Realistic Modeling of Wireless Communication Graphs
for the Design of Efficient Sensor Network Routing Protocols

by

Marco Antonio Zúñiga Zamalloa

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(Electrical Engineering)

October 2006

Copyright 2006 Marco Antonio Zúñiga Zamalloa
Dedication

To my brother Carlo, my sisters Dany and Caroline, my nephew Sedrik,
and specially papá y mamá, Federico y Hortensia.

And to the wife and kids that I do not have yet, but some day I will.
Acknowledgements

I see my years as a graduate student at the University of Southern California as the most rewarding experience of my life. I have had the great opportunity to interact not only with outstanding researchers but specially superb people.

Before joining the program I heard that the relationship with your advisor is perhaps the single most important one that you have as a PhD student, and I am grateful to have had the opportunity to work with Professor Bhaskar Krishnamachari. Bhaskar, thank you for showing me what a great mentor and a great leader is. I would also like to thank the support and encouragement of Professor Ahmed Helmy and Professor Ali Zahid throughout my studies. And, I am grateful to have met Professor Chen Avin whose insightful comments and guidance contributed significantly to the results presented in chapter 5. Thank you also to the members of my Qualifying and Defense committees Professors Urbashi Mitra, Professor Konstantinos Psounis, and Professor Ramesh Govindan whose comments and suggestions improve significantly the quality of the work presented in this thesis.

My fellow students played a central role during my studies. I would like to thank specially Shyam, Kiran, Sundeep and Avinash for their valuable comments on my work and for their unconditional friendship. Members of the ANRG and NOMADS group have also been a major source of support like Dong-Jin, Amitabha, Ganesh, Jaber, Nara, Gang, Hua, Pai-Han, Joon, Yi, Divyesh, Fan, Yang and Karim.
## Contents

Dedication ii

Acknowledgements iii

List Of Tables vii

List Of Figures viii

Abstract x

1 Introduction 1

1.1 Overview ...................................................... 1

1.2 Research Contributions ........................................ 2

1.2.1 Transitional Region Analysis ............................... 2

1.2.2 Impact of Lossy Links in Geographic Routing .............. 3

1.2.3 Performance of Random Walks on Heterogeneous Networks 4

1.3 Scope .................................................................... 4

2 Related Work 6

2.1 Link Layer Modeling ............................................. 7

2.2 Geographical Routing ............................................. 8

2.3 Random Walk Based-Queries .................................... 11

3 Analysis of the Transitional Region 13

3.1 Overview .......................................................... 13

3.2 Impact of Channel Multi-Path ................................... 16

3.2.1 Channel and Radio Receiver Models ....................... 16

3.2.2 Impact on Link Reliability (Extent of Transitional Region) 19

3.2.3 Expectation and Variance of Packet Reception Rate .......... 24

3.2.4 Comparison With Available Link Models .................. 30

3.3 Impact of Hardware Variance ................................... 31

3.3.1 Model .......................................................... 32

3.3.2 Impact on Asymmetric Links ................................. 32

3.3.3 Impact on Extent of Transitional Region ................... 35

3.4 Experiments with Motes ......................................... 36

3.4.1 Channel and Radio Parameters ............................... 37

3.4.2 Chain Topologies .............................................. 40

3.5 Summary .......................................................... 43
4 Impact of Lossy Links on Geographic Routing

4.1 Overview ................................................. 46
4.2 Scope, Assumptions and Metrics .................................................. 48
4.3 Analytical Model .................................................. 50
   4.3.1 Problem Description .............................................. 50
   4.3.2 Analysis for ARQ case ........................................... 52
   4.3.3 Analysis for the No-ARQ case .................................. 54
4.4 Geographic Forwarding Strategies for Lossy Networks ......................... 56
   4.4.1 Distance-based Forwarding ....................................... 57
   4.4.2 Reception-based Forwarding ..................................... 57
   4.4.3 PRR×d ...................................................... 57
4.5 Comparison of Different Strategies .............................................. 58
   4.5.1 PRR×d ...................................................... 58
   4.5.2 Absolute Reception-Based ....................................... 60
   4.5.3 Distance-Based .................................................. 61
   4.5.4 Comparison ..................................................... 61
4.6 Experiments with Motes .................................................... 64
4.7 Summary ........................................................ 66

5 Performance of Random Walks on Heterogeneous Networks ....................... 67

5.1 Overview .................................................... 67
5.2 Enhancing Random Walks for Heterogeneity ...................................... 69
5.3 Analytical Results .................................................. 70
   5.3.1 Parameters of Line Topology ..................................... 71
      5.3.1.1 Analysis of region 1 ......................................... 73
      5.3.1.2 Analysis of region 2 ......................................... 75
      5.3.1.3 Analysis of region 3 ......................................... 76
   5.3.2 Local Minimum for Maximum Hitting Time .................................. 77
   5.3.3 Local Minimum for Expected Hitting Time ................................ 78
5.4 Numerical and Experimental Results .............................................. 80
   5.4.1 Line Topologies .................................................. 81
   5.4.2 Regular Grids and Random Geometric Graphs ............................. 84
      5.4.2.1 Grids ........................................................ 84
      5.4.2.2 Random Geometric Graphs ................................... 86
   5.4.3 Low-Power Wireless Graphs ......................................... 87
5.5 Experiments with Motes .................................................... 88
5.6 Summary ........................................................ 90

6 Conclusions ...................................................... 91
6.1 Analysis of the Transitional Region .............................................. 91
6.2 Impact of Lossy Links on Geographic Routing ................................... 92
6.3 Performance of Random Walks on Heterogeneous Networks ...................... 93

Appendix A
Models from Communication Theory .................................................... 95
   A.1 Log-Normal Path Loss Model ............................................ 95
   A.2 Encoding and Modulation ............................................... 96
   A.3 Noise Floor ....................................................... 97
Appendix B
Random Walks ........................................................................ 98
B.1 Resistance Method ................................................................. 98
B.2 Time to Absorption in Markov Chains ................................. 99

Appendix C
The Radio Irregularity Model ...................................................... 101

Appendix D
The Transitional Region in MicaZ Motes ................................. 103
## List Of Tables

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Mathematical Notation</td>
<td>17</td>
</tr>
<tr>
<td>3.2</td>
<td>Empirical Channel Parameters</td>
<td>37</td>
</tr>
<tr>
<td>3.3</td>
<td>Empirical Radio Parameters</td>
<td>38</td>
</tr>
<tr>
<td>3.4</td>
<td>Comparison of Total Variance and Channel Variance</td>
<td>40</td>
</tr>
<tr>
<td>3.5</td>
<td>Analytical Extent of Transitional Region</td>
<td>42</td>
</tr>
<tr>
<td>3.6</td>
<td>Theoretical Models for the Link Layer</td>
<td>45</td>
</tr>
<tr>
<td>4.1</td>
<td>Mathematical Notation</td>
<td>51</td>
</tr>
<tr>
<td>4.2</td>
<td>Empirical Results for Different Forwarding Strategies</td>
<td>64</td>
</tr>
<tr>
<td>5.1</td>
<td>Mathematical Notation</td>
<td>71</td>
</tr>
<tr>
<td>5.2</td>
<td>Minimum Maximum and Minimum Average Hitting Times per Region</td>
<td>77</td>
</tr>
<tr>
<td>5.3</td>
<td>Random Geometric Graphs</td>
<td>86</td>
</tr>
<tr>
<td>5.4</td>
<td>Grid Deployments in Realistic Environments</td>
<td>87</td>
</tr>
<tr>
<td>5.5</td>
<td>Random Deployments in Realistic Environment</td>
<td>88</td>
</tr>
</tbody>
</table>
List Of Figures

3.1 (a) A receiver response where $\psi_\ell$ and $\psi_h$ determine different regions of link quality, (b) Interaction of $\gamma_\ell$ and $\gamma_h$ with the channel to determine the transitional region. 20

3.2 Analytical representation of equations 3.6 and 3.7. 21

3.3 Impact of $\sigma$ and $\eta$ on extent of transitional region. (a) $\Gamma$ for different values of $\eta$ and $\sigma$, (b) solid curves represent average power decay, and dotted lines the $[-2\sigma, 2\sigma]$ interval of the variance. 23

3.4 Impact of perfect receiver threshold on extent of transitional region, (a) Analytical representation, (b) An instance of PRR vs distance. 24

3.5 cdfs for packet reception rate for receivers in different regions in a specific environment ($\eta = 3, \sigma = 3$). 25

3.6 Linear approximation of receiver threshold and Gaussian SNR, the mean of the Gaussian depends on the transmitter-receiver distance. 26

3.7 Comparison of $E[\Psi]$ and $\text{Var}[\Psi]$ with their linear approximations, $E[\Psi_L]$ and $\text{Var}[\Psi_L]$. 29

3.8 Comparison of cdfs between the Gaussian model (black curves) and our analytical model (dotted curves) for receivers in different regions. 30

3.9 Impact of hardware variance on asymmetric links. 33

3.10 Impact of channel multi-path and hardware variance on extent of transitional region, (a) Impact of channel multi-path (real channel + identical non-variant hardware), (b) Impact of hardware variance (ideal channel + hardware variance), (d) Combined impact of channel multi-path and hardware variance. 36

3.11 Correlation between output power and noise floor. 37

3.12 Simulation results for the relation between in-degree and out-degree, (a) positive correlation, (b) negative correlation. 39

3.13 Empirical correlation between in-degree and out-degree for different power levels, (a) indoor environment, (b) outdoor environment. 40
3.14 Kolmogorov-Smirnov test between the empirical and model-generated packet re-ception rate for medium and high output powers in a grass area. 41

3.15 Comparison of empirical measurements and instances of analytical model, (a) Empirical indoor $P_t=-7$dB, (b) Empirical outdoor $P_t=-7$dB, (c) Empirical outdoor $P_t=5$dB, (d), (e) and (f) are the analytical counterparts. 42

4.1 Impact of channel multi-path on $E[\xi_{d,\text{ARQ}}]$, (a) impact of path loss exponent $\eta$, (b) impact of channel variance $\sigma$. 53

4.2 Energy efficiency metric for the ARQ case. The transitional region often has links with good performance as per this metric. 54

4.3 Impact of different parameters on $q_d$, (a) $\tau$, (b) $\eta$, (c) $\sigma$, (d) $P_t$. 59

4.4 Performance of PRR blacklisting. 62

4.5 Performance of distance blacklisting. 63

5.1 Examples of line topologies. The big circles denote cluster-heads and the dashed lines their coverage $k$, all nodes within $k$ hops from the cluster-head are its neighbors. 73

5.2 Resistance calculation for regions 2 and 3. 75

5.3 $C_{\text{query}}$, $C_{\text{event}}$ and $C_{\text{total}}$ vs the number of clusters for a line topology with 121 nodes and different values of $k$ (6, 8 and 10). 83

5.4 Maximum hitting time for a line topology with 121 nodes and values $k$ ranging from 5 to 10. 83

5.5 $C_{\text{query}}$ (expected hitting time) for a line topology with 121 nodes and values $k$ ranging from 5 to 10. 84

5.6 (a) $C_{\text{event}}$, (b) $C_{\text{query}}$ and (c) $C_{\text{total}}$ for a grid topology with 169 nodes and values of $k$ ranging from 2 to 6. 85

5.7 Empirical study of random walks on degree-heterogeneous graphs, (a) presents the degree of the nodes, (b) presents the query cost for graphs with different number of clusters. 89

D.1 Comparison of empirical measurements for channel, radio and link between mica2 and micaZ motes, $P_t=-10$ dBm for both type of motes, (a) channel mica2, (b) radio mica2, (c) link mica2, (d), (e), (f) are their micaZ counterparts. 104
Abstract

Recent advances in low-power processor technology, radios, sensors and actuators will allow us to monitor and instrument the physical world with unprecedented granularity and precision. These new types of systems, often referred to as wireless sensor networks, present unique challenges compared to current networked and embedded systems. One of the most important challenges is the design of efficient routing protocols considering the particular characteristics of the underlying communication graph.

Sensor network deployments are characterized by a number of non-idealities such as spatio-temporal variation in wireless link quality, link asymmetry, hardware variance, node heterogeneity and randomized placement. Our thesis is that a realistic modeling of the underlying communication graph incorporating these effects is necessary to design highly efficient routing mechanisms for sensor networks. We substantiate this thesis by developing a realistic communication graph model and incorporating it into the design of a geographic routing mechanism and a random walk-based querying mechanism.

First, we provide an in depth analysis of the so-called transitional region in multi-hop wireless networks, and propose a more realistic link layer model. Then, based on this model we present analytical results on the impact of channel multi-path on geographical routing. Finally, we use our model to study random walk-based queries and show that their performance can be enhanced by exploiting the node degree heterogeneity present in real wireless sensor networks.
Chapter 1

Introduction

1.1 Overview

Recent technological advances in MEMS\(^1\) and chip design have resulted in the emergence of a novel networking paradigm where networks of several wireless sensors can interact with the physical world. As in any other network, one of the main tasks of these new wireless sensor networks (WSN) is the communication of information among nodes. Given that in the end the underlying communication graph determines to a great extent the quality and quantity of communication of the network, it is important to have an accurate representation of the graph in order to design efficient routing protocols.

Unfortunately, most of the research up-to-date has assumed homogeneous networks using the ideal binary model\(^2\). While this ideal set-up is sufficient to capture limiting bounds on some important properties of WSN such as connectivity and capacity [33, 35], it does not capture other important properties of real deployments. For example, it does not capture the effect of channel multi-path on link reliability, the impact of hardware variance on link asymmetry, nor the cluster-head characteristic of heterogeneous deployments [32].

---

1. Micro-Electro-Mechanical Systems
2. In this model, a node receives a packet only if it is within the circular transmission range of the sender.
Recent empirical studies have shown that the differences mentioned above can considerably degrade the performance of protocols designed under the ideal model. Ganesan et. al. [28] have shown that the behavior of even the simplest flooding mechanism can be significantly affected due to asymmetric and occasional long-distance links. It was observed that in real deployments the flooding tree presents an important presence of cluster-heads. Zhou et. al. [72] report that unreliable links can have a negative impact on routing protocols, specially location-based routing protocols, such as geographic routing. Other works [22, 62, 58] have proposed mechanisms to leverage on the particular properties of wireless links. These studies show that traditional minimum hop-count metrics perform poorly in terms of throughput and energy efficiency, and that routing metrics based on required-number-of-transmissions have a better performance.

The previous works motivate the following thesis:

**A realistic modeling of the underlying communication graph of wireless sensor networks is necessary to design efficient routing protocols.**

### 1.2 Research Contributions

The proposed thesis is supported through three studies. First, we provide an in-depth understanding of the root causes of unreliability and asymmetry of wireless links and propose a realistic link layer model. Based on this model, the second study presents optimal local forwarding metrics to minimize energy consumption in geographical routing. Finally, we present an algorithm that improves the performance of random walk based-queries by exploiting node degree heterogeneity.

#### 1.2.1 Transitional Region Analysis

Several empirical studies [28, 71, 62] have revealed that real links deviate to a large extent from the ideal binary model. These studies have identified the existence of three distinct reception regions in the wireless link: connected, transitional, and disconnected. In the connected region links are
often of good quality; and the disconnected region present no practical links for communication. On the other hand, the transitional is generally characterized by high-variance in reception rates and asymmetric connectivity. Because of its inherent unreliability and extent, the transitional region has a major impact on the performance of upper-layer protocols.

In this study, we present an in-depth analysis of the transitional region for static and low-dynamic environments (no significant time-variance was considered). Our analysis shows how channel multi-path affects the extent of the transitional region, and quantifies the impact of hardware variance on link asymmetry. One of the major contributions of this work is the derivation of a more realistic link layer model.

1.2.2 Impact of Lossy Links in Geographic Routing

In geographic routing, each node forwards a packet to the neighbor whose location is closest to the destination. However, the existence of unreliable links exposes a key weakness in greedy forwarding. Neighbors that are closest to the destination (also likely to be farthest from the forwarding node) may have poor links with the current node, these “weak links” can result in increased energy wastage due to dropped packets. On the other hand, if the forwarding mechanism attempts to maximize per-hop reliability by forwarding only to close neighbors with good links, it may cover only a small geographic distance at each hop, which would also result in greater energy expenditure due to the need for more transmissions.

Based on the model derived in this thesis, we present an study of the trade-off between distance-hop and energy described above. Our analysis, simulations and experiments all show that the product of the packet reception rate (PRR) and the distance traversed toward the destination is the optimal local forwarding metric for systems using automatic repeat request (ARQ).
1.2.3 Performance of Random Walks on Heterogeneous Networks

Random walks are an important approach for querying in unstructured systems. In random walks, nodes are visited sequentially in a random order with successive nodes being neighbors in the graph. However, most of the literature is focused on simple classes of deterministic graphs (such as 2D Torii), and a major property of real-life networks, degree heterogeneity, is left out of discussion.

Heterogeneity on the degree distribution is a characteristic observed in real large-scale WSN for different reasons, ranging from hardware variance to recent proposals [32] suggesting that heterogeneous networks consisting of cluster-heads and regular nodes is a convenient direction to scale WSN.

In this study, we propose the use of a simple algorithm to exploit the heterogeneity of the communication graph to enhance the performance of random-walk-based queries. We present analytical results on linear topologies and numerical results on 2D topologies based on the link layer model proposed in this thesis. Our results show that a small percentage of nodes being cluster-heads can lead to significant improvements in performance.

1.3 Scope

It is important to consider that as technology evolves, future generation radios may reduce multi-path effects\(^3\), and applications using these radios may not be considerably affected by multi-path and hardware variance effects. Nevertheless, the large-scale scenarios targeted by wireless sensor networks pose major constraints on the cost of the radios, leading to resource constrained devices in terms of radio, energy and processing capabilities. Hence, given that even resource-rich scenarios such as cellular networks still face significant shortcomings, the extra constraints posed

---

\(^3\)Technology evolution has also led to virtually error-free wired links on the internet, from initial middle-of-the-road links.
by WSN (chiefly among them cost and energy) would further complicate the achievement of ideal communication channels.

Also, even though this work is focused on WSN, similar link properties have been observed in other multi-hop wireless networks such as Mobile Ad Hoc Networks (MANETS), and hence, some of our results may be applicable to such networks as well.

The remainder of the thesis is organized in the following way. Chapter 2 presents the related work on the areas covered by our studies. In chapter 3 we present a more realistic link layer model for WSN. The proposed model is used in chapters 4 and 5 to enhance the performance of geographic routing and random walk-based queries, respectively. Our conclusions are presented in chapter 6. And the interested reader can find some introductory material on the analytical tools used in this thesis in Appendices A and B.
Chapter 2

Related Work

Recent empirical results [28, 62, 32] have shown the striking discrepancy between the performance of WSN routing protocols in real scenarios versus their behavior under ideal settings. These studies have lead the WSN community to have an increased understanding of the need for realistic link layer models.

While the related work on the specific area of routing protocol design on real WSN communication graphs is not extensive, the areas of channel modeling and routing protocols have been extensively studied independently. On one hand, during the last several years a number of routing protocols for wireless sensor networks have been proposed ([38, 37, 16, ?] just to name a few), however most of these protocols were studied based on ideal assumptions. On the other hand, the area of communication theory has developed in the last decades several interesting models to capture the random behavior of the wireless channel. However, most of these efforts were targeted for a different types of networks, mainly satellite and cellular.

In this thesis we use some of the rich tools developed in communication theory to propose a more realistic link layer model for WSN. Then, we use this model to study the performance of geographic routing and random walk-based queries on real scenarios. In the next sections, we present the related work on the areas covered by this thesis.
2.1 Link Layer Modeling

Years of research in wireless communications, particularly cellular networks, provide a rich set of models and tools for analyzing the physical layer [56]. However, most of the research has been focused on improving physical layer parameters, mainly the bit error rate (BER). From a networking perspective, an ideal abstraction to design protocols is an underlying communication graph where edges are tagged with the probability of successfully receiving a packet (packet reception rate). Initially, the wireless sensor network community adopted the ideal binary model as the communication premise to build these graphs.

Unfortunately, the ideal model does not capture important phenomena of the wireless link. For instance, unreliable and asymmetric links are entirely ignored and other important associated characteristics such as the node degree distribution can be completely misleading. Recent empirical studies have reported that the performance of routing protocols designed under ideal settings can differ significantly when deployed in real scenarios.

Kotz et al. [41] shows that packet losses lead to different connectivity graphs, and coverage ranges that are neither circular nor convex and are often noncontiguous. In one of the earliest works [28], Ganesan et al. present empirical results on the behavior of a simple flooding in a dense sensor network. They found that the flooding tree exhibits a high clustering behavior, in contrast to the more uniformly distributed tree obtained with the ideal binary model. Couto et al. [22] and Woo et al. [62], report that when the real channel characteristics are taken into account, the minimum hop-count metric has poor performance. They show that in real scenarios a cost-based routing using a minimum expected transmission metric shows good performance. Zhou et al. [72] report that radio irregularity has a significant impact on routing protocols, but a relatively small impact on MAC protocols.

The previous studies made clear to the community that the design of WSN routing protocols require the need for more realistic communication graphs. In order to help overcome this problem
some tools and models have been proposed recently. In [62], the authors derive a packet loss model based on aggregate statistical measures such as mean and standard deviation of packet reception rate. The model assumes a Gaussian distribution of the packet reception rate for a given transmitter-receiver distance, which as it will be shown in Chapter 3, is not accurate. Using the SCALE tool [12], Cerpa et al. [13] use several statistical techniques to provide a spectrum of models of increasing complexity and increasing accuracy. A more recent model, the Radio Irregularity Model (RIM), was proposed in [72]. Based on experimental data, RIM provides a radio model that takes into account both the non-isotropic properties of the propagation media and the heterogeneous properties of devices to build a richer link model.

While the described models are important steps toward a realistic link quality model, they do not provide significant mathematical insight on how the channel and radio parameters affect link unreliability and asymmetry. Also, some of these models [62, 13] do not provide a systematic way to generalize (i.e., extend their validity and accuracy) beyond the specific radio and environment conditions of the experiments from which the models are derived.

In chapter 3 we use some tools from communication theory, the log-normal path loss model and the BER expressions of various modulation and encoding schemes, to present an in-depth analysis on unreliable and asymmetric links and provide simple and more realistic analytical models for the link layer.

### 2.2 Geographical Routing

Geographic routing [26, 38] is a key paradigm that is quite commonly adopted for information delivery in wireless ad-hoc and sensor networks where the location information of nodes is available. The main component of geographic routing is usually a greedy forwarding mechanism whereby each node forwards a packet to the neighbor that is closest to the destination.
Geographic routing is attractive for these new types of wireless networks because of its low overhead and the minimal state required at each node. Several works have presented mechanisms to overcome some of the problems of geographic routing such as dead-ends [8, 38, 43], and it has been proven to be an efficient, low-overhead method of data delivery if it is reasonable to assume (i) sufficient network density, (ii) accurate localization and (iii) high link reliability independent of distance within the physical radio range.

However, while assuming highly dense sensor deployment and reasonably accurate localization may be acceptable in some classes of applications, under the assumptions of (iii) concerning highly reliable and deterministic links is unlikely to be valid in any realistic deployment.

Given the attractiveness of geographic routing some recent works have explored their performance on real scenarios. Kim et. al. [40] use simulations and test-bed measurements to show that the differences between the ideal model and real links cause persistent failures in geographic routing, even on static topologies. These differences can cause three kinds of pathologies in the planarization process required to avoid dead-ends: a link in the planar subgraph is removed when it should not be; nodes at the two ends of an asymmetric link disagree on whether or not the link belongs in the planar graph; or a pair of crossed links remain in the supposedly planar subgraph. The authors propose the Cross-Link Detection Protocol (CLDP) as the solution for these problems. On the same line of work, Zhou et. al. [72] used empirical data to show that geographic forwarding perform worse in the presence of radio irregularity than on-demand protocols, such as AODV and DSR. The authors propose a technique called Symmetric Geographic Forwarding to alleviate the negative impact of link asymmetry. In this technique, beacon messages are allowed to contain not only the node’s ID and position, but also the IDs of all its neighbors. When a node receives a beacon message it registers the sender as its neighbor. If the receiver finds its own ID in the neighbor list it marks the communication link as symmetric, otherwise, it marks the link as asymmetric. Whenever a node needs to forward a packet, it selects only those neighboring nodes with which it is connected through symmetric links.
Similarly to the studies described above, our work aims to overcome challenges posed by the unreliable wireless links, but in a different domain. In chapter 4 we show that the distance-greedy mechanism of geographic routing can significantly degrade its performance, and we propose a new greedy metric to maximize the energy efficiency of geographic routing in real scenarios. This metric considers not only distance but also the quality of the link in order to minimize the number of transmissions. Our work is related to the studies done in [22] and [62], where a minimum expected transmission metric is proposed to optimize routing. However, while the minimum expected transmission metric is a global path metric, our work provides an optimal local metric suitable for scalable routing protocols such as geographic routing.

Our work has sparked interest in the community on optimal geographic forwarding strategies on real low-power wireless links, and some works have followed-up on our initial study. In [45], the authors propose a new metric called *normalized advance* (NADV), which also studies the *distance-hop trade-off* and provides some flexibility in terms of the metric to be optimized, such as energy or delay. Li et al. [47] insert power control to our proposed PRR×d metric. In [70], studies the PRR×d, among other metrics, in 802.11b networks and suggest that the link quality should be tested using data traffic.

Finally, it is important to describe the research done in the wireless communication area. Traditionally, wireless communication have focused only on pure physical layer techniques to overcome the effects of multi-path, for example Rake receivers [9, 48] and multiple input multiple output (MIMO) systems [27, 19]. However, some recent studies have explored the interaction between cooperative diversity techniques, in the physical layer, and routing, in the network layer. In [18], the authors consider in a unified fashion the effects of cooperative communication via transmission diversity and multi-hopping as well as optimal power allocation schemes in fading channels. Khandani et. al. [39] propose a mechanism based on omni-directional antennas to optimize the energy efficiency on the transmission of a single message from a source to destination through sets of nodes acting as cooperating relays. They present a solution to optimally allocate
power for a set of source nodes to a set of destination nodes. Our work, presented in chapter 4, differs from the previous works in that it uses only techniques at the network layer with inexpensive radios that do not require any extra functionality at the physical layer.

2.3 Random Walk Based-Queries

Random walks on graphs have been studied mathematically, and there is a substantial-yet-growing body of theoretical literature on the subject [3, 7, 11]. Recently, they are finding increasing use in a wide range of protocols in the context of several networked distributed systems. For instance, they have been used in Grid-aware operating systems [25], in unstructured P2P Networks [20, 1, 17], for hybrid application overlays [68], for group membership services in mobile ad hoc networks [23, 6], for distributed model checking [67], and for index quality determination for the world-wide web [36].

Specifically in the context of unstructured wireless sensor networks, different variants of random-walk-based protocols have been proposed and analyzed by several research groups. Servetto and Barrenechea [65] proposed and analyzed the use of constrained random walks on a grid for performing load-balanced routing between two known nodes. Avin and Brito [5] have argued that even simple random walks can be used for efficient and robust querying because they are inherently load-balanced and their partial cover times show good scaling behavior. The ACQUIRE protocol [64] provides a tunable look-ahead parameter to combine random walks with controlled floods and show that such random-walk-based hybrids can outperform flooding and even expanding-ring-based approaches in the presence of replicated data. The rumor routing algorithm [10] is a hybrid push-pull mechanism that advocates the use of multiple random walks from the events as well as the sinks, so that their intersection points can be used to provide a rendezvous point. Shakkottai [66] has analyzed different variants of random-walk-based query mechanisms and concludes that source and sink-driven sticky-searches (similar to rumor routing)
provide a rapid increase of query success probability with the number of steps. Most recently, Alanyali et al. [2] have proposed the use of random walks in energy-constrained networks to perform efficient distributed computation of a class of decomposable functions (useful in computing certain kinds of aggregates).

However, most of the research up-to-date has been based on simple classes of deterministic graphs (such as grids). Due to its increasing attractiveness, it is important to evaluate the performance of random walks on real scenarios. An important characteristic that has not been considered in the studies described above is degree heterogeneity. Degree heterogeneity is a highly likely characteristic WSN that has been already identified in some empirical works [28, 13], and the model derived in chapter 3 captures this characteristic to some extent.

The impact of degree heterogeneity on random walks have been explored by Gkantsidis [31] et. al., who report that random walks achieves better results than flooding for searching in Peer-2-Peer networks when the overlay topology is clustered. In chapter 5, we use results relating random walks and electrical networks to show that the insertion of cluster-heads (degree heterogeneity) in line topologies can enhance considerably the performance of random walk-based queries in WSN. Related to our analytical result is the work presented by Ghosh [29] et.al. where they present an optimization approach to reduce the effective resistance between two vertices, which in turn can reduce the commute time.
Chapter 3

Analysis of the Transitional Region

Experimental studies have demonstrated that the behavior of real links in low-power wireless networks (such as wireless sensor networks) deviates to a large extent from the ideal binary model used in several simulation studies [71, 62]. In particular, there is a large transitional region in wireless link quality that is characterized by significant levels of unreliability and asymmetry. In this chapter, we provide a comprehensive analysis of the root causes of unreliability and asymmetry. In particular, we derive expressions for the distribution, expectation, and variance of the packet reception rate as a function of distance, and for the location and extent of the transitional region. These expressions incorporate important channel and radio parameters such as the path loss exponent and variance of the channel, and the modulation, encoding, and hardware variance of the radios.

3.1 Overview

Wireless sensor network protocols are often evaluated through simulations that make simplifying assumptions about the link layer, such as the ideal binary model. In this model, packets are received only within the circular radio range of the transmitter. However, the real characteristics of low-power wireless links differ greatly from those on the ideal model, chiefly among these differences are the unreliable and asymmetric nature of real links. The significant differences
between the ideal model and the real behavior can lead to erroneous performance evaluation of upper-layer protocols (network layer and above).

Several studies ([28, 71, 62]) have classified low-power wireless links in three distinct reception regions: connected, transitional, and disconnected. In the connected region, links are often of good quality, stable and symmetric. On the other hand, the transitional region is characterized by the presence of unreliable and asymmetric links; and the disconnected region presents no practical links for transmission. Unfortunately, the transitional region is often quite significant in size, and in dense deployments such as those envisioned for sensor networks, a large number of the links in the network (even higher than 50% [71]) can be unreliable.

Recent studies have shown that unreliable and asymmetric links can have a major impact on the performance of upper-layer protocols. In [28], it is shown that the behavior of even the simplest flooding mechanism can be significantly affected due to asymmetric and occasional long-distance links. In [41], it is argued that the routing structures formed taking into account unreliable links can be significantly different from the structures formed based on the simple binary model. Similarly, the authors of [72] report that such unreliable links can have a negative impact on routing protocols, particularly geographic forwarding schemes.

Other works ([62, 22]) have proposed mechanisms to take advantage of nodes in the transitional region. For instance, the authors of [22] found that protocols using the traditional minimum hop-count metric perform poorly in terms of throughput, and that a new metric called ETX (expected number of transmissions), which uses nodes in the transitional region, has a better performance.

The significant impact of real link characteristics on the performance of upper-layer protocols has created an increased understanding of the need for realistic link layer models for wireless sensor networks. In order to address this need, some recent works ([62, 72, 13]) have proposed new link models based on empirical data. However, a shortcoming of these models is that they do not provide enough mathematical insight into how channel multi-path and hardware variance
affect link unreliability and asymmetry. Also, some of these works ([13, 62]) are valid only for the specific channel and radio parameters used in the deployment.

In this chapter, we use analytical tools from communication theory, simulations and experiments to present an in-depth analysis of unreliable and asymmetric links in low-power multi-hop wireless networks. The main contributions of this work are twofold. First, it allows us to quantify the impact of the wireless environment and radio characteristics on link reliability and asymmetry. And second, we propose a systematic way to generalize models for the link layer that can be used to facilitate the design of efficient routing protocols.

We also derive expressions for the packet reception rate as a function of distance, and for the size of the transitional region. These expressions incorporate several radio parameters such as modulation, encoding, output power, frame size, receiver noise floor and hardware variance; as well as important channel parameters, namely, the path loss exponent and the log-normal variance.

The Chapter is organized as follows. Section 3.2 studies the impact of multi-path on link reliability. First, we present a model for the packet reception rate as a function of distance in subsection 3.2.1. Based on this model, in subsection 3.2.2 we study the impact of channel and radio parameters on link reliability by analyzing their effect on the extent of the transitional region. Then, in subsection 3.2.3 we present approximate expressions for the expectation and variance of the packet reception rate as a function of distance. The section ends with a comparison of available link models with the one proposed in this thesis (subsection 3.2.4).

We study the impact of hardware variance in section 3.3. Hardware variance has already been identified as the cause of link asymmetry [13], in addition, we also show that it can play a significant role on the extent of the transitional region. In subsection 3.3.1, we present a model for hardware variance. Based on this model, the impact of hardware variance on link asymmetry and reliability is quantified in subsections 3.3.2 and 3.3.3, respectively. Finally, in section 3.4 we
present empirical measurements based on a test-bed of mica2 motes which validate some analytical insights of sections 3.2 and 3.3. A summary is presented in section 3.5.

Before proceeding we present the scope of our work. Our study is focused on static and low-dynamic environments and it does not consider interference effects nor the non-isotropic property of radio coverage. However, our work can be complemented with other research efforts to incorporate these properties. For instance, in [61] the authors focus on the study of interference in wireless sensor networks, Cerpa et. al. [14] study some temporal properties and [72] provides an interesting model for the non-isotropic characteristic of radio coverage, the models presented in these works can be used to complement ours. Appendix C presents some guidelines on how to combine the non-isotropic RIM model [72] with our work.

### 3.2 Impact of Channel Multi-Path

The extent of the transitional region is the result of placing specific devices, for example mica2 motes, in a specific environment, like the aisle of a building. If the characteristics of one of these elements is altered (radio or channel) then the extent of the transitional region is also altered. With the intent of analyzing how the channel and the radio determine this extent; first, we define models for both elements, and subsequently study their interaction.

#### 3.2.1 Channel and Radio Receiver Models

From the network-layer perspective, a desired abstraction for link quality is the packet reception rate as a function of distance. This abstraction can be derived by composing the channel model, which provides the received signal strength (RSS) as a function of distance, with the radio-receiver model, which provides the packet reception rate (PRR) as a function of the signal to noise ratio (SNR).
Packet Reception Rate Parameters
- packet reception rate (PRR) \( \Psi \)
- a specific PRR value in the range of \( \Psi \) \( \psi \)
- high PRR \( \psi_h \)
- low PRR \( \psi_l \)

Signal to Noise Ratio Parameters
- signal to noise ratio (SNR) \( \Upsilon \)
- a specific SNR value in the range of \( \Upsilon \) \( \gamma \)
- SNR value corresponding to \( \psi_h \) \( \gamma_h \)
- SNR value corresponding to \( \psi_l \) \( \gamma_l \)
- mean of SNR (Gaussian) for distance \( d \) \( \mu(d) \)
- bit error rate as a function of SNR \( \beta \)
- bit error rate as a function of SNR in dB \( B \)

Channel Parameters
- path loss exponent \( \eta \)
- standard deviation \( \sigma \)
- output power \( P_t \)
- received power \( P_r \)
- noise floor \( P_n \)
- Gaussian random variable \( N \)

Transitional Region Parameters
- transitional region coefficient \( \Gamma \)
- beginning of transitional region \( d_b \)
- end of transitional region \( d_e \)

<table>
<thead>
<tr>
<th>Description</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet Reception Rate Parameters</td>
<td></td>
</tr>
<tr>
<td>- packet reception rate (PRR)</td>
<td>( \Psi )</td>
</tr>
<tr>
<td>- a specific PRR value in the range of ( \Psi )</td>
<td>( \psi )</td>
</tr>
<tr>
<td>- high PRR</td>
<td>( \psi_h )</td>
</tr>
<tr>
<td>- low PRR</td>
<td>( \psi_l )</td>
</tr>
<tr>
<td>Signal to Noise Ratio Parameters</td>
<td></td>
</tr>
<tr>
<td>- signal to noise ratio (SNR)</td>
<td>( \Upsilon )</td>
</tr>
<tr>
<td>- a specific SNR value in the range of ( \Upsilon )</td>
<td>( \gamma )</td>
</tr>
<tr>
<td>- SNR value corresponding to ( \psi_h )</td>
<td>( \gamma_h )</td>
</tr>
<tr>
<td>- SNR value corresponding to ( \psi_l )</td>
<td>( \gamma_l )</td>
</tr>
<tr>
<td>- mean of SNR (Gaussian) for distance ( d )</td>
<td>( \mu(d) )</td>
</tr>
<tr>
<td>- bit error rate as a function of SNR</td>
<td>( \beta )</td>
</tr>
<tr>
<td>- bit error rate as a function of SNR in dB</td>
<td>( B )</td>
</tr>
<tr>
<td>Channel Parameters</td>
<td></td>
</tr>
<tr>
<td>- path loss exponent</td>
<td>( \eta )</td>
</tr>
<tr>
<td>- standard deviation</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>- output power</td>
<td>( P_t )</td>
</tr>
<tr>
<td>- received power</td>
<td>( P_r )</td>
</tr>
<tr>
<td>- noise floor</td>
<td>( P_n )</td>
</tr>
<tr>
<td>- Gaussian random variable</td>
<td>( N )</td>
</tr>
<tr>
<td>Transitional Region Parameters</td>
<td></td>
</tr>
<tr>
<td>- transitional region coefficient</td>
<td>( \Gamma )</td>
</tr>
<tr>
<td>- beginning of transitional region</td>
<td>( d_b )</td>
</tr>
<tr>
<td>- end of transitional region</td>
<td>( d_e )</td>
</tr>
</tbody>
</table>

Table 3.1: Mathematical Notation

In the remainder of the chapter, the SNR function is denoted by \( \Upsilon \) and the PRR function by \( \Psi \). Also, the lowercase greek letters: \( \gamma = \Upsilon(.) \) and \( \psi = \Psi(.) \), represent values taken by \( \Upsilon \) and \( \Psi \) for specific points in their respective domains. Table 3.1 presents a summary of the notation used in this chapter.

**Channel:** One of the most common radio propagation models is the log-normal path loss model [56]. This model can be used for large and small coverage systems [59]. Furthermore, empirical studies have shown that the log-normal model provides more accurate multi-path channel models than Nakagami and Rayleigh for indoor environments [53].

According to this model the received power \( (P_r) \) in dB is given by:
\[ P_r(d) = P_t - PL(d_0) - 10 \eta \log_{10}\left(\frac{d}{d_0}\right) + \mathcal{N}(0, \sigma) \tag{3.1} \]

Where \( P_t \) is the output power, \( \eta \) is the path loss exponent that captures the rate at which signal decays with respect to distance, \( \mathcal{N}(0, \sigma) \) is a Gaussian random variable with mean 0 and variance \( \sigma \) (standard deviation due to multi-path effects), and \( PL(d_0) \) is the power decay for the reference distance \( d_0 \).

Equation 3.1 does not consider non-isotropic transmission, which is an important characteristic of low-power wireless links. In Appendix C we present some guidelines on how to incorporate these non-isotropic effects in our model by using the expressions derived in the RIM model [72]. Appendix C also presents some information on how to include the path-loss effect caused by obstacles.

Radio Receiver: The receiver response is given by the packet reception rate as a function of the SNR. The packet reception rate can be derived from bit-error-rates expressions that are widely available in the wireless communication literature.

For a modulation \( M \), the packet reception rate (\( \Psi \)) is defined in terms of the bit-error-rate (\( \beta_M \)) as \(^1\):

\[
\Psi(\gamma) = (1 - \beta_M(\gamma))^f
\tag{3.2}
\]

Where \( f \) is the number of bits transmitted, and step 3 in Table 3.6 presents expressions of \( \beta_M \) for some common narrowband modulation schemes.

\( \beta_M \) is a function of the SNR, which can be obtained from equation 3.1 and is given by:

---

\(^1\)For ease of explanation, the encoding is assumed to be NRZ. Table 3.6 presents expressions for other encoding techniques.
\[ \Upsilon(d) = P_r(d) - P_n \]
\[ = \mathcal{N}(\mu(d), \sigma) \] (3.3)

Where \( \mathcal{N}(\mu(d), \sigma) \) is a Gaussian random variable with mean \( \mu(d) \), variance \( \sigma^2 \) and \( P_n \) is the noise floor. \( \mu(d) \) can be derived by inserting equation 3.1 in equation 3.3, which leads to:

\[ \mu(d) = P_t - PL(d_0) - 10 \eta \log_{10} \left( \frac{d}{d_0} \right) - P_n \] (3.4)

Given that the SNR in equation 3.3 is in dB, let us redefine the packet reception rate in equation 3.2 as a function of the SNR in dB. Denoting \( \omega(x) = 10^{x/10} \) and the bit-error-rate for SNR in dB as \( \mathcal{B}_M(\gamma_{dB}) = \beta_M(\omega(\gamma_{dB})) \), the packet reception rate \( \Psi \) can be redefined as:

\[ \Psi(\gamma_{dB}) = (1 - \mathcal{B}_M(\gamma_{dB}))^f \] (3.5)

While the previous equation is general and valid for any modulation \( M \), the figures in this section assume Non-Coherent FSK (NCFSK) modulation. The figures are for illustrative purposes and any modulation would serve that purpose. NCFSK was chosen because the empirical evaluation presented in section 3.4 uses NCFSK radios (the CC1000 equipped mica2 motes).

### 3.2.2 Impact on Link Reliability (Extent of Transitional Region)

In this subsection our aim is to quantify the impact of channel multi-path on the extent of the transitional region. Given that the channel model is a function of the SNR vs distance and the receiver response is a function of the PRR vs SNR, we can derive the behavior of the PRR vs distance by linking both expressions through the SNR metric. First, we derive the SNR values that determine which links are good or unreliable in the receiver response, and then we use these SNR values to obtain the beginning and end of the transitional region in the channel model.
Figure 3.1: (a) A receiver response where $\psi_\ell$ and $\psi_h$ determine different regions of link quality, (b) Interaction of $\gamma_\ell$ and $\gamma_h$ with the channel to determine the transitional region.

Even though there are no strict definitions for the different regions in the literature, one valid definition is the following:

*Definition 1:* In the connected region links have a high probability ($> p_h$) of having high packet reception rates ($> \psi_h$).

*Definition 2:* In the disconnected region links have a high probability ($> p_\ell$) of having low packet reception rates ($< \psi_\ell$).

Where $p_h$ and $p_\ell$ can be chosen as any numbers close to 1 and 0 respectively.

Letting $B_M^{-1}(\cdot)$ be the inverse\(^2\) of $B_M(\gamma dB)$, and $\Psi^{-1}(\psi) = B_M^{-1}(1 - \psi^{1/f})$ be the inverse of $\Psi$; the PRR values $\psi_h$ and $\psi_\ell$, from the definitions above, can be mapped to their corresponding SNR values in dB: $\gamma_h = \Psi^{-1}(\psi_h)$ and $\gamma_\ell = \Psi^{-1}(\psi_\ell)$. These SNR values determine the beginning and end of the transitional region.

---

\(^2\)BER functions are injective, hence, while there might not be a closed-form expression for their inverse function, the SNR in the domain can always be obtained numerically.
Figure 3.2: Analytical representation of equations 3.6 and 3.7.

Figure 3.1 (a) shows how $\psi_h$ and $\psi_\ell$ determine three different regions for link quality in the radio-receiver response (equation 3.5), and Figure 3.1 (b) shows how $\gamma_h$ and $\gamma_\ell$ interact with the channel (equation 3.3) to determine the extent of the connected, transitional and disconnected regions.

According to Definition 1 the beginning of the transitional region ($d_b$) satisfies the following condition:

$$p(\Psi > \psi_h) = p_h, \quad \therefore \Psi \text{ is injective}$$

$$\Rightarrow p(\Upsilon > \gamma_h) = p_h, \quad \therefore \Upsilon \text{ is Gaussian} \quad (3.6)$$

And according to Definition 2 the end of transitional region ($d_e$) satisfies:

$$p(\Psi < \psi_\ell) = p_\ell, \quad \therefore \Psi \text{ is injective}$$

$$\Rightarrow p(\Upsilon < \gamma_\ell) = p_\ell,$$

$$\Rightarrow p(\Upsilon \geq \gamma_\ell) = (1 - p_\ell), \quad \therefore \Upsilon \text{ is Gaussian}$$

$$\Rightarrow Q\left(\frac{\gamma_\ell - \mu(d_\ell)}{\sigma}\right) = (1 - p_\ell) \quad (3.7)$$

Where $Q(.)$ is the tail integral of a unit Gaussian probability density function (pdf) and $\mu(.)$ is given by equation 3.4. Figure 3.2 depicts an analytical representation of the previous equations.
This figure shows how the interaction between the channel and the receiver response determine the extent of the transitional region. Finally, \( d_b \) and \( d_e \) can be derived from equations 3.6 and 3.7:

\[
\begin{align*}
    d_b &= 10^{\frac{\gamma h - \sigma Q^{-1}(p_h) - P_t + P_n + PL(d_0)}{-10n}} \\
    d_e &= 10^{\frac{\gamma \ell - \sigma Q^{-1}(1-p_\ell) - P_t + P_n + PL(d_0)}{-10n}} \\
\end{align*}
\]  

(3.8)

While equation 3.8 provides absolute values for the extent of the different regions, it may not be useful to compare the link-quality of different scenarios. With that aim, we define the \textit{transitional region coefficient} \( \Gamma \) which is the ratio of the extent of the transitional with respect to the extent of the connected region.

\[
\begin{align*}
    \Gamma &= \frac{d_e - d_b}{d_b} \\
           &= 10^{\frac{(\gamma h - \gamma \ell) + \sigma(Q^{-1}(1-p_\ell) - Q^{-1}(p_h))}{10n}} - 1 \\
\end{align*}
\]  

(3.9)

The lower the coefficient \( \Gamma \) the smaller the transitional region compared to the connected one. For example, for the ideal binary model, where \( \gamma h = \gamma \ell \) and \( \sigma = 0 \), the coefficient \( \Gamma = 0 \). Notice that \( \Gamma \) is independent of the noise floor \( P_n \) and output power \( P_t \); a higher output power would increase the connected region, but it would increase the transitional region as well, keeping a constant ratio.

Equation 3.9 predicts the impact of the channel on the transitional region. Given that \( p_h \) and \( p_\ell \) are high probabilities, \((Q^{-1}(1-p_\ell) - Q^{-1}(p_h))\) is positive, and hence, while a small \( \sigma \) decreases the relative extent of the transitional region, a small \( \eta \) increases it. Therefore, scenarios with high \( \eta \) and low \( \sigma \) reduce the relative size of the transitional region. Figure 3.3 (a) presents \( \Gamma \) for different values of \( \eta \) and \( \sigma \), where \( p_\ell = p_h = 0.9 \), \( \gamma h = 10.23 \) dB and \( \gamma \ell = 8.20 \) dB \(^3\).

Figure 3.3 (b) depicts analytically the impact of \( \eta \) and \( \sigma \) on the extent of the transitional region. The SNR bounds on the radio receiver (\( \gamma h \) and \( \gamma \ell \)) are fixed and independent of the

\(^3\gamma h \) and \( \gamma \ell \) were obtained for a NCFSK radio with Manchester encoding and a frame size of 100 bytes. Different modulations, encoding and packet sizes do not have a significant impact on \( \Gamma \), and the results are not presented due to space constraints. Some of these results are available in [73].
environment. When \( \sigma \) increases from 1 to 2 the signal values (y-axis) have a higher probability of entering the transitional region at closer distances from the transmitter and leaving it at farther distances, which results in a larger transitional region. When \( \eta \) is increased (left arrow), the faster decay of the signal strength decreases the width of the transitional region.

Equation 3.9 also predicts the impact of the receiver. The sharper the receiver threshold, the smaller \((\gamma_h - \gamma_\ell)\) and the smaller the \(\Gamma\) coefficient. However, even with a perfect threshold receiver \((\gamma_h = \gamma_\ell)\), as the one used on the ideal model, the transitional region would still exist due to channel multi-path \((\sigma)\). Figure 3.4 (a) depicts analytically the behavior of a perfect threshold receiver in a real channel, and Figure 3.4 (b) shows an instance of the link behavior. Notice that in this hypothetical scenario the transitional region would consist only of 0/1 links.

The model also allows to provide the cumulative distribution function (cdf) of the packet reception rate as a function of distance. According to equation 3.5:

\[
F(\psi) = p(\Psi < \psi) \\
= p(\Upsilon < \Psi^{-1}(\psi)) \\
= 1 - Q\left(\frac{\Psi^{-1}(\psi) - \mu(d)}{\sigma}\right) \tag{3.10}
\]
Where $\mu(d)$ is the average SNR decay (equation 3.4). Figure 3.5 shows an example of the cumulative distribution $F(\psi)$ for $\eta = 3$ and $\sigma = 3$. Three different transmitter-receiver distances are shown: end of connected region, middle of transitional and beginning of disconnected region. We can notice that independent of the region where the receiver is, the link has a higher probability of being either good or bad (above 0.9 or below 0.1 PRR) than being unreliable (between 0.9 and 0.1). For instance, in the middle of the transitional region the link has a 30% probability of being unreliable; and the probability of observing unreliable links at the end of the connected region or at the beginning of the disconnected region is small ($< 5\%$). Empirical measurements in [13, 73] agree with the analytical $cdf$ in equation 3.10.

It is important to remark that the obtained $cdfs$ are valid only for the scope of this work (static and low-dynamic environments); highly dynamic environments add a new dimension of time to the $cdfs$.

### 3.2.3 Expectation and Variance of Packet Reception Rate

Even though a longer distance does not necessarily imply a lower packet reception rate, the expected value of the packet reception rate does decrease monotonically with distance in a given
propagation direction\(^4\). In this subsection, we present approximate expressions for the expectation and variance of the packet reception rate \(\Psi\). These expressions are important because they confirm mathematically that the transitional region has a higher variability in PRR than the connected region.

First we present the general expressions for the expectation and variance in equations 3.11 and 3.12. These expressions depend on the PRR versus distance function (receiver response given in equation 3.5) and the probability density function (pdf) of the SNR for a given distance \(d\) (which is log-normally distributed). Given the mathematical complexity of dealing with the receiver response and the pdf, we derive approximate expressions for the expectation and variance of the packet reception rate \(\Psi\).

In general, the first two moments of \(\Psi\) are defined by:

\[
E[\Psi] = \int_{-\infty}^{\infty} \Psi(\gamma_{dB}) f(\gamma_{dB}, d) \, \delta\gamma_{dB}
\]

\[
E[\Psi^2] = \int_{-\infty}^{\infty} \Psi^2(\gamma_{dB}) f(\gamma_{dB}, d) \, \delta\gamma_{dB}
\]

\(^4\)The radio model used in this work is isotropic, but this is not true of practical antennas. By linearity of expectation, since \(E[\Psi_a(d)]\) is monotonic with distance for a given propagation direction \(a\), it can be shown that the expected PRR averaged over all angles is also monotonic with distance; however, it should be kept in mind that expected PRR values at different angles may show non-distance-monotonic behavior with respect to each other.
Figure 3.6: Linear approximation of receiver threshold and Gaussian SNR, the mean of the Gaussian depends on the transmitter-receiver distance.

Where \( f(\gamma_{dB}, d) \) represents the pdf of SNR (a Gaussian random variable with parameters \( \mu(d) \) and \( \sigma \)).

The sharp thresholds of \( \Psi \) and \( \Psi^2 \) permit linear approximations:

\[
\Psi(\gamma) \approx \Psi_L(\gamma) = \begin{cases} 
0, & \gamma \leq \gamma_{0e} \\
m_e \gamma + b_e, & \gamma_{0e} < \gamma < \gamma_{1e} \\
1, & \gamma \geq \gamma_{1e}
\end{cases} \quad (3.13)
\]

\[
\Psi^2(\gamma) \approx \Psi^2_L(\gamma) = \begin{cases} 
0, & \gamma \leq \gamma_{0v} \\
m_v \gamma + b_v, & \gamma_{0v} < \gamma < \gamma_{1v} \\
1, & \gamma \geq \gamma_{1v}
\end{cases} \quad (3.14)
\]

Where \( m_e, m_v \) and \( b_e, b_v \) are the slopes and y-intercepts of the linear approximations \( \Psi_L \) and \( \Psi^2_L \), and \( \gamma \) is in dB. Figure 3.6 shows the approximation procedure for \( \Psi_L \); the procedure for \( \Psi^2_L \) is similar. The mechanism to obtain the slopes, y-intercepts and limit points of equations 3.13 and 3.14 is presented later.

The linear models lead to the following approximations of equations 3.11 and 3.12:
\[ E[\Psi] \approx \int_{\gamma_0}^{-} \Psi L(\gamma dB) f(\gamma dB, d) \delta \gamma dB \]
\[ = \int_{\gamma_0}^{-} (m_e \gamma + b_e) f(\gamma dB, d) \delta \gamma dB \]
\[ + Q\left(\frac{\gamma_e - \mu(d)}{\sigma}\right) \]
\[ (3.15) \]

\[ E[\Psi^2] \approx \int_{\gamma_0}^{-} \Psi^2 L(\gamma dB) f(\gamma dB, d) \delta \gamma dB \]
\[ = \int_{\gamma_0}^{-} (m_v \gamma + b_v) f(\gamma dB, d) \delta \gamma dB \]
\[ + Q\left(\frac{\gamma_0 - \mu(d)}{\sigma}\right) \]
\[ (3.16) \]

In the above approximations \( f(\gamma dB, d) \) is evaluated separately on intervals \([\gamma_0, \gamma_1e]\) and \([\gamma_0, \gamma_1v]\) for \( E[\Psi] \) and \( E[\Psi^2] \), respectively. Both intervals represent the linear approximations of the sharp thresholds of \( \Psi \) and \( \Psi^2 \), and these thresholds are narrow compared to the \([\mu - 4\sigma, \mu + 4\sigma]\) domain of \( f(\gamma dB, d) \), hence, linear approximations can be used as well for \( f(\gamma dB, d) \) in \([\gamma_0, \gamma_1e]\) and \([\gamma_0, \gamma_1v]\).

Let us denote \( f_\Psi(\gamma dB, d) \) and \( f_{\Psi^2}(\gamma dB, d) \) as the linear approximations of \( f(\gamma dB, d) \) for intervals \([\gamma_0, \gamma_1e]\) and \([\gamma_0, \gamma_1v]\):

\[ f_\Psi(\gamma dB, d) = m_{ge} \gamma + b_{ge} \]
\[ f_{\Psi^2}(\gamma dB, d) = m_{gv} \gamma + b_{gv} \]
\[ (3.17) \]
\[ (3.18) \]

where:

\[ m_{ge} = \frac{f(\gamma_{1e}, d) - f(\gamma_{0e}, d)}{\gamma_{1e} - \gamma_{0e}} \quad b_{ge} = \frac{f(\gamma_{0e}, d) \gamma_{1e} - f(\gamma_{1e}, d) \gamma_{0e}}{\gamma_{1e} - \gamma_{0e}} \]
\[ m_{gv} = \frac{f(\gamma_{1v}, d) - f(\gamma_{0v}, d)}{\gamma_{1v} - \gamma_{0v}} \quad b_{gv} = \frac{f(\gamma_{0v}, d) \gamma_{1v} - f(\gamma_{1v}, d) \gamma_{0v}}{\gamma_{1v} - \gamma_{0v}} \]

\(^5\)While the domain of a Gaussian random variable is \([-\infty, +\infty]\), the interval \([\mu - 4\sigma, \mu + 4\sigma]\) contains most of the probability space (.999), and it is wide compared to the sharp threshold of the receiver for common values of \( \sigma \) ([60]).
Figure 3.6 shows the approximation procedure for $f_\Psi(\gamma dB, d)$ (Gaussian SNR curve for $E[\Psi]$); the procedure for $f_{\Psi^2}(\gamma dB, d)$ is similar.

Finally, based on equations 3.15 and 3.16, the first and second moment approximations of the packet reception rate are given by:

$$E[\Psi] \approx \int_{\gamma_0}^{\gamma_1} (m_e \gamma + b_e) f_\Psi(\gamma dB, d) \delta\gamma dB$$
$$+ Q(\frac{\gamma_1 - \mu(d)}{\sigma})$$
$$= \int_{\gamma_0}^{\gamma_1} (m_e \gamma + b_e) (m_{ge} \gamma + b_{ge}) \delta\gamma dB$$
$$+ Q(\frac{\gamma_1 - \mu(d)}{\sigma})$$
$$= ((m_e + m_{ge}) \gamma^2 + (b_e m_{ge} + b_{ge} m_e) \gamma^2 + b_e b_{ge} \gamma)\gamma_0 + Q(\frac{\gamma_1 - \mu(d)}{\sigma})$$

$$E[\Psi^2] \approx \int_{\gamma_0}^{\gamma_1} (m_v \gamma + b_v) f_{\Psi^2}(\gamma dB, d) \delta\gamma dB$$
$$+ Q(\frac{\gamma_1 - \mu(d)}{\sigma})$$
$$= \int_{\gamma_0}^{\gamma_1} (m_v \gamma + b_v) (m_{gv} \gamma + b_{gv}) \delta\gamma dB$$
$$+ Q(\frac{\gamma_1 - \mu(d)}{\sigma})$$
$$= ((m_v + m_{gv}) \gamma^2 + (b_v m_{gv} + b_{gv} m_v) \gamma^2 + b_v b_{gv} \gamma)\gamma_0 + Q(\frac{\gamma_1 - \mu(d)}{\sigma})$$

In general, the parameters of $\Psi_L$ and $\Psi^2_L$ (slopes, y-intercepts and limit points of equations 3.13 and 3.14) can be obtained by curve-fitting $\Psi$ and $\Psi^2$ through least squares regression techniques, nevertheless, our studies suggest that choosing a line that passes through points A and B with PRRs of 0.1 and 0.9 provides an accurate approximation. Hence, A and B defined as $(\Psi^{-1}(0.1), 0.1)$ and $(\Psi^{-1}(0.9), 0.9)$ can be used to obtain the different parameters of $\Psi_L$:

$^6$Actually, no significant differences were found if points A and B are chosen in intervals [0.01, 0.2] and [0.8, 0.99], respectively.
Figure 3.7: Comparison of $E[\Psi]$ and $\text{Var}[\Psi]$ with their linear approximations, $E[\Psi_L]$ and $\text{Var}[\Psi_L]$. 

\[
m_c = \frac{0.9 - 0.1}{\gamma_B - \gamma_A} \quad b_c = \frac{0.1 \gamma_B - 0.9 \gamma_A}{\gamma_B - \gamma_A}
\]

\[
\gamma_{0v} = \frac{-b_v}{m_v} \quad \gamma_{1v} = \frac{1-b_v}{m_v}
\]

Where $\gamma_A = \Psi^{-1}(0.1)$ and $\gamma_B = \Psi^{-1}(0.9)$, both in dB. For $\Psi_L^2$, points A and B are $(\Psi^{-1}(\sqrt{0.1}), 0.1)$ and $(\Psi^{-1}(\sqrt{0.9}), 0.9)$. 

Figure 3.7 shows an example of numerically calculated curves for the expectation and variance (from equations 3.11 and 3.12), and their approximations through equations 3.19 and 3.20 for $\eta = 3$ and $\sigma = 3$. In general, the error depends on the parameters of $f(\gamma_{dB}, d)$ (pdf of SNR). The smaller $\sigma$, the larger the error because the width of the receiver threshold starts to be comparable with the width of the bell of the Gaussian curve which leads to a less accurate linearization. However, for common values of $\sigma$ [60] the bell is significantly wider than the receiver threshold and the approximation errors are not significant. Also, while the expectation decreases monotonically with distance, the variance has a bell shape whose maximum lies in the transitional region; this behavior agrees with empirical observations in [62].
3.2.4 Comparison With Available Link Models

Some popular wireless network simulators [46, 30] and recent studies [62] had been using a Gaussian random variable to represent the packet reception rate. The PRR function based on the Gaussian model (\( \Psi_G \)) has the following form:

\[
\Psi_G = \begin{cases} 
1, & X > 1 \\
x, & 0 \leq X \leq 1 \\ 
0, & X < 0
\end{cases}
\]  
(3.21)

Where \( X \) is a Gaussian random variable with parameters \( \mu = E[\Psi] \) and \( \sigma^2 = Var(\Psi) \). The Gaussian model leads to the following \( cdf \) \( F_G \):

\[
F_G(\psi) = \begin{cases} 
1 - Q\left(\frac{E[\Psi]}{\sqrt{Var[\Psi]}}\right), & \psi = 0 \\
1 - Q\left(\frac{\psi - E[\Psi]}{\sqrt{Var[\Psi]}}\right), & 0 < \psi < 1 \\
1, & \psi = 1
\end{cases}
\]  
(3.22)

Figure 3.8 shows a comparison between the \( cdfs \) of the Gaussian model (equation 3.22) and our analytical model (equation 3.10) for receivers in the connected, transitional and disconnected states.
regions. Contrary to the analytical cdf, where links have higher probability of being either good or bad (above 0.9 or below 0.1 PRR), the Gaussian model leads to links that have a high probability of being between 0.9 and 0.1; 60% for the node in the transitional region and 40% for the node in the connected region, which may lead to misleading results in protocol testing. The results shown are for $\eta = 3$, $\sigma = 3$ and a non-coherent FSK radio, but similar trends are obtained for different parameters.

3.3 Impact of Hardware Variance

In the previous section it was assumed that all radios have the same output power $P_t$ and noise floor $P_n$, however, hardware variance induces some fluctuation around the output power set by the user and around the average noise floor. This variance problem is partially solved during the manufacturing process, where radios with a low output power and/or a high noise floor (low sensitivity) are usually discarded. However, no upper-bound is used in the filtering process and hardware variance remains as a problem. As stated in [54]: This filtering process is justifiable, since radios that are more powerful or more sensitive are generally desirable.

Hardware variance has already been identified as the cause of asymmetric links [28]. In this section, we not only quantify the effect of hardware variance on link asymmetry, but we also show that hardware variance can have a significant impact on the extent of the transitional region.

It is important to notice that while the output power variance can be calibrated to the same value for all radios, the noise floor variance cannot be eliminated through calibration since it depends on the thermal noise generated by the underlying solid state structure.
3.3.1 Model

Hardware variance causes Gaussian distributions (in dB) in the output power and noise floor [54]. In order to capture these effects let us redefine equation 3.3 by denoting $SNR_{AB}$ as the signal-to-noise ratio measured at B for the output power of A, then $SNR_{AB} (\Upsilon_{AB})$ is given by:

$$\Upsilon_{AB} = P_{t,A} - PL(d) - P_{n,B}$$

$$= N(P_t, \sigma_{tx}) - PL(d) - N(P_n, \sigma_{rx}) \quad (3.23)$$

Where $\sigma_{tx}^2$ are $\sigma_{rx}^2$ are the variances of the output power and the noise floor respectively, and $PL(d) = PL(d_0) + 10 \eta \log_{10}(\frac{d}{d_0}) + N(0, \sigma)$ is the channel path loss (which is identical in both directions: $A \rightarrow B$ and $B \rightarrow A$).

Empirical measurements (Section 3.4) show that there is some correlation between the output power and noise floor within the same radio. Our model captures this correlation by representing the output power and noise floor as a multivariate Gaussian distribution, as shown below:

$$\left( \begin{array}{c} T \\ R \end{array} \right) \sim N \left( \left( \begin{array}{c} P_t \\ P_n \end{array} \right), \left( \begin{array}{cc} S_T & S_TR \\ S_RT & S_R \end{array} \right) \right) \quad (3.24)$$

Where $P_t$ is the nominal output power, $P_n$ is the average noise floor, $S$ the covariance matrix between the output power and noise floor; and $T$ and $R$ are the actual output power and noise floor of a specific radio, respectively.

3.3.2 Impact on Asymmetric Links

When the output power level of all the nodes is set to the same value, radios with identical non-variant hardware ($\sigma_{tx} = 0, \sigma_{rx} = 0$) lead to the same SNR in both directions ($\Upsilon_{AB} = \Upsilon_{BA}$ according to equation 3.23), which in turn leads to the same packet reception rate (symmetric links).
For radios with hardware variance, $\Upsilon_{AB}$ can be different from $\Upsilon_{BA}$. Figure 3.9 shows the effect of $\Upsilon_{AB} - \Upsilon_{BA}$ on link asymmetry. Due to the sharp threshold of the receiver, a small value of $\Upsilon_{AB} - \Upsilon_{BA}$ ($\sim 3.2$ dB) may lead to significantly different packet reception rates in both directions (1.0 and 0.4).

$\Upsilon_{AB} - \Upsilon_{BA}$ is a random variable and the larger the variance of this difference, the higher the probability of link asymmetry. In order to quantify the impact of hardware variance on link asymmetry we will analyze the variance of $(\Upsilon_{AB} - \Upsilon_{BA})$.

Letting $(T_A, R_A)$ and $(T_B, R_B)$ be the respective output power and noise floor of radios A and B, then:

$$
\Upsilon_{AB} - \Upsilon_{BA} = (T_A - PL(d) - R_B) - (T_B - PL(d) - R_A)
$$

(3.25)

$$(T_A + R_A) - (T_B + R_B)$$

$(T_A + R_A)$ and $(T_B + R_B)$ are gaussian random variables representing the sum of the output power and noise floor of different radios ($A$ and $B$), and can be assumed to be independent\(^7\).

\(^7\)The manufacturing process can create some correlation among different radios if different batches are produced from special high (low) quality materials, but we assume that all radios belong to the same process.
\((T_A + R_A)\) and \((T_B + R_B)\) are generated from the same multivariate Gaussian distribution and can be represented by \((T + R)\), hence, \(Var(\Upsilon_{AB} - \Upsilon_{BA}) = 4 \times Var(T + R)^8\), and:

\[
Var(T + R) = E[(T + R)^2] - E^2[T + R] \\
+ 2(E[TR] - E[T]E[R]) \\
= Var(T) + Var(R) + 2Cov(T, R) \\
= S_T + S_R + 2S_{TR}
\]  

Which leads to:

\[
Var(\Upsilon_{AB} - \Upsilon_{BA}) = 4(S_T + S_R + 2S_{TR}) \tag{3.27}
\]

Where \(S_T\), \(S_R\) and \(S_{TR}\) are elements of the covariance matrix in equation 3.24.

Equation 3.27 shows that a positive correlation (positive \(S_{TR}\)) between the output power and noise floor of a radio leads to a high variance of \(\Upsilon_{AB} - \Upsilon_{BA}\) (higher probability of link asymmetry), while a negative correlation (negative \(S_{TR}\)) reduces the variance (lower probability of link asymmetry). Notice that a negative correlation implies that nodes with output powers higher than \(P_t\) (better transmitter) will usually have a noise floor lower than \(P_n\) (better receiver), and vice versa.

Hence a negative correlation between the output power and noise floor leads to the lowest probability of link asymmetry, followed by zero correlation and positive correlation.

\textsuperscript{8}This is derived from the facts that for a random variable \(X\), \(Var(X) = Var(-X)\); and for i.i.d random variables \(X_i\), \(Var(\sum_i X_i) = \sum_i Var(X_i)\)
3.3.3 Impact on Extent of Transitional Region

In equation 3.3, the randomness of the SNR was due uniquely to multi-path effects, but the variance of the output power and noise floor introduces two other sources of randomness. The combined effect of output power variance, channel multi-path and noise floor variance led to a new expression for the SNR (equation 3.23). Based on this equation the SNR $\Upsilon$ is given by:

$$
\Upsilon = \mathcal{N}(P_t, \sigma_{tx}) - PL(d) - \mathcal{N}(P_n, \sigma_{rx})
$$

(3.28)

Where $\sigma^2_{hw} = \sigma^2_{tx} + \sigma^2_{rx}$. Finally, $\Upsilon$ is given by:

$$
\Upsilon = \mathcal{N}(P_t - P_n, \sigma_{hw}) - PL(d)
$$

(3.29)

Where $PL(d_0) = PL(d_0) + 10 \eta \log_{10}(\frac{d}{d_0})$, and the total variance of the system ($\sigma_t$) is given by:

$$
\sigma^2_t = \sigma^2_{ch} + \sigma^2_{tx} + \sigma^2_{rx}
$$

(3.30)

The hardware variance generates a pseudo-path loss variance ($\sigma_{hw}$). Equation 3.9 showed that the larger the variance, the larger the extent of the transitional region; hence, radios with hardware variance will always increase the extent of the transitional region.

To obtain accurate results for the extent of the transitional region, $\sigma$ should be replaced by $\sigma_t$ in all corresponding equations in Section 3.2. The impact of hardware variance on the extent of the transitional region can be observed in Figure 3.10, which presents simulated link qualities for $\eta = 3$, $\sigma_{ch} = 3$ and $\sigma_{hw} = 3$. Figure 3.10 (a) shows the transitional region when invariant hardware is placed in a real channel (effect of $\sigma^2_{ch}$), Figure 3.10 (b) presents a hypothetical scenario where variant hardware is placed in an ideal scenario (no multi-path effects), we observe that even
in the absence of multi-path effects a transitional region is observed due to the pseudo-variance $\sigma^2_{hw}$. Finally, Figure 3.10 (c) presents the combined effects of $\sigma_{ch}$ and $\sigma_{hw}$, showing a larger transitional region than in Figures 3.10 (a) and (b).

### 3.4 Experiments with Motes

We now present empirical results conducted in static and low-dynamic environments to validate our analytical results on the impact of $\eta$, $\sigma_{ch}$ and $\sigma_{hw}$ on the extent of the transitional region. We will also observe that the correlation between output power and noise floor in mica2 motes is negative, which is the least damaging in terms of link asymmetry among the different correlations (positive, zero, negative).

We considered two environments, an indoor environment (aisle of a building), and an outdoor environment (football field). All the measurements were made using mica2 motes. These devices use Non-Coherent FSK modulation at 915 MHz with Manchester encoding and provide data rates of 38.4 Kbaud.
3.4.1 Channel and Radio Parameters

**Channel:** Two motes were used to measure the path loss exponent ($\eta$), the variance ($\sigma_{ch}^2$) and the initial decay $PL(d_0)$ of the channel. Table 3.2 presents the values for $\eta$ and $\sigma_{ch}$. The reference distance ($d_0$) of the log-normal model was set to 1m and its corresponding power decay was found to be 55 dB.

**Radio:** One mote was selected as a common receiver and sender to capture the variance of the output power $P_t$ and noise floor $P_n$. The measurements were done in an isolated empty room, where each mote had the same source power and was placed at the same physical position with respect to the reference mote. Figure 3.11 presents the empirical measurements, which shows a negative correlation between output power and noise floor. From our experiments, the resultant covariance matrix is given by:

$$
S = \begin{pmatrix}
6.0 & -3.3 \\
-3.3 & 3.7
\end{pmatrix}
$$

<table>
<thead>
<tr>
<th>environment</th>
<th>$\eta$ (95% conf. bounds)</th>
<th>$\sigma_{ch}$ (95% conf. bounds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>outdoor</td>
<td>4.7 (4.3 - 5.1)</td>
<td>3.2 (2.6 - 3.8)</td>
</tr>
<tr>
<td>indoor</td>
<td>3.3 (2.1 - 4.5)</td>
<td>5.5 (4.6 - 6.8)</td>
</tr>
</tbody>
</table>

Table 3.2: Empirical Channel Parameters
Table 3.3: Empirical Radio Parameters

<table>
<thead>
<tr>
<th></th>
<th>(95% conf. bounds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>output power $\sigma_{tx}$</td>
<td>2.3 (1.7 - 3.5)</td>
</tr>
<tr>
<td>noise floor $\sigma_{rx}$</td>
<td>1.9 (1.4 - 3.1)</td>
</tr>
</tbody>
</table>

The standard deviations of the output power ($\sigma_{tx}$) and noise floor ($\sigma_{rx}$) are presented in table 3.3, these values lead to $\sigma_{hw} = 3.0$. Different power levels were tested and all levels showed a similar variance.

The negative correlation of mica2 motes is due to several factors. Nowadays chip implementation is moving toward single chip design, and hence, the performance of the transmitter and receiver is determined by the common underlying solid-state structure. Board implementation and antenna gains further enhance this correlation since a common path goes from the antenna to the chip. Hence, a radio with a good solid-state structure (low thermal noise) and a high path-antenna gain will lead to higher output powers (good transmitter), and a lower noise floor (good receiver). Also, while our measurements where done in controlled scenarios, characteristics of real deployments such as the remaining output power of batteries, may further enhance the correlation.

It is important to observe that this negative correlation leads to some nodes being good transmitters and receivers which may create some cluster behavior, as observed in some empirical studies [28, 13].

The negative correlation has a direct impact on the relation between the out-degree and in-degree of nodes\(^9\). In section 3.3.2 we had stated that the negative correlation between the output power and noise floor leads to the lowest level of link asymmetry which implies that the in-degree and out-degree of nodes will be more similar than for positive and zero correlations. Figure 3.12 shows simulation results for the relation between in-degree and out-degree for positive and negative correlations between the output power and noise floor. It can be observed the close

\(^9\)In-degree is the number of neighbors that can communicate with a specific node, out-degree is the number of neighbors that a specific node can communicate with.
Figure 3.12: Simulation results for the relation between in-degree and out-degree, (a) positive correlation, (b) negative correlation.

relation between in-degree and out-degree for the negative correlation. Figure 3.13 shows the empirical in-degree/out-degree relation of all nodes for all the tested power levels\(^{10}\), and we can observe that the empirical trend agrees with the simulation results. This close relation between in-degree and out-degree is highly desirable given the strong dependence that several medium access and network layer protocols have on symmetric links.

**Noise Floor:** Given that our work does not consider interference, the noise floor can be obtained by the well-known thermal noise equation [56], which leads to a value of -115 dBm for the parameters of the radio chip [75]. However, our measurements showed that the average noise floor is approximately -105 dBm, the 10 dB difference is mainly due to losses from the output-pin of the chip to the antenna, which are not considered in the thermal noise equation. These losses depend on board implementation and are beyond the scope of this work. Hence, for the model, the average noise floor \(P_n\) will be set to -105 dBm.

Finally, it is important to consider that bit-error-rate expressions are usually given in terms of \(\frac{E_b}{N_0}\) (known as SNR per bit), however, most commercially available radios provide only RSSI

\(^{10}\)Links were considered valid if they had a PRR above 10%. The same trend is observed for any blacklisting threshold.
measurements, which can be converted to SNR per packet (ϒ). Γ has a simple relation with \( \frac{E_b}{N_0} \):

\[
\Gamma = \frac{E_b}{N_0} \frac{R}{B_N},
\]

for mica2 motes \( R = 19.2 \text{ kbps} \) (data rate) and \( B_N = 30 \text{ kHz} \) (noise bandwidth). Hence, all RSSI measurements can be converted to \( \frac{E_b}{N_0} \) values.

### 3.4.2 Chain Topologies

For each environment, a chain topology of 21 motes was deployed with nodes spaced 1 meter apart. The frame size was 50 bytes with a preamble of 28 bytes. A simple TDMA protocol was implemented to avoid collisions. Every mote transmitted 100 packets at a rate of 5 packets/sec. Upon reception of a packet the sequence number and the received signal strength \( P_r \) were stored; simultaneously, the noise floor was sampled. The average SNR and packet reception rate were measured for all the links in the network.

According to equation 3.30, the expected total variance is the sum of the channel and radio variances (sum of variances of tables 3.2 and 3.3). Table 3.4 shows the expected \( \sigma_t \), the measured

<table>
<thead>
<tr>
<th></th>
<th>Expected ( \sigma_t )</th>
<th>Measured ( \sigma_t )</th>
<th>Measured ( \sigma_{ch} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>indoor</td>
<td>6.3</td>
<td>6.1</td>
<td>5.5</td>
</tr>
<tr>
<td>outdoor</td>
<td>4.8</td>
<td>5.1</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Table 3.4: Comparison of Total Variance and Channel Variance
σt and the σch. The expected σt is derived from the σch and σhw of tables 3.2 and 3.3, and the measured σt is obtained from the chain topology experiment. We observe that the expected and measured σt are similar. It can also be observed that σch is smaller than σt (specially for the outdoor environment) confirming that hardware variance contributes to the total variance, and consequently to the extent of the transitional region.

In order to validate our model we present two comparisons. The first one is a formal method based on the Kolmogorov-Smirnov (K-S) test and the second one is a comparison of the packet reception rates versus distance between empirical and simulated data. Figure 3.14 presents the cumulative distribution of the packet reception rate for the chain topology described above. Two power levels, -7 dBm and 5 dBm, are presented for the outdoor environment. For the number of links in the chain topology (420) and a confidence interval of 10% the K-S table has a threshold value of 0.06. For practical purposes let us define links with PRR above 0.9 as reliable and neglect links with PRR below 0.1, then the distance D of the K-S test is considered between 0.1 and 0.9. We observe that low density networks such as the medium power case would not pass the test (0.17 > 0.06), but high density networks such as the high power case pass the test (0.05 < 0.06),

---

*A very reasonable assumption considering that links below 0.1 would incur in significant losses or in high number of retransmissions*
i.e. for the high power case both distributions can be considered similar (empirical and simulated).

It is also important to notice that the empirical data shows that most of the probability mass is either above 0.9 or below 0.1 for any output power as it was shown in Section 3.2.4. The next comparisons will illustrate why low-dense networks are predicted less accurately than high dense networks.

Figure 3.15 shows the empirical packet reception rates versus distance compared to their analytical counterparts. It can be observed that the model provides a reasonable approximation of the real behavior. Table 3.5 shows the expected beginning and end of the transitional region

<table>
<thead>
<tr>
<th></th>
<th>beginning (m)</th>
<th>end (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>indoor $P_t = -7\text{dB}$</td>
<td>4.8</td>
<td>23.4</td>
</tr>
<tr>
<td>outdoor $P_t = -7\text{dB}$</td>
<td>3.4</td>
<td>8.1</td>
</tr>
<tr>
<td>outdoor $P_t = 5\text{dB}$</td>
<td>6.1</td>
<td>14.6</td>
</tr>
</tbody>
</table>

Table 3.5: Analytical Extent of Transitional Region
according to equation 3.8 for the scenarios in Figure 3.15. We can observe that this equation provides reasonable predictions for the empirical observations. Several power levels were tested, from -20 dBm to 5 dBm in steps of 1 dBm, and all power levels showed a similar behavior to the proposed model. A comparison of Figures 3.15 (b) and (e) provides a better understanding of why the K-S test fails to assess both distributions as similar. In the simulated case (Figure 3.15 (e)), the model tends to classify more links as good links than the empirical data. For instance, the simulated data shows that all links under a distance of 2 meters are considered to be 1.0, while the empirical data shows some unreliable links at a distance of 2 meters. Also, for distances 5 and 6 some links are good (> 0.9) in the simulated data, while in the empirical data most of them are below 0.9. Since the transitional region is narrow, the disagreement on a few links leads to approximately 15% of the mass probability shifting from bad links to good links, which causes the failure of the K-S test. Nevertheless, it is also worth considering that the major disagreement is only on the differences at the extremes (good and bad links), the slope for the unreliable links is similar, showing that both empirical and modeled data have a close approximation on the number of unreliable links. It is important to mention that our model is not meant to be an exact replica of the environment but an approximation to it.

3.5 Summary

This chapter presented an analysis on the unreliability and asymmetry of low-power wireless links. The analysis presented allow us to provide analytical expressions for the boundaries of the transitional region and a systematic approach to obtain mathematical link layer models.

The analysis also allowed us to provide some important insights about the impact of channel multi-path and hardware variance on the transitional region. First, the relative size of the transitional region (Γ coefficient) is higher for lower path loss exponents and higher variances. Second, hardware variance induces a pseudo-log-normal variance, which increases the size of the
transitional region. Third, a negative correlation between the output power and noise floor leads to nodes that are good transmitters and receivers, which helps to explain the clustering behavior observed in previous works [28, 13]. And fourth, even with a perfect-threshold radio, the transitional region still exists as long as there are multi-path effects.

Even though the simulations and empirical validation in this chapter were based on radios using NC-FSK modulation and Manchester encoding, the model can be easily extended to other radio characteristics. Table 3.6 presents the steps required for other common modulation techniques and encoding schemes. It is important to highlight that while different modulations, encoding and packet sizes lead to different sizes of the regions, they do not significantly affect the transitional region $\Gamma$ coefficient, some of the results are available in [73].

The work presented in this chapter contributes to a better understanding of the behavior of low-power wireless links but is not exhaustive. It can be complemented with other studies to capture other important phenomenon present in real scenarios; for instance, contention models from [61], temporal properties from [14] and correlations due to direction of propagation from [72].

The next two chapters use the expressions and model derived in this chapter to evaluate and enhance the performance of geographic routing and random walk-based querying mechanisms.
<table>
<thead>
<tr>
<th>STEP 0 : Radio</th>
<th>Obtain output power and noise floor for all nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Can use Cholesky decomposition to generate multivariate r.v. equation 3.24</td>
</tr>
<tr>
<td></td>
<td>For mica2: $-20 \text{ dBm} &lt; P_t &lt; 5 \text{ dBm}$, $P_n = -105 \text{ dBm}$</td>
</tr>
<tr>
<td>STEP 1 : Channel</td>
<td>Obtain channel parameters $PL(d_0), n, \sigma$</td>
</tr>
<tr>
<td></td>
<td>Can be obtained through own empirical measurements, or from some published results [60]</td>
</tr>
<tr>
<td>STEP 2 : SNR</td>
<td>Obtain SNR in dB ($\gamma_{dB}$) as a function of distance $d$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{dB}(d) = T - PL(d_0) - 10\log_{10}(\frac{d}{d_0}) - N(0, \sigma) - R$</td>
</tr>
<tr>
<td>STEP 3 : Modulation</td>
<td>Select modulation and insert $\gamma(d)$ from previous step, but not in dB (i.e. $\gamma = 10 \frac{\gamma_{dB}}{10}$)</td>
</tr>
<tr>
<td></td>
<td>Convert from $\frac{E_b}{N_0}$ to RSSI by inserting appropriate bit data rate $R$ and noise bandwidth $B_N$</td>
</tr>
<tr>
<td></td>
<td>According to modulation select appropriate BER ($P_e$)</td>
</tr>
<tr>
<td>ASK noncoherent</td>
<td>$\frac{1}{2}[\exp^{-\gamma(d) \frac{B_N}{R}} + Q(\sqrt{\gamma(d) \frac{B_N}{R}})]$</td>
</tr>
<tr>
<td>ASK coherent</td>
<td>$Q(\sqrt{\gamma(d) \frac{B_N}{R}})$</td>
</tr>
<tr>
<td>FSK noncoherent</td>
<td>$\frac{1}{2} \exp^{-\gamma(d) \frac{B_N}{R}}$</td>
</tr>
<tr>
<td>FSK coherent</td>
<td>$Q(\sqrt{\gamma(d) \frac{B_N}{R}})$</td>
</tr>
<tr>
<td>PSK binary</td>
<td>$\frac{1}{2} \exp^{-\gamma(d) \frac{2B_N}{R}}$</td>
</tr>
<tr>
<td>PSK differential</td>
<td>$\frac{1}{2} \exp^{-\gamma(d) \frac{2B_N}{R}}$</td>
</tr>
<tr>
<td>STEP 4 : Encoding</td>
<td>Select packet reception rate</td>
</tr>
<tr>
<td></td>
<td>Select according to encoding scheme, then insert frame, preamble length and $P_e$ obtained in previous step</td>
</tr>
<tr>
<td>NRZ</td>
<td>$(1 - P_e)^{8\ell} (1 - P_e)^{8(f-\ell)}$</td>
</tr>
<tr>
<td>4B5B</td>
<td>$(1 - P_e)^{8\ell} (1 - P_e)^{8(f-\ell)1.25}$</td>
</tr>
<tr>
<td>Manchester</td>
<td>$(1 - P_e)^{8\ell} (1 - P_e)^{8(f-\ell)2.0}$</td>
</tr>
<tr>
<td>SECDED</td>
<td>$(1 - P_e)^{8\ell} ((1 - P_e)^8 + 8P_e(1 - P_e)^7)^{f-\ell}3.0$</td>
</tr>
</tbody>
</table>

Table 3.6: Theoretical Models for the Link Layer
Chapter 4

Impact of Lossy Links on Geographic Routing

Geographic routing [38] is a popular mechanism that uses location information to deliver packets in multi-hop wireless networks. With this mechanism, nodes need to know only the location information of their neighbors and the location of the final destination. The low state and overhead of geographic routing makes it an attractive mechanism for packet delivery in WSN.

Geographic routing commonly employs a maximum-distance greedy technique to forward packets. At each hop, packets are delivery to the neighbor geographically closest to the destination. This forwarding technique performs well in ideal conditions such as the binary model, where a link exist as long as the receiver is within the transmission range. However, in the previous chapter we showed that as the transmitter-receiver distance increases the probability of encountering an unreliable link also increases. Hence, in real scenarios the distance-greedy mechanism of geographic routing is likely to select lossy links, which would degrade its performance. In this chapter we used the model presented in chapter 3 to identify and illustrate this weak-link problem, and present optimal local forwarding techniques to maximize the energy efficiency of geographic routing.

4.1 Overview

Geographic routing is a key paradigm that is quite commonly adopted for information delivery in wireless ad-hoc and sensor networks where the location information of the nodes is available (either
a-priori or through a self-configuring localization mechanism). Geographic routing protocols are efficient in wireless networks for several reasons. For one, nodes need to know only the location information of their direct neighbors in order to forward packets and hence the state stored is minimum. Further, such protocols conserve energy and bandwidth since discovery floods and state propagation are not required beyond a single hop.

The main component of geographic routing is usually a greedy forwarding mechanism whereby each node forwards a packet to the neighbor that is closest to the destination. This can be an efficient, low-overhead method of data delivery if it is reasonable to assume (i) sufficient network density, (ii) accurate localization and (iii) high link reliability independent of distance within the physical radio range.

However, while assuming highly dense sensor deployment and reasonably accurate localization may be acceptable in some classes of applications, it is clear that assumption (iii) concerning highly reliable links is unlikely to be valid in any realistic deployment. The existence of such unreliable links exposes a key weakness in greedy forwarding that we refer to as the weakest link problem. At each step in greedy forwarding, the neighbors that are closest to the destination (also likely to be farthest from the forwarding node) may have poor links with the current node. These “weak links” would result in a high rate of packet drops, resulting in drastic reduction of delivery rate or increased energy wastage if retransmissions are employed.

This observation brings to the fore the concept of neighbor classification based on link reliability. Some neighbors may be more favorable to choose than others, not only based on distance, but also based on loss characteristics. This suggests that a blacklisting/neighbor selection scheme may be needed to avoid ‘weak links’. But, what is the most energy-efficient forwarding strategy and how does such strategy draw the line between ‘weak’ and ‘good’ links?

We articulate the following energy trade-off between distance per hop and the overall hop count, which we simply refer to as the distance-hop energy trade-off for geographic forwarding. If the geographic forwarding scheme attempts to minimize the number of hops by maximizing the
geographic distance covered at each hop (as in greedy forwarding), it is likely to incur significant
energy expenditure due to retransmission on the unreliable long weak links. On the other hand,
if the forwarding mechanism attempts to maximize per-hop reliability by forwarding only to close
neighbors with good links, it may cover only a small geographic distance at each hop, which
would also result in greater energy expenditure due to the need for more transmission hops for
each packet to reach the destination. We will show in this chapter that the optimal forwarding
choice is generally to neighbors in the transitional region.

In this chapter, our goal is to study the energy and reliability trade-offs pertaining to geo-
graphic forwarding analytically under the realistic packet loss model presented in Chapter 3. We
emphasize, however, that the framework, fundamental results and conclusions of this chapter are
quite robust and not limited by the specific characteristics of this model.

The rest of the chapter is organized as follows. In section 4.2, we present the scope, assumptions
and metrics of our analysis. Then, we provide a mathematical analysis of the optimum distance
in the presence of unreliable links in section 4.3. A set of tunable geographic forwarding strategies
is presented in section 4.4, and in section 4.5, we evaluate the performance of these strategies.
The effectiveness of the metrics presented in our analysis is validated through experiments with
motes in section 4.6. Finally, we present a summary of the chapter in section 4.7.

4.2 Scope, Assumptions and Metrics

The analysis of geographic routing is based on the link layer model presented in Chapter 3 and
it uses parameters that resemble the radio in mica2 motes: non-coherent frequency shift keying
for the modulation technique and Manchester for the encoding (Table 3.6). We will also use the
expressions for the packet reception rate Ψ (equation 3.5), the distance denoting the end of the
transitional region $d_e$ (equation 3.8), the cumulative distribution function $F(Ψ)$ (equation 3.10)
and the expectation $E[Ψ]$ (equation 3.19).
Scope: Our work presents techniques to reduce the energy consumption of geographic routing during communication events (transmission and reception of packets). Nevertheless, we should offer some caveats regarding the scope of our work. Our models do not consider other means of energy savings such as sleep/awake cycles, transmission power control, nor other sources of energy consumption such as processing or sensing. We also do not address MAC-layer behavior such as contention/interference. The work in this chapter is aimed for low traffic scenarios where interference does not play a significant role. Scenarios with medium or high traffic would require a different analysis.

Assumptions: Our analysis is based on the following assumptions:

- Nodes know the location and the link quality (PRR) of their neighbors.
- Nodes know the location of the sink (final destination).
- Nodes are deployed in a chain.

Metrics: From the end-user perspective, an efficient sensor network should provide as much data as possible utilizing as little energy as possible. Hence, in order to evaluate the energy efficiency of different strategies we use the following metrics:

- Delivery Rate ($r$): percentage of packets sent by the source that reach the sink.
- Total Number of Transmissions ($t$): total number of packets sent by the network to attain delivery rate $r$.
- Energy Efficiency ($\xi$): number of packets delivered to the sink for each unit of energy spent by the network in communication events.

$\xi$ can be derived from the delivery rate $r$ and the total number of transmissions $t$. Let $p_{src}$ be the number of packets sent by the source, $e_{tx}$ and $e_{rx}$ the amount of energy required by a node to

---

1Li et al. present an interesting extension of our work in [47], which includes power control.
transmit and receive a packet\textsuperscript{2}. Therefore, the total amount of energy consumed by the network for each transmitted packet is constant and it is given by:

\[ e_{\text{total}} = e_{\text{tx}} + e_{\text{rx}} \]  \hspace{1cm} (4.1)

Hence, the total energy consumed due to communication events is \( e_{\text{total}} \times t \), and \( \xi \) is given by:

\[ \xi = \frac{P_{\text{src}} r}{e_{\text{total}} t} \]  \hspace{1cm} (4.2)

Table 4.1 presents the notation used in this Chapter.

4.3 Analytical Model

Given a realistic link layer model, akin to the one described in Chapter 3, our goal is to explore the distance-hop trade-off in order to maximize the energy efficiency of the network during communication events.

4.3.1 Problem Description

This sub-section describes the notation and set-up used in the analysis. We assume that nodes are placed every \( \tau \) meters in a chain topology\textsuperscript{3}. A node is considered to have neighbors up to a distance of \( 2[d_e] \), where \( d_e \) is the end of the transitional region (equation 3.8), the set of distances to the neighbors is given by \( \varphi = \{ \tau, 2\tau, 3\tau, ..., 2[d_e] \} \textsuperscript{4} \), and the distance between source and sink is denoted by \( d_{\text{src-sink}} \).

\textsuperscript{2}The listening cost of passive neighbors (early rejection) in random deployments can be easily inserted and the interested reader is refereed to our initial work [58].

\textsuperscript{3}A non-constant distance between nodes can be also chosen. However, a constant distance \( \tau \) allows a fair comparison of the different regions (connected, transitional, disconnected).

\textsuperscript{4}The selection of \( 2[d_e] \) as a “nominal range” does not affect the results of this work. Even though other distances can be considered, \( 2[d_e]\tau \) was selected because the expectation and variance of PRR (equations 3.19 and 3.20) it can be derived that nodes beyond this distance have a small probability of having active links.
<table>
<thead>
<tr>
<th>Description</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet Reception Rate Parameters</td>
<td></td>
</tr>
<tr>
<td>- packet reception rate (PRR) [Random Process]</td>
<td>$\Psi$</td>
</tr>
<tr>
<td>- packet reception rate for a distance $d$ [Random Variable]</td>
<td>$\Psi_d$</td>
</tr>
<tr>
<td>- cumulative distribution function of $\Psi_d$</td>
<td>$F(\Psi_d)$</td>
</tr>
<tr>
<td>- expected packet reception rate</td>
<td>$E[\Psi]$</td>
</tr>
<tr>
<td>- a specific PRR value in the range of $\Psi$</td>
<td>$\psi$</td>
</tr>
<tr>
<td>- blacklisting threshold</td>
<td>$\psi_{th}$</td>
</tr>
<tr>
<td>Signal to Noise Ratio Parameters</td>
<td></td>
</tr>
<tr>
<td>- signal to noise ratio (SNR)</td>
<td>$\Upsilon$</td>
</tr>
<tr>
<td>- a specific SNR value in the range of $\Upsilon$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>- SNR value corresponding to $\psi_{th}$</td>
<td>$\gamma_{th}$</td>
</tr>
<tr>
<td>Channel Parameters</td>
<td></td>
</tr>
<tr>
<td>- path loss exponent</td>
<td>$\eta$</td>
</tr>
<tr>
<td>- standard deviation</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>- output power</td>
<td>$P_t$</td>
</tr>
<tr>
<td>Transitional Region Parameters</td>
<td></td>
</tr>
<tr>
<td>- end of transitional region</td>
<td>$d_e$</td>
</tr>
<tr>
<td>Energy Efficiency Parameters</td>
<td></td>
</tr>
<tr>
<td>- end-to-end delivery rate</td>
<td>$r$</td>
</tr>
<tr>
<td>- end-to-end number of transmissions</td>
<td>$t$</td>
</tr>
<tr>
<td>- energy efficiency</td>
<td>$\xi$</td>
</tr>
<tr>
<td>- energy spent by network for one transmission</td>
<td>$c_{total}$</td>
</tr>
<tr>
<td>- optimal forwarding distance</td>
<td>$d_{opt}$</td>
</tr>
<tr>
<td>- distance between source and sink</td>
<td>$d_{src-snk}$</td>
</tr>
<tr>
<td>- number of packets transmitted by source</td>
<td>$P_{src}$</td>
</tr>
<tr>
<td>- number of hops</td>
<td>$h$</td>
</tr>
<tr>
<td>- set of distances to neighbors</td>
<td>$\phi$</td>
</tr>
<tr>
<td>- probability that distance $d$ has the highest energy efficiency</td>
<td>$q_d$</td>
</tr>
</tbody>
</table>

Table 4.1: Mathematical Notation
Let $\xi_d$ be the random variable that denotes the energy efficiency obtained if a distance $d$ is traversed at each hop, then, the optimal forwarding distance $d_{\text{opt}}$ is the one that maximizes the expected value of $\xi_d$:

$$d_{\text{opt}} = \arg \max_{d \in \varphi} E[\xi_d] \quad (4.3)$$

In the next subsections we derive optimal forwarding metrics for the ARQ and No-ARQ cases.

### 4.3.2 Analysis for ARQ case

We assume no a-priori constraint on the maximum number of retransmissions (i.e. $\infty$ retransmissions can be performed), therefore, $r$ is equal to 1, and according to equation 4.2 the energy efficiency is given by:

$$\xi_{\text{ARQ}} = \frac{p_{\text{src}}}{k t} \quad (4.4)$$

Letting $\Psi_d$ be the random variable representing the PRR for a transmitter-receiver distance $d$, the expected number of transmissions at each hop is $\frac{p_{\text{src}}}{\Psi_d}$. The number of hops $h$ is equal to $\frac{d_{\text{src}} - d_{\text{sink}}}{d}$, therefore, the total number of transmission $t$ is given by:

$$t = \frac{d_{\text{src}} - d_{\text{sink}}}{d} \frac{p_{\text{src}}}{\Psi_d} \quad (4.5)$$

Substituting $t$ in equation 4.4, we obtain the energy efficiency metric for a transmitter-receiver distance $d$:

$$\xi_{d_{\text{ARQ}}} = \frac{d \Psi_d}{c_{\text{total}}^{\Psi_d}} \quad (4.6)$$

$d$ is defined (constant) for $\Psi_d$, therefore, the expected value of $\xi_{d_{\text{ARQ}}}$ is given by:
Figure 4.1: Impact of channel multi-path on $E[\xi_{d_{ARQ}}]$, (a) impact of path loss exponent $\eta$, (b) impact of channel variance $\sigma$.

\[ E[\xi_{d_{ARQ}}] = \frac{dE[\Psi_d]}{e_{total}d_{src-sink}} \]  

(4.7)

c_{total}$ and $d_{src-sink}$ are constants and an approximate expression for $E[\Psi_d]$ was derived in equation 3.19. Hence, in order to maximize the energy efficiency of systems with ARQ we need to maximize $dE[\Psi_d]$ (PRR×distance product).

Next, we evaluate $E[\xi_{d_{ARQ}}]$ for all $d \in \varphi$ (equation 4.3). Unfortunately, the computation of $E[\Psi_d]$ involves the $Q$ function (tail-integral of the Gaussian distribution) for which no closed-form expressions are known. Hence, we evaluate equation 4.3 numerically.

Figures 4.3 (a) and (b) depict the impact of channel multi-path on $dE[\xi_{d_{ARQ}}]$. In these figures the following parameters were used as the basis of comparison $\tau=1$m, $\eta=3$, $\sigma=3$, $P_t=-10$ dBm and $f=100$; and the x-axis represent the transmitter-receiver distance $d$ normalized with respect to the end of the transitional region, which is approximately 20 meters for the parameters given above. The beginning and end of the transitional region are depicted by vertical lines.

Figure 4.3 (a) presents the impact of the path loss exponent $\eta$. We observe that for a higher $\eta$ the optimal forwarding distance shifts left. This is due to the fact that for a higher path loss
Figure 4.2: Energy efficiency metric for the ARQ case. The transitional region often has links with good performance as per this metric.

exponent the received signal strength decays faster, which in turn reduces the packet reception rate. Figure 4.3 (b) presents the impact of the channel variance $\sigma$. In this case the optimal forwarding distance shifts right, which is due to the fact that a higher $\sigma$ increases the probability of finding good links farther away from the sender but also decreases the probability of finding good links close to the sender. In both figures it is important to highlight that the optimal forwarding distance lies in the transitional region, showing the distance-hop energy trade-off.

In real environments the packet reception rate takes instance of the r.v. $\Psi_d$, hence, the optimal local forwarding metric for a node is the one that maximizes the product of the PRR of the link and the distance to the neighbor. Figure 4.2 shows simulations for the PRR$\times$d metric in a line topology, where for each neighbor, the PRR obtained was multiplied by its distance. It can be observed that nodes in the transitional region usually have the highest value for this metric.

4.3.3 Analysis for the No-ARQ case

In systems with ARQ, at each step a node transmits the same amount of data as the source ($r = 1$), this characteristic allowed us to do the analysis independently of $d_{\text{src-sink}}$. On the other hand, in systems without ARQ the amount of data decreases at each hop, hence in order to maintain an
acceptable delivery rate, the longer the \( d_{\text{src-sink}} \) the higher the PRR of the chosen links should be. The analysis in this section explains this behavior.

Letting \( i \in [1, 2, ..., \lceil h \rceil] \) be the hop counter, we denote \( \Psi^i_d \) as the r.v. representing the packet reception rate for the distance \( d \) traversed at each hop \( i \). \( \Psi^i_d \) are i.i.d \( \forall i \in [1, 2, ..., \lceil h \rceil] \). This notation allow us to define the delivery rate \( r \) for systems without ARQ traversing a distance \( d \) at each hop:

\[
r = p_{\text{src}} \prod_{i=1}^{\lceil h \rceil} \Psi^i_d
\]

(4.8)

The number of packet transmissions required at each hop \( i \) \((t^i)\) is given by:

\[
t^i = p_{\text{src}} \prod_{j=1}^{i-1} \Psi^j_d
\]

(4.9)

Where \( \Psi^0_d = 1 \), to accommodate for the number of transmissions required at the source (equal to \( p_{\text{src}} \)). The total number of transmissions \( t \) is the sum of \( t_i \), \( \forall i \in [1, 2, ..., \lceil h \rceil] \). Therefore, \( t \) is given by:

\[
t = p_{\text{src}} \sum_{i=1}^{\lceil h \rceil} \left( \prod_{j=1}^{i-1} \Psi^j_d \right)
\]

(4.10)

Then \( \xi_{\text{woARQ}} \) is given by:

\[
\xi_{\text{woARQ}} = \frac{\prod_{i=1}^{\lceil h \rceil} \Psi^i_d}{\sum_{i=1}^{\lceil h \rceil} \left( \prod_{j=1}^{i-1} \Psi^j_d \right)}
\]

(4.11)

In real scenarios each link will take an instance of the random variable, letting \( \psi \) be an instance of the PRR for a given link, the local calculation of the delivery rate would be \( r = p_{\text{src}} \psi^{\lceil h \rceil} \) and the number of transmissions would be sum given by:
\[ t = p_{src} \sum_{i=1}^{[h]} \psi^{(i-1)} = p_{src} \frac{(\psi)^h - 1}{\psi - 1} \quad (4.12) \]

Which leads to a metric of:

\[ \xi_{\text{dwoARQ}} = \frac{(\psi)^h(\psi-1)}{e_{\text{total}}(1-(\psi)^h)} = \frac{(\psi)^h(1-\psi)}{e_{\text{total}}(1-(\psi)^h)} \quad (4.13) \]

Given that the PRR of a link is in the interval \((0,1)\), \(\frac{(1-\psi)}{1-(\psi)^h} < 1\), and for large number of hops \((\psi)^h\) in the numerator decreases exponentially while \(1 - (\psi)^h\) in the denominator increases. Therefore, equation 4.13 shows that in systems without ARQ, specially for large number of hops, nodes should choose links with high PRRs. Otherwise for long distances the delivery rate, and hence the energy efficiency, will tend to zero.

### 4.4 Geographic Forwarding Strategies for Lossy Networks

In this section, we present some forwarding strategies that will be compared with the \(\text{PRR} \times d\) metric. The aim of these strategies is to avoid the weakest link problem, and they are classified into two categories: distance-based and reception-based. In distance-based policies nodes need to know only the distance to their neighbors, while in reception-based policies, in addition to the distance, nodes need to know also the link’s PRR of their neighbors. All the strategies use greedy-like forwarding, in that first a set of neighbors is blacklisted based on a certain criteria and then the packet is forwarded to the node closest to the destination among the remaining neighbors.
4.4.1 Distance-based Forwarding

**Original Greedy:** Original greedy is similar to the current forwarding policy used in common geographic routing protocols. Original greedy is a special case of the coming blacklisting policies, when no nodes are blacklisted.

**Distance-based Blacklisting:** In this case, each node blacklists neighbors that are above a certain distance from itself. In this chapter the “nominal” radio range is defined as $2d_e$. For example if the radio range is considered to be 40 m and the blacklisting threshold is 20%, then the farthest 20% of the radio range (8 m) is blacklisted and the packet is forwarded through the neighbor closest to the destination from those neighbors within 32 m.

4.4.2 Reception-based Forwarding

**Absolute Reception-based Blacklisting:** In absolute reception-based blacklisting, each node blacklists neighbors that have a reception rate below a certain threshold. For example, if the blacklisting threshold is 20%, then only neighbors closer to the destination with a reception rate above 20% are considered for forwarding the packet.

**Best Reception Neighbor:** Each node forwards to the neighbor that has the highest PRR and is closer to the destination. This strategy is ideal for systems without ARQ.

4.4.3 PRR $\times$ d

This is the metric shown in our analysis and it can be observed as a mixture of the distance ($d$) and reception (PRR) based. For each neighbor, that is closer to the destination, the product of the PRR and distance.
4.5 Comparison of Different Strategies

The analytical model derived in section 4.3 provides the optimal forwarding distance. Nevertheless, in order to accurately evaluate the distance-hop trade-off we need to quantify the amount of energy saved by choosing the best candidate according to the optimal metric with respect to other methods. In this section, we compare analytically the energy efficiency of the different strategies presented in the previous section for systems with ARQ in a chain topology.

In order to compare the different strategies we require their expected energy efficiency \( E[\xi] \). In general, a strategy \( S \) has an expected energy efficiency \( E[\xi_S] \) given by:

\[
E[\xi_S] = \sum_{d\in\varphi} E[\xi_S|d_f = d] \ p(d_f = d) = \sum_{d\in\varphi} E[\xi_S|d_f = d] \ q_d
\] (4.14)

Where \( \varphi \) is the set of distances to neighbors, \( d_f \) is the distance traveled at each hop, and \( q_d \) is the probability that \( \xi_d > \xi_\ell, \forall \ell \in \varphi, \ell \neq d \). In the remainder of this section we denote the random variable \( \xi|d_f = d \) as \( \xi_d \). The next subsections provide \( E[\xi] \) for different strategies.

4.5.1 PRR×d

For the \( PRR \times d \) metric \( q_d \) is given by:

\[
q_d = \int_0^\infty P((x < \xi_d < x + dx) \land (\xi_j < x, \forall j\in\varphi, j \neq d)) \ dx
\] (4.15)

The energy efficiency of different distances can be considered independent:\footnote{The link quality (PRR) is a function of the SNR which is the sum of many contributions, coming from different locations, with random phases [56].}

\[
q_d = \int_0^\infty P(x < \xi_d < x + dx) P(\xi_j < x, \forall j\in\varphi, j \neq d) \ dx
\] (4.16)
Finally, \( q_d \) given by:

\[
q_d = \int_{0}^{\infty} f_{\xi_d}(x) \prod_{\forall j \in \Phi, j \neq d} F_{\xi_j}(x) \, dx \tag{4.17}
\]

Where \( f_{\xi_d}(x) \) and \( F_{\xi_d}(x) \) are the pdf and cdf of the metric \( \xi_d \). Given that these expressions depend on the \( Q \) function we provide numerical solutions in Figure 4.3 for \( q_d \). This figure shows the impact of different parameters on \( q_d \).

Figures 4.3 (a) and (b) show that when \( \tau \) and \( \eta \) increase the probability \( q_d \) shifts left, closer to the connected region. On the other hand, when \( \sigma \) and \( P_t \) increase \( q_d \) shifts right, closer to the end of the transitional region. These behavior is explained by the change in the number of neighbors (node density with respect to the coverage range).

The higher the number of neighbors, the higher the probability of discovering neighbors with good links (high PRR) that are closer to the destination (long \( d \)), which increases \( q_d \). Keeping all
the parameters constants, a larger $\tau$ or a higher $\eta$ (faster signal decay) reduces the density. On the other hand a higher $P_t$ increases the coverage range, and higher $\sigma$ increases the probability of finding good links farther away from the sender. Hence, the higher the density (number of neighbors), the higher $q_d$.

The expected energy efficiency of the PRR×d metric for a distance $d$ is given by equation 4.7. Hence, according to equation 4.14 the expected energy efficiency for systems with ARQ using the PRR×d metric is given by:

$$E[\xi_{PRR \times d}] = \sum_{d \in \phi} \frac{dE[\Psi_d]}{e_{total}d_{src-sink}} \cdot q_d$$

(4.18)

### 4.5.2 Absolute Reception-Based

Let us define $\psi_{th}$ as the blacklisting threshold of absolute reception, where valid links have PRR values on the interval $[\psi_{th}, 1)$. In order to choose $d$ as the forwarding distance, links with distances longer than $d$ should have a $\text{PRR} < \psi_{th}$, and the link at distance $d$ should have a $\text{PRR} \geq \psi_{th}$.

Hence, $q_d$ for absolute reception-based (ARB) blacklisting is given by:

$$q_{d_{ARB}} = p(\Psi_d \geq \psi_{th}) \prod_{d_w \in \phi, d_w > d} p(\Psi_{d_w} < \psi_{th})$$

(4.19)

Given that a link is considered valid if $\Psi_d \geq \psi_{th}$, the expected number of transmissions at each hop is $\frac{p_{src}}{E[\Psi_d|\Psi_d > \psi_{th}]}$. Hence, the expected value of the energy efficiency conditioned on the fact that $\Psi_d > \psi_{th}$ is given by:

$$E[\xi_{d_{ARB}}] = \frac{d}{e_{total}d_{src-sink}} E[\Psi_d|\Psi_d > \psi_{th}]$$

(4.20)

Denoting $\gamma = \Psi^{-1}(\psi)$ and $\gamma_{th} = \Psi^{-1}(\psi_{th})$ the probability density function of the packet reception rate conditioned on $\Psi_d > \psi_{th}$ is $f(\psi|\Psi_d > \psi_{th})$, which can be mapped to SNR values as $f(\gamma|Y_d > \gamma_{th})$, then:
\[ E[\Psi_d|\Psi_d > \psi_{th}] = \int_{\psi_{th}}^{1} \psi f(\Psi|\Psi > \psi_{th})d\psi \]
\[ = \int_{-\gamma_{th}}^{+\infty} \Psi(\gamma) f(\gamma|\Upsilon > \gamma_{th})d\gamma \] (4.21)

Combining the previous two equations we obtain the expected energy efficiency for absolute reception base (ARB):

\[ E[\xi_{ARB}] = \sum_{d \in \varphi, d \leq d_{th}} \frac{dE[\Psi_d|\Psi_d > \psi_{th}]}{e_{total} d_{src-snk}} q_{d_{ABR}} \] (4.22)

### 4.5.3 Distance-Based

When the blacklisting is based on distance the energy efficiency of the forwarding distance \( d (\xi_d) \) is the same as equation 4.7. Denoting \( d_{th} \) as the distance blacklisting threshold, distance based blacklisting will select a distance \( d \) the neighbor at distance \( d \) has a PRR > 0 and the neighbors with distances longer than \( d \) have a PRR = 0. The probability \( q_d \) of distance based (DB) blacklisting is given by:

\[ q_{d_{DB}} = p(\Psi_d > 0) \prod_{d_{w} \epsilon \varphi, d_{w} < d_{th}} p(\Psi_{d_{w}} = 0) \] (4.23)

Finally, the expected energy efficiency is given by:

\[ E[\xi_{DB}] = \sum_{d \in \varphi, d \leq d_{th}} \frac{dE[\Psi_d]}{e_{total} d_{src-sink}} q_{d_{DB}} \] (4.24)

### 4.5.4 Comparison

Figures 4.4 and 4.5 show the comparison of energy efficiency in a chain topology for different channel, radio and deployment parameters, as in section 4.3. The figures show the relative performance of the different strategies with respect to the PRR \( \times d \) metric, i.e. the y axis show the how
Figure 4.4: Performance of PRR blacklisting

much extra energy is required to attain the same delivery rate as PRR×d. Similarly to section 4.3, the base model of comparison have parameters $\tau=1$, $\eta=3$, $\sigma=3$, $P_t=-10$ dBm and $f=100$. Original greedy is a specific case of distance-based blacklisting, when no distance is blacklisted; and best reception is a specific case of absolute reception-based when a high blacklisting threshold is selected.

Figure 4.5 confirms the significant energy expenditure of original greedy, but there are other important insights from these comparisons. First, $\tau$, $\eta$, $\sigma$ and $P_t$ have an important impact on the relative performance of the different metrics due to its influence in the number of neighbors (node density per coverage range) and the expected energy efficiency. An increase in $\tau$ or $\eta$, or a decrease in $P_t$ leads to a lower node density, which implies that the strategies will start to choose the same nodes, given the lack of options, and the energy efficiency will be more similar among them. When $\sigma$ is increased, it improves the performance of absolute reception-based and
Figure 4.5: Performance of distance blacklisting

decreases the one of distance-based. This is due to the fact that $\sigma$ increases the probability of both, encountering good links at farther distances and bad links at shorter distances. Second, blacklisting links with PRR below 10% improves significantly the performance of reception-based. This is due to the observation done in Chapter 3 (Figure 3.5) with respect to the $cdf$ of the PRR, where it was noted that most of the links are either “good” or “bad”, hence, by blacklisting links below 1% we eliminate most of the bad links. Third, reception-based strategies perform better than distance-based. This due to the fact that reception-base takes advantage of good quality links in the transitional region (farther away from the transmitter), on the other hand, distance-based blacklist potential good links, furthermore, the closer the distance does not necessarily imply better links, and distance-based is still vulnerable to select bad-quality links at medium distances. Fourth, it is important to consider that while some thresholds of distance and absolute reception based strategies show close performance to that of PRR$\times$dist, these values change according to
the channel, radio and deployment parameters requiring a pre-analysis of the scenario, on the other hand, \( \text{PRR} \times d \) is a local metric that does not require any \textit{a priori} configuration. Finally, the results show that Best Reception is also a good metric and it can be good candidate for systems without ARQ given that these systems require to select good quality links.

### 4.6 Experiments with Motes

In order to validate our methodology and conclusions, we undertook an experimental study on motes. Twenty-one (21) mica2 motes were deployed in a chain topology spaced every 60 cm (~2 feet). The source (node 0) and sink (node 20) were placed at opposite extremes of the chain. The power level was set to -20 dBm and the frame size was 50 bytes. Three different forwarding strategies were tested:

- **OG**: neighbor closer to the sink whose PRR > 0.
- **BR**: neighbor with highest PRR. In case two or more neighbors have the same PRR, the one closer to the sink is chosen.
- **PRR\( \times d \)**: neighbors are classified according to the PRR\( \times d \) metric.

<table>
<thead>
<tr>
<th>( r ) (%)</th>
<th>sce1</th>
<th>sce2</th>
<th>sce3</th>
<th>sce4</th>
<th>sce5</th>
<th>sce6</th>
</tr>
</thead>
<tbody>
<tr>
<td>OG</td>
<td>0</td>
<td>2</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>94</td>
</tr>
<tr>
<td>BR</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>PRR( \times d )</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>82</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( t )</th>
<th>sce1</th>
<th>sce2</th>
<th>sce3</th>
<th>sce4</th>
<th>sce5</th>
<th>sce6</th>
</tr>
</thead>
<tbody>
<tr>
<td>OG</td>
<td>70</td>
<td>110</td>
<td>312</td>
<td>78</td>
<td>121</td>
<td>858</td>
</tr>
<tr>
<td>BR</td>
<td>652</td>
<td>407</td>
<td>730</td>
<td>754</td>
<td>701</td>
<td>903</td>
</tr>
<tr>
<td>PRR( \times d )</td>
<td>563</td>
<td>425</td>
<td>632</td>
<td>547</td>
<td>560</td>
<td>883</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relative ( \xi ) (%)</th>
<th>sce1</th>
<th>sce2</th>
<th>sce3</th>
<th>sce4</th>
<th>sce5</th>
<th>sce6</th>
</tr>
</thead>
<tbody>
<tr>
<td>OG</td>
<td>inf</td>
<td>inf</td>
<td>+54</td>
<td>inf</td>
<td>inf</td>
<td>+3</td>
</tr>
<tr>
<td>BR</td>
<td>+16</td>
<td>-4</td>
<td>+16</td>
<td>+14</td>
<td>+25</td>
<td>+2</td>
</tr>
<tr>
<td>PRR( \times d )</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 4.2: Empirical Results for Different Forwarding Strategies.
First, the motes exchange test packets to measure the PRR of the links and populate their routing tables accordingly. Afterwards, the source sends 50 packets to the sink for each of the 3 different strategies (150 total). A maximum of 5 transmissions (1 transmission + 4 retransmissions) are allowed at each hop, if the packet is not received after the fifth attempt, it is dropped.

Six different scenarios are studied: a football field, an indoor-building environment and four different outdoor-urban areas. The channel characteristics of some scenarios are significantly different, and hence, instead of providing a cumulative result, we present the results for each one of them.

Table 4.2 shows the delivery rate \( r \), the number of transmission \( t \), and the energy efficiency \( \xi \) for the different scenarios. BR have an \( r \) of 100% in all scenarios, and PRR\( \times d \) have 100% for all scenarios except scenario-4 (82%). Greedy performs poorly in most of them with zero or close to zero \( r \) in most cases.

With regard to the number of transmissions, BR requires more transmissions than PRR\( \times d \) in all scenarios except scenario 2 where BR performs better. Given that \( r \) is similar for BR and PRR\( \times d \), the difference in the energy efficiency is determined by the number of transmissions. BR consumes between 2 to 25% more energy than PRR\( \times d \), only in scenario-2 performed 4% worse. On the other hand, in the two scenarios where scenarios where Greedy has a non-zero \( r \), it consumed 3% and 54% more energy than PRR\( \times d \). It is interesting to observe that the energy “wasted” by greedy forwarding depends on where the first weak link is encountered. In some scenarios the first weak link is at the beginning of the chain and hence the energy wasted is not significant, however, in other scenarios the weak link is present at the middle or end of the chain which caused a greater energy waste.

Although this experimental study is limited in size, it provides two important conclusions. First, it does serve to confirm and validate our earlier findings from the analytical and simulation studies regarding the PRR\( \times d \) metric. And second, it shows that the best reception metric is also a good metric for real deployments. Based on the insights of section 4.5, we believe that higher
densities will lead to bigger savings in terms of energy for the PRR\times d metric with respect to other strategies.

4.7 Summary

We have presented a detailed study of geographic routing in the context of the model presented in chapter 3. We have provided a mathematical analysis of the optimal forwarding metric for both ARQ and No-ARQ scenarios. We have also validated some of our approaches using real experiments on motes.

Key results from our study indicate that the common greedy forwarding approach would result in very poor packet delivery rate. Efficient geographic forwarding strategies do take advantage of links in the high variance transitional region for energy-efficiency. An important forwarding metric that arose from our analysis, simulations and experiments is PRR\times d, particularly in high-density networks where ARQ is employed. Our results also show that reception-based forwarding strategies are generally more efficient than distance-based strategies.

Finally, it is important to highlight that the PRR\times d metric is recommended in static or low dynamic environments, such as environmental monitoring. In highly dynamic environments the link quality can change drastically throughout time and hence stable estimates of PRR might not be possible. In this scenarios, best reception is probably the best forwarding strategy.
Chapter 5

Performance of Random Walks on Heterogeneous Networks

In this chapter we analyze the performance of random walk-based queries in heterogeneous sensor networks. A random walk on a graph is a discrete-time stochastic processes that starts from any vertex and at each step selects an adjacent vertex uniformly at random [49]. Due to its simplicity random walks have attracted considerable attention as querying mechanisms for WSN [64, 65, 5, 10, 66]. However, most of these studies have been focused on simple deterministic graphs such as regular grids that do not consider degree heterogeneity. As observed in the model presented in chapter 3 degree heterogeneity is an important property of real-life networks.

In this chapter we show that random walk-based queries can considerably enhance their performance by exploiting degree heterogeneity. We demonstrate that by using a simple algorithm and including a few high-degree nodes (10% <), query cost can be reduced between 30% and 70%.

5.1 Overview

An important approach for querying in unstructured systems is the use of random walks. This approach is gaining popularity in the networking community since random walks are intuitively simple — nodes are visited sequentially in a random order with successive nodes being neighbors in the graph [65, 5, 64, 10]. There is also a significant body of theoretical literature on random walks as querying mechanisms [3, 7, 11]. However, while this literature provides much insight
into the scaling behavior of random walks on simple classes of deterministic graphs (such as 2D Torii), a major property of real-life networks, heterogeneity, was left out of discussion.

Heterogeneity on the degree distribution is a highly likely characteristic in real large-scale wireless systems, such as sensor networks, for a number of reasons. First, as shown in the model proposed in this thesis, in real scenarios, channel multi-path and hardware variance lead to significant changes in degree distribution compared to ideal scenarios. Second, random deployments are inherently non-regular graphs. Third, empirical studies [?, 74] have revealed that hardware variance on the sensitivity and output power of radios lead to nodes with significantly higher degree than the average (cluster-heads). Fourth, some works [32] have recently proposed that for wireless sensor networks to scale, heterogenous networks consisting of highly capable and low capable devices are required.

In this work we propose the use of a push-pull mechanism to exploit the heterogeneity of the underlying communication graph to enhance the performance of random-walk-based queries. We take advantage of a well known property of the simple random walks: its stationary distribution \( \pi(v) = d(v)/2m \), where \( d(v) \) denotes the degree of node \( v \) and \( m \) the number of edges in the graph. This means that nodes with higher degree are visited more frequently by the random walk.

The idea we use is intuitively simple, events are pushed towards high-degree nodes (cluster-heads) and pulled from the cluster-heads by performing a random-walk-based query. In this scenario some important questions arise: What is the impact of heterogeneity on performance? and How much heterogeneity is needed?

We use two theoretical tools to explore these questions. Our analytical results are based on the direct connection between a random walk on a graph and the resistance of the electrical network obtained from the graph by viewing each edge as a unit resistor [24, 15]. We bound the influence of cluster-heads on the resistance which in turn bound on the query cost. Our second tool is the use of absorption states in the transitional probability matrix of a graph \( G \) to obtain the expected number of steps in a query.
The main contribution of this work is to show that heterogeneity allows random-walk-based queries to enhance their performance. In particular, the striking result we obtain is that for a line topology where cluster-heads have a coverage $k$ (cover $k$ nodes to the right and $k$ nodes to the left) and are uniformly distributed (evenly spaced), a fraction of $\frac{4}{5k}$ nodes being cluster-heads can offer a reduction in query cost of $O(1 - \frac{1}{k^2})$ by using a simple distributed algorithm. In intuitive terms this translates to requiring less than 10% of the nodes being cluster-heads to obtain two orders of magnitude improvement in query cost. Through numerical analysis, we present results showing that in 2D networks also a small percentage of nodes being cluster-heads ($\sim 10\%$) can lead to significant improvements in performance (between 30% and 70% depending on the coverage of the high-degree nodes).

5.2 Enhancing Random Walks for Heterogeneity

In this section we present the event-push query-pull algorithm used in our work. We consider a network where two type of nodes are available: i) nodes with limited communication capability (low degree) and ii) nodes with higher communication capabilities (high degree, cluster-heads). Our focus is on infrastructure-less networks (no location nor GPS capabilities), nodes are able to communicate only with their neighbors, and they are aware of their neighbors degree (if neighbors are cluster-heads or not).

Our query mechanism is built from two parts: first an event $e$ is generated randomly at any node in the network, and second, a random-walk-based query is issued in order to find the event. The event $e$ can either remain at the node where it was generated or move to a cluster-head (if one exists). In a network where all nodes have the same degree, the event remains at the node where it appears. When cluster-heads are present, upon detection of an event, the event follows Algorithm 1:
**Algorithm 1** Event forwarding in Heterogeneous Network

**Require:** event $e$ at node $v_i$

1: while node $v_i$ is not a cluster-head do
2: if there is a neighbor $v_j$ with degree($v_j$) > degree($v_i$) then
3: forward $e$ to the neighbor with the highest degree;
   break ties uniformly at random among candidates
4: else
5: forward $e$ to a random neighbor
6: end if
7: end while

In order to find the event, a query is issued through a predefined sink node $s$. The query follows a *simple random walk* where the next node is chosen uniformly from the neighbors, until it finds the event $e$.

In the scenario described above there will be two costs: i) the cost of moving the event to a cluster-head, $C_{\text{event}}$ and the cost of the query (i.e random walk) to find the event, $C_{\text{query}}$. The total cost is the sum of both: $C_{\text{total}} = C_{\text{event}} + C_{\text{query}}$.

Denoting $L$ as the number of low-degree nodes and $H$ as the number of high-degree nodes, how does $C_{\text{total}}$ vary with the ratio $\frac{H}{H+L}$? Is there a value of $\frac{H}{H+L}$ that will reduce $C_{\text{total}}$ significantly?

### 5.3 Analytical Results

In this section we focus on line topologies and we are interested in analyzing the impact of cluster-heads on cost of event discover ($C_{\text{total}}$). The main result is as follows:

**Theorem 1.** Consider a line topology with $(n+1)$ nodes, where cluster-heads have a degree $2k$ and are uniformly distributed. The first local minima for the maximum hitting time and for the query cost is obtained when the fraction of high-degree nodes is $\frac{4}{2k}$. And for this fraction a reduction in query cost of $\Theta(1 - \frac{1}{k})$ is obtained.

As mentioned earlier, this result is obtained using bounds on the resistance of an electrical circuit related to the network graph. The remainder of the section is dedicated to the proof of
this theorem, and Table 5.1 presents the notation used for the proof, which explained in detail in the next two subsections.

### 5.3.1 Parameters of Line Topology

We consider the undirected graph $L(n, k, d) = G(V, E)$ with the following parameters. An element of $V$ or $E$ is represented by a lowercase, a subset or array is represented by a bold lowercase and the complement of a set or element will be denoted with an upper-bar, for example $e$ represents an element, $e$ an array (subset) and $\bar{e}$ and $\bar{e}$ represent the complements of $e$ and $e$, respectively.

For a source $s \in V$ and the subset of cluster-head nodes $e \subseteq V$, the average hitting time from $s$ to $e$ is:

$$h_{se} = \frac{\sum_{e \in e} h_{se}}{|e|} \quad (5.1)$$

The commute time $C_{uv}$ is the expected time taken by a random walk starting at $u$ to reach $v$ and come back to $u$, i.e. $C_{uv} = h_{uv} + h_{vu}$. $C_{uv}$ is given by:

$$C_{uv} = 2mR_{uv} \quad (5.2)$$

Where $m$ is the number of edges in the graph and $R_{uv}$ is the effective resistance between nodes $u$ and $v$. In case of symmetry $h_{uv} = h_{vu}$, which implies that the commute time is two times the hitting time. We will use this property in the analysis of the hitting time for line topologies.

<table>
<thead>
<tr>
<th>$(n + 1)$</th>
<th>number of nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>cluster coverage</td>
</tr>
<tr>
<td>$d$</td>
<td>inter cluster-heads distance</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>overlapping of cluster-heads coverage (region 2)</td>
</tr>
<tr>
<td>$(s + 1)$</td>
<td>number of clusters</td>
</tr>
</tbody>
</table>

Table 5.1: Mathematical Notation
In the line topology $L(n, k, d) = G(V, E)$, the set of nodes is $V = v_0, v_1, \ldots v_n$, since the index goes from 0 to $n$, the number of nodes is $|V| = n + 1$, we will also denote $n$ as the number of edges when no cluster-head has been added (when nodes communicate only with their immediate right and left neighbors). In addition to the initial $n$ edges, each cluster-head has edges to all nodes which are less than or equal to $k$ nodes away from it on the line. $d$ is the inter cluster-head distance, hence, a node $v_i$ is a cluster-head iff $i \mod d = 0$. The set of edges $E$ for $L(n, k, d)$ is defined as follows:

$$E = \{v_iv_j \mid (i \mod d = 0 \text{ and } |i - j| \leq k) \text{ or } |i - j| = 1\}$$

We will consider the case where $n \gg k > 1$. Since we are interested in the limit $n \to \infty$ we will consider only the cases where $n = sd$, where $s$ is an integer. Then, the total number of clusters is given by $(s + 1)$.

For a given $n$ and $k$ we are interested in studying the impact of $d$ on $C_{\text{total}}$ via the resistance, and the analysis is divided in different regions according to the inter cluster-heads distance $d$:

$$\text{regions} = \begin{cases} 
\text{region 1}, & 2k \leq d \leq n/2 \\
\text{region 2}, & k < d < 2k \\
\text{region 3}, & 1 \leq d \leq k 
\end{cases} \quad (5.3)$$

Figure 5.1 presents examples of line topologies for the 3 different regions. The big circles denote cluster-heads and the dashed lines their coverage $k$. Notice that $k$ is constant for all three topologies.

Later in this section we will show that $C_{\text{total}}$ depends mainly on $C_{\text{query}}$, specially for $d < 2(k+1)$ where $C_{\text{event}}$ will be shown to be less than 1. For this reason, the resistance analysis will be focused on $C_{\text{query}}$. 
Figure 5.1: Examples of line topologies. The big circles denote cluster-heads and the dashed lines their coverage $k$, all nodes within $k$ hops from the cluster-head are its neighbors.

For the different regions we will first obtain expressions for the number of edges $m$ and the effective resistance $R$ between the nodes at the extremes of the line. Then, in subsections 5.3.2 and 5.3.3 we use these expressions, together with symmetry, to obtain the maximum hitting time and average query cost, $C_{\text{query}}$.

5.3.1.1 Analysis of region 1

Recalling that the number of nodes in $\mathcal{L}$ is $(n + 1)$, and that when no clusters are present the initial number of edges is $n$. Besides $n$, each cluster-head contributes with $2(k - 1)$ new edges. Then the total number of edges is given by:
\[ m_1 = n + 2(k - 1)s \]
\[ = n + 2(k - 1)n/d \]
\[ = \frac{n}{d}(d + 2(k - 1)) \]  

(5.4)

The coverage on each side of a cluster-head can be represented by an effective resistance \( r(k) \) (Figure 5.1). \( r(k) \) can be derived based on \( k \) using techniques to reduce resistors in parallel and series:

\[ k = 2 \Rightarrow r(2) = \frac{2}{3} \]
\[ k = 3 \Rightarrow r(3) = \frac{5}{8} \]
\[ k = 4 \Rightarrow r(4) = \frac{13}{21} \]

It can be derived that the numerator and denominator follows the fibonacci series \( \text{fib}(.), \) hence:

\[ r(k) = \frac{\text{fib}(2k - 1)}{\text{fib}(2k)} \]  

(5.5)

Since every cluster-head covers its right and left \( k \)-neighbors and the extreme vertices cover only one side, the effective resistance \( R_1 \) between the extremes of \( \mathcal{L} \) is given by:

\[ R_1 = (2r(k) + d - 2k)s \]
\[ = (2r(k) + d - 2k)n/d \]  

(5.6)

Finally, using equation 5.2 and symmetry, the hitting time for nodes at the extreme of the line is given by:

\[ h_1 = (\frac{n}{d})^2 (d + 2(k - 1)) (d + 2(r(k) - k)) \]  

(5.7)
5.3.1.2 Analysis of region 2

The number of edges is the same as equation 5.4 since every cluster-head that is added brings \(2(k-1)\) new edges. However, the effective resistance between 2 cluster-heads is different. In this subsection we provide upper and lower bounds for this resistance.

Letting \(\alpha_k\) be the overlap between the coverage of two neighboring cluster-heads, where \(0 < \alpha < 1\), then \(d = k(2-\alpha)\). The resistance circuit is equivalent to the one shown in Figure 5.2, where \(r(x) = r(k(1-\alpha))\) (the effective resistance for the non-overlapping part). Given that \(r()\) converges to the inverse of the golden ratio, \(1/2 < r(x) < 1\), further, considering Rayleigh’s monotonicity law we can provide upper and lower bounds for the resistance as shown in Figure 5.2:

\[
\frac{2}{\alpha k+2} < r_2 < \frac{2}{\alpha k+1} \\
\frac{2}{\alpha k+2} s < R_2 < \frac{2}{\alpha k+1} s \\
\frac{2}{\alpha k+2} \frac{n}{d} < R_2 < \frac{2}{\alpha k+1} \frac{n}{d}
\]  

(5.8)

Finally, the hitting time is bounded by:
\[ 2 \left( \frac{n}{d} \right)^2 \left( \frac{4k - k\alpha - 2}{\alpha k + 2} \right) < h_2 \]
\[ < 2 \left( \frac{n}{d} \right)^2 \left( \frac{4k - k\alpha - 2}{\alpha k + 2} \right) \]

**(5.9)**

### 5.3.1.3 Analysis of region 3

For this region, we will analyze only the point where \( d = k \).

Contrary to the case of regions 1 and 2, neighboring cluster-heads have an edge connecting each other, hence each cluster-head brings \( (2(k - 1) - 1) \) new edges. And the total number of edges is given by:

\[
m_3 = n + (2k - 3)s
\]
\[
= n + (2k - 3)n/d
\]
\[
= 3 \frac{n}{k} (k - 1)
\]

**Equation (5.10)**

The resistance between two cluster-heads can be transformed to an equivalent circuit as shown in Figure 5.2, which leads to an equivalent resistance of \( \frac{2}{k+1} \) between two neighboring cluster-heads. Hence, the effective resistance between the extremes of the line is given by:

\[
R_3 = \frac{2}{k+1} g
\]
\[
= \frac{2}{k+1} \frac{n}{k}
\]

**Equation (5.11)**

Finally, the hitting time for \( d = k \) is given by:

\[
h_3(d = k) = 6 \left( \frac{n}{k} \right)^2 \frac{k-1}{k+1}
\]

**Equation (5.12)**
Table 5.2: Minimum Maximum and Minimum Average Hitting Times per Region

<table>
<thead>
<tr>
<th>Region</th>
<th>Maximum Hitting Time ( (h) ) [subsection 5.3.2]</th>
<th>Average Hitting Time ( (C_{\text{query}}) ) [subsection 5.3.3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1 (lower bound)</td>
<td>( \left( \frac{n}{k} \right)^2 \left(k - \frac{1}{2} \right) )</td>
<td>( \frac{(2k-1)n(n+k)}{k^2} )</td>
</tr>
<tr>
<td>Region 2 (upper bound)</td>
<td>( 2 \left( \frac{2n}{5k} \right)^2 \left( \frac{13k-8}{3k+4} \right) )</td>
<td>( \frac{4n(13k-8)(5n+5k)}{3(3k+4)(5k)^2} )</td>
</tr>
<tr>
<td>Region 3</td>
<td>( 6 \left( \frac{n}{k} \right)^2 \left( \frac{k}{k+1} \right) )</td>
<td>( \frac{(k-1)(2d+k)n}{(k+1)k^2} )</td>
</tr>
</tbody>
</table>

### 5.3.2 Local Minimum for Maximum Hitting Time

In this section we analyze the hitting time between the sink \( (v_0) \) and the last cluster-head on the line \( (v_n) \). Table 5.2 shows the minimum value of \( h_1 \), \( h_2 \) and \( h_3 \). In region 3, we analyzed only one point \((d = k)\), hence, for region 3 the value in Table 5.2 is the one obtained in equation 5.12.

For region 1, the minimum value is obtained for \( d = 2k \), we further assume a lower bound on \( h_1 \) by setting \( r(k) = 0.5 \).

In the case of region 2, the hitting time depends on the overlapping \( \alpha \). Even though \( \alpha k \) should take only integer values, we assume \( \alpha \) to be real in order to differentiate \( h_2 \) (equation 5.9) and obtain the value of \( \alpha \) that minimizes \( h_2 \). The values of \( \alpha \) for the upper and lower bounds are given by \((\alpha_U \text{ and } \alpha_L)\):

\[
\alpha_U = \frac{12nk-7n+4k-16k^2}{2(2n-4k+1)k} - \sqrt{8nk^2 - 88n^2k + 96nk^2 - 128nk^2 + 17n^2 + 48nk - 16n} \tag{5.13}
\]

\[
\alpha_L = \frac{-8k^2 + 6nk - 4n - 2\sqrt{-8k^3n + 5n^2k^2 - 4n^2k + 8nk}}{2k(n-2k)} \tag{5.14}
\]

For large \( n \) and \( k \), \( \alpha_U = \alpha_L = 0.7639 \), hence \( \alpha_{\text{opt}} \approx \frac{3}{4} \), Table 5.2 shows the upper bound of \( h_2 \) for \( \alpha_{\text{opt}} \).

From Table 5.2 we observe that \( h_1 = \Theta(n^2/k) \), and \( h_2 \) and \( h_3 \) are \( \Theta(n^2/k^2) \), hence for large \( k \) the minimum is either on \( h_2 \) or \( h_3 \). Since \( n \) and \( k \) are given, we can compare directly upper bound of \( h_2 \) and \( h_3 \) given in Table 5.2, which leads to:
\[ \lim_{k \to \infty} \frac{h_3}{h_2} = 1.0817 \quad (5.15) \]

Hence, the first local minima is in region 2 and the inter cluster-head distance that attains this minimum is \( d = \frac{2}{3}k \), which corresponds to a ratio \( r \) between high-degree nodes and the total number of nodes:

\[ r = \left( \frac{4n}{5k} \right) \left( \frac{1}{n+1} \right) \approx \frac{4}{5k} \quad (5.16) \]

It is important to notice that the global minima might be in region 3, however, the reduction obtained by moving from the first local minima to the global minima is not significant, as it will be shown numerically in the next section.

### 5.3.3 Local Minimum for Expected Hitting Time

For regions 1, 2 and region 3 (when \( d = k \)), cluster-head \( i \) is visited only after cluster-head \( i - 1 \) has been visited. This property allows us to discard all nodes beyond a given cluster-head \( i \) when we are interested in obtaining the hitting time to \( i \). Hence, for cluster-head \( i \) we can consider a new line topology, which is a shorter version of the original one, where the sink is at one extreme and cluster-head \( i \) is at the end of the new line. This behavior allows us to use directly the expressions derived in the previous subsections with the only change to be made in the initial number of edges \( n \).

Recalling that the distance set between the sink and cluster-heads is given by:

\[ \{0, d, 2d, 3d, \ldots (s-1)d, sd\} \]

\( C_{query} \) is the average hitting time (\( E[h_i] \)) to all clusters and for region 1 is given by:
\[
E[h_1] = \sum_{i=0}^{s} h_1(n = id) \frac{\sum_{i=1}^{s} \left(\frac{R_i}{d}\right)^2 (d + 2(k - 1))(d + 2(r(k) - k))}{s+1} \]

\[
= \frac{(d+2(k-1))(2r(k)+d-2k)\sum_{i=1}^{s} i^2}{s+1}
\]

\[
= (d + 2(k - 1))(2r(k) + d - 2k) \frac{s(s+1)}{6}
\]

(5.17)

For Region 2, using the upper bound of equation 5.9 and recalling that \(d = k(2 - \alpha)\), the expected hitting time is given by:

\[
E[h_2] = \sum_{i=0}^{s} h_2(n = id) \frac{\sum_{i=1}^{s} 2\left(\frac{n}{d}\right)^2 \left(\frac{4k - k\alpha - 2}{\alpha k + 1}\right)}{s+1}
\]

\[
= \frac{2(4k - k\alpha - 2)}{(\alpha k + 1)(s+1)} \sum_{i=1}^{s} i^2
\]

\[
= \frac{n(2n+1)(4k - k\alpha - 2)}{3(\alpha k + 1)(k(2 - \alpha))^2}
\]

(5.18)

The derivative of this equation with respect to \(\alpha\) leads to an expression that is considerably more complicated than equation 5.13 and it is not presented. However, the results are the same as the obtained for the maximum hitting time, i.e. \(\alpha_U = \alpha_L = 0.7639\) and \(\alpha_{opt} \approx 0.7\). The closed-form expressions and values can be easily obtained by using mathematical software such as Matlab or Mathematica.

For Region 3, when \(d = k\), the expected hitting time is given by:

\[
E[h_3(d = k)] = \sum_{i=0}^{s} \frac{6\left(\frac{n}{k}\right)^2 k - 1}{k + 1} \frac{1}{s+1}
\]

\[
= \frac{6(k-1)}{(s+1)(k+1)} \sum_{i=0}^{s} i^2
\]

\[
= \frac{k-1}{k+1}s(2s + 1)
\]

(5.19)
Table 5.2 also presents the minimum value of $C_{\text{query}}$ for the 3 regions: $d = 2k$, $r(k) = 0.5$ for region 1 (lower bound), $d = \frac{5}{4}k$ for region 2 (upper bound) and $d = k$ for region 3 (only point analyzed). The comparisons lead to the same result as the ones obtained for the maximum hitting time: the order of $h_1$ is greater than the order of $h_2$ and $h_3$, and as $n$ and $k$ goes to infinity the ratio of $h_3$ over $h_2$ goes to 1.0817.

Now we have all the elements to prove Theorem 1

**Proof of Theorem 1:**

The optimization of the maximum and average hitting times (equations 5.9 and 5.18) leads to $d = \frac{5}{4}k$, using Table 5.2 we proved that the first local minima occurs in region 2 for both the maximum and average hitting times, and that their cost are $\Theta(\frac{n^2}{k^2})$. It is known that random walks on regular line topologies have a cost of $\Theta(n^2)$ [3], hence, a line topology $L(n, k, d)$ with $d = \frac{5}{4}k$ leads to a cost reduction of $\Theta(1 - \frac{1}{k^2})$ for both the maximum and average hitting times. □

### 5.4 Numerical and Experimental Results

In this section we use Markov numerical methods and simulations to study the impact of heterogeneity on the performance of random walk-based queries. First, we present numerical results on a line topology that validates the analytical contributions of Section 5.3. Then, we present numerical results on regular grids and random geometric graphs. Finally, we present simulation results on realistic graphs for wireless sensor networks.

In Chapter ?? we provide an introduction to obtain hitting times in Markov Chains. Let us consider that the event is in node $e \in V$ and the sink is in node $s \in V$, then, the array $Q_e$ representing the expected hitting time from each node to $e$ (except $e$) is given by:

$$Q_e = (I - M_e)^{-1} \mathbf{1}$$  \hspace{1cm} (5.20)
Where $I$ is the identity matrix, $1$ is a column vector of ones and $M_e$ is the matrix resulting from deleting the row and column corresponding to $e$ in $M$, where $M$ is the transition probability matrix defining the Markov chain in $G$. Hence, in our case $h_{se} = Q_e(s)$, where $Q_e(s)$ represents the $s$’th element of the array.

5.4.1 Line Topologies

For a given line topology $L(n, k, d)$ with transition probability matrix $M$, where the sink is the extreme left vertices $v_0$. Clusters-heads are positioned at $v_{id}$ where $i=0, 1, ..., s$. The hitting time from $v_0$ to a specific cluster-head is $Q_{id}(v_0)$, where $Q_{id}$ is given by:

$$Q_{id} = (I - M_{id})^{-1}1$$

(5.21)

Based on equation 5.1, $C_{query}$ is the average hitting time over all cluster-heads and is given by:

$$C_{query} = \frac{d}{2n}(Q_0(v_0) + Q_{sd}(v_0)) + \frac{d}{n} \sum_{i=1}^{s-1} Q_{id}(v_0)$$

$$= \frac{d}{2n}(Q_{sd}(v_0)) + \frac{d}{n} \sum_{i=1}^{s-1} Q_{id}(v_0)$$

(5.22)

In the previous equation all cluster-heads have the same weight except for the extreme ones ($v_0$ and $v_{sd}$), this is due to the fact that at the extremes, the expected number of stored events is half of those stored at the intermediate cluster-heads. However, for large number of cluster-heads the weight can be considered similar and:

$$C_{query} \approx \frac{\sum_{i=1}^{s} Q_{id}(v_0)}{s+1}$$

(5.23)

$C_{event}$ will be derived according to the regions defined in 5.3 and algorithm 1. There are $s + 1$ cluster-heads for which there is no need to move the event, hence the cost is zero.
When \( d < 2(k + 1) \), all nodes will be directly connected to a cluster-head and \( C_{\text{event}} = \frac{n-s}{n+1} \).

However, when \( d \geq 2(k + 1) \) (most of region 1), there will be orphan nodes between any pair of consecutive cluster-heads. Due to symmetry, the cost will be the same for any subset of nodes between any two neighboring cluster-heads and for simplicity we consider cluster-heads \( v_0 \) and \( v_d \).

For these clusters, nodes between \((k + 1)\) and \((d - k - 1)\) are orphan. Let us define \( a = (k + 1) \), \( b = (d - k - 1) \) and \( M_{(a:b)} \) as the \( M \)'s sub-matrix which includes only the rows and columns between \( a \) and \( b \). For \( M_{a:b} \), \( Q_{a:b} = (I - M_{a:b})^{-1} 1 \) is the array containing the expected number of steps of each orphan node to reach the closest node directly connected to a cluster-head. Hence, the cost of moving orphan nodes between \( a \) and \( b \) to a cluster-head is given by:

\[
C_{\text{orphan}} = \sum_{i \in Q} (Q_{a:b}(i) + 1) 
\]

(5.24)

Where the constant 1 represents the cost of moving the event from a node connected to a cluster-head to the cluster-head.

Finally, recalling that \((n + 1)\) is the total number of nodes in \( L \), \( C_{\text{event}} \) is given by:

\[
C_{\text{event}} = \begin{cases} 
\frac{2k(s+1)+sC_{\text{orphan}}}{n+1}, & 2(k + 1) \leq d \leq n/2 \\
\frac{n-s}{n+1}, & 2k \leq d < 2(k + 1) \\
\frac{n-s}{n+1}, & \text{region 2} \\
\frac{n-s}{n+1}, & \text{region 3} 
\end{cases} 
\]

(5.25)

Figure 5.3 shows \( C_{\text{query}}, C_{\text{event}} \) and \( C_{\text{total}} \) vs the number of clusters for a line topology with 121 nodes and different values of \( k \). The solid lines represent the cost obtained using Markov numerical analysis (equations 5.23 and 5.25), and the dotted lines represent simulation results, it can be observed that the Markov method provides an accurate representation of the cost. Also it must be noted that the query cost accounts for most of the total cost, which validates the focus of the analytical section on \( C_{\text{query}} \).
Figures 5.4 and 5.5 compare the maximum and expected hitting time between the Markov analysis (full lines) and the expressions obtained through the resistance method (dotted lines) for a line topology with 121 nodes and values of $k$ ranging from 5 to 10. The dotted lines show that the bounds get tighter for higher values of $k$ in both figures. The figures also shows a line with circle markers depicting the number of clusters required to reach the first local minima according to our analysis (Theorem 1). It is important to notice that the analytical values for the number of clusters are not necessarily integers and hence they may not match exactly the numerical ones, specially for low values of $k$. However, for large values of $k$ and $n$ the floor or ceiling of the
Figure 5.5: $C_{\text{query}}$ (expected hitting time) for a line topology with 121 nodes and values $k$ ranging from 5 to 10.

analytical value will not incur in significant differences. It is also important to notice that the first cluster-heads added leads to a significant cost reduction (region 1 and part of region 2), adding even more high-degree nodes beyond this point provides diminishing returns or in some cases even degrades the performance.

5.4.2 Regular Grids and Random Geometric Graphs

Grids and Random Geometric Graphs (RGG) are common models to study various properties and protocols for wireless systems. In the previous section, we showed that in line topologies the addition of cluster-heads can greatly reduce the cost of random-walk-based queries, do cluster-heads have the same significant effect on 2-dimensional topologies?

5.4.2.1 Grids

We assume that the number of cluster-heads is a perfect square and they are uniformly distributed on the grid, i.e. the grid is divided in the same number of cells as cluster-heads and each cluster-head is positioned in the node at the center of each cell.
Figure 5.6: (a) $C_{\text{event}}$, (b) $C_{\text{query}}$ and (c) $C_{\text{total}}$ for a grid topology with 169 nodes and values of $k$ ranging from 2 to 6.

According to the algorithm presented in 1, events appearing in a cluster-head has a cost of 0, events appearing in nodes directly connected to a cluster-head have a cost of 1 and events appearing in orphan nodes perform a simple random walk until it hits a node that is directly connected to a cluster-head. Denoting $M$ as the transitional probability matrix, $w \subseteq V$ as the subset of vertices containing all cluster-heads and all the nodes directly connected to them and $\bar{w} \subseteq V$ as its complement; we define $M_w$ as the sub-matrix where all the rows and columns of the vertices in $w$ have been removed. Hence, the hitting time for orphan nodes is given by:

$$h_{\bar{w}w} = \sum_{i=0}^{\mid\bar{w}\mid} (Q_w(i) + 1)$$

(5.26)

And $C_{\text{event}}$ cost for the grid topology is given by:

$$C_{\text{event}} = \frac{(\mid w \mid - (s + 1)) + h_{\bar{w}w}}{n}$$

(5.27)

In our analysis the sink is located at the bottom-left corner of the grid ($v_0$). The query cost $C_{\text{query}}$ is the average hitting time from the sink to the set of cluster-heads and it is given by combining equations 5.1 and 5.20.
Table 5.3: Random Geometric Graphs

<table>
<thead>
<tr>
<th>Transmission Range</th>
<th>Clustering</th>
<th>No Clustering</th>
<th>Savings (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.12</td>
<td>679.7</td>
<td>833.0</td>
<td>18.4</td>
</tr>
<tr>
<td>0.18</td>
<td>414.2</td>
<td>296.2</td>
<td>-39.8</td>
</tr>
<tr>
<td>0.24</td>
<td>171.4</td>
<td>225.0</td>
<td>23.8</td>
</tr>
<tr>
<td>0.30</td>
<td>88.0</td>
<td>202.7</td>
<td>56.6</td>
</tr>
<tr>
<td>0.36</td>
<td>52.9</td>
<td>193.5</td>
<td>72.7</td>
</tr>
</tbody>
</table>

Figure 5.6 presents results for $C_{event}$, $C_{query}$ and $C_{total}$ ((a), (b) and (c) respectively) for 169 nodes deployed on a 2D grid. The y axis represents the cost and the x axis the number of clusters. There are two important observations: i) contrary to the line topology, the case when all nodes are high-degree perform significantly worse, ii) similar to the line topology, the first cluster-heads account for most of the savings (greater than 30%) and the higher $k$ the higher the savings. Also note that, as $k$ increases, the case where all nodes are high-degree approaches a complete graph.

Another important difference with respect to the line topology is that the event cost plays a significant role in the total cost, while the query cost does not have the same significant impact. This may be due to several reasons one of them is that for the same number of nodes the diameter of a line ($\Theta(n)$) topology is significantly larger than a grid ($\Theta(\sqrt{n})$) topology. From the resistance method perspective, in grids we can not eliminate the edges beyond a given cluster, all edges should be considered and hence according to equation 5.2 this would make the query cost for different clusters more similar.

### 5.4.2.2 Random Geometric Graphs

The procedure for getting event and query costs in random geometric graphs is the same as for grids (equation 5.27, and a combination of equations 5.1 and 5.20). However, the interesting case in random geometric graphs is that even when only low-degree nodes are deployed there are some inherent cluster-heads due to some favorable geographical position. We further enhance the inherent cluster-heads formed by increasing their transmission range. According to the algorithm presented in 1, in these scenarios the event moves in a greedy way towards the local cluster-head.
<table>
<thead>
<tr>
<th>output power (dBm)</th>
<th>clustering</th>
<th>no clustering</th>
<th>savings (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-14</td>
<td>263.1</td>
<td>428.7</td>
<td>38.6</td>
</tr>
<tr>
<td>-13</td>
<td>211.7</td>
<td>370.4</td>
<td>42.8</td>
</tr>
<tr>
<td>-12</td>
<td>174.6</td>
<td>333.4</td>
<td>47.6</td>
</tr>
<tr>
<td>-11</td>
<td>152.6</td>
<td>311.4</td>
<td>51.0</td>
</tr>
<tr>
<td>-10</td>
<td>132.5</td>
<td>278.1</td>
<td>52.3</td>
</tr>
</tbody>
</table>

Table 5.4: Grid Deployments in Realistic Environments

Table 5.3 presents results for 169 nodes deployed randomly on a 1x1 square area. The results are the average over 50 runs. The initial radius is 0.12 which for this density gives a connectivity probability of $\approx 0.5$. The table has two columns named “clustering” and “no clustering”. Due to the random deployment some nodes will end up being local-clusters (their degree is higher or equal than their neighbors). For the “clustering” column, we enhance these local clusters by increasing their transmission range to the value given in the “transmission range” column, while the nodes that are not local cluster-heads have a range of 0.12. For the “no clustering” column all nodes have the transmission range given the "transmission range" column, but events stay in the nodes where they appear. We observe that in random geometric graphs clustering also have a significant impact on the performance of random-walk-based queries (except for $r=0.18$ where "no clustering" is better). It is important to mention that for the initial $r=0.12$ approximately $\sim 11\%$ of the nodes end up being local clusters.

### 5.4.3 Low-Power Wireless Graphs

Using the link layer model presented in 3, we evaluate through simulations the effectiveness of cluster-heads in realistic graphs, which are characterized by the presence of unreliable and asymmetric links. In order to guarantee the survival of the random walk we implemented a 3-way handshake protocol. A node with the random walk issues a request to the next neighbor to receive the random walk, upon reception of the packet the neighbor acknowledges the reception of the random walk, finally upon reception of the acknowledgment, the original node sends a release packet which ends the transfer of the random walk.
<table>
<thead>
<tr>
<th>output power (dBm)</th>
<th>clustering</th>
<th>no clustering</th>
<th>savings (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-14</td>
<td>367.2</td>
<td>557.7</td>
<td>34.2</td>
</tr>
<tr>
<td>-13</td>
<td>250.4</td>
<td>432.9</td>
<td>42.2</td>
</tr>
<tr>
<td>-12</td>
<td>243.8</td>
<td>380.8</td>
<td>36.0</td>
</tr>
<tr>
<td>-11</td>
<td>207.9</td>
<td>332.9</td>
<td>37.6</td>
</tr>
<tr>
<td>-10</td>
<td>169.9</td>
<td>294.4</td>
<td>42.3</td>
</tr>
</tbody>
</table>

Table 5.5: Random Deployments in Realistic Environment

Tables 5.4 and 5.5 present the results for grid and random deployments. The presentation is similar to Table 5.3, the “clustering” column represents networks were only the inherent local cluster-heads are enhanced by increasing their output power, while in the “no clustering” column all nodes increase their output power but events remain in the nodes where they appear. We can observe that clustering plays a significant role in reducing the cost of random walk-based queries (between 30% and 50%). On grid deployments approximately 12% of the nodes are inherent cluster-heads, while on random deployments about 8% of nodes are cluster-heads.

5.5 Experiments with Motes

In this section we present some empirical results. Thirty-one (31) micaZ motes were deployed in a chain topology, where nodes were spaced every 1 meter. The high-degree nodes had a higher output power than the low degree nodes. And the low degree nodes increased their output power only to reply to the high degree nodes. Each mote sent 100 test packets to measure the PRR of the links.

Communication graphs were obtained for a chain topology where all the nodes had the same low output power, for a chain where all nodes had the same high output power and for a mixture of both. In the graph were both type of nodes are present, the high degree nodes were evenly spaced on the chain. We simulated 300 random walk based queries on each of these graphs. The event followed Algorithm 1, and the query started at the left-most part of the chain (position 0) and followed a random walk until it hit the event. The results are presented in Figure 5.7.
Figure 5.7: Empirical study of random walks on degree-heterogeneous graphs, (a) presents the degree of the nodes, (b) presents the query cost for graphs with different number of clusters.

Figure 5.7 (a) presents the nodes’ degree. Four curves are presented: for the low-power graph, for the graphs with 1 and 5 clusters, and for the high power graph. Small horizontal arrows depict the position of the high degree nodes. We observe that the average high-degree is around 15 ($k = 7$). Figure 5.7 (b) present the average cost of finding the event. The x-axis represent the number of high-degree nodes (cluster-heads) and the y-axis the cost. $\theta(S)$ represents the case where the events in the low-power graph (0 cluster-heads) are static and remain in the nodes where they appear. $31(S)$ represents static events for the high-power graph (all nodes are cluster heads.). $\theta$ and $31$ represent the low-power and high-power graphs where events follow Algorithm 1.

It is interesting to observe by solely using Algorithm 1, without the presence of cluster-heads, we can obtain significant gains in both the low-power graph ($\theta(S) \rightarrow \theta : \sim 25\%$) and the high-power graph ($31(S) \rightarrow 31 : \sim 45\%$). This observation was also made in previous section by simulating the random walk based-query on a realistic graph. It is also important to highlight that the first clusters account for most of the savings, adding more cluster either leads to diminishing returns or consumes more energy. This behavior was predicted by our analysis. It also interesting to observe that even though the degree of the low-power nodes is higher than 2, the ratio $\frac{4}{5k}$ leads to approximately 3.4 cluster-heads to hit the first local minimum, which is a decent prediction.
Finally, it is important to state that the empirical studies presented in this section are limited and are shown only as a proof-of-concept. Large scale networks need to be tested in order to accurately assess the impact of the contributions of this Chapter on degree-heterogenous networks.

5.6 Summary

This chapter focused on the impact of node degree heterogeneity on random walk-based queries. As observed in the model presented in chapter 3, degree heterogeneity is a common property of WSN communication graphs, and we proposed a simple push-pull algorithm to exploit this property and reduce the query cost of random walks.

Using connections between random walks and electrical resistance, and Markov chain analysis of the time to absorption we provide some important conclusions. First, our work provides interesting theoretical results for line topologies showing that when cluster-heads have a coverage \( k \) (cover \( k \) nodes to the right and left) and are uniformly distributed, a fraction of \( \frac{4}{5k} \) nodes being cluster-heads can offer a reduction in query cost of \( O(1 - \frac{1}{k^2}) \). Second, in 2D random and grid deployments we presented numerical and simulation results showing that with a small percentage of high-degree nodes in the network (< 10%), significant cost savings can be obtained — between 30% and 70% depending on the coverage of the high-degree nodes.

Finally, we presented a limited set of empirically-derived results on micaZ motes that confirms the considerable impact of degree-heterogeneity on real WSN communication graphs.
Chapter 6

Conclusions

In this thesis, we have argued that the design of efficient WSN routing protocols require a communication graph that incorporates the non-idealities present in common deployments. We have substantiated this thesis by presenting three studies. In our first study, we have analyzed the behavior of the transitional region, and we have also proposed a more realistic link layer model. Based on this link-layer model, in our second and third studies we have proposed mechanisms to improve the performance of geographic routing protocols and random walk-based queries, respectively.

6.1 Analysis of the Transitional Region

In the first study (Chapter 3), we have presented an in-depth analysis of unreliable and asymmetric links in low-power multi-hop wireless networks. The main contributions of this work are twofold. First, it quantifies the impact of the wireless environment and radio characteristics on link reliability and asymmetry. And second, we propose a systematic way to generalize models for the link layer that can be used for the efficient design of routing protocols for WSN.

We have also derived expressions for the packet reception rate as a function of distance, and for the size of the transitional region. These expressions incorporate several radio parameters such
as modulation, encoding, output power, frame size, receiver noise floor and hardware variance; and channel parameters such as the path loss exponent and the log-normal variance.

The expressions we have derived provide some important insights into the impact of channel multi-path and hardware variance on the link behavior. First, the relative size of the transitional region is higher for lower path loss exponents and higher variances. Second, hardware variance induces a pseudo-log-normal variance, which increases the size of the transitional region. Third, a negative correlation between the output power and noise floor leads to nodes that are good transmitters and receivers, which helps to explain the clustering behavior observed in previous empirical studies [28, 13]. And fourth, even with a perfect-threshold radio, the transitional region still exists as long as multi-path effects exist.

Our work contributes to a better understanding of the behavior of low-power wireless links but is not exhaustive. It can be complemented with other studies to capture other important phenomenon present in real scenarios; for instance, contention models from [61], temporal properties from [14] and correlations due to direction of propagation from [72] (Appendix C). Also, from preliminary results (Appendix D), we have found that even spread spectrum radios show transitional region effects; we therefore believe there is value in extending this work to other settings.

6.2 Impact of Lossy Links on Geographic Routing

In our second study (Chapter 4), we have studied the performance of geographic routing under the model derived in Chapter 3. In real scenarios, the distance-greedy forwarding mechanism of geographic routing leads to significant wastage of energy resources due to dropped packets. We have shown that the optimal forwarding choice is generally to neighbors in the transitional region, which denotes that efficient geographic forwarding strategies do take advantage of the high variance in packet reception rate of this region.
The most important contribution of this work has been the derivation of local optimal forwarding metrics to balance the distance-hop energy trade-off for both ARQ and No-ARQ scenarios. Specifically, for ARQ systems, our analysis, simulations and empirical observations have shown that the product of the packet reception rate and the distance towards destination ($P_{RR} \times d$) is the optimal local metric. Our analysis has also shown that reception-based forwarding strategies are generally more efficient than distance-based strategies.

While the work on this thesis has been focused on chain topologies, we also have results [58] that show that the $P_{RR} \times d$ metric outperforms other metrics in 2-D random deployments. It is also important to highlight that the $P_{RR} \times d$ metric is recommended for static or low dynamic environments, such as environmental monitoring. In highly dynamic environments the link quality can change drastically with time, and hence, stable estimates of PRR might not be possible. In these scenarios, best reception is probably the best forwarding strategy.

6.3 Performance of Random Walks on Heterogeneous Networks

In our last study (Chapter 5), we have analyzed the impact of node degree heterogeneity on random walk-based queries. As shown by some empirical studies [28, 13], and further proved by our analysis on hardware variance, degree heterogeneity is a common property of WSN communication graphs (even on homogeneous networks deployed on a regular grid [28]).

The main contribution of this study has been to show that by using a simple distributed push-pull algorithm and having a small percentage of high-degree nodes in the network (10% <), significant cost savings can be obtained — between 30% and 70% depending on the coverage of the high-degree nodes. Our work has also provided interesting theoretical results for line topologies showing that when cluster-heads have a coverage $k$ (cover $k$ nodes to the right and left) and are uniformly distributed, a fraction of $\frac{4}{5k}$ nodes being cluster-heads can offer a reduction in query cost of $\Theta(1 - \frac{1}{k^2})$.  

93
It is also important to mention that one of the drawbacks of random walks is the significant delay that they encounter. In our work, we have achieved simultaneous reduction in energy cost and delay by minimizing the required number of steps on the random walk. However, an accurate quantification of the delay reduction should include specific characteristics of the protocol used at the MAC layer.
Appendix A

Models from Communication Theory

This section gives an overview of the log-normal path loss model, encoding and modulation.

A.1 Log-Normal Path Loss Model

When an electromagnetic signal propagates, it may be diffracted, reflected and scattered. These effects have two important consequences on the signal strength. First, the signal strength decays exponentially with respect to distance. And second, for a given distance \( d \), the signal strength is random and log-normally distributed about the mean distance dependent value.

Due to the unique characteristics of each environment, most radio propagation models use a combination of analytical and empirical methods. One of the most common radio propagation models is the log-normal path loss model [56]. This model can be used for large and small coverage systems [59]. Furthermore, empirical studies have shown that the log-normal path loss model provides more accurate multi-path channel models than Nakagami and Rayleigh for indoor environments [53].

According to this model the received power (\( P_r \)) in dB is given by:

\[
P_r(d) = P_t - PL(d_0) - 10 \eta \log_{10}(\frac{d}{d_0}) + N(0, \sigma)
\]  

(A.1)
Where $P_t$ is the output power, $\eta$ is the path loss exponent which captures the rate at which signal decays with respect to distance, $\sigma$ is the standard deviation due to multi-path effects and $PL(d_0)$ is the power decay for the reference distance $d_0$.

### A.2 Encoding and Modulation

Usually, the radio does not send binary data directly (‘0s’ and ‘1s’), but it encodes them into bauds. This process is called encoding, and one baud can be less, more or equal to 1 bit. After encoding, the radio uses a modulation mechanism to send these bauds over the wireless channel. The radio have several options; it can modulate the amplitude, frequency or phase of the carrier frequency.

Both, the encoding and modulation used, play an important role on the link behavior of WSN. Now, we will provide an example to obtain the packet reception rate ($\Psi$) for radios using Non-Coherent FSK (NCFSK) modulation and Non Return to Zero (NRZ) encoding (where 1 bit = 1 baud).

In the presence of additive white gaussian noise (AWGN) the probability of bit error $P_e$ of the receiver is given by:

$$P_e = \frac{1}{2} \exp^{-\gamma}$$  \hspace{1cm} (A.2)

Where $\gamma$ is the signal to noise ratio (SNR). A frame is received correctly if all its bits are received correctly, hence, for a frame\(^1\) of length $f$ (in bytes) the probability of successfully receiving a packet is:

$$\Psi = (1 - P_e)^{8f}$$  \hspace{1cm} (A.3)

\(^1\)A frame consists of: preamble, network payload (packet) and CRC
Finally, by inserting equation A.2 in equation A.3, the PRR $\Psi$ is defined as:

$$\Psi = (1 - \frac{1}{2} \exp^{-\frac{\gamma}{2}})^{8f} \quad (A.4)$$

Table 3.6 provides expressions for various encoding and modulation techniques.

### A.3 Noise Floor

Another important element that determines the link behavior is the noise floor. The temperature of the environment influences the thermal noise generated by the radio components (noise figure)$^2$. When the receiver and the antenna have the same ambient temperature the noise floor is given by [56]:

$$P_n = (F + 1)kT_0B \quad (A.5)$$

Where $F$ is the noise figure, $k$ the Boltzmann’s constant, $T_0$ the ambient temperature and $B$ the equivalent bandwidth.

$^2$Interfering signals can further influence the noise floor, but they are out of the scope of this thesis.
Appendix B

Random Walks

In this section we present two mathematical tools used to derive hitting times in random walks.

B.1 Resistance Method

For a graph \( G(V, E) \) where \(|V|\) is the number of nodes and \(|E|\) is the number of edges, the following notation will be used. An element of \( V \) or \( E \) is represented by a lowercase, a subset or array is represented by a bold lowercase and the complement of a set or element will be denoted with an upper-bar, for example \( e \) represents an element, \( e \) an array (subset) and \( \bar{e} \) and \( \bar{e} \) represent the complements of \( e \) and \( e \), respectively.

The hitting time \( h_{uv} \) is the expected time taken by a simple random walk starting at \( u \) to reach \( v \) for the first time [49].

For a source \( s \in V \) and the subset of cluster-head nodes \( e \subseteq V \), the average hitting time from \( s \) to \( e \) is:

\[
h_{se} = \frac{\sum_{e \in e} h_{se}}{|e|} \quad \text{(B.1)}
\]

The commute time \( C_{uv} \) is defined as the expected time taken by a random walk starting at \( u \) to reach \( v \) and come back to \( u \). Note that by definition \( C_{uv} = h_{uv} + h_{vu} \), but in general \( h_{uv} \neq h_{vu} \).
In a seminal work, Doyle and Snell [24] explored the connection between a random walk on a graph $G$ and the resistance of an electrical network obtained from $G$ by viewing each edge as a unit resistor. In [15], Chandra et al. extended this work and proved the following equality that relates the commute time $C_{uv}$ and the effective resistance $R_{uv}$ of the electrical network of $G$:

$$C_{uv} = 2mR_{uv} \quad (B.2)$$

Where $m$ is the number of edges in the graph. Notice that in case of symmetry\(^1\) $h_{uv} = h_{vu}$, which implies that the commute time is two times the hitting time. We use this property in the analysis of the hitting time for line topologies.

### B.2 Time to Absorption in Markov Chains

In this subsection, we briefly describe a key result in the analysis of Markov chains with absorbing states. Quite often it is of interest to find out what is the expected time it takes for a chain starting at a transient state $i$ to reach the absorption state. This time is called the time to absorption from state $i$.

Given a graph $G(V, E)$, a random walk on the graph can be defined by a Markov chain $\mathcal{M}$, where $\mathcal{M}$ represents the transition probability matrix. The notation used for elements and subsets of $V$ and $E$ is the same as the one used in the previous subsection.

Let us consider that the event is in node $e \in V$ and the sink is in node $s \in V$. As presented in [63], absorption states can be used to obtain hitting times. Let $\mathcal{M}_e$ be the matrix resulting from deleting the row and column corresponding to $e$ in $\mathcal{M}$, and let $Q_e$ be:

$$Q_e = (I - \mathcal{M}_e)^{-1}1 \quad (B.3)$$

\(^1\)The graph $G'$ where we name $u$ as $v$ and vice versa is isomorphic to $G$
Where $I$ is the identity matrix and $\mathbf{1}$ is a column vector of ones. $Q_e$ is an array representing the expected hitting time from each node to $e$ (except $e$). Hence, in our case $h_{se} = Q_e(s)$, where $Q_e(s)$ represents the $s$’th element of the array.
Appendix C

The Radio Irregularity Model

This appendix provides some steps to include non-isotropic properties of the transmission coverage, and the impact of obstacles.

In the RIM model [72], the authors present the degree of irregularity (DOI) coefficient as a mean to capture the variation per unit degree change in the direction of radio propagation. In that work the received power is given by:

\[
P_r(d) = P_t - DOI\text{AdjustedPathLoss} + \mathcal{N}(0, \sigma)
\]

(\text{C.1})

Where \(K_i\) is a coefficient to represent the difference in path loss in different directions, and the method to obtain it is presented in the RIM model. Hence, denoting \(PL(d_0) = PL(d_0) + 10 \eta \log_{10}(\frac{d}{d_0})\), equation 3.1 can be modified to include non-isotropic effects:

\[
P_r(d) = P_t - PathLoss \times K_i + \mathcal{N}(0, \sigma)
\]

(C.2)

The effect of obstacles can be included by inserting a new variable on the previous equation. Let us denote \(\Omega_{vw}\) as the path loss in dB due to an obstacle between nodes \(v\) and \(w\), for example a wall. Then, letting \(v\) be the transmitter, the received power at \(w\) is given by:

\[
P_r(d) = P_t - \frac{PL(d_0)}{d_0} \times K_i + \mathcal{N}(0, \sigma)
\]

(C.2)
\[ P_{rw}(d) = P_t - (PL(d_0) + \Omega_{vw}) \times K_i + \mathcal{N}(0, \sigma) \] (C.3)

Hence, if the layout of the environment is provided, the previous equation can be used to include additional path loss for each pair of nodes according to the obstacles between them.
Appendix D

The Transitional Region in MicaZ Motes

Some preliminary empirical evaluations were done with micaZ devices. These motes have a 2.4 GHz IEEE 802.15.4/ZigBee\(\text{™}\) RF transceiver, which uses DSSS modem with 2 Mchips/s and 250 kbps effective data rate. A chain topology with the same methodology as Chapter 3.4.2 was deployed in the same indoor environment as mica2 motes.

Figure D.1 presents empirical measurements for the channel, radio and links for mica2 and micaZ. The nominal output power for both types of motes was -10 dBm. We observe that the transitional region still has a significant extent. However, for the same output power micaZ radios seem to have a larger connected and transitional regions.

No major differences were found in the shadowing standard deviation for both deployments (around 6.1 for both). However, the path loss exponent for micaZ measurements is 1.94 which is smaller than the corresponding value for mica2 in Table 3.2 \(\eta = 3.3\) . According to equation 3.8 a smaller \(\eta\) increases the size of both regions, which provides some intuition as to why the extent of the regions are larger for micaZ motes for the same output power.

The spread spectrum techniques seem to partially combat multi-path by decreasing \(\eta\) and consequently providing a larger coverage for the same output power, however, as stated in equation 3.9 a lower \(\eta\) implies a larger transitional region which increases the number of unreliable
Figure D.1: Comparison of empirical measurements for channel, radio and link between mica2 and micaZ motes, $P_t = -10$ dBm for both type of motes, (a) channel mica2, (b) radio mica2, (c) link mica2, (d), (e), (f) are their micaZ counterparts.

and asymmetric links. An in-depth study of the impact of low-cost spread spectrum radios in the transitional region is part of our future work.
Reference List


