end-to-end latency, T_{flow} , and the end-to-end packet error rate, PER_{e2e} , are also found using the similar approach in Chapter 5 and the extension discussed above. Next, we formalize our optimization problem for packet size optimization in WSN.

6.5 Packet Size Optimization

It can be observed from (108-112) that the energy consumption of a flow is mainly affected by the packet size and the SNR threshold parameter, ψ_{Th} . This is also true for end-to-end packet error rate and end-to-end delay. Consequently, these two parameters affect the utility functions in (102-104). In Fig. 43(a) and Fig. 43(b), the energy per useful bit is shown as a function of payload length l_D for three different values of SNR threshold ψ_{Th} for ARQ and FEC with $t=5$, respectively. It is also evident from these figures that the optimum value for packet size for a specific objective function depends on the routing decisions and hence ψ_{Th} .

The effect of the SNR threshold value ψ_{Th} can be explained as follows. ψ_{Th} controls the minimum quality of the wireless channel at each hop since the routes are constructed according to this value. Moreover, the average hop length increases for a lower ψ_{Th} value [79]. This has two consequences in overall energy consumption of the communication. If low quality channel is chosen, the energy consumption may increase due to retransmissions or packet drops. On the other hand, since longer hops are constructed, the overall energy consumption can still be decreased. For a low ψ_{Th} value, low quality links may be chosen, which necessitates smaller packet sizes to maintain an acceptable reliability. However, this causes inefficiency due to increased affect of header and trailer of the packet and may decrease the energy efficiency. As a result, packet size optimization is significantly affected by the routing decisions. Therefore, the value of ψ_{Th} is also considered in our optimization framework.

In Fig. 43(a) and Fig. 43(b), it is observed that the energy consumption first decreases with increasing payload size. This decrease is due to many reasons. As discussed in Section 6.3.2, longer packet sizes result in lower contention in the wireless channel. As a result, the energy consumed to transmit a single packet at one hop decreases. Moreover, as shown in Fig. 42(a), packet error rate is not affected when the packet size is increased up to a specific value. As a result, the efficiency of

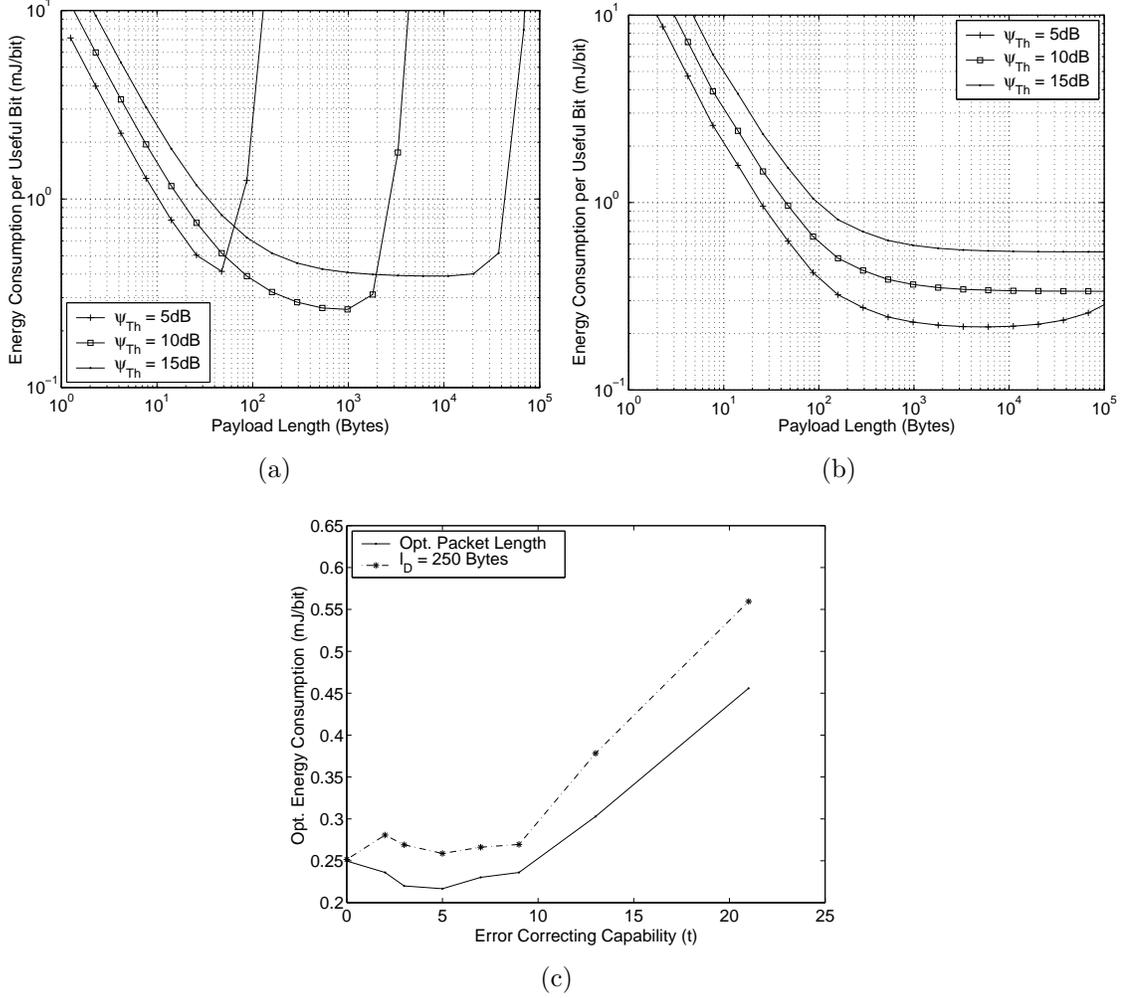


Figure 43: Energy per useful bit (103) vs. payload length for (a) ARQ and (b) BCH with $t=5$. (c) Energy consumption per bit with optimum packet length and $l_D = 250$ bytes.

the communication improves. However, the utility function has a minimum at a particular payload length value and increases above this value. This is mainly due to the increasing packet errors and hence retransmissions or packet drops. It can also be observed that the choice of SNR threshold value, ψ_{Th} , determines the minimum of the energy consumption and also the optimum packet size associated with this value.

It is clear that the choice of both packet size and the SNR threshold value, ψ_{Th} , is crucial to minimize/maximize the utility functions (102-104). Hence, in our optimization problem, our goal is to find both ψ_{Th} and packet size l_D . Consequently, our

Table 3: Optimal l_D and ψ_{Th} values

Problem	ECC Type	l_D (byte)	ψ_{Th} (dB)	\mathcal{U}_{eng} (mJ/bit)	T_{flow} (ms)	\mathcal{U}_{tput} (kbps)	p_{flow}
\mathbf{P}_{max}^{tput}	ARQ (N=4)	152	7.5	0.2659	76.7	15.8	0.9996
	FEC (t=5)	1103	4.5	0.2204	1419	6.1	0.9813
	FEC (t=7)	546	3	0.2450	944	4.5	0.9674
	FEC (t=9)	1133	2.5	0.2389	2482	3.6	0.9840
\mathbf{P}_{min}^{eng}	ARQ (N=4)	473	9	0.2497	248.3	15.2	0.9994
	FEC (t=5)	4933	5	0.2167	6521	5.9	0.9849
	FEC (t=7)	2915	3.5	0.2302	5106	4.4	0.9719
	FEC (t=9)	5342	3	0.2360	12019	3.5	0.9910
\mathbf{P}_{min}^{res}	ARQ (N=7)	25	5	0.5190	17.5	11.2	0.9998
	FEC (t=5)	53	3.5	0.5028	93.9	4.4	0.9797
	FEC (t=7)	39	2	0.5917	90.7	3.2	0.9495
	FEC (t=9)	26	1	0.7496	85.4	2.3	0.9437

optimization problems become:

\mathbf{P}_{max}^{tput} : Packet throughput maximization

$$\textit{Given} : D, \eta, \sigma, n, k, t \quad (115)$$

$$\textit{Find} : \psi_{Th}, l_D \quad (116)$$

$$\textit{Maximize} : \mathcal{U}_{tput} = \frac{l_D(1-PER_{e2e})}{T_{flow}} \quad (117)$$

\mathbf{P}_{min}^{eng} : Energy consumption per useful bit minimization

$$\textit{Given} : D, \eta, \sigma, n, k, t \quad (118)$$

$$\textit{Find} : \psi_{Th}, l_D \quad (119)$$

$$\textit{Minimize} : \mathcal{U}_{eng} = \frac{E_{flow}}{l_D(1-PER_{e2e})} \quad (120)$$

\mathbf{P}_{min}^{res} : Resource utilization minimization

$$\textit{Given} : D, \eta, \sigma, n, k, t \quad (121)$$

$$\textit{Find} : \psi_{Th}, l_D \quad (122)$$

$$\textit{Minimize} : \mathcal{U}_{res} = \frac{E_{flow}T_{flow}}{l_D(1-PER_{e2e})} \quad (123)$$

The optimum values of l_D and ψ_{Th} have been found using the optimization toolbox of MATLAB. In Table 3, the results are shown for the three optimization problems, \mathbf{P}_{max}^{tput} , \mathbf{P}_{min}^{eng} , and \mathbf{P}_{min}^{res} . The columns l_D , ψ_{Th} , \mathcal{U}_{eng} , T_{flow} , \mathcal{U}_{tput} , and p_{flow} refer to optimal payload length, optimal SNR threshold, energy consumption per useful bit, end-to-end latency, packet throughput, and end-to-end success rate, respectively. When the throughput maximization problem, \mathbf{P}_{max}^{tput} , is concerned, ARQ scheme with a payload length of 152 bytes achieves the highest throughput, \mathcal{U}_{tput} . Note that for this optimization problem, FEC schemes result in lower throughput. This is due to the encoding and decoding latency incurred by the FEC schemes as well as increased packet length and the increased collisions and transmission latency associated with it.

The energy per bit minimization problem, \mathbf{P}_{min}^{eng} , results in favor of FEC schemes with higher packet sizes, l_D . It can be observed that the FEC can accommodate longer packet sizes without affecting the energy efficiency. However, an interesting result is that very high packet sizes are determined for this problem. This is due to the fact that longer packet sizes are more efficient when a sufficient link quality is guaranteed. This result is evident when the optimal ψ_{Th} values are compared for \mathbf{P}_{max}^{tput} and \mathbf{P}_{min}^{eng} . However, this selection results in very high end-to-end latency values as shown in the T_{flow} column. Furthermore, optimal packet values may not be feasible in current wireless sensor nodes. As an example, the ZigBee standard defines 250 byte as the maximum packet length [37]. In Fig. 43(c), the optimal energy consumption per bit is shown for both the optimum l_D values and a fixed 250 byte packet size for various error correcting capability, t . The values for $t = 0$ correspond to the ARQ case. It is evident that when high packet lengths can not be accommodated, energy efficiency decreases as much as 20%.

The results for \mathbf{P}_{min}^{res} show how our proposed resource utilization metric leverages energy consumption and end-to-end latency performance. By compromising from

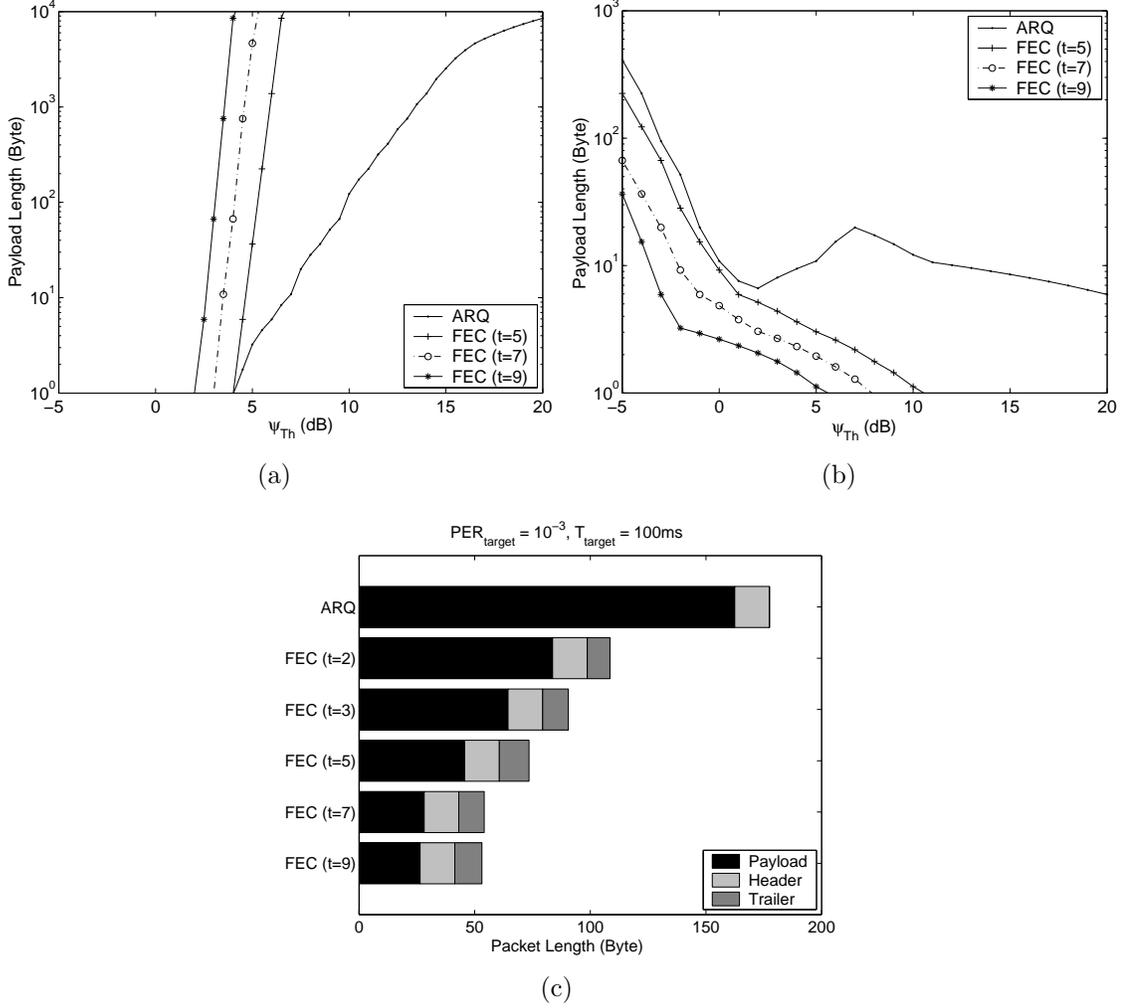


Figure 44: Boundaries for the acceptable ξ_{Th} - packet size region subject to (a) end-to-end packet error rate and (b) end-to-end delay. (c) Optimal packet length for different error control techniques.

energy consumption per bit, \mathcal{U}_{end} , (2-fold increase), end-to-end latency, T_{flow} , can be significantly decreased (15 times decrease) leading to optimal payload length, l_D , values in the range of 25-50 bytes. In this case, when energy efficiency is concerned, FEC with $t = 5$ and payload length of $l_D = 53$ bytes is optimal. If throughput, \mathcal{U}_{tput} , and end-to-end success, p_{flow} are also important, ARQ scheme with packet length of $l_D = 25$ bytes is the optimal value with a slight increase in energy consumption per bit, \mathcal{U}_{eng} .

Table 4: Overview of Results

Energy.	Application Requirement			Error Control	l_D (byte)
	Throughput.	Delay	Reliability		
-	High	-	-	ARQ	152
Low	-	-	-	FEC	Max*
Low	-	Low	-	ARQ	25
Low	-	Low	High	ARQ	162

As shown in Table 3, minimizing energy consumption per bit, \mathcal{U}_{eng} , independently leads to possibly unacceptable end-to-end latency, T_{flow} , and success rate, p_{flow} , values. Since long payload lengths, l_D , lead to both degradation of success rate, p_{flow} , and increase in latency, T_{flow} , the optimization problem should also take this into account. Furthermore, certain WSN applications that focus mainly on real-time monitoring require strict end-to-end latency and success rate (reliability) constraints. Based on these observations, we need to develop an optimization framework for energy minimization subject to delay and reliability constraints.

The optimization problem with delay and reliability constraints can be cast as follows:

$\mathbf{P}_{min,con}^{eng}$: Energy consumption per useful bit minimization with delay and reliability constraints

$$Given : D, \eta, \sigma, n, k, t, P_{target}, T_{target} \quad (124)$$

$$Find : \psi_{Th}, l_D \quad (125)$$

$$Minimize : \mathcal{U}_{eng} = \frac{E_{flow}}{l_D(1-PER_{e2e})} \quad (126)$$

$$Subject\ to : T_{flow} \leq T_{target} \quad (127)$$

$$PER_{e2e} \leq P_{target} \quad (128)$$

where T_{target} and P_{target} are end-to-end delay and end-to-end packet error rate constraints, respectively.

Considering an end-to-end packet error rate of $P_{target} = 10^{-3}$ and end-to-end latency of $T_{target} = 100ms$, in Fig. 44(a) and Fig. 44(b), the boundaries for the feasible $\psi_{Th} - l_D$ region for the optimization is shown for various error control techniques, respectively. In Fig. 44(a), the feasible region is the area that falls to the right and below each curve. Similarly in Fig. 44(b), the $\psi_{Th}-l_D$ values that fall to the left and below each curve satisfy the end-to-end delay requirement, T_{target} . Consequently, the composite of each curve defines the feasible region for the optimization problem $\mathbf{P}_{min,con}^{eng}$. In Fig. 44(c), the optimum packet sizes including the header and trailer for each error control technique are shown. Note that the values for FEC with $t > 9$ are not shown since the end-to-end delay requirements, T_{target} , cannot be satisfied for FEC codes with very high error correcting capabilities. This observation also agrees with our results in Chapter 5. An important observation is that when end-to-end delay and reliability constraints are used, the optimal packet sizes range between 25-160 bytes. When end-to-end delay, T_{flow} , and success rate, p_{flow} , is important, our results show that ARQ with optimal payload length of $l_D = 162$ bytes result in optimal performance. An overview of the result for the cross-layer packet size optimization is shown in Table 4.

6.6 Packet Size Optimization in Wireless Underwater and Underground Sensor Networks

In this section, we will extend our analysis of packet size optimization for underwater and underground sensor networks. Considering the unique properties of these networks compared to terrestrial wireless sensor networks, the optimization framework will be extended and the optimal packet sizes for each of these networks will be found. We first present the channel models for these networks and discuss the unique properties of these networks. Then, we provide the results for our optimization problems in these challenging environments.

6.6.1 Underwater Channel Model

Wireless Underwater Sensor Networks are characterized by an acoustic communication channel [5]. The acoustic channel is characterized by the Urlick path loss formula which is given below:

$$TL(d, f) = \chi \cdot \log(d) + \alpha(f) \cdot d + A \quad (129)$$

where the path loss, $TL(d, f)$ is shown in dB as a function of internode distance d and operating frequency f . The term χ is the geometric spreading which can be spherical for deep water and cylindrical for shallow water. The last term A is the transmission anomaly and accounts for the degradation of the acoustic intensity caused by multiple path propagation, refraction, diffraction, and scattering of sound. Moreover, the propagation delay in underwater acoustic channel varies between 1460 m/s and 1520 m/s.

Moreover, we model the randomness in the channel using the Rayleigh fading channel model [35], where the envelope of the signal is modeled as a Rayleigh distributed random variable, α . Consequently, the received energy per bit per noise power spectral density is given by $\gamma = \alpha^2 E_b / N_o$, which has a distribution as follows:

$$f_{\Gamma}(\gamma) = \frac{1}{\gamma_0} \exp\left(-\frac{\gamma}{\gamma_0}\right), \quad (130)$$

where $\gamma_0 = E[\alpha^2] E_b / N_o$. E_b / N_o can be directly found from the signal-to-noise ratio (SNR) of the channel, which is given in dB in underwater acoustic channels as

$$\psi_{0,dB}(d, f) = SL_{dB\text{re}uPa} - TL_{dB}(d, f) - NL_{dB\text{re}uPa} \quad (131)$$

where $SL_{dB\text{re}uPa}$ and $NL_{dB\text{re}uPa}$ are the signal level at the transmitter and the noise level given in dB with reference to μPa . Then, $E_b / N_o = \psi B_N / R$, where $\psi = 10^{\psi_{dB}(d,f)/10}$, B_N is the noise bandwidth, and R is the data rate. The signal level, SL , in (131) can be related to the intensity, I_t and, hence, the transmit power, P_t of

the transceiver as follows:

$$I_t = \frac{P_t}{2\pi 1m H}, \quad (132)$$

$$SL = 10\log\left(\frac{I_t}{0.67 \times 10^{-18}}\right), \quad (133)$$

where P_t is the transmit power in Watts, and H is the depth in m .

An important observation about underwater communication is the effect of surface reflected rays as shown through experiments in [18]. More specifically, it has been observed that the received signal follows a 2-path Rayleigh model, where the direct path and the surface reflected path contributes to the received signal strength. Moreover, the signal from each path can be modeled as an independent Rayleigh distributed signal. Consequently, the bit error rate experienced at a node is a result of combination of these signals.

In underwater acoustic modems, generally, binary non-coherent FSK is used. Therefore, the bit error rate in underwater acoustic networks is given by:

$$p_b^{FSK} = \frac{1}{2 + \gamma_0} \quad (134)$$

Finally, (106) for underwater acoustic channel is given by

$$P\{\psi_j > \psi_{Th}\} = \exp\left(-\frac{\psi_{Th}}{\psi_0}\right), \quad (135)$$

where ψ_0 is given in (131).

In order to apply our results to the underwater channel, the bit error rate and the probability that the received signal to noise ratio (SNR) is higher than a threshold needs to be found. Since the 2-path Rayleigh channel model does not have a closed for expression for SNR distribution, we have performed simulations to find these values.

In our simulations, 5000 independent pairs of Rayleigh distributed random variables are generated. The received signal strength for each location is then found for these independent trials. Consequently, the empirical cdf found as a result of these trials is used for (106). This value is used to find the expected bit error rate, energy consumption, and latency at each point in the simulations.

6.6.2 Underground Channel Model

The channel model for underground wireless communication has been developed in [45]. This model follows a 2-path location-based Rayleigh fading channel model. While the details of this model can be found in [45], here, we provide an overview of the underground channel model.

The path loss, L_p , in an underground environment is given as follows:

$$L_p = L_0 + L_{m1} + L_\alpha \quad (136)$$

where L_0 is the path loss due to attenuation in free space, L_{m1} is the attenuation loss due to the difference of the wavelength of the signal in soil, λ , compared to the wavelength in free space, λ_0 , and L_α is the transmission loss caused by attenuation. Consequently, L_{m1} and L_α are given by:

$$L_{m1} = 154 - 20\log(f)(Hz) + 20\log(\beta) , \quad (137)$$

$$L_\alpha = 8.69\alpha d , \quad (138)$$

respectively, where f is the operating frequency, $alpha$ is the attenuation constant in $1/m$ and $beta$ is the phase shifting constant in $radian/m$.and $beta$. Given that the path loss in free space is $L_0 = 20\log(4\pi\lambda_0)$, the path loss of an EM wave in soil is given by:

$$L_p = 6.4 + 20\log(d) + 20\log(\beta) + 8.69\alpha d , \quad (139)$$

Note that the path loss, L_p , in (139) depends on the attenuation constant, α , and the phase shifting constant, β . The values of these parameters depend on the dielectric properties of soil. Using Peplinski's principle [53], the dielectric properties of soil in the 0.3-1.3 GHz band can be calculated as follows:

$$\epsilon = \epsilon' - j\epsilon'' , \quad (140)$$

$$\epsilon' = 1.15\left[1 + \frac{\rho_b}{\rho_s}(\epsilon_s^{\alpha'}) + m_v^{\beta'} \epsilon_{f_w}^{\alpha'} - m_v\right]^{1/\alpha'} - 0.68, \quad (141)$$

$$\epsilon'' = [m_v^{\beta''} \epsilon_{f_w}^{\alpha'}]^{1/\alpha'} , \quad (142)$$

where ϵ_m is the relative complex dielectric constant of the mixture of soil and water, m_v is the volumetric water content (VWC) of the soil, ρ_b is the bulk density in grams per cubic centimeter, $\rho_s = 2.66g/cm^3$ is the specific density of the solid soil particles, $\alpha' = 0.65$ is an empirically determined constant, and β' and β'' are empirically determined constants, dependent on soil-type and given by

$$\beta' = 1.2748 - 0.519S - 0.152C , \quad (143)$$

$$\beta'' = 1.33797 - 0.603S - 0.166C , \quad (144)$$

respectively, where S and C stand for the mass fractions of sand and clay, respectively. ϵ'_{f_w} and ϵ''_{f_w} are the real and imaginary parts of the relative dielectric constant of water. The Peplinski principle [53] governs the value of the complex propagation constant of the EM wave in soil, which is given as $\gamma = \alpha + j\beta$ with

$$\alpha = \omega \sqrt{\frac{\mu\epsilon'}{2} [\sqrt{1 + (\frac{\epsilon''}{\epsilon'})^2} - 1]} , \quad (145)$$

$$\beta = \omega \sqrt{\frac{\mu\epsilon'}{2} [\sqrt{1 + (\frac{\epsilon''}{\epsilon'})^2} + 1]} , \quad (146)$$

where $\omega = 2\pi f$ is the angular frequency, μ is the magnetic permeability, and ϵ' and ϵ'' are the real and imaginary parts of the dielectric constant as given in (141) and (142), respectively.

As we have shown in [45], if the sensors are buried at a depth less than 2 m, i.e., *low depth*, the influence of the wave reflection by ground surface should be considered. Consequently, the total path loss of two-path channel model can be deduced as follows:

$$L_f(dB) = L_p(dB) - V_{dB} , \quad (147)$$

where L_p is the path loss due to the single path given in (139) and V_{dB} is the attenuation factor due to the second path in dB, i.e., $V_{dB} = 10 \log V$.

Consider the case where two sensors are buried at a depth of H_1 and H_2 , respectively, with a horizontal distance of L , and an end-to-end distance of d . Then, the attenuation factor, V , can be deduced as follows [15]:

$$\begin{aligned}
V^2 &= 1 + (\Gamma \cdot \exp(-\alpha\Delta(r)))^2 \\
&\quad - 2\Gamma \exp(-\alpha\Delta(r)) \\
&\quad \times \cos\left(\pi - \left(\phi - \frac{2\pi}{\lambda}\Delta(r)\right)\right), \tag{148}
\end{aligned}$$

where, Γ and ϕ are the amplitude and phase angle of the reflection coefficient at the reflection point P, $\Delta(r) = r - d$, is the difference of the two paths and α is the attenuation constant given in (145).

Finally, assuming 2PSK is used as the modulation, the bit error rate (BER) is shown as

$$p_b = \frac{1}{2} \operatorname{erfc}(\sqrt{SNR}), \tag{149}$$

where $\operatorname{erfc}(\cdot)$ is the error function and SNR is given by

$$\psi = P_t - L_f - P_n, \tag{150}$$

where P_t is the transmit power, L_f is the total path loss given in (147), and P_n is the energy of noise, which is found to be -103 dBm. Next, we present the results of our packet size optimization for underwater and underground sensor networks.

6.6.3 Results

In this section, we present the results for packet size optimization for underwater and underground environment. More specifically, the optimum packet size is found by solving three different optimization problems, i.e., \mathbf{P}_{min}^{eng} , \mathbf{P}_{max}^{tput} , and \mathbf{P}_{Min}^{eng} defined in Section 6.4. The solutions of these problems apply to specific application requirements of both underwater and underground wireless sensor networks. Furthermore, we present the results for constrained optimization problem, where the optimum packet

size is determined subject to reliability and latency constraints. We also discuss the effects of error control techniques, and environmental effects on the optimum packet size. Our results focus on two major communication environments, i.e., underwater (Section 6.6.3.1) and underground (Section 6.6.3.2). For the underwater case, we consider a deep water network, the sensor nodes are deployed at a depth so that reflections from surface or the bottom is negligible. These networks can be used for water quality measurement as well as drift observations. We also consider a deep water network, where the sensors are deployed close to the surface. For this case we use the two-ray underwater channel model explained in Section 6.6.1. These networks are used for ocean bottom measurements and surveillance. For the shallow water case, we consider a network deployed close to the surface of the water, where reflections from the sea surface needs to be considered. Finally, we investigate the packet size optimization problem in wireless underground sensor networks according to the channel model presented in Section 6.6.2 and [45]. In particular, we present the effects of bury depth and volumetric water content on the optimum packet size.

6.6.3.1 Wireless Underwater Sensor Networks

As explained before, for the underwater case, we investigate both the deep water and the shallow water sensor networks. It is important to note that the differences between propagation characteristics for deep and shallow water result in significantly different optimum packet lengths for these environments. Next, we present the results for packet size optimization in deep water and shallow water environment.

Deep Water Environment Communication in deep water can be mainly characterized by two different architectures. First, the sensors can be deployed in the middle of the ocean where the effects of reflections from surface and bottom are negligible. Consequently, the communication is characterized by a single ray communication as explained in Section 6.6.1. The results for the energy consumption minimization is

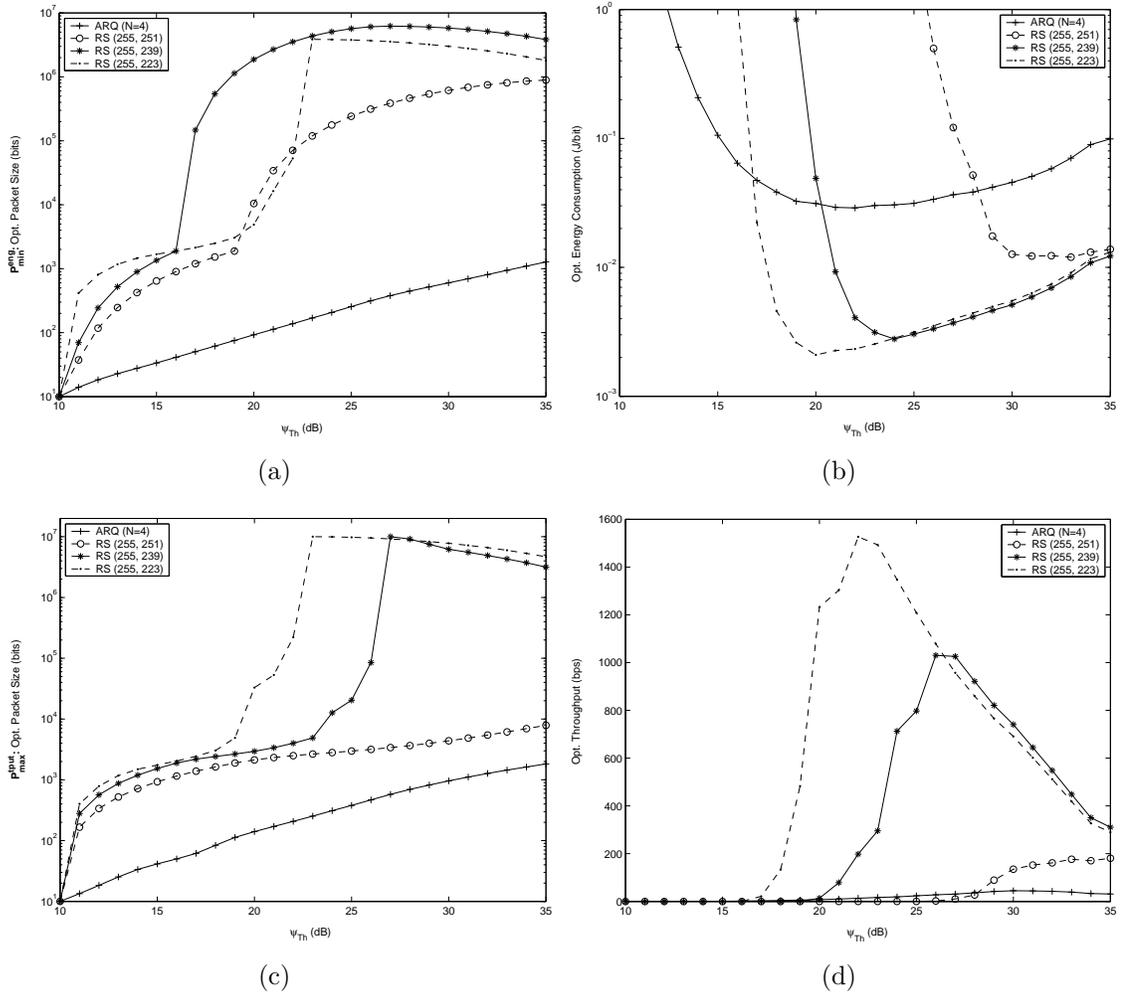


Figure 45: Optimum energy consumption vs. ψ_{Th} for (a) deep water and (b) shallow water, and optimum packet size for (c) deep water and (d) shallow water.

shown in Fig. 45(a) and Fig. 45(b), where the optimum packet sizes and the optimum energy consumption is shown as a function of the received SNR threshold, ψ_{Th} . We investigate four different error control schemes: ARQ, RS (255,251,2), RS (255,239,8), and RS (255,223,16). It can be observed from Fig. 45(a) that the optimum packet size increases with ψ_{Th} . This related to the increased channel quality for higher values of ψ_{Th} . When routes are constructed with higher ψ_{Th} value, longer packet sizes can be accommodated without errors. Consequently, the optimum packet size increases. Furthermore, while ARQ scheme results in packet sizes up to 125 bytes, longer packet

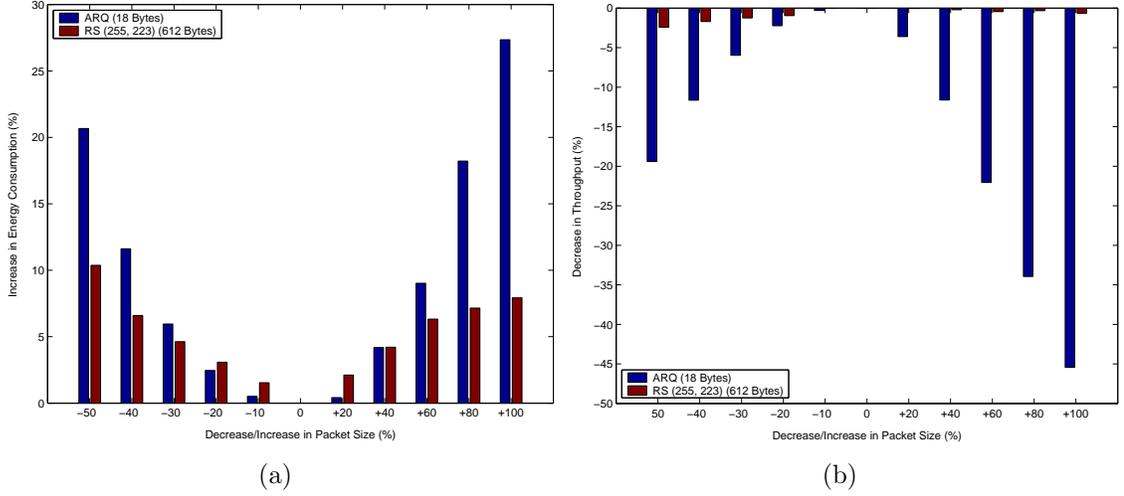


Figure 46: Sensitivity of (a) energy consumption and (b) throughput to packet size for ARQ and FEC in deep water.

sizes can be accommodated via RS codes. This is related to the error resiliency of these schemes in the underwater channel, which is characterized by high bit error rates. In Fig. 45(b), the optimum energy consumption is shown as a function of the received SNR threshold, ψ_{Th} , which shows that the minimum energy consumption is a function of both the packet size and ψ_{Th} . Moreover, ARQ scheme is highly energy inefficient when compared to the three RS codes. Since the underwater channel is characterized by high bit error rates and high propagation delay, the retransmissions required by the ARQ protocol significantly decreases the efficiency. This is further amplified since shorter packets can be supported by the ARQ scheme.

The results of the throughput maximization problem, $\mathbf{P}_{\max}^{\text{put}}$, which is defined in Section 6.5, are shown in Fig. 45(c) and Fig. 45(d). Figure 45(d) clearly shows that ARQ scheme is inefficient in terms of packet throughput since it provides values up to several bps. On the other hand, the packet throughput of the RS schemes increase for higher error correction capability, t . Similarly, the value of ϕ_{Th} that leads to the optimum value of packet throughput is lower for RS codes with higher t . This is particularly important since choosing a lower ϕ_{Th} value relaxes the constraints in route

construction. Furthermore, in Fig. 45(b), it can be observed that RS (255,223,16) is also more energy efficient when compared to the other schemes. As a result, lower deployment density can be possible for more powerful error correction codes without hampering the energy consumption.

The results of our optimization problem reveals that packet sizes of 18 bytes and 612 bytes are optimal for ARQ and RS (255,223,16) schemes, respectively. In Fig. 46(a) and Fig. 46(b), the sensitivity analysis of our optimization solutions are shown. More specifically, in both figures, the x -axis shows the deviation, in percentage, from the optimum packet size. We consider the cases where the optimum packet size is decreased up to the half of the optimum value and increased up to the twice the optimum value. In both figures, it can be observed that the ARQ protocol is more sensitive to the deviation from the optimum value. Furthermore, for RS (255,223,16), the energy consumption minimization problem is more sensitive to the deviations from the optimum value than the throughput maximization problem. As an example, doubling the optimum packet size leads to an increase of 7% in energy consumption, while a decrease of less than 1% in throughput. The sensitivity analysis is of particular importance since the actual value of the packet size that will be used in the network may be different from the optimum value because of hardware constraints. Hence, packet size values can be chosen according to this analysis.

The second case for the deep water analysis is ocean bottom deployment, where reflections from the ocean bottom is effective in communication. The optimization results for this case are shown in Figs. 47. In Fig. 47(a), the optimum packet size for the three optimization problems, \mathbf{P}_{min}^{eng} , \mathbf{P}_{max}^{tput} , and \mathbf{P}_{Min}^{eng} are shown as a function of the error correction capability, t . Note that $t = 0$ corresponds to the ARQ scheme. The corresponding optimum values for energy consumption, throughput, and resource utilization are also shown in Fig. 47(c) - Fig. 47(e). Since the resource utilization minimization problem also considers end-to-end latency, the optimum packet size

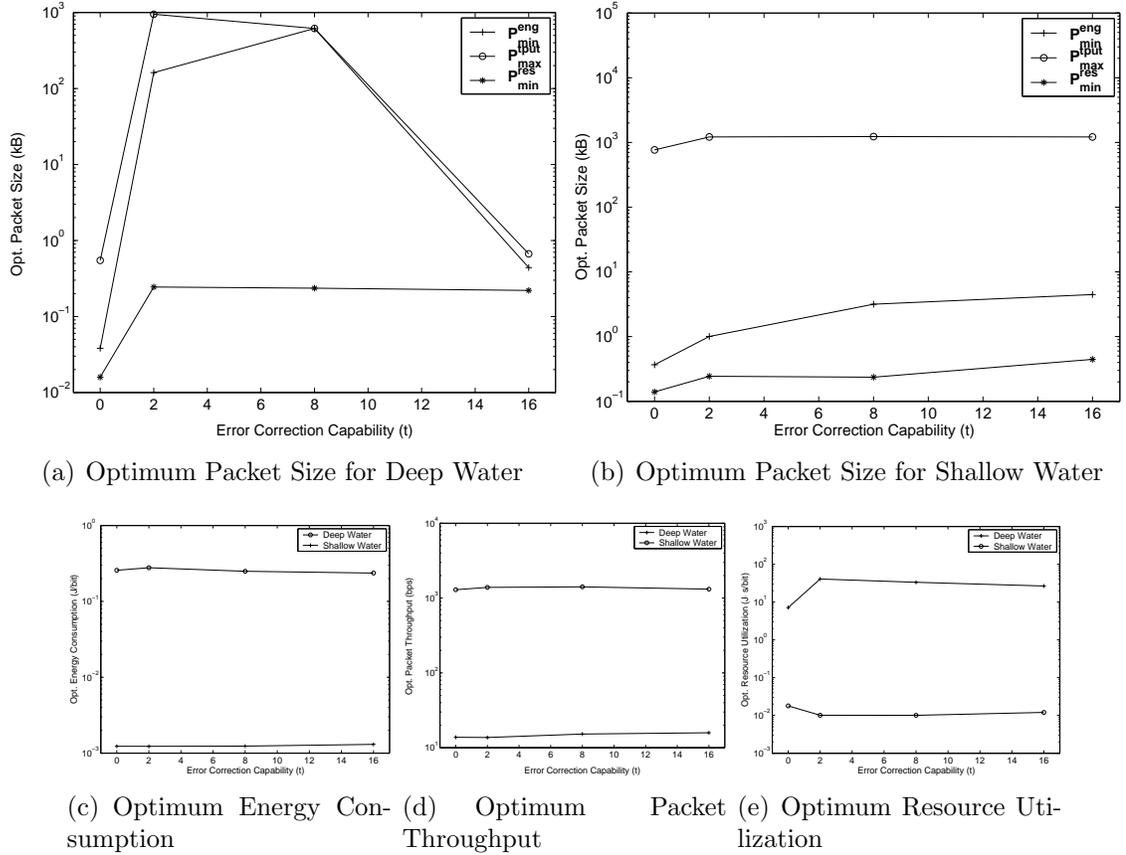


Figure 47: Optimization results for P_{min}^{eng} , P_{max}^{eng} , and P_{Min}^{eng} .

found for this problem is significantly smaller than that found for the other two problems as shown in Fig. 47(a). Moreover, for each problem, the ARQ scheme can accommodate smaller packet sizes than the FEC codes. As an example, the optimum packet size for the throughput maximization problem is 547 bytes for ARQ scheme and 616 kbytes for RS (255,239,8) code. Despite this significant increase in optimum packet size for the RS (255,239,8) code, the throughput is also increased 9% compared to the ARQ scheme. Note also that the reflection from the ground surface provides higher channel quality for a given node distance and results in a higher throughput for the ARQ scheme, which can be seen by comparing Fig. 45(d) and Fig. 47(d). It can be observed from Fig. 47(c) that the energy consumed per successfully received bit is very similar (0.25 J/bit for ARQ and 0.24 J/bit for RS (255,239,8), which shows that

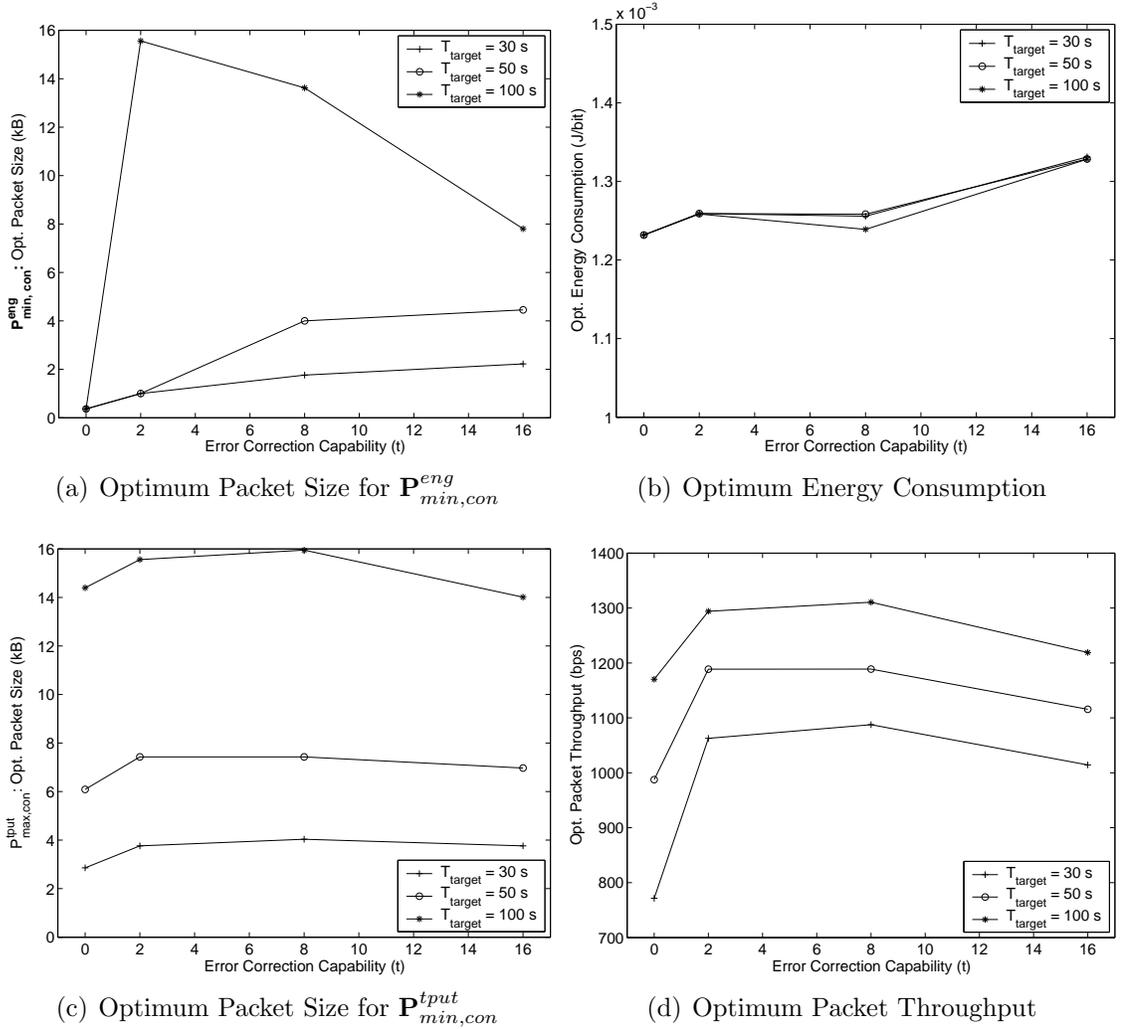


Figure 48: Constrained optimization results as a function of error correction capability, t , for different delay bounds in shallow water.

forward error correction code provide significant data transmission capability without hampering the energy consumption.

Shallow Water Environment Shallow water communication is characterized by a lower propagation loss compared to deep water because of both the cylindrical spreading and reflection from the water surface. In this section, we present our optimization results for the shallow water environment. The optimum packet size values for each of the optimization problems are shown in Fig. 47(b). Compared

to the deep water environment, in shallow water environment, channel errors are not severe and hence, the ARQ scheme also results in optimum packet size values comparable to the FEC schemes. Moreover, optimum packet size does not vary significantly for different error correction capabilities of FEC codes. As an example, the energy maximization problem results in optimum packet sizes between 300 bytes - 4.4 kbytes. Similarly, the resource utilization problem results in optimum packet sizes between 140-440 bytes. Furthermore, as shown in Fig. 47(c) and Fig. 47(d), RS (255,251,2) code is the most energy efficient and leads to highest throughput. Moreover, it is observed from Fig. 47(e) that RS codes are more resource efficient than the ARQ schemes.

We also present the results for the constrained optimization problems, $\mathbf{P}_{\min, \text{con}}^{\text{eng}}$ and $\mathbf{P}_{\min, \text{con}}^{\text{tput}}$, in Figs. 48. In these optimization results, we use $P_{\text{target}} = 10^{-3}$ and three different values for the end-to-end delay as shown in Figs. 48. An important results of the constrained optimization is the significant decrease in optimum packet size for both of these problems. Since the end-to-end delay is bounded, packet sizes between 2-16 bytes are optimal depending on the error correction capability. The optimal packet sizes and the resulting energy consumption for the constrained energy consumption minimization problem are shown in Fig. 48(a) and Fig. 48(b), respectively. While increasing the end-to-end delay increases the optimum packet size for the RS codes, the energy consumption is slightly decreased for larger T_{target} . Moreover, the optimum packet size for the ARQ scheme is not affected by the delay constrained. On the other hand, end-to-end delay constraint has an important impact on the packet throughput as shown in Fig. 48(d). An increase in T_{target} from 30s to 100s increases the packet throughput 20%.

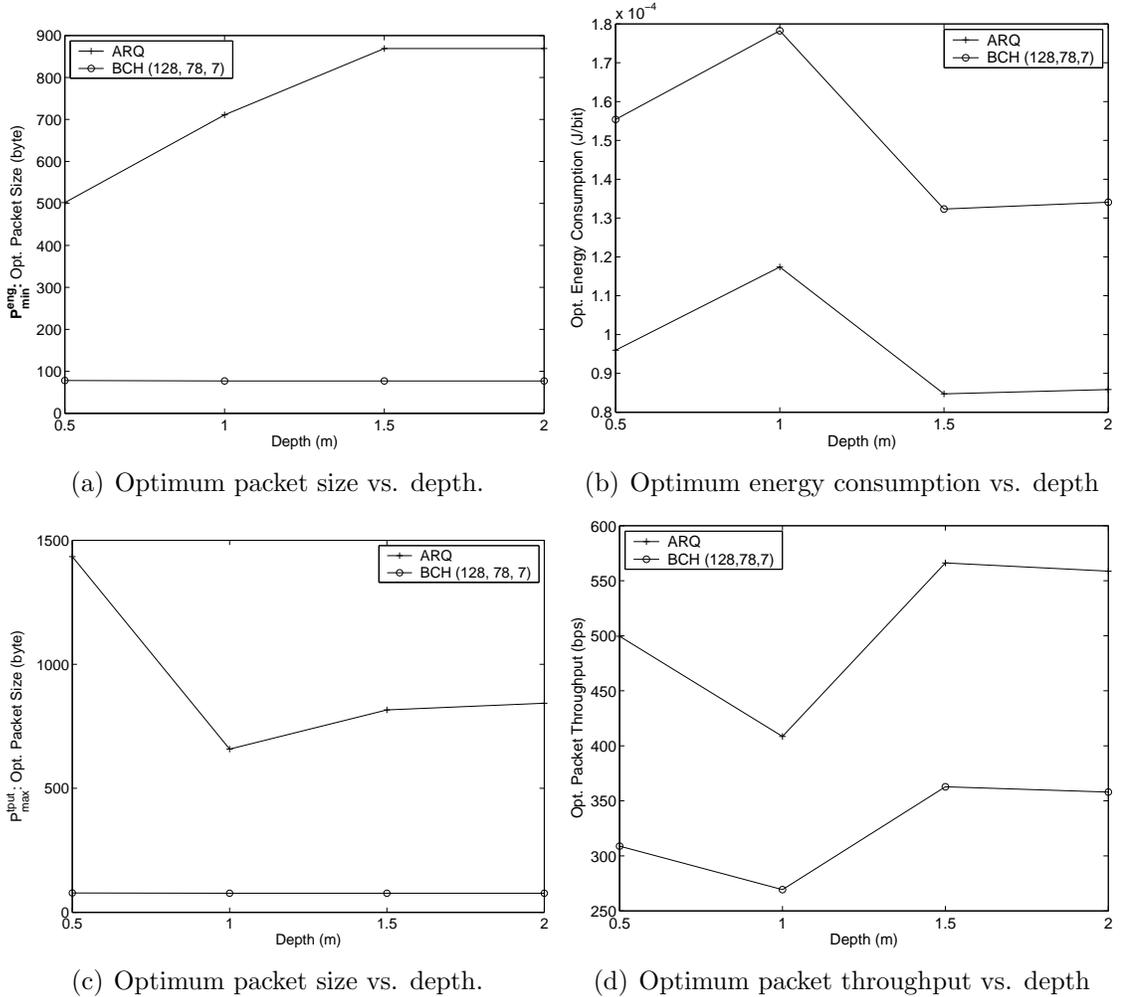
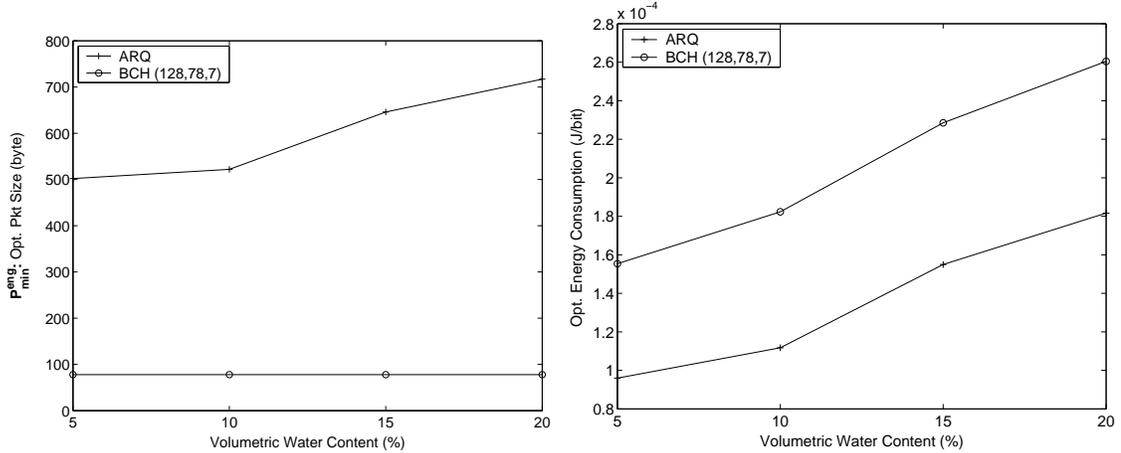


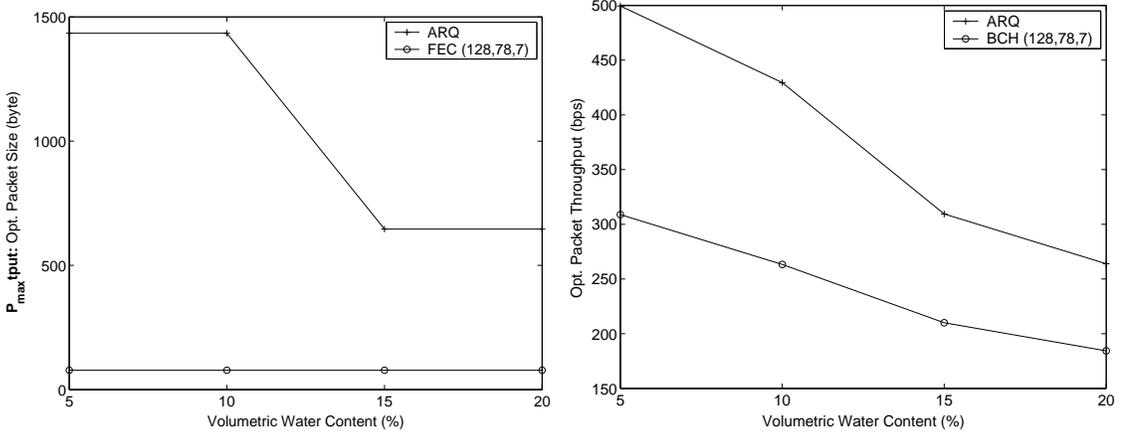
Figure 49: Optimization results for P_{min}^{eng} and P_{max}^{tput} in underground environment as a function of depth.

6.6.3.2 Wireless Underground Sensor Networks

In this section, we present the optimization results for underground sensor networks. As we have investigated in [45], the underground communication is characterized by the composition as well as the volumetric water content of the soil. Furthermore, the burial depth of the wireless sensors is an important factor in communication underground. Consequently, we investigate the effects of burial depth and the volumetric water content on optimal packet size in underground sensor networks. We consider the ARQ scheme and BCH (128,78,7) FEC code for our results. Moreover, we use the



(a) Optimum packet size vs. volumetric water content. (b) Optimum energy consumption vs. volumetric water content.



(c) Optimum packet size vs. volumetric water content. (d) Optimum packet throughput vs. volumetric water content.

Figure 50: Optimization results for P_{min}^{eng} and P_{max}^{tput} in underground environment as a function of volumetric water content.

typical operation parameters for the Mica2 nodes [22], which are shown to be feasible for wireless underground sensor networks at frequencies 400 MHz [45].

Energy consumption is of particular importance in wireless underground sensor networks since the deployment of nodes poses significant problems compared to the terrestrial case. Therefore, it is difficult, if not impossible, to recharge or re-deploy sensors in an WUSN after deployment. Optimizing the packet size for longer network lifetime, i.e., smaller energy consumption is, hence, important in WUSN. In Figs. 49,

the results for the problems, P_{\min}^{eng} and P_{\max}^{tput} , are shown as a function of the burial depth. The optimum packet size for the BCH code is not affected by the burial depth. However, it can be seen that, burial depth influences the energy consumption significantly. More specifically, the energy consumption is higher for burial depth of 1m. This is because of the destructive effects of the reflected rays from the ground surface at this depth, which is also discussed in [45]. Furthermore, energy consumption is lowest for burial depth of 1.5m and increases slightly above this value. The reason for this increase is the increased error rate at higher depth because of the higher attenuation underground. Similar observations can also be made for the throughput maximization problem. The packet throughput decreases at depth 1m as shown in Fig. 49(d). Note that for both problems, ARQ scheme is preferable with packet size values between 500bytes and 1.5kB.

The effect of volumetric water content is shown in Figs. 50. The change in optimum packet size for energy consumption minimization and packet throughput maximization is different as shown in Fig. 50(a) and Fig. 50(c), respectively. Increase in volumetric water content results in higher packet sizes for the energy consumption minimization problem, where the optimum energy consumption also increases for higher values of volumetric water content. On the other hand, the optimum packet size decreases for the packet throughput maximization problem. More specifically, smaller packets are used to combat the increased bit error rate. As shown in Fig. 50(d), packet throughput decreases for higher values of volumetric water content. As an example, the ARQ scheme results in a decrease of 47% when the volumetric water content is increased from 5% to 20%.

Finally, in Fig. 51, the results of the constrained optimization problem for the underground environment are shown. More specifically, the optimum packet sizes as well as the optimum energy consumption and packet throughput associated with them are shown as a function of end-to-end delay bound for different burial depth and

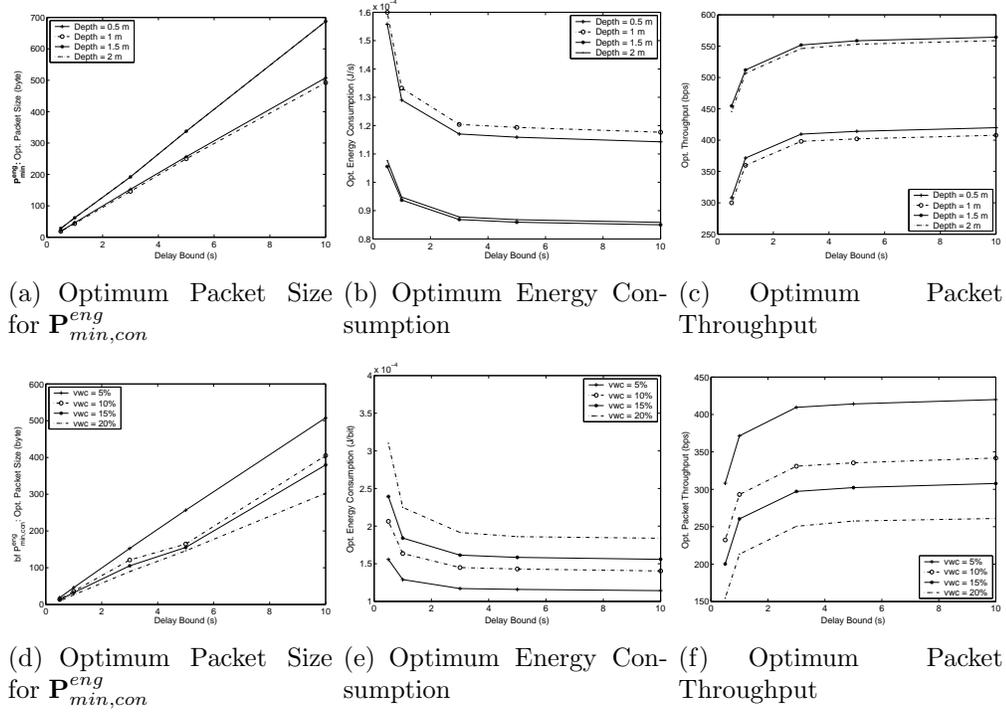


Figure 51: Constrained optimization results as a function of end-to-end delay bound in underground environment for (a-c) burial depth and (d-f) volumetric water content.

volumetric water content (VWC) values. In Fig. 51(a) and Fig. 51(d), it is shown that the optimum packet size increases with increasing delay bound. Moreover, the effect of burial depth on energy consumption and packet throughput can be observed in Fig. 51(b) and Fig. 51(c), respectively. Consistent with our results in Fig. 49(b), the optimum energy consumption is highest for burial depth of 1m and lowest for 1.5m. An important observation is that relaxing the end-to-end delay bound cannot provide lower energy consumption higher than a specific value, i.e., 5s. More specifically, the optimum energy consumption stays constant above this value of the end-to-end delay constraint.

It is shown in Fig. 51(d) that an increase in volumetric water content decreases the optimum packet size. Furthermore, this has an important effect on both energy consumption and packet throughput. In particular, increasing the volumetric water content from 5% to 20% increases energy consumption by 60% and decreases packet

throughput by 37%. This significant dependence on the volumetric water content necessitates the communication protocols be tailored to account for changes in the water content of the soil.

CHAPTER VII

CONCLUSIONS AND FUTURE WORK

7.1 Research Contributions

7.1.1 Theory of Correlation in Wireless Sensor Networks

In addition to the collaborative nature of the WSN, the existence of spatial and temporal correlations among the sensor observations are significant and unique characteristics of the WSN. The correlation in WSN can be considered in developing new energy-efficient networking protocols specifically tailored for WSN paradigm. These protocols utilizing the correlation to conserve energy resources may drastically enhance the overall network performance. In Chapter 2, a theoretical framework is developed to capture the spatial and temporal correlations in wireless sensor networks. More specifically, theoretical analysis of spatio-temporal correlation characteristics of point and field sources in WSN is performed. This analytical work provides tools for finding the feasible operating region in terms of spatial and temporal resolution for a specific distortion constraint considering spatio-temporal correlation, signal properties, and network variables in WSN. Furthermore, our theoretical framework constitutes a basis for the development of energy-efficient communication protocols for WSN. Moreover, based on our framework, we discuss possible efficient medium access and reliable event transport approaches taking advantage of the spatial and temporal correlations in WSN, respectively. We show via mathematical analysis, their results, case studies and discussions that correlation in WSN can be exploited to significantly improve the energy-efficiency in WSN. Extensions of this work include the study of spatio-temporal characteristics of point and field sources with different correlation models. Moreover, different reconstruction models such as filtering or nearest

node representation can be investigated. Furthermore, the effect of the impurities of the wireless channel will be investigated. Finally, the comparison of the correlation models with empirical data will be conducted.

7.1.2 Spatial Correlation-based Collaborative Medium Access Control in WSN

The theoretical framework developed in Chapter 2 provides important tools for the design of energy-efficient communication protocols. Among these, exploiting spatial correlation at the MAC layer is a powerful means of reducing the energy consumption in WSN under collective performance limits. This can be achieved by collaboratively regulating medium access so that redundant transmissions from correlation neighbors are suppressed. By allowing only a subset of sensor nodes to transmit their data to the sink, the proposed MAC protocol not only conserves energy, but also minimizes unnecessary channel access contention and thereby improves the packet drop rate without compromising the event detection latency. This is in contrast to the energy-latency tradeoffs that have been the main focus of many energy efficient MAC proposals for WSN [74, 59, 83].

The spatial Correlation-based Collaborative MAC (CC-MAC) protocol proposed in Chapter 3 is designed for distributed implementation and has two components: *Event MAC (E-MAC)* that filters out the correlation in source records and *Network MAC (N-MAC)* that prioritizes the transmission of route-thru packets over other packets. Route-thru packets are representative of an entire correlation region and hence given higher priority on their way to the sink. Using simulations, the performance of the CC-MAC protocol is investigated and significant performance gains in terms of energy consumption, latency and packet drop rate are shown.

7.1.3 XLM: Cross-layer Module for Efficient Communication in WSN

The unique characteristics of the WSN phenomenon such as energy consumption limitation and limited processing capabilities necessitate highly efficient communication protocols for long lasting, robust WSNs. Recently, cross-layering in designing a communication stack such that state information flows throughout the stack has been investigated. Recent work on WSNs [61], [75], [81] also reveals that cross-layer integration techniques result in significant energy gains. Although the literature shows the advantages of this approach, previous work focused on the joint design of two to three layers only, such as the physical, media access control (MAC), and routing layers. In Chapter 4, a unified cross-layer module (XLM) is designed for unified event transmission in WSNs.

XLM is a cross layer communication module for WSNs, which replaces the entire traditional layered protocol architecture that has been used so far in WSNs. The design principle of XLM is complete unified cross-layering such that both the information and the functionalities of traditional communication layers, i.e., transport, network, and MAC layers, are melted in a single module. The protocol operation of XLM is governed by the new concept of initiative determination. Based on this concept, XLM performs receiver-based contention, initiative-based forwarding, local congestion control, and distributed duty cycle operation in order to realize efficient and reliable communication in WSNs. In a cross-layer simulation platform, the state-of-the-art layered protocol configurations have been implemented along with XLM to provide a complete evaluation. Analytical performance evaluation and simulation experiment results show that XLM significantly improves the communication performance and outperforms the traditional layered protocol architectures in terms of both network performance and implementation complexity.

7.1.4 Cross-layer Analysis of Error Control in WSN

In Chapter 5, a cross-layer analysis of error control schemes is presented. Forward error control (FEC) coding improves the error resiliency by sending redundant bits through the wireless channel. It is shown that this improvement can be exploited by *transmit power control* or *hop length extension* through channel-aware cross-layer routing protocols in WSNs. The results of our cross-layer analysis reveal that hop length extension decreases both energy consumption and end-to-end latency for certain FEC codes when compared to ARQ. On the other hand, transmit power control can be exploited in situations where energy consumption is of paramount importance and can be traded off for end-to-end latency. In Chapter 5, the effects of hybrid ARQ schemes are also investigated and a comprehensive comparison of these three error control schemes are presented. Moreover, it has been shown that the selection of suitable error control scheme depends on the physical architecture of the sensor nodes as well as the end-to-end distance and target PER. Finally, FEC and hybrid ARQ schemes are shown to significantly improve the end-to-end latency performance of WSNs through hop length extension without hampering the energy efficiency and the end-to-end PER. This makes these schemes important candidates for delay sensitive traffic in WSN when used in combination with retransmissions in hybrid ARQ schemes.

7.1.5 Cross-layer Packet Size Optimization

In Chapter 6, a cross-layer optimization framework for packet size optimization in WSN is developed. This framework considers medium access collisions, routing decisions, as well as wireless channel affects to determine the optimal packet length. A key result of this analysis is that contrary to conventional wireless networks, longer packets improve the performance of medium access control in WSN. On the other hand, the packet size is limited in terms of energy efficiency due to wireless channel

errors. Our cross-layer framework clearly reveals this tradeoff. In this framework, three objective functions are used to investigate the various performance metrics such as throughput, energy consumption per bit, latency, and packet error rate. The results of our analysis show that the routing choices significantly affect overall performance and hence the optimal packet length in WSN due to the multi hop nature of the network. Optimal packet length is found considering this important factor in WSN. The results of this analysis highlight that the optimal packet length is closely affected by the error control technique and the requirements of the application. As an example, when high throughput is required where energy consumption, latency, and reliability are not important, a payload length of 152 bytes is necessary. Moreover, for applications with only energy efficiency requirements where delay and reliability can be traded off, our results show that the maximum available packet length should be used. The resource utilization objective function and the constraint optimization framework can be utilized to determine optimal packet length values for applications that leverage energy consumption, delay, and reliability such as real-time monitoring. Furthermore, the cases for underwater and underground sensor networks are investigated to determine the optimal packet size in these challenged environments. The developed optimization framework provides a flexible tool to determine the optimal packet size for different application requirements and network topologies.

7.2 Future Research Directions

The investigation of correlation in the context of cross-layer communication in wireless sensor networks provide many research areas in various fields of wireless networking.

7.2.1 Correlation in Multimedia Sensor Networks:

Recent advances in the CMOS technology have enabled the realization of cost effective cameras that can be attached to wireless sensor nodes. This development has led to research in wireless multimedia sensor networks (WMSNs). The correlation

characteristics of these networks are fundamentally different than WSNs since the information content depends on the directivity of the cameras of the sensors instead of their positions. Exploiting this correlation in WMSNs has potential improvements in the fields of telecommunication, networking, and video encoding.

7.2.2 Vast Area Sensor Network Testbed

Design and implementation of communication protocols for WSNs, so far, have mainly focused on simulations. However, recent studies reveal that cross-layer effects such as wireless channel and sensor node imperfections significantly affect the performance of the proposed protocols. While research in WSNs has recently started considering testbed evaluations and practical protocol implementation, most of these studies have limited scope in terms of network size. To overcome this, design and deployment of a vast area sensor network testbed that consists of heterogeneous wireless sensor nodes is necessary. Such a large scale network will enable the comprehensive testing of the scalability of communication protocols, highlight novel cross-layer interactions that are not foreseen through theoretical work, and lead to many novel WSN paradigms such as wireless sensor network management, data mining, and network health monitoring.

7.2.3 Cross-layer Communication in Cognitive Radio Networks

Cognitive radio networks are characterized by the vast amount of available spectrum and the spectrum sensing capabilities of cognitive radios. Thus, communication in cognitive radio network necessitates a cross-layer approach to exploit the available spectrum for the applications. This necessitates a design of a complete communication suite for cognitive radio networks that focuses on cross-layer design including spectrum sensing, spectrum characterization, spectrum sharing, and spectrum mobility.

7.2.4 Integration of Sensor Networks and Wireless Internet

The evolution in wireless technology has enabled the realization of various network architectures for different applications such as cognitive radio networks, mesh networks, and wireless sensor networks. In order to extend the applicability of these architectures and provide useful information anytime and anywhere, integration of these networks with Internet is an important challenge. So far, research has progressed in each of these areas separately, however, realization of these networks will require tight integration and interoperability. In this respect, it is crucial to develop location- and spectrum-aware cross-layer communication protocols as well as heterogeneous network management tools for integration of wireless sensor networks, cognitive radio networks, mesh networks, and Internet.

REFERENCES

- [1] “The network simulator (ns-2).” Available at <http://www.isi.edu/nsnam/ns/>. Last accessed on June 2007.
- [2] “Tinyos.” Available at <http://www.tinyos.net/>. Last accessed on June 2007.
- [3] “Vector quantization.” Available at <http://www.dice.ucl.ac.be/lee/software/vq/main.html>. Last accessed on June 2007.
- [4] AKAN, . B. and AKYILDIZ, I. F., “Event-to-sink reliable transport in wireless sensor networks,” *IEEE/ACM Transactions on Networking*, vol. 13, pp. 1003–1016, Oct. 2005.
- [5] AKYILDIZ, I. F., POMPILI, D., and MELODIA, T., “Underwater acoustic sensor networks: research challenges,” *Journal of Ad Hoc Networks (Elsevier)*, vol. 3, pp. 257–279, March 2005.
- [6] AKYILDIZ, I. F. and STUNTEBECK, E., “Wireless underground sensor networks: Research challenges,” *Ad Hoc Networks Journal, (Elsevier)*, vol. 4, pp. 669–686, Dec. 2006.
- [7] AKYILDIZ, I. F., SU, W., SANKARASUBRAMANIAM, Y., and CAYIRCI, E., “Wireless sensor networks: a survey,” *Computer Networks (Elsevier) Journal*, vol. 38, pp. 393–422, March 2002.
- [8] AKYILDIZ, I. F., VURAN, M. C., and AKAN, O. B., “On exploiting spatial and temporal correlation in wireless sensor networks,” in *Proc. WiOpt’04: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, pp. 71–80, March 2004.
- [9] AKYILDIZ, I. F., VURAN, M. C., and AKAN, O. B., “A cross-layer protocol for wireless sensor networks,” in *Proc. Conference on Information Science and Systems (CISS ’06)*, (Princeton, NJ), March 2006.
- [10] ARISHA, K. A., YOUSSEF, M. A., and YOUNIS, M. Y., “Energy-aware tdma-based mac for sensor networks,” *Computer Networks Journal (Elsevier)*, vol. 43, pp. 539–694, Dec. 2003.
- [11] BANDYOPADHYAY, S. and COYLE, E. J., “Spatio-temporal sampling rates and energy efficiency in wireless sensor networks,” in *Proc. IEEE INFOCOM 2004*, vol. 3, pp. 1728 – 1739, March 2004.

- [12] BAROWSKI, Y., BIAZ, S., and AGRAWAL, P., “Towards the performance analysis of IEEE 802.11 in multi-hop ad-hoc networks,” in *Proc. IEEE Wireless Communications and Networking Conference*, vol. 1, pp. 100 – 106, March 2005.
- [13] BERGER, J. O., DE OLIVIERA, V., and SANZO, B., “Objective bayesian analysis of spatially correlated data,” *J. American Statistical Assoc.*, vol. 96, pp. 1361–1374, Dec. 2001.
- [14] BIANCHI, G., “Performance analysis of the IEEE 802.11 distributed coordination function,” *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535 – 547, March 2000.
- [15] CHENG, D. K., *Field and wave electromagnetics*. Addison-Wesley Publishers, 2nd ed., Jan. 1989.
- [16] CHIANG, M., “To layer or not to layer: Balancing transport and physical layers in wireless multihop networks,” in *Proc. IEEE INFOCOM '04*, vol. 4, pp. 2525 – 2536, March 2004.
- [17] CHIANG, M., “Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control,” *IEEE JSAC*, vol. 23, pp. 104–116, Jan 2005.
- [18] CHITRE, M., POTTER, J., and HENG, O., “Underwater acoustic channel characterisation for medium-range shallow water communications,” in *Proc. OCEANS'04 MTS/IEEE/TECHNO-OCEAN'04*, vol. 1, (Kobe, Japan), pp. 40–45, Nov. 2004.
- [19] CHOI, J., SO, J., and KO, Y., “Numerical analysis of IEEE 802.11 broadcast scheme in multihop wireless ad hoc networks,” in *Proc. International Conference on Information Networking*, pp. 1–10, 2005.
- [20] CLOUQUEUR, T., PHIPATANASUPHORN, V., RAMANATHAN, P., and SALUJA, K., “Sensor deployment strategy for target detection,” in *Proc. WSNA'02*, (Atlanta, USA), September 2002.
- [21] CORP., A., “Atmega128 datasheet.” <http://www.atmel.com>. Last accessed on June 2007.
- [22] CORP., C., “Mica2 datasheet.” <http://www.xbow.com>. Last accessed on June 2007.
- [23] CORP., C., “Micaz datasheet.” <http://www.xbow.com>. Last accessed on June 2007.
- [24] CRISTESCU, R. and BEFERULL-LOZANO, B., “Lossy network correlated data gathering with high-resolution coding,” in *Proc. IPSN 2005*, pp. 218–224, April 2005.

- [25] CRISTESCU, R., BEFERULL-LOZANO, B., and VETTERLI, M., “On network correlated data gathering,” in *Proc. IEEE INFOCOM 2004*, vol. 4, pp. 2571–2582, March 2004.
- [26] CRISTESCU, R. and VETTERLI, M., “On the optimal density for real-time data gathering of spatio-temporal processes in sensor networks,” in *Proc. IPSN 2005*, April 2005.
- [27] CUI, S., MADAN, R., GOLDSMITH, A. J., and LALL, S., “Joint routing, MAC, and link layer optimization in sensor networks with energy constraints,” in *Proc. IEEE ICC '05*, May 2005.
- [28] DE, S., QIAO, C., PADOS, D. A., CHATTERJEE, M., and PHILIP, S. J., “An integrated cross-layer study of wireless CDMA sensor networks,” *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 1271 – 1285, Sep 2004.
- [29] FELLER, S. and ET AL., “Tracking and imaging humans on heterogeneous infrared sensor array for tactical applications,” in *Proc. SPIE Aerosense 2002*, April, 2002.
- [30] GANESAN, D., CRISTESCU, R., and BEFERULL-LOZANO, B., “Power-efficient sensor placement and transmission structure for data gathering under distortion constraints,” in *Proc. IPSN 2004*, pp. 142 – 150, April 2004.
- [31] GASTPAR, M. and RIMOLDI, B., “Source-channel communication with feedback,” in *Proc. IEEE Information Theory Workshop 2003*, pp. 279–282, March 2003.
- [32] GASTPAR, M. and VETTERLI, M., “Power, spatio-temporal bandwidth, and distortion in large sensor networks,” *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 745 – 754, April 2005.
- [33] GOBLICK, T. J., “Theoretical limitations on the transmission of data from analog sources,” *IEEE Trans. Info. Theory*, vol. IT-11(4), pp. 558 –567, October 1965.
- [34] HAAPOLA, J., SHELBY, Z., POMALAZA-REZ, C., and MHNEN, P., “Cross-layer energy analysis of multi-hop wireless sensor networks,” in *Proc. European Workshop on Wireless Sensor Networks*, Feb 2005.
- [35] HAYKIN, S., *Communication Systems*. Wiley Publishers, 3 ed., 1994.
- [36] IEEE, “Wireless LAN medium access control (MAC) and physical layer (PHY) specifications,” *ANSI/IEEE Standard 802.11*, 1999.
- [37] “Ieee 802.11g: Wireless mac and phy specifications: Further higher- speed physical layer extension in the 2.4 ghz band,” June 2003.

- [38] INTANAGONWIWAT, C., GOVINDAN, R., ESTRIN, D., HEIDEMANN, J., and SILVA, F., “Directed diffusion for wireless sensor networking,” *IEEE/ACM Trans. on Networking*, vol. 11, pp. 2 – 16, February 2003.
- [39] KARP, B. and KUNG, H. T., “GPSR: Greedy perimeter stateless routing for wireless networks,” in *Proc. ACM/IEEE MobiCom '00*, (Boston, MA), pp. 243 – 254, Aug 2000.
- [40] KARVONEN, H., SHELBY, Z., and POMALAZA-RAEZ, C., “Coding for energy efficient wireless embedded networks,” in *Proc. Int. Workshop on Wireless Ad-Hoc Networks*, pp. 300 – 304, June 2004.
- [41] KUHN, F., WATTENHOFER, R., ZHANG, Y., and ZOLLINGER, A., “Geometric ad-hoc routing: Of theory and practice,” in *Proc. ACM PODC*, (Boston, MA), July 2003.
- [42] KUSUMA, J., DOHERTY, L., and RAMCHANDRAN, K., “Distributed compression for sensor networks,” in *Proc. IEEE Image Processing 2001*, vol. 1, pp. 82–85, Oct 2001.
- [43] LANDOLSI, M. A. and STARK, W. E., “On the accuracy of gaussian approximations in the error analysis of DS-CDMA with OQPSK modulation,” *IEEE Trans. Comm.*, vol. 50, pp. 2064–2071, Dec. 2002.
- [44] LETTIERI, P. and SRIVASTAVA, M. B., “Adaptive frame length control for improving wireless link throughput, range, and energy efficiency,” in *Proc. IEEE INFOCOM 1998*, vol. 2, pp. 564 –571, April 1998.
- [45] LI, L., VURAN, M. C., and AKYILDIZ, I. F., “Characteristics of underground channel for wireless underground sensor networks,” in *Proc. Med-Hoc-Net'07*, (Corfu, Greece), June 2007.
- [46] LIN, S. and D. J. COSTELLO, J., *Error control coding: fundamentals and applications*. Prentice-Hall, 1983.
- [47] LIN, X. and SHROFF, N. B., “The impact of imperfect scheduling on cross-layer congestion control in wireless networks,” *IEEE/ACM Trans. on Networking*, vol. 14, pp. 302–315, April 2006. (invited for fast-track review from IEEE INFOCOM'06).
- [48] LIN, X., SHROFF, N. B., and SRIKANT, R., “A tutorial on cross-layer optimization in wireless networks,” *IEEE Journal on Selected Areas in Communications on “Non-Linear Optimization of Communication Systems”*, vol. 8, pp. 1452–1463, June 2006.
- [49] LINDE, Y., BUZO, A., and GRAY, R. M., “An algorithm for vector quantizer design,” *IEEE Trans. Comm.*, vol. COM-28, pp. 84 –95, Jan. 1980.

- [50] MADAN, R., CUI, S., LALL, S., and GOLDSMITH, A., “Cross-layer design for lifetime maximization in interference-limited wireless sensor networks,” in *Proc. IEEE INFOCOM '05*, vol. 3, pp. 1964 – 1975, March 2005.
- [51] MEGUERDICHIAN, S. and ET AL., “Coverage problems in wireless ad-hoc sensor networks,” in *Proc. IEEE INFOCOM 2001*, vol. 3, (Alaska), pp. 1380–1387, April 2001.
- [52] “Mica motes and sensors.” Available at <http://www.xbow.com>. Last accessed on June 2007.
- [53] PEPLINSKI, N., ULABY, F., and DOBSON, M., “Dielectric properties of soils in the 0.3-1.3-GHz range,” *IEEE Trans. Geoscience and Remote Sensing*, vol. 33, pp. 803–807, May 1995.
- [54] POOR, V., *An Introduction to signal detection and estimation*. Springer-Verlag, 2 ed., 1994.
- [55] POPPE, F., VLEESCHAUWER, D. D., and PETIT, G. H., “Choosing the UMTS air interface parameters, the voice packet size and the dejittering delay for a voice-over-IP call between a UMTS and a PSTN party,” in *Proc. IEEE INFOCOM 2001*, vol. 2, pp. 805–814, April 2001.
- [56] PRADHAN, S. S., KUSUMA, J., and RAMCHANDRAN, K., “Distributed compression in a dense microsensor network,” *IEEE Signal Processing Magazine*, vol. 19, pp. 51–60, March 2002.
- [57] PRADHAN, S. S., PURI, R., and RAMCHANDRAN, K., “n-channel symmetric multiple descriptions- part i: (n,k) source-channel erasure code,” *IEEE Transactions on Information Theory*, vol. 50, pp. 47– 61, January 2004.
- [58] PRADHAN, S. S. and RAMCHANDRAN, K., “Distributed source coding: symmetric rates and applications to sensor networks,” in *Proc. Data Compression Conference, 2000*, pp. 363 –372, April 2000.
- [59] RAJENDRAN, V., OBRACZKA, K., and GARCIA-LUNA-ACEVES, J. J., “Energy-efficient, collision-free medium access control for wireless sensor networks,” in *Proc. ACM SenSys 2003*, (Los Angeles, CA), November 2003.
- [60] ROSSI, M. and ZORZI, M., “Cost efficient localized geographical forwarding strategies for wireless sensor networks,” in *Proc. TIWDC 2005*, (Sorrento, Italy), 2005.
- [61] SAFWAT, A., HASSANEIN, H., and MOUFTAH, H., “ECPS and E2LA: new paradigms for energy efficiency in wireless ad hoc and sensor networks,” in *Proc. IEEE GLOBECOM 2003*, vol. 6, pp. 3547–3552, Dec. 2003.

- [62] SANKARASUBRAMANIAM, Y., AKYILDIZ, I. F., and MCLAUGHLIN, S. W., “Energy efficiency based packet size optimization in wireless sensor networks,” in *Proc. IEEE Internal Workshop on Sensor Network Protocols and Applications*, pp. 1–8, May 2003.
- [63] SAVVIDES, A., HAN, C., and STRIVASTAVA, M., “Dynamic fine-grained localization in ad-hoc networks of sensors,” in *Proc. ACM MobiCom 2001*, (Rome, Italy), pp. 166–179, July 2001.
- [64] SCAGLIONE, A. and SERVETTO, S., “On the interdependence of routing and data compression in multi-hop sensor networks,” in *Proc. MOBICOM '02*, (Atlanta, GA), September 2002.
- [65] SCHWIEBERT, L., GUPTA, S. K. S., and WEINMANN, J., “Research challenges in wireless networks of biomedical sensors,” in *Proc. ACM/IEEE MOBICOM '01*, pp. 151–165, 2001.
- [66] SCHWIEGER, K., KUMAR, A., and FETTWEIS, G., “On the impact of the physical layer on energy consumption in sensor networks,” in *Proc. EWSN '05*, pp. 13 – 24, Feb. 2005.
- [67] SEADA, K., ZUNIGA, M., HELMY, A., and KRISHNAMACHARI, B., “Energy-efficient forwarding strategies for geographic routing in lossy wireless sensor networks,” in *Proc. ACM Sensys '04*, November 2004.
- [68] SHIH and ET AL., E., “Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks,” in *Proc. ACM Mobicom'01*, pp. 272–286, July 2001. Rome, Italy.
- [69] SKRABA, P., AGHAJAN, H., and BAHAI, A., “Cross-layer optimization for high density sensor networks: Distributed passive routing decisions,” in *Proc. Ad-Hoc Now '04*, (Vancouver), July 2004.
- [70] STANN, F. and HEIDEMANN, J., “RMST: Reliable data transport in sensor networks,” in *Proc. IEEE SNPA '03*, (Anchorage, Alaska), pp. 102–112, April 2003.
- [71] STEERE, D., BAPTISTA, A., MCNAMEE, D., PU, C., and WALPOLE, J., “Research challenges in environmental observation and forecasting systems,” in *Proc. ACM/IEEE MOBICOM '00*, (Boston, MA), August 2000.
- [72] STUBER, G. L., *Principles of Mobile Communication*. Kluwer Academic Publisher, 2001.
- [73] V. SHNAYDER, E., “Simulating the power consumption of large-scale sensor network applications,” in *Proc. ACM SenSys'04*, (Baltimore, MD), Nov. 2004.

- [74] VAN DAM, T. and LANGENDOEN, K., “An adaptive energy-efficient MAC protocol for wireless sensor networks,” in *Proc. ACM SenSys 2003*, (Los Angeles, CA), November 2003.
- [75] VAN HOESEL, L., NIEBERG, T., WU, J., and HAVINGA, P. J. M., “Prolonging the lifetime of wireless sensor networks by cross-layer interaction,” *IEEE Wireless Communications*, vol. 11, pp. 78–86, Dec. 2004.
- [76] VURAL, S. and EKICI, E., “Analysis of hop-distance relationship in spatially random sensor networks,” in *Proc. ACM MobiHoc '05*, pp. 320–331, May 2005.
- [77] VURAN, M. C. and AKAN, O. B., “Spatio-temporal characteristics of point and field sources in wireless sensor networks,” in *Proc. IEEE ICC '06*, (Istanbul, Turkey), June 2006.
- [78] VURAN, M. C., AKAN, O. B., and AKYILDIZ, I. F., “Spatio-temporal correlation: Theory and applications for wireless sensor networks,” *Computer Networks Journal (Elsevier)*, vol. 45, pp. 245–259, June 2004.
- [79] VURAN, M. C. and AKYILDIZ, I. F., “Cross layer analysis of error control in wireless sensor networks,” in *IEEE SECON '06*, (Reston, VA), September 2006.
- [80] VURAN, M. C. and AKYILDIZ, I. F., “Spatial correlation-based collaborative medium access control in wireless sensor networks,” *IEEE/ACM Transactions on Networking*, vol. 14, pp. 316–329, April 2006.
- [81] VURAN, M. C., GUNGOR, V. C., and AKAN, O. B., “On the interdependency of congestion and contention in wireless sensor networks,” in *Proc. ICST SenMetrics '05*, 2005.
- [82] WICKER, S. B. and BARTZ, M. J., “Type-II hybrid-ARQ protocols using punctured MDS codes,” *IEEE Transactions on Communications*, vol. 42, pp. 1431–1440, Feb, March, April 1994.
- [83] YE, W., HEIDEMANN, J., and ESTRIN, D., “Medium access control with coordinated adaptive sleeping for wireless sensor networks,” *IEEE/ACM Transactions on Networking*, vol. 12, pp. 493–506, June 2004.
- [84] ZHONG, L., SHAH, R., GUO, C., and RABAEY, J., “An ultra-low power and distributed access protocol for broadband wireless sensor networks,” in *Networld+Interop: IEEE Broadband Wireless Summit*, (Las Vegas), May 2001.
- [85] ZORZI, M., “A new contention-based MAC protocol for geographic forwarding in ad hoc and sensor networks,” in *Proc. IEEE ICC 2004*, vol. 6, pp. 3481–3485, June 2004.
- [86] ZORZI, M. and RAO, R., “Geographic random forwarding (GeRaF) for ad hoc and sensor networks: energy and latency performance,” *IEEE Trans. Mobile Computing*, vol. 2, pp. 349–365, Oct. - Dec. 2003.

- [87] ZORZI, M. and RAO, R., “Geographic random forwarding (GeRaF) for ad hoc and sensor networks: multihop performance,” *IEEE Trans. Mobile Computing*, vol. 2, pp. 337–348, Oct.-Dec. 2003.
- [88] ZUNIGA, M. and KRISHNAMACHARI, B., “Analyzing the transitional region in low power wireless links,” in *Proc. IEEE SECON '04*, pp. 517 – 526, Oct 2004.

VITA

Mehmet Can Vuran was born on July 28, 1980, in Istanbul, Turkey. He received his B.Sc. degree in Electrical and Electronics Engineering from Bilkent University, Ankara, Turkey in 2002. He received his M.S. and Ph.D. degree under the supervision of Prof. Ian F. Akyildiz in Electrical and Computer Engineering from Georgia Institute of Technology, Atlanta, GA. in 2004 and 2007, respectively. He is currently an Assistant Professor in Computer Science and Engineering Department at the University of Nebraska-Lincoln. In 2006, he received the BWN Lab Researcher of the Year 2006 Award. He also received the 2007 ECE Graduate Research Assistant Excellence Award from School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA.

INDEX

- A**
angle-based routing 88
- C**
CC-MAC 46, 57
 event MAC 60
 network MAC 64
channel model
 log-normal 124
 underground 171
 underwater 169
correlation 13
 field source 30, 38
 model 18
 neighbor 56
 point source 27, 36
 radius 56
 spatial 14, 20, 32
 spatio-temporal 27, 36, 38
 temporal 14, 24, 34
covariance models 20
- E**
error control
 automatic repeat request (ARQ)
 117
 forward error correction (FEC) 118
 hybrid ARQ (HARQ) 119
expected energy consumption 123
expected hop distance 123
- H**
hop length extension 119, 132
- I**
initiative function 83
iterative node selection 52
- M**
MAC failure rate 152
Mehmet Can Vuran 200
- N**
node duty
 router 85, 91
 source 85, 90
- P**
packet throughput 157
Peplinski principle 172
- R**
resource utilization 158
- T**
taxonomy function 136
transmit power control 119, 138
- U**
Urick path loss 169
- V**
Voronoi regions 53