6

Intelligent Antenna Sharing in Cooperative Diversity Wireless Networks

by

Aggelos Anastasiou Bletsas

Diploma, Electrical and Computer Engineering,
Aristotle University of Thessaloniki (1998)
S.M., Media Arts and Sciences,
Massachusetts Institute of Technology (2001)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2005

Author_____

Program in Media Arts and Sciences
August 5, 2005

Certified by_____

Andrew B. Lippman
Principal Research Scientist, Program in Media Arts and Sciences
MIT Media Lab
Thesis Supervisor

Accepted by_____

Andrew B. Lippman
Chairman, Departmental Committee on Graduate Studies
Program in Media Arts and Sciences

Intelligent Antenna Sharing in Cooperative Diversity Wireless Networks

by

Aggelos Anastasiou Bletsas

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on August 5, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

Cooperative diversity has been recently proposed as a way to form virtual antenna arrays that provide dramatic gains in slow fading wireless environments. However, most of the proposed solutions require simultaneous relay transmissions at the same frequency bands, using distributed space-time coding algorithms. Careful design of distributed space-time coding for the relay channel is usually based on global knowledge of some network parameters or is usually left for future investigation, if there is more than one cooperative relay.

We propose a novel scheme that eliminates the need for space-time coding and provides diversity gains on the order of the number of relays in the network. Our scheme first selects the best relay from a set of M available relays and then uses this "best" relay for cooperation between the source and the destination. Information theoretic analysis of outage probability shows that our scheme achieves the same diversity-multiplexing gain tradeoff as achieved by more complex protocols, where coordination and distributed space-time coding for M relay nodes is required. Additionally, the proposed scheme increases the outage and ergodic capacity, compared to non-cooperative communication with increasing number of participating relays, at the low SNR regime and under a total transmission power constraint.

Coordination among the participating relays is based on a novel timing protocol that exploits local measurements of the instantaneous channel conditions. The method is distributed and allows for fast selection of the best relay as compared to the channel coherence time. In addition, a methodology to evaluate relay selection performance for any kind of wireless channel statistics is provided. Other methods of network coordination, inspired by natural phenomena of decentralized time synchronization, are analyzed in theory and implemented in practice.

It was possible to implement the proposed, virtual antenna formation technique in a custom network of single antenna, half-duplex radios.

Thesis Supervisor: Andrew B. Lippman
Title: Principal Research Scientist, Program in Media Arts and Sciences, MIT Media Lab

Intelligent Antenna Sharing in Cooperative Diversity Wireless Networks

by

Aggelos Anastasiou Bletsas

The following people served as readers for this thesis:

Thesis Reader_____

Joseph A. Paradiso
Associate Professor
Program in Media Arts and Sciences, MIT

Thesis Reader_____

Moe Win
Associate Professor
Laboratory for Information and Decision Systems (LIDS), MIT

Acknowledgements

Ph.D. theses commonly seem like an endless journey. My voyage to the unknown seas of research started with a brand new ship. That ship endured storms, survived wild tempests and enjoyed subsequent sunshine and calmness. Now, the ship has arrived to its destination and even though it is not brand new, the difficulties it sustained during the journey made it more capable for future endeavors: the ship is ready to sail again, more experienced and adventurous than before, to its new destinations.

For my Ph.D. journey, I am grateful to my advisor Andy Lippman. He gave me the ticket for my Ph.D. voyage and freedom plus all necessary means to explore. At the same time he kept me on track towards the final destination. This thesis was shaped under great guidance and effort from Andy.

Joe Paradiso, the second member of my committee, had always a good word for my work. He was always motivating and he and his student Josh Lifton provided support in several dimensions. Moe Win, the third member of my committee, had his office door always open for me for encouragement and he was always ready to answer hard research questions. I am very proud that I collaborated with all the above people.

David Reed had always a calm way to explain his stand on my research and conversations with him always sparked interesting new ideas. Ashish Khisti helped me a lot during the final stretch of my thesis. His sharp comments on my thesis proposal helped me explore additional dimensions and multiply the quality of my work. I am grateful for all his help. Thucydides (Duke) Xanthopoulos helped me at the beginning of my graduate studies and

endlessly discussed with me interesting research problems and even more interesting solutions. His creative thinking was a good model for me.

My Undergraduate Researchers, Vimal Bhalodia, Amanda Lechman and Marios Michalakis that I advised during my Ph.D. years, spent numerous hours on assisting me in the preparation of the group demos. I thank them all for their hard work and I hope they learned (something) from me.

My office-mates, Dean Christakos, Ilia Mirkin and Dimitris Vyzovitis helped me in various software related problems and provided me with three examples of how a real *MIT hacker* looks like.

My Group Mates, Jamie Cooley, Jeff Hallbig, Casey Muller, Hector Yuen, Arthur Petron helped in various stages of my research, especially during the preparations for the Lab's sponsor open houses and they showed me what *MIT group collaboration* is about. I thank them all, as well as the rest of the Viral Comm Gang, Kwan Hong Lee and Fulu Li for all their efforts.

The Media Lab Network and Computer Systems Group (NecSys), Facilities Group as well as several other people in the lab like Stacie Slotnick, Tamara Hearn, Betsy Cimento and Sandy Sener provided me with a working environment, where I had to worry only about research problems and nothing else.

Special thanks I would like to give to my executive, media lab *aunts* Deborah Widener and Polly Guggenheim who made sure that I was always in good shape and health. They always satisfied my requests with a beautiful smile and clearly demonstrated to me professionalism, diligence and care.

My crazy roomates, Maziar Tavakoli-Dastjerdi, Saul Griffith, Yael Maguire, John Maloney and Noah Vawter endured my experiments in greek cuisine cooking (even though they didn't always know *that dish was an experiment...*) and provided a fun place to live and (sometimes) party.

My many good friends provided support during my graduate studies at MIT. They all believed in me, encouraged and motivated me. My close friends at Crete: Yiannis Kallitsakis, Dia Koskina, Olga Kolokitha, Andonis Plevrakis, Costas Gazis, Ioanna Koutsoupaki, Mihalis Tzatzanis. My close friends here in Boston (including Greek Mafia): Thodoros Konstantakopoulos, Anna Stefanidou, Ioannis Kitsopanidis, Nikol Papadopoulou, Duke Xanthopoulos, Margarita Dekoli, Vasilis Ntziachristos, Christina Benou, Eirini Iliaki, Ioannis Kizanis, Anna Kondyli, Yiannis Zacharakis, George Themelis, Dimitris Rovas, Christina Samaraki, Georgia Konstadinopoulou, Karrie Karrahalios, Paris Smaragdis, Costas Pelekanakis, Thais Aleluia, Petros Bufunos, Thodoros Akiskalos, Christi Electris, Constadinos Caramanis, George Constadinidis, Maria-Katerina Nikolinakou, Angelina Aessopou, Wei Chai and those I forgot.

Special thanks I would like to give to my two Boston buddies, Thodoros Konstantakopoulos and Yiannis Kitsopanidis for their olympian patience to put up with me all those years in Boston. I met them during my graduate studies and our connection all those years is probably the most important gift I received from MIT.

I would not have come to MIT if I didn't have a family that found education as one of the greatest virtues and encouraged its children to continuously study. Therefore, I would like to thank my brother and his family here in Boston, my sister and her family in Bergen, Norway, Eleftheria and her family in Thessaloniki, Greece, and of course my parents and my grandaunt in Chania, Crete.

"Tolerating Ambiguity is a sign of maturity", Stephen A. Benton (1941-2003). The ship is now ready to sail again.

…στην Ελευθερία, στον Θοδωρή, στην Αιμιλία-Αναστασία,
στον Κωστή και στην Χριστίνα…

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the era of pervasive computing and communications, another thesis on wireless communication and networking might seem obsolete or outdated. However, we have all experienced *bad reception* while using our cell phone (also known as poor quality of service), we have all forgotten to recharge the device during the night and subsequently become unable to use it during the day (energy/battery problems), and we have waited for too long for cellular technology to mature until we could start exchanging pictures or videos with our friends, using our cell phones. Even in that case, data speed (throughput) is significantly less than the speed of Wi-Fi wireless technology we have been using in our homes. Finally, we have all failed to talk to our friends using our cell phones in large venues such as the celebration of 4th of July in front of Media Lab, when thousands of people alongside Charles river gather to enjoy the spectacular fireworks but overload the statistically provisioned cellular network.

Can we enhance the quality of service (QoS), increase the data speed (throughput) and/or reduce the required energy (therefore increase battery life), without overusing common resources such as spectrum or scarce resources like the available battery energy? Can we further reduce the transmission power levels of every base station and therefore minimize public health risks due to electromagnetic radiation? Can we create wireless networking architectures that scale with increasing number of users and, if possible, perform better as

the users in the system increase?

Recent developments on multi-antenna transceivers (also known as Multi-Input Multi Output systems) show that for the same bandwidth and power[1] resources compared to traditional single-antenna communication, MIMO systems increase throughput (*multiplexing gain*) and/or increase reliability of communication (*diversity gain*). The extra degree of freedom (apart from time and frequency) results from space, by exploiting the possible statistical independence between the transmitting-receiving antenna pairs. The statistics of the multi-antenna wireless channel provide independent, parallel spatial communication channels, at the same carrier frequency and at the same time. In other words, MIMO systems exploit space and statistical properties of the wireless channel and typically need intensive signal processing computation for channel estimation and information processing. Apart from extensive computation requirements, engineering and physical limitations preclude the utilization of many antennas at the mobile terminal (typically no more that two antennas at the cordless phone) and therefore, multi-antenna transceivers are typically utilized at the base station side.

What happens when multiple antennas belong to different users? Can we exploit multiple observations of the same information signal, from users distributed in space, given the broadcast nature of the wireless medium? Can we earn the benefits of traditional MIMO theory when the antennas belong to different users? In other words, this thesis explores users in a network as an additional degree of freedom, apart from time, frequency and space, in combination with the intrinsic properties of the wireless channel (figure 1-1).

The problem of user cooperation in wireless communication poses exciting challenges: a) computation (processing) capabilities of cooperating users are limited, since we assume they are mobile, with fixed computation capacity and energy consumption b) cooperation basically means that one user will use their own battery to relay information destined to a different user, while the receiver will exploit the direct and the relayed transmission. Therefore, strong incentives should be inherent in any cooperative scheme and c) coordination at

[1] Average energy and power will be used equivalently, since they are different by a multiplying factor, the information symbol duration.

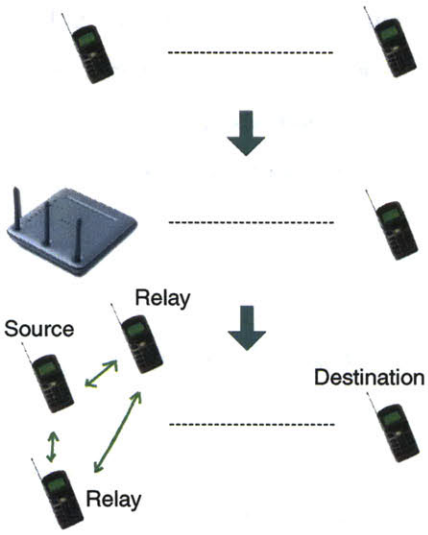

Figure 1-1: Multiple antenna transceivers improve the efficiency of wireless communication. What if the antennas belonged to different terminals? This thesis studies this problem and proposes a practical scheme for existing, low cost radios.

the network level among the cooperative nodes should be facilitated, requiring important modifications in existing communication stacks. This is due to the fact that existing networking has been structured for point-to-point, non-cooperative communication links that act as wires.

We are interested in practical schemes that address all the above issues and are applicable using existing RF hardware architectures. To investigate performance, we have done theoretical analysis and we have also implemented proposed solutions using low cost embedded radios. Cooperation could lead to substantial total (network) transmission power savings or increased spectral efficiency (in bits per second per hertz) under certain conditions. The goal of this thesis is to provide distributed and adaptive cooperation algorithms that can be applied in practice.

We will extensively study coordination algorithms required for user cooperative communication. The notion of cooperation can be extended to other important problems: if users in a network have strong incentives to cooperate for efficient wireless communication, then they could use cooperative strategies for network time keeping and positioning. We

will show that cooperative communication networks could autonomously maintain a global clock (time keeping), using local computation. Therefore the network becomes the timing system with specific accuracy and precision performance, again as a function of number of users. Efficient communication and autonomous timing are considered important problems, in future wireless sensor networks.

1.1 Approach and Related Work

In this thesis, we introduce and analyze a practical scheme that forms a virtual antenna array among single antenna terminals, distributed in space. The setup includes a set of cooperating relays that are willing to forward received information towards the destination. The proposed method features a distributed algorithm that selects the most appropriate relay to forward information towards the receiver. The decision is based on the end-to-end *instantaneous* wireless channel conditions for all nodes and the algorithm is distributed among the cooperating wireless terminals with minimal overhead.

The best relay selection algorithm lends itself naturally into cooperative diversity protocols [69, 70, 44, 33], which have been recently proposed to improve reliability in wireless communication systems using distributed virtual antennas. The key idea behind these protocols is to create additional paths between the source and destination using intermediate relay nodes. In particular, Sendonaris, Erkip and Aazhang [69], [70] proposed a way of *beamforming* where source and a cooperating relay, assuming knowledge of the forward channel, adjust the phase of their transmissions so that the two copies can add coherently at the destination. Beamforming requires considerable modifications to existing RF front ends that increase complexity and cost. Laneman, Tse and Wornell [44] assumed no Channel State Information (CSI) at the transmitters and therefore assumed no beamforming capabilities and proposed the analysis of cooperative diversity protocols under the framework of diversity-multiplexing tradeoff. Their basic setup included one sender, one receiver and one intermediate relay node and both analog as well as digital processing at the relay node were considered. The diversity-multiplexing tradeoff of cooperative diversity protocols with

multiple relays was studied in [43, 2]. While [43] considered the case of orthogonal transmission[2] between source and relays, [2] considered the case where source and relays could transmit simultaneously. It was shown in [2] that by relaxing the orthogonality constraint, a considerable improvement in performance could be achieved, albeit at a higher complexity at the decoder. These approaches were however information theoretic in nature and the design of practical codes that approach these limits was left for further investigation.

Such code design is difficult in practice and an open area of research: while space time codes for the Multiple Input Multiple Output (MIMO) link do exist [21] (where the antennas belong to the same central terminal), more work is needed to use such algorithms in the relay channel, where antennas belong to different terminals distributed in space. The relay channel is fundamentally different than the point-to-point MIMO link since information is not *a priori* known to the cooperating relays but rather needs to be communicated over noisy links. Moreover, the number of participating antennas is not fixed since it depends on how many relay terminals participate and how many of them are indeed useful in relaying the information transmitted from the source. For example, for relays that decode and forward, it is necessary to decode successfully before retransmitting. For relays that amplify and forward, it is important to have a good received SNR, otherwise they would forward mostly their own noise [58].

Therefore, the number of participating antennas in cooperative diversity schemes is in general random and variable and space-time coding invented for fixed number of antennas should be appropriately modified. It can be argued that for the case of orthogonal transmission studied in this thesis (i.e. transmission during orthogonal time or frequency channels) codes can be found that maintain orthogonality in the absence of a number of antennas (relays). That was pointed in [43] where it was also emphasized that it remains to be seen how such codes could provide residual diversity without sacrifice of the achievable rates.

Additionally, proposed amplify and forward distributed space-time coding [35] usually as-

[2]Note that in that scheme the relays do not transmit in mutually orthogonal time/frequency bands. Instead they use a space-time code to collaboratively send the message to the destination. Orthogonality refers to the fact that the source transmits in time slots orthogonal to the relays. Throughout this work we will refer to Laneman's scheme as orthogonal cooperative diversity.

sumes that the receiver knows the channel conditions between initial source and all participating relays. Even though such an assumption is convenient for analysis purposes, it is far from practical in actual implementations, since the receiver has no way to estimate those channel conditions: wireless channel between source and relays could be estimated at the relays and such estimate could be communicated to the destination. Such overhead might be prohibitive in actual implementations.

In short, providing for practical space-time codes for the cooperative relay channel is fundamentally different than space-time coding for the MIMO link channel and is still an open and challenging area of research.

Apart from practical space-time coding for the cooperative relay channel, the formation of virtual antenna arrays using individual terminals distributed in space, requires significant amount of coordination. Specifically, the formation of cooperating groups of terminals involves distributed algorithms while synchronization at the packet level is required among several different transmitters. Those additional requirements for cooperative diversity demand significant modifications to almost all layers of the communication stack (up to the routing layer) which has been built according to "traditional", point-to-point (non-cooperative) communication.

In fig. 1-2 a transmitter transmits its information towards the receiver while all neighboring nodes are in listening mode. For a practical cooperative diversity in a three-node setup, the transmitter should know that allowing a relay at location (B) to relay information, would be more efficient than repetition from the transmitter itself. This is not a trivial task and such an event depends on the wireless channel conditions between transmitter and receiver as well as between transmitter-relay and relay-receiver. What if the relay is located in position (A)? This problem also manifests in the multiple relay case, when one attempts to simplify the physical layer protocol by choosing the best available relay. In [83] it was suggested that the best relay be selected based on location information with respect to source and destination based on ideas from geographical routing proposed in [85]. Such schemes require knowledge or estimation of distances between all relays and destination and therefore require either a) infrastructure for distance estimation (for example GPS receivers at each terminal) or b)

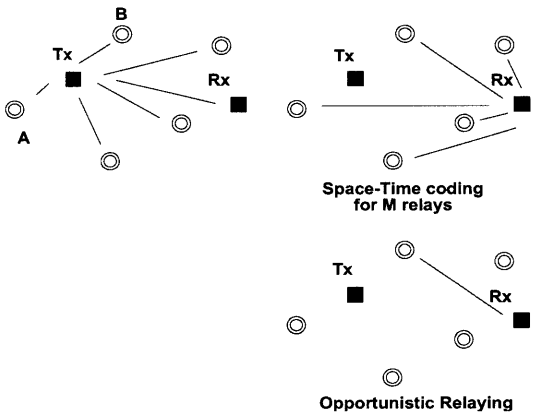

Figure 1-2: A transmission is overheard by neighboring nodes. Distributed Space-Time coding is needed so that all overhearing nodes could simultaneously transmit. In this work we analyze "Opportunistic Relaying" where the relay with the strongest transmitter-relay-receiver path is selected, among several candidates, in a distributed fashion using instantaneous channel measurements.

distance estimation using expected SNRs which is itself a non-trivial problem and is more appropriate for static networks and less appropriate for mobile networks, since in the latter case, estimation should be repeated with substantial overhead.

In contrast, we introduce a novel scheme that selects the best relay between source and destination based on *instantaneous* channel measurements. The proposed scheme requires no knowledge of the topology or its estimation. The technique is based on signal strength measurements rather than distance and requires a small fraction of the channel *coherence time.* Specifically, a single relay is chosen in a *proactive* manner, before the actual information is transmitted. Then that single relay is used to forward information towards the destination, as opposed to simultaneous transmissions from all relays (at the same frequency band) that would require space-time coding and expensive RF front ends (figure 1-2). All these features make the design of such a scheme highly challenging and the proposed solution becomes non-trivial. Additionally, the algorithm itself provides for the necessary coordination in time and group formation among the cooperating terminals.

The three-node reduction of the multiple relay problem we consider, greatly simplifies the physical layer design. In particular, the requirement of space-time codes is completely

eliminated if the source and relay transmit in orthogonal time-slots. We further show that there is essentially no loss in performance in terms of the diversity-multiplexing tradeoff as compared to the transmission scheme in [43] which requires simultaneous transmissions and distributed space-time coding across the relays that successfully decode the source message. We also note that our scheme can be used to simplify the non-orthogonal multiple relay protocols studied in [2]. Intuitively, the gains in cooperative diversity do not come from using complex schemes, but rather from the fact that we have enough relays in the system to provide sufficient diversity.

The simplicity of the technique allows for immediate implementation in existing radio hardware. An implementation of the scheme using custom radio hardware is described in section 2.3. Its adoption could provide improved flexibility (since the technique addresses coordination issues), reliability and efficiency (since the technique inherently builds upon diversity) in future 4G wireless systems, down to low-cost sensor networks.

1.1.1 Key Contributions

One of the key contribution of this work is to propose and analyze a simplification of user cooperation protocols at the physical layer by using a smart relay selection algorithm at the network layer. We take the following steps, towards this end:

- We suggest and analyze a new protocol for selection of the "best" relay between the source and destination. This protocol has the following features:
  - The protocol is distributed and each relay only makes local channel measurements.
  - Relay selection is based on *instantaneous* channel conditions in slow fading wireless environments. No prior knowledge or estimation of topology is required.
  - Relay selection is *proactive* i.e. selection is performed before information transmission, allowing for non-participating relays to enter sleep mode. In that way, reception energy is saved while diversity benefits of a large number of relays is realized.

- The amount of overhead involved in selecting the best relay is minimal. It is shown that there is a flexible tradeoff between the time incurred in the protocol and the resulting error probability.

- The impact of smart relaying on the performance of user cooperation protocols is studied. In particular, it is shown that there is no loss in performance (in terms of the diversity-multiplexing tradeoff) if only the best relay participates in cooperation. Intelligent relaying as introduced in this thesis, provides an alternative solution with a very simple physical layer to conventional cooperative diversity protocols that rely on space-time codes. The scheme could be further used to simplify space-time coding in the case of non-orthogonal transmissions.

- Outage and ergodic[3] capacity of the proposed protocol grow with increasing number of relays, under a total transmission power constraint. Existing work in the field usually assumes that total power scales when relay nodes are added into the system. In this work, the benefits of relaying come from the nature of the introduced technique that allows the cooperating relays to act as *wireless channel sensors*, not as power adders. Total transmission power is kept constant with increased number of relays at the analysis of chapters 3,4.

- This thesis introduces, analyzes in theory and implements in hardware centralized and decentralized time keeping algorithms. Time keeping can be viewed as a useful way to provide coordination and simplify scheduling in wireless cooperative networks.

Since the communication scheme exploits the wireless channel at its best, via distributed cooperating relays, we naturally called it *opportunistic relaying.* The term "opportunistic" has been widely used in various different contexts. In [5], it was used in the context of repetitive transmission of the same information over several paths, in 802.11b networks. In our setup, we do not allow repetition since we are interested in providing diversity without sacrificing the achievable rates, which is a characteristic of repetition schemes. The term

[3]throughout this work, we will assume *slow* fading where outage rates are appropriate for discussion. However, we will also consider *fast* fading to further explore ergodic rates.

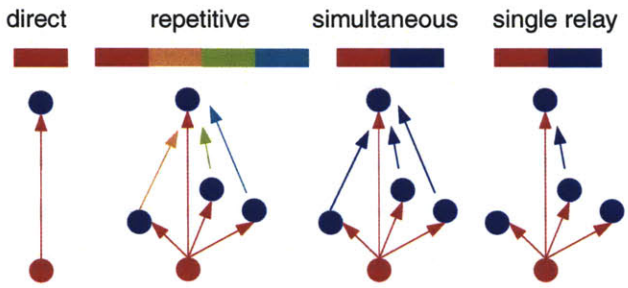

Figure 1-3: A transmission is overheard by neighboring nodes. Relays could repeat one after the other, or simultaneous transmit at the same frequency band using a Distributed Space-Time code. Alternatively, a single relay could be elected to retransmit. Under opportunistic relaying, the last two approaches become equivalent from a diversity-multiplexing tradeoff point-of-view.

"opportunistic" has also been used in the context of efficient *flooding* of signals in multi-hop networks [68], to increase communication range and therefore has no relationship with our work. We first encountered the term "opportunistic" in the work by Viswanath, Tse and Laroia [80], where the base station always selects the best user for transmission in an artificially induced fast fading environment. In our work, a mechanism of multi-user diversity is provided for the relay channel, in single antenna terminals. Our proposed scheme, resembles *selection diversity* that has been proposed for centralized multi-antenna receivers [55]. In our setup, the single antenna relays are distributed in space and attention has been given in selecting the "best" possible antenna, well before the channel changes again, using minimal communication overhead.

1.1.2 Additional Relevant Work

It is important to note that under no Channel State Information (CSI) at the transmitters, cooperation can not in general increase the *ergodic* capacity between a source and a destination, assisted by a single intermediate cooperating relay. That was demonstrated in [42]. In this thesis, we show that under opportunistic relaying, outage as well as ergodic capacity are increased provided that more than one cooperating relays are employed, even when no

CSI is available at the source.

The authors in [43] showed that when source and relays transmit in subsequent time slots, then there is an inherent loss in terms of the achievable communication speed (rate). That loss can range from a factor of $1/2$ when relays simultaneously transmit via a distributed space-time code or a factor of $1/M$ when M relays repeat, one after the other, the received information (figure 1-3). An interesting alternative occurs when the relays retransmit compressed versions of the original message in a round-robin fashion. In such cases, the rate loss is expected to be less than $1/M$. Analysis of a single compress and forward relay can be found in [37]. References there summarize mostly information theoretic work on the relay channel.

In [7], we provided a practical scheme in the context of Orthogonal Frequency Division Multiplexing (OFDM) wireless networks (like in 802.11a), where cooperative communication could be employed without further sacrifice of rate: direct and relayed transmission could happen within the same symbol period, due to the special structure of OFDM symbols and properties of over-sampling. Therefore, benefits of cooperation can be experienced without additional delay or reduced rate, at the cost of increased computation at each node.

Splitting a long range communication into a set of shorter range transmissions through intermediate nodes, decreases interference to neighboring streams and improves overall throughput laws of wireless networking. That was demonstrated in [29]: n nodes are randomly placed in a unit area disk, and throughput per node drops as $1/\sqrt{n\ log\ n}$ instead of $1/n$ as one might expect. That result is under the assumption of fixed radius transmission range. Moreover, if nodes are placed carefully on the disk, individual rates can drop even slower with $1/\sqrt{n}$ and the total distance-throughput of the network can scale with $\sqrt{n}$ (in meters times bits per second). This surprising result, is based on perfect scheduling of information routing and non-realistic wireless channel assumptions. Therefore it serves as an upper bound of "best performance" in a wireless network. Subsequent research validated this result and examples of such work can be found in [76], [38]. Subsequent research pointed out that there is always a throughput-delay tradeoff and multihop communication can NOT optimize both (see for example the discussion in [59], [77]). This drawback is important

for delay limited applications, such as voice over IP. In this thesis, we tried to improve *local* spectral efficiency (in terms of bps/Hz) without utilizing several hops and without increasing the end-to-end communication delay.

Spread spectrum communication is employed in [71] and it is shown that when it is combined with local scheduling based on local time synchronization, then the network can sustain considerable data traffic. The argument there is that spread spectrum communication, in contrast to TDMA/FDMA medium access schemes, could *survive* concurrent transmissions up to a level where Signal-to-Noise-and-Interference-Ratio (SNIR) is not severely degraded. In this work we adhere to single frequency networks and we manage to show that improved communication is achieved via smart cooperation among nodes and not via smart processing at each intermediate node. Additionally, we provide concrete time synchronization algorithms, either centralized or decentralized.

We need to mention work on antenna selection diversity in traditional MIMO systems and its performance as it is summarized in [55]. Antenna selection, where high-SNR signals are utilized while lower-SNR signals are discarded, could provide tools and intuition to study antenna sharing among different users in wireless networking. References in [55] provide state-of-art work, in MIMO systems in general.

Finally, we mention work in [81], [58] where schemes like amplify-and-forward are analyzed using ergodic capacity as the performance measure. Practical signal processing schemes that could realize the calculated capacity bounds were not given. Furthermore, the most difficult challenges of network coordination were not addressed while assumptions regarding global knowledge of channel state information were made. Treatment of cooperative diversity as a physical layer scheme only, is very common in the literature. Unfortunately, cooperative diversity involves different users and therefore requires significant work not only at the physical layer, but also at the link layer (layer 2) and at the routing layer (layer 3). Concentrating in one layer only, fails to provide for a practical scheme, overestimates the performance benefits and underestimates associated overheads.

All the above references are a just a small snapshot of the overall interest in the research community, the last four years. Additional relevant work can be found in [51] and its

references. In this thesis, we try to resolve all practical problems and address cooperative diversity as a communication scheme that needs special care in all layers, from routing down to physical layer.

1.2 Basic Assumptions and Thesis Goals

For convenience, we summarize the assumptions throughout this work:

- Simple RF-front ends:
  - Half-duplex radios. Radios cannot receive and transmit simultaneously. Analysis is performed in terms of bps/Hz while full duplex radios usually employ different frequency bands for transmit and receive.
  - No rate adaptation. Radios can not adapt the transmission rate to the wireless channel conditions. That is either because there is no channel state information at the source or because each radio already operates at the minimum transmission speed.
  - No phased arrays (No *beamforming*). Distributed phased arrays require carrier phase stability and control that heavily increases the cost of RF-front ends. Feasibility of distributed[4] beamforming is questionable at frequencies around 2GHz.
- We assume a flat fading, discrete signal model, where a_{sd} denotes the channel coefficient between source s and destinations d while n_d is the noise term at the destination. We further assume (mostly) Rayleigh fading unless otherwise mentioned.

$$y_d = a_{sd}\, x_s + n_d \tag{1.1}$$

[4] *distributed* here refers to antennas that belong to different radio terminals, with obviously different oscillators.

- Neighboring *interfering* streams = noise. The cooperation scheme is evaluated in the presence of noise, that models thermal noise at the receiver, as well as interference from neighboring streams.

- The *average* received signal drops exponentially with distance d between source and destination, with an exponent v ranging between $2-4$.

$$\mathcal{E}[|a_{sd}|^2] \propto 1/d^{\nu} \tag{1.2}$$

- Slow Fading. The wireless channel does not change very often. Analysis is performed in terms of the channel *coherence time* that describes how often the channel changes. Under no CSI at the transmitter, there is no transmission rate that can sustain reliable communication. In other words, the Shannon capacity is zero and reliable communication is a challenging problem.

Thesis goals can be summarized in the following list:

1. Invention of algorithms that react to the physics of the environment: Cooperative nodes in the network adapt their behavior to instantaneous wireless channel realizations. The algorithms ought to a) scale with increasing number of cooperating users and b) require deterministic time to converge to a solution, well before the channel changes (well below the channel *coherence time*). Therefore, the network reacts to the physics of the environment in real time, using measured characteristics of that space.

2. Invention of distributed algorithms with unknown network topology: There is no central point of control that has global knowledge of the network (for example, there is no knowledge on how many nodes cooperate in the network). There is no knowledge regarding the topology of the network or distances to neighboring nodes.

3. Implementation in radio hardware: we tried to provide signal processing techniques, as well as modulation, transmission and coordination techniques that could be applied in existing hardware. Moreover, we will not assume that any transmitter could fix

its transmission radius to a given distance. Communication range is a function of transmission power as well as wireless channel characteristics, which are not user-defined.

The above components differentiate our work from existing approaches in the field, since prior research has focused on a subset of the above elements. In this work, we devise system level solutions that would provide for distributed, infrastructure-free networks where communications are improved with increasing number of cooperating nodes, using simple hardware and intelligent algorithms applicable in practice.

1.3 Thesis Roadmap

In the following chapter, we present our proposal for a practical cooperative diversity scheme and describe its implementation, in a custom, low-cost and embedded wireless network. In chapter 3, we calculate the diversity-multiplexing tradeoff of our scheme and show that our scheme incurs no performance loss, when compared to more involved schemes that require simultaneous transmissions and space-time coding. We calculate outage and ergodic capacity as a function of participating relay nodes and show the performance benefits, in comparison with non-cooperative wireless communication. Power gains, are also discussed. Coordination performance among the relays is quantified, with an analysis that applies for various wireless channel statistics. In chapter 4, we analyze our scheme as an RF scheduling algorithm and show that its power allocation results in superior performance, compared to prior art. In chapter 5, we present network coordination, based on centralized or decentralized time keeping inspired from biological phenomena. We summarize our findings, in chapter 6.

Chapter 2

Opportunistic Relaying

After underlining a few facts about wireless propagation, we describe in more detail *opportunistic relaying* and discuss our implementation in low cost hardware.

2.1 The closer is not always the better...

Imagine inserting relays, literally anywhere near the receiver or transmitter. The goal is to find one relay in a "hot spot" that receives the signal well. If that relay is simultaneously in a hot spot with respect to the ultimate recipient of the information, then this relay can effectively support the communication. The more relays there are, the more likely we can find such intermediate.

Let's start with a simple scenario: in figure 2-1, a transmitter is placed close to a conductive wall and the received electromagnetic field amplitude is calculated at the far field region, approximately one hundred wavelengths away. We assume transmission of a single carrier and we observe that the received amplitude is not constant since there is destructive or constructive addition of the direct and reflected signals. In this simple scenario, there might be locations in space where the field amplitude might be larger than that in locations closer to the transmitter (observe the circled point in figure 2-1). Moving from constructive

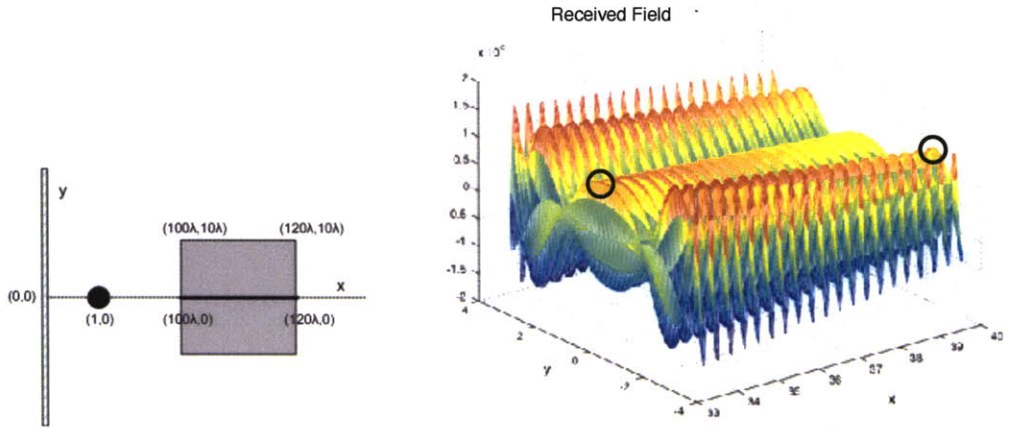

Figure 2-1: LEFT: A transmitter is placed close to a *perfect reflector*, that could be a conductive wall. Assuming no absorption from the wall (perfect reflection), we calculate the electromagnetic field amplitude at specific region, at the far field. RIGHT: The calculated field amplitude as a function of space, for the case depicted in the previous picture. Depending on the phase difference between the direct signal and the signal reflected by the wall, there are locations far away from the transmitter, that have stronger field amplitude than locations closer to the transmitter. Observe, for example the circled points.

to destructive addition of the two rays, involves small physical movements in space, on the order of a quarter of a wavelength. Antenna sharing techniques described in this thesis, exploit in a distributed and decentralized way, cooperating users located at those points where the wireless channel is as "good" as possible. Therefore, the more cooperating users, the higher the probability that we can find one of them in a "hot spot".

In reality, wireless propagation is much more complex than the two-ray model described above. The wireless channel typically involves many reflectors, scatterers and obstructions. It changes at a rate interval (*coherence time*) that depends on wavelength and mobility. A large number of reflectors corresponds to a complex fading channel coefficient (2-dimensional since there are in-phase and quadrature-phase components) with a normal distribution[1]. The amplitude of a circularly symmetric complex Gaussian random variable corresponds to a Rayleigh-distributed random variable and the amplitude squared, corresponds to an exponential random variable.

[1] according to the central limit theorem.

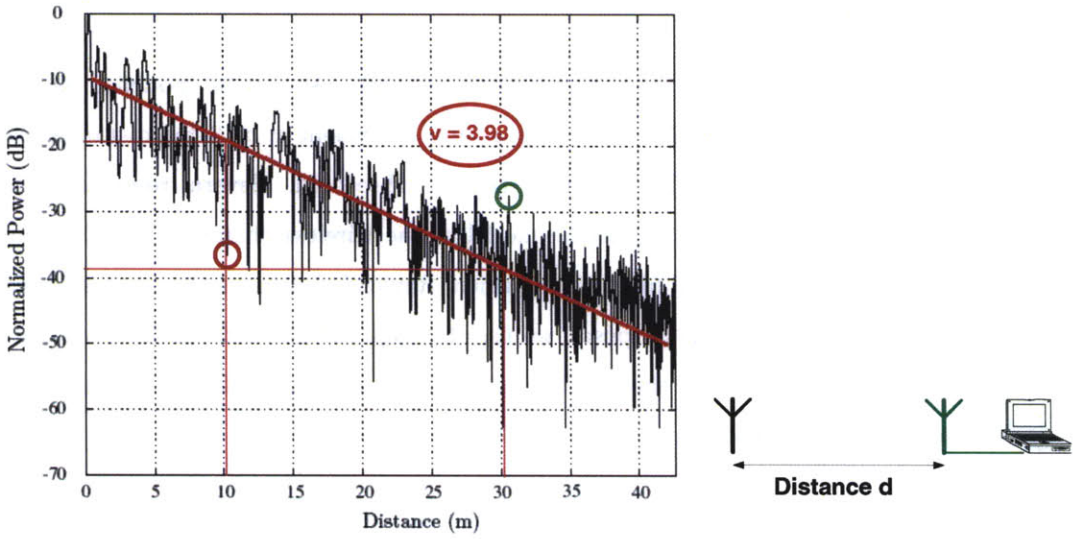

Figure 2-2: Measurement of the received power profile as function of distance at 916 MHz for an indoor environment [26].

If a_{ij} is the (complex) fading coefficient between transmitter i and receiver j, then from fig. 2-2 we can make an estimate of $E[|a_{ij}|^2]$, as a function of distance. We assume that received power $P_r \propto E[|a|^2] \propto 1/d_{ij}^v$, where v is the propagation coefficient and shows how quickly power decreases, as a function of distance. In free space, since electromagnetic field drops as $1/d$, the received power drops as $1/d^2$ and $v = 2$. In practice, there is no free space and v is greater than two, in highly reflective environments or even less than two, when RF propagation is *waveguided*. For example, using the linear markers in figure 2-2, we can estimate v close to 4 ($v = 3.98$).

From figure 2-2, we can also see that an *instantaneous* measurement of the signal strength as a function of distance reveals *hot spots* where received signal is stronger with increased distance from the transmitter (observe for example the markers revealing a factor of 10 stronger signal when distance is doubled). If several measurements are repeated for each distance then the linear marker of figure 2-2 would depict the *average* channel measurement, for that particular environment. Relay cooperative selection ought to exploit such phenomena and concentrate on signal strength rather than distance-based criteria. Consecutively, the focus in this thesis is on instantaneous channel measurements instead of average, for

improved performance.

Cassioli, Win and Molisch [14] have shown that the path loss can be modelled as a two slope function in a log-log scale, with propagation coefficient $v \simeq 2$ for distances close to the transmitter and $v \simeq 7$ for distances above a threshold. Several researchers have suggested Lognormal fading as a realistic model of the wireless channel power loss while others have suggested Nakagami fading from which, Rayleigh fading can be seen as a special case. For the discussion in this proposal, we use Rayleigh fading with various propagation coefficients v, since Rayleigh is the baseline model used in communication research and a good approximation of reality.

It is interesting to note that in free-space, where transmit and receive antennas are placed at different heights, received power drops faster than $1/d^2$ for large d, due to phase difference between direct signal and signal reflected by the ground. Therefore, $v = 4$ is a very realistic assumption for both indoor or outdoor environments.

2.2 Opportunistic Relaying Description

According to opportunistic relaying, a single relay among a set of M relay nodes is selected, depending on which relay provides for the "best" end-to-end path between source and destination (fig. 1-2, 2-3). The wireless channel coefficient a_{si} between source and each relay i, as well as the channel coefficient a_{id} between relay i and destination affect performance. These parameters model the propagation environment between any communicating terminals and change over time, with a rate that macroscopically can be modelled as the *Doppler shift*, inversely proportional to the channel *coherence time*. Opportunistic selection of the "best" available relay involves the discovery of the most appropriate relay, in a distributed and "quick" fashion, well before the channel changes again. We will quantify the speed of relay selection in the following section.

The important point to make here is that under the proposed scheme, the relay nodes monitor the *instantaneous* channel conditions towards source and destination, and decide in

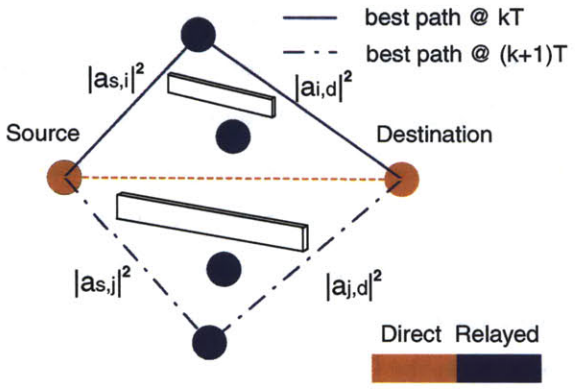

Figure 2-3: Source transmits to destination and neighboring nodes overhear the communication. The "best" relay among M candidates is selected to relay information, via a distributed mechanism and based on instantaneous end-to-end channel conditions. For the diversity-multiplexing tradeoff analysis, transmission of source and "best" relay occur in orthogonal time channels. The scheme could be easily modified to incorporate simultaneous transmissions from source and "best" relay.

a distributed fashion which one has the *strongest* path for information relaying, well before the channel changes again. In that way, topology information at the relays (specifically location coordinates of source and destination at each relay) is not needed. The selection process *reacts* to the physics of wireless propagation, which are in general dependent on several parameters including mobility and distance. By having the network select the relay with the strongest end-to-end path, *macroscopic* features like "distance" are also taken into account. Moreover, the proposed technique is advantageous over techniques that select the best relay *a priori*, based on distance toward source or destination, since distance-dependent relay selection neglects well-understood phenomena in wireless propagation such as *shadowing* or *fading*: communicating transmitter-receiver pairs with similar distances might have enormous differences in terms of received SNRs. Furthermore, average channel conditions might be less appropriate for mobile terminals than static. Selecting the best available path under such conditions (zero topology information, "fast" relay selection well bellow the coherence time of the channel and minimum communication overhead) becomes

non-obvious and it is one of the main contributions of this work.

More specifically, the relays overhear a single transmission of a Ready-to-Send (RTS) packet and a Clear-to-Send (CTS) packet from the destination. From these packets, the relays assess how appropriate each of them is for information relaying. The transmission of RTS from the source allows for the estimation of the instantaneous wireless channel a_{si} between source and relay i, at each relay i (fig. 2-3). Similarly, the transmission of CTS from the destination, allows for the estimation of the instantaneous wireless channel a_{id} between relay i and destination, at each relay i, according to the reciprocity theorem[66][2]. Note that the source does not need to listen to the CTS packet[3] from the destination.

Since communication among all relays should be minimized for reduced overall overhead, a method based on time is selected: as soon as each relay receives the CTS packet, it starts a timer from a parameter h_i based on the instantaneous channel measurements a_{si}, a_{id}. The timer of the relay with the best end-to-end channel conditions will expire first. That relay transmits a short duration *flag* packet, signaling its presence. All relays, while waiting for their timer to reduce to zero (i.e. to expire) are in listening mode. As soon as they hear another relay to flag its presence or forward information (the best relay), they back off.

For the case where all relays can listen to source and destination, but they are "hidden" from each other (i.e. they can not listen to each other), the best relay notifies the destination with a short duration *flag* packet and the destination notifies all relays with a short broadcast message.

The channel coefficients a_{si}, a_{id} at each relay, describe the quality of the wireless path between source-relay-destination, for each relay i. Since the two hops are both important for end-to-end performance, each relay should quantify its appropriateness as an active relay, using a function that involves the link quality of both hops. Two functions are used in this work: under policy I, the minimum of the two is selected (equation (2.1)), while

[2]We assume that the forward and backward channels between the relay and destination are the same from the reciprocity theorem. Note that these transmissions occur on the same frequency band and same coherence interval.

[3]The CTS packet name is motivated by existing MAC protocols. However unlike the existing MAC protocols,the source does not need to receive this packet.

under policy II, the harmonic mean of the two is used (equation (2.2)). Policy I selects the "bottleneck" of the two paths while Policy II balances the two link strengths and it is a smoother version of the first one.

Under policy I:

$$h_i = \min\{|a_{si}|^2, |a_{id}|^2\} \tag{2.1}$$

Under policy II:

$$h_i = \frac{2}{\frac{1}{|a_{si}|^2} + \frac{1}{|a_{id}|^2}} = \frac{2\ |a_{si}|^2\ |a_{id}|^2}{|a_{si}|^2 + |a_{id}|^2} \tag{2.2}$$

The relay i that maximizes function h_i is the one with the "best" end-to-end path between initial source and final destination. After receiving the CTS packet, each relay i will start its own timer with an initial value T_i, inversely proportional to the end-to-end channel quality h_i, according to the following equation:

$$T_i = \frac{\Lambda}{h_i} \tag{2.3}$$

Here Λ is a constant. The units of Λ depend on the units of h_i. Since h_i is a scalar, Λ has the units of time. For the discussion in this work, Λ has simply values of $\mu secs$.

$$h_b = \max\{h_i\}, \iff \tag{2.4}$$

$$T_b = \min\{T_i\},\ i \in [1..M]. \tag{2.5}$$

Therefore, the "best" relay times out first (since it started from a smaller initial value, according to equations (2.3)-(2.5)). This is the relay that participates in forwarding information from the source. The rest of the relays, will overhear the "flag" packet from the best relay (or the destination, in the case of hidden relays) and back off.

After the best relay has been selected, then it can be used to forward information towards the destination. Whether that "best" relay will transmit simultaneously with the source or not, is completely irrelevant to the relay selection process. However, in the diversity-multiplexing tradeoff analysis in the next chapter, we strictly allow only one transmission

at each time and therefore we can view the overall scheme as a two-step transmission: one from source and one from "best" relay, during a subsequent (orthogonal) time channel (fig. 2-3).

A flowchart of the algorithm, as performed at each relay, is given in fig. 2-4.

2.2.1 How well is the selection performed?

The probability of having two or more relay timers expire precisely "at the same time" is zero. However, the probability of having two or more relay timers expire within the same time interval un is non zero and can be explicitly evaluated, given knowledge of the wireless channel statistics.

The only case where opportunistic relay selection fails is when one relay can not detect that another relay is more appropriate for information forwarding. Note that we have already assumed that all relays can listen initial source and destination, otherwise they do not participate in the scheme. We will assume two extreme cases: a) all relays can listen to each other b) all relays are hidden from each other (but they can listen source and destination). In that case, the flag packet sent by the best relay is received from the destination which responds with a short broadcast packet to all relays. Alternatively, other schemes based on "busy tone" (secondary frequency) control channels could be used, requiring no broadcast packet from the destination and partly alleviating the "hidden" relays problem.

In fig. 2-5, collision of two or more relays can happen if the best relay timer T_b and one or more other relay timers expire within $[t_L,\ t_C]$ for the case of no hidden relays (case (a)). This interval depends on the radio switch time from receive to transmit mode d_s and the propagation times needed for signals to travel in the wireless medium. In custom low-cost transceiver hardware, this switch time is typically on the order of a few $\mu secs$ while propagation times for a range of 100 meters is on the order of $1/3$ $\mu secs$. For the case of "hidden" relays the uncertainty interval becomes $[t_L,\ t_H]$ since now the duration of the flag packet should be taken into account, as well as the propagation time towards the destination and back towards the relays and the radio switch time at the destination. The

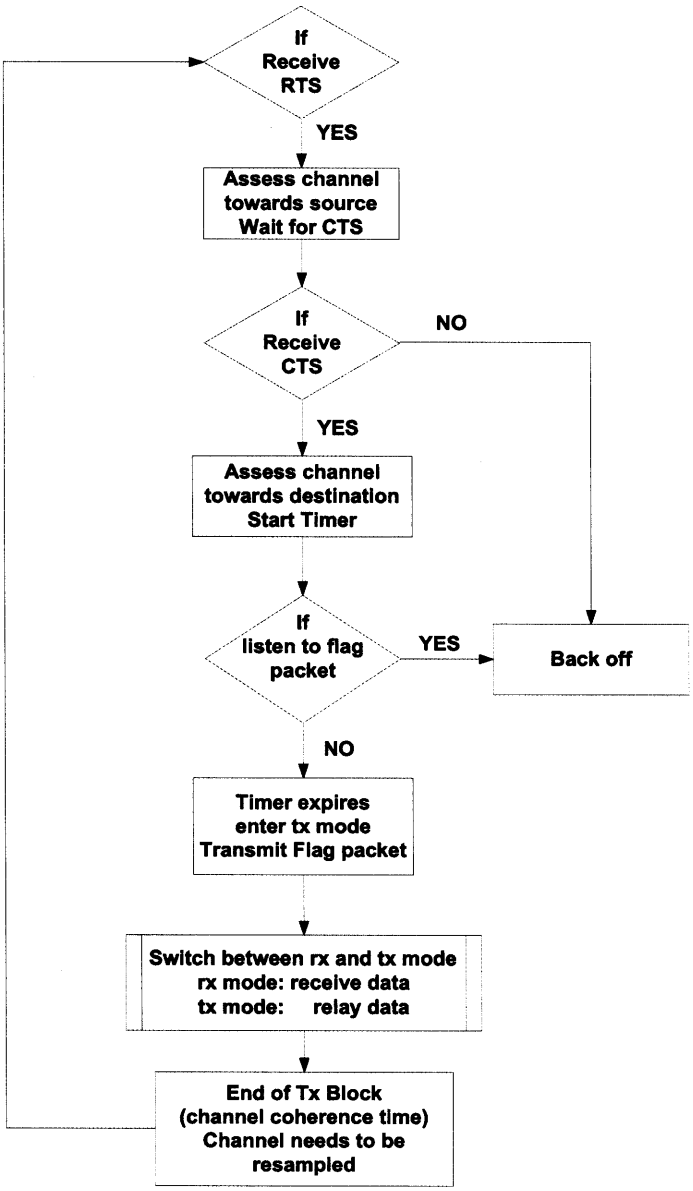

Figure 2-4: Flowchart of the algorithm as performed at each relay.

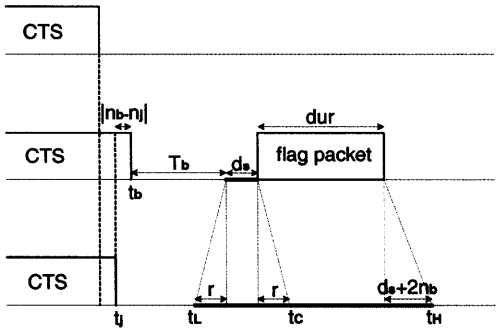

Figure 2-5: The middle row corresponds to the "best" relay. Other relays (top or bottom row) could erroneously be selected as "best" relays, if their timer expired within intervals when they can not hear the best relay transmission. That can happen in the interval $[t_L, t_C]$ for case (a) (No Hidden Relays) or $[t_L, t_H]$ for case (b) (Hidden Relays). t_b, t_j are time points where reception of the CTS packet is completed at best relay b and relay j respectively.

duration of the flag packet can be made small, even one bit transmission could suffice. In any case, the higher this uncertainty interval un, the higher the probability of two or more relay timers to expire within that interval. That's why we will assume maximum values of un, so that we can assess worst case scenario performance.

(a) No Hidden Relays:

$$un = r_{max} + |n_b - n_j|_{max} + d_s \tag{2.6}$$

(b) Hidden Relays:

$$un = r_{max} + |n_b - n_j|_{max} + 2d_s + dur + 2n_{max} \tag{2.7}$$

where:

- n_j: propagation delay between relay j and destination. n_{max} is the maximum.
- r: propagation delay between two relays. r_{max} is the maximum.
- d_s: receive-to-transmit switch time of each radio.
- dur: duration of flag packet, transmitted by the "best" relay.

In any case, the probability of having two or more relays expire within the same interval un, out of a collection of M relays, can be described by the following expression:

$$\begin{aligned} Pr(Collision) &\leq Pr(any\ T_j < T_b + un \mid j \neq b) \\ where\ T_b &= \min\{T_j\},\ j \in [1, M]\ and\ un > 0. \end{aligned} \tag{2.8}$$

Notice that we assume failure of relay selection when two or more relays collide. Traditional CSMA protocols would require the relays to sense that collision, back-off and retry. In that way collision probability could be further reduced, at the expense of increased latency overhead for relay selection. We will analyze the collision probability without any contention resolution protocol and further improvements are left for future work.

In the next chapter we provide an analytic way to calculate a close-form expression of equation (2.8) for any kind of wireless fading statistics. We also discuss how it can be made arbitrarily small.

2.2.2 A note on Time Synchronization

In principle, the RTS/CTS transmissions between source and destination, existent in many Medium Access Control (MAC) protocols, is only needed so that all intermediate relays can assess their connectivity paths towards source and destination. The reception of the CTS packet triggers at each relay the initiation of the timing process, within an uncertainty interval that depends on different propagation times, identified in detail in the previous section. Therefore, an explicit time synchronization protocol among the relays is not required. Explicit time synchronization would be useful between source and destination, only if there was no direct link between them. In that case, the destination could not respond with a CTS to a RTS packet from the source, and therefore source and destination would need to *schedule* their RTS/CTS exchange by other means. In such cases "crude" time synchronization would be useful. Accurate synchronization schemes, server-based [10] or decentralized [11], do exist and have been studied elsewhere. We will assume that source

and destination are in communication range and therefore no synchronization protocols are needed.

2.2.3 A note on Multi-hop extension

It is important to emphasize that since the RTS/CTS exchange is needed only at the relays, the overall scheme can easily be generalized at the case where source and destination are not in communication range. A solution based on time synchronization was described above. Alternatively, another simple protocol modification could be devised: the relays, upon reception of the RTS packet contend for the channel so that one of them could notify the destination that the relays await for a CTS packet. The contention resolution could follow the same timer-based approach. Then the destination responds with a CTS packet. From that point, the algorithm proceeds as described, selecting the relay with the best "end-to-end" path.

2.2.4 A note on Channel State Information (CSI)

CSI at the relays, [in the form of link strengths (not signal phases)], is used at the network layer for "best" relay selection. CSI is not required at the physical layer and is exploited neither at the source nor the relays. The wireless terminals in this work do not exploit CSI for *beamforming* and do not adapt their transmission rate to the wireless channel conditions, either because they are operating in the minimum possible rate or because their hardware does not allow multiple rates. We will emphasize again that no CSI at the physical layer is exploited at the source or the relays, during the diversity-multiplexing tradeoff analysis, in the following chapter.

2.2.5 Comparison with geometric approaches

As can be seen from the above equations, the scheme depends on the instantaneous channel realizations or equivalently, on received *instantaneous* SNRs, at each relay. An alternative

approach would be to have the source know the location of the destination and propagate that information, alongside with its own location information to the relays, using a simple packet that contained that location information. Then, each relay, assuming knowledge of its own location information, could assess its proximity towards source and destination and based on that proximity, contend for the channel with the rest of the relays. That is an idea, proposed by Zorzi and Rao [85] in the context of fading-free wireless networks, when nodes know their location and the location of their destination (for example they are equipped with GPS receivers). Their objective was to study geographical routing and study the average number of hops needed under such schemes. All relays are partitioned into a specific number of geographical regions between source and destination and each relay identifies its region using knowledge of its location and the location of source and destination. Relays at the region closer to the destination contend for the channel first using a standard CSMA splitting scheme. If no relays are found, then relays at the second closest region contend and so on, until all regions are covered, with a typical number of regions close to 4. The latency of the above distance-dependent contention resolution scheme was analyzed in [86].

Zorzi and Rao's scheme of distance-dependent relay selection was employed in the context of Hybrid-ARQ, proposed by Zhao and Valenti [83]. In that work, the request to an Automatic Repeat Request (ARQ) is served by the relay closest to the destination, among those that have decoded the message. In that case, code combining is assumed that exploits the direct and relayed transmission (that's why the term *Hybrid* was used)[4]. Relays are assumed to know their distances to the destination (valid for GPS equipped terminals) or estimate their distances by measuring the expected channel conditions using the ARQ requests from the destination or using other means.

We note that our scheme of opportunistic relaying differs from the above scheme in the following aspects:

- The above scheme performs relay selection based on geographical regions while our

[4]The idea of having a relay terminal respond to an ARQ instead of the original source, was also reported and analyzed in [44] albeit for repetition coding instead of hybrid code combining.

scheme performs selection based on instantaneous channel conditions. In wireless environment, the latter choice could be more suitable as relay nodes located at similar distance to the destination could have vastly different channel gains due to effects such as fading.

- The above scheme requires measurements to be only performed once, if there is no mobility among nodes but requires several rounds of packet exchanges to determine the average SNR. On the other hand opportunistic relaying requires only three packet exchanges in total to determine the instantaneous SNR, but requires that these measurements be repeated in each coherence interval. We show in a subsequent section that the overhead of relay selection is a small fraction of the coherence interval with collision probability less than 0.6%.

- We also note that our protocol is a proactive protocol since it selects the best relay before transmission. The protocol can easily be made to be reactive (similar to [83]) by selecting the relay after the first phase. However this modification would require all relays to listen to the source transmission which can be energy inefficient from a network sense.

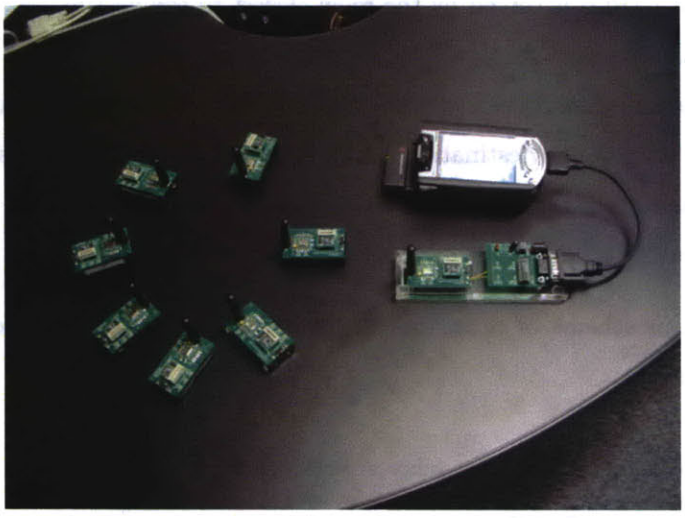

Figure 2-6: Low cost embedded radios at 916 MHz, built for this work.

2.3 Hardware Implementation

Simplicity of the proposed cooperative diversity scheme was a design prerequisite, so that it could be implemented using existing low cost radio hardware. The main problem with current approaches is that they require simultaneous transmissions (at the same frequency band and at the same time). It is well known that electromagnetic waves add in a highly non-linear way, vector-wise, where amplitude, carrier frequency as well as phase are important. In order for simultaneous transmissions to be effective, all the above parameters need to be controlled and adjusted, among the participating radios, distributed in space. Most of the cooperative diversity approaches neglect the above implementation difficulties and focus on a simplified baseband analysis. In that sense, cooperative diversity demonstration had been left as a future exercise.

From the above we can understand that simultaneous transmissions require radio front ends that depart from the conventional norm. Even though such endeavor is not impossible, and research efforts are underway, we choose to devise cooperative diversity protocols that exploit *existing* radios and therefore are cost-effective *today*. We further show in the next chapter that there is no performance loss, from a diversity-multiplexing tradeoff point of view.

We were interested in a portable demonstration and therefore we designed the simplest possible hardware: we used a 916MHz on-off keying radio module from RF Monolithics, with 115kHz bandwidth and part-15 compliance. The module can tranmsit/receive continuous digital waveforms and it is the duty of the design engineer to built the necessary protocol on top of this functionality. We interfaced that module to a low-cost 8051 microcontroller (MCU), driven by a 22.1184MHz crystal oscillator. The mcu/crystal board was designed by J. Lifton, a fellow colleague and friend, in the context of *Pushpin Computing* [50]. We chose the specific MCU since it had a detailed and well-written specification manual. We designed a new printed circuit board (PCB) using Protel, interfacing the pushpin MCU with the radio module and wrote all the necessary software functions for bit/byte/frame/packet transmission, synchronization and reception. We added an RS232 interface so that the

embedded network was connected to PDAs and the rest of the digital world. A picture of the hardware built is given in Fig. 2-6.

In order to demonstrate the benefits of cooperative diversity we created a room size demo. Text information was transmitted from one side of the room to a receiver connected to a large, alphanumeric store display at the other side of the room (Fig. 2-7). Relays at the vicinity of communication would provide for additional reliability, in the presence of people moving in the room. Received information would be presented at the store display, demonstrating that errors would be decreased when opportunistic relaying was used.

2.3.1 Signal Structure

Information was sent periodically, in blocks corresponding to 16 characters of information, since that was the selected message length that could be displayed at the receiver display. The message would scroll from left to right with a duration of 2-3 seconds. Therefore, messages of 16 characters were sent with that period.

Before every message transmission, "best" relay selection would be performed, according to the described algorithm. Then, 16 frames were transmitted from the source, corresponding to the 16 characters of the message. Each frame (out of those 16 frames) were repeated from the best relay, provided that it had been correctly decoded. That is why the signal structure shown in Fig. 2-9 (second row, first picture), has empty slots destined for the transmission of the best relay. Each frame included the necessary synchronization preamble, followed by 4 bytes (32 bits) that included header information (source id, destination id, sequence id), data information as well as a *Cyclic Redundancy Check (CRC)* for error detection purposes. CRC information was required so that the relay could find out whether it had correctly decoded the message. The destination received information from the source as well as information from the best relay and decided about the original message. Even though we could use a *Maximum Ratio Combiner (MRC)*, we chose to further simplify the receiver structure: the receiver decoded both messages and kept the one with the correct message (assertion made with the help of the CRC field).

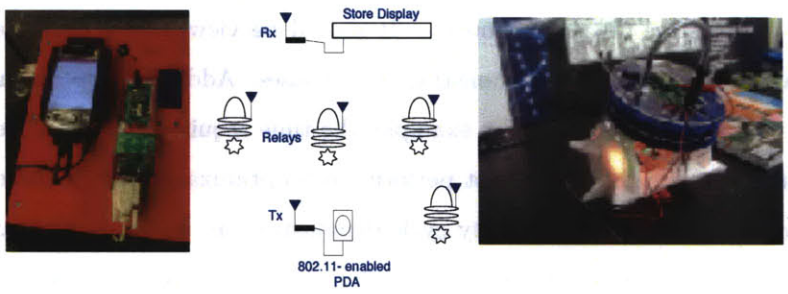

Figure 2-7: Distributed selection of "best" relay path. The intermediate relay nodes overhear the handshaking between Tx and Rx. Based on the method of distributed timers, the relay that has the best signal path from transmitter to relay and relay to receiver is picked with minimal overhead. The receiver combines direct and relayed transmission and displays the received text on a store display. The "best" relay signals with an orange light. The transmitter transmits weather information coming from a 802.11-enabled pda.

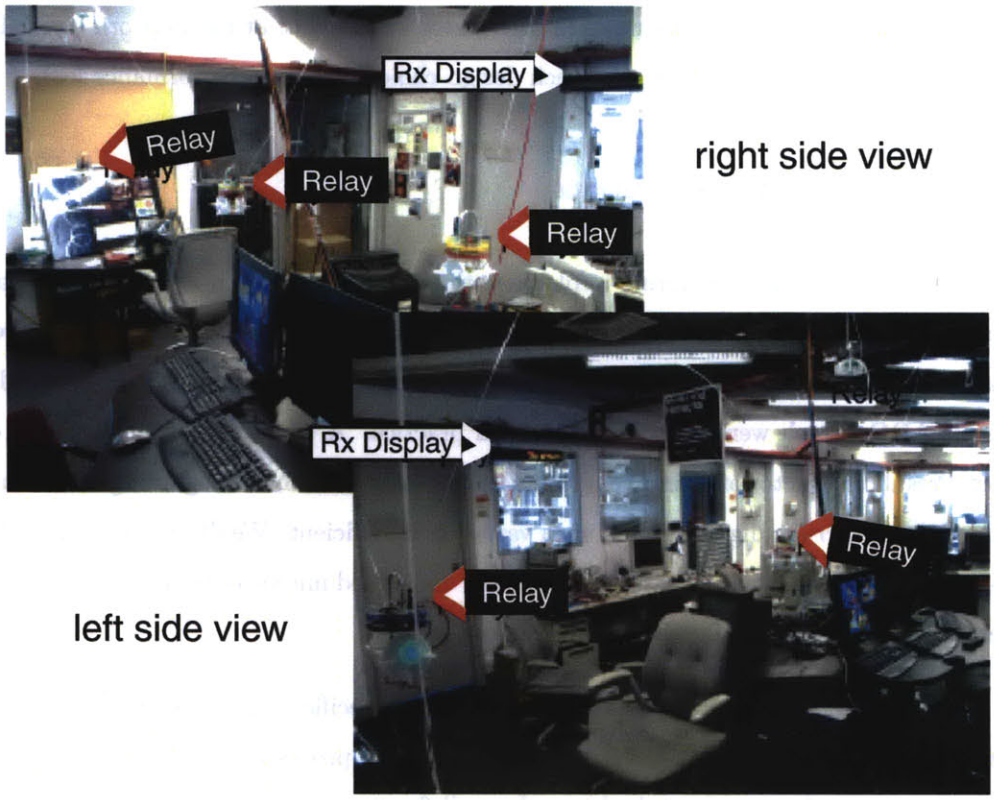

Figure 2-8: Laboratory demonstration. Relays and destination are depicted.

The signal structure used (fig. 2-9) is a specific example of how opportunistic relaying can be used in cooperative diversity contexts. It should be viewed as a concrete example for a specific application, built for demonstration purposes. Additional optimization could be performed if that was necessary. For example, the time required for "best" relay selection, could be further reduced. We did not perform such optimization, since there was no such need in our slow bit-rate and low duty cycle demonstration. In the performance section of the next chapter, such optimization is explored. Additionally, our embedded radios did not have much computation power given the 8-bit processor structure. More complex receiver structures, like a Maximum Ratio Combiner receiver or an advanced error correcting Code Combiner receiver require more powerful computation and could be used if we had selected a more powerful microprocessor for each embedded radio. Note however that increased complexity at each receiver increases the necessary required reception energy [53], having a significant impact on the overall energy budget. We chose to keep the individual nodes as simple as possible and rather create *distributed intelligence* at the network layer. In the following section, we will show that such design choice incurs no performance loss.

2.3.2 Practical Considerations

One of our main concerns during implementation was the limited resolution of the analog-to-digital converter (ADC) at each relay radio, during the evaluation of the signal strength path, towards source and destination. Fortunately, the pushpin micro-controller used, had 12-bit ADCs which were proven adequate in practice. A slight movement in space could easily result to a factor of 10 in strength fluctuation as we saw in the beginning of this chapter and therefore, *crude* digitization of such variation is sufficient. We did not experiment with 4-bit ADCs, which can be easily found in many embedded micro-controllers. Probably, such resolution would be insufficient for our scheme.

A second concern during implementation was about specific channel estimation algorithms at the relays, given the 8-bit architecture of each micro-processor which resulted to limited computation performance: we had to reduce all floating-point calculations in order to improve accuracy and speed. That was the main limitation from a hardware perspective and

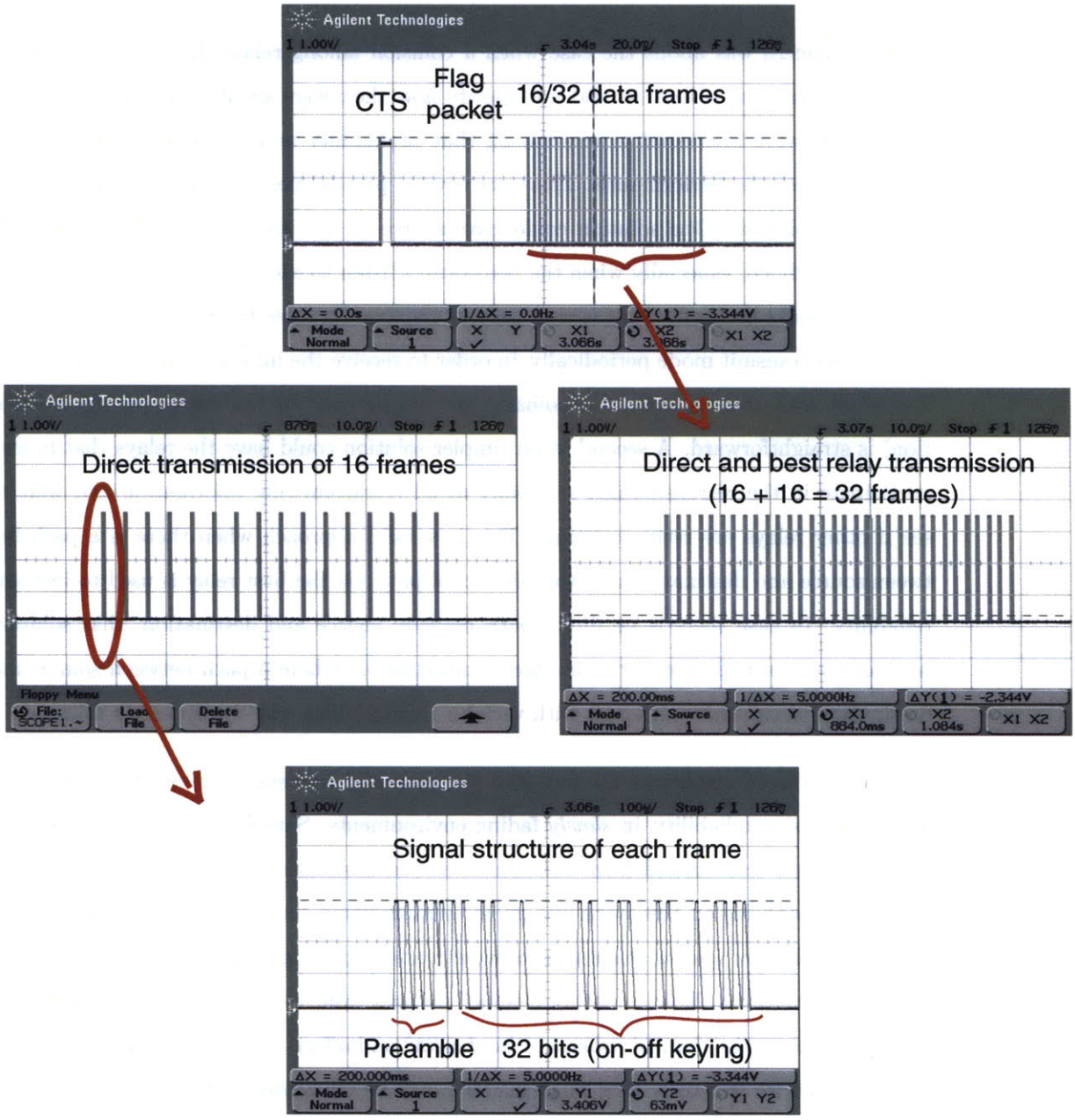

Figure 2-9: Baseband signal structure at the digital output of the receiver . The waveforms are measured at the receiver's radio using a digital oscilloscope. Notice that the time resolution is the same for the two plots at the middle row.

future implementations should facilitate more powerful micro-controllers.

A third concern was about the case when a collision among relays did occur. In the algorithm description above, we did not specify how the relays should act after a collision, since a) we were interested in minimizing that probability and b) we were interested in worst case scenario performance. There are several possibilities on what the relays should do after a collision. One solution could have source or destination notify the relays that they have collided, especially when the relays can't listen to each other: in that case one of the relays could back off. That is very easy to implement, since the relays switch between receive and transmit mode periodically, in order to receive the information from the source. Therefore, a control-bit indicating collision and transmitted by the source (or the destination) is straightforward. A second, even simpler solution could have the relays that indeed participate in the retransmission randomly avoid retransmitting information and wait to see if other relays are retransmitting. This is a valid approach when there is a path between source and destination and the additional path via the best relay is used to increase reliability. In case there is no direct path between source and destination, that solution is clearly suboptimal. For our room-size demonstration, where a path between source and destination is available (although with variable quality) that was the approach followed.

Finally, we should emphasize the fact that the proposed cooperative diversity technique is about increasing reliability, in *slowly* fading environments. Sampling of space, in the form of pilot signals transmitted from the source or the destination, needs to be periodically repeated. Emphasis on this work was given in minimizing the overhead time required for *best* relay selection and no assumptions were made regarding *smoothness* of the wireless channel fluctuations. How smoothly channel changes from *coherence time* interval to the next, depends on the channel *coherence bandwidth*[5] and affects wireless channel estimation. In section 3.4 we will show that relay selection can be efficiently performed within a time interval, two to three orders of magnitude smaller than the channel coherence time. Therefore, sampling of space can be repeated with small overhead more often, within intervals smaller than the coherence interval. Such sampling would be sufficient even at cases where

[5]inversely proportional to the time delay spread of the transmitted signal.

the wireless channel fluctuated in a discontinuous and abrupt fashion.

Chapter 3

Performance

In this chapter, we quantify the performance of opportunistic relaying. Using tools from multiple antenna theory, we show that opportunistic relaying at the network layer, is as efficient as the most complex space-time coding algorithms at the physical layer, from a diversity-multiplexing gain point of view. Specifically, in section 3.1 we use the elegant tool of diversity-multiplexing gain tradeoff, popularized in [84] and show that the balance between reliability and communication speed of a single, "best" relay, is as good as having all relays transmit at the same frequency/time channels.

This rather surprising finding suggests that opportunistic relaying is a simple way to implement cooperative diversity schemes, without performance loss. In section 3.2 we derive the outage capacity of opportunistic relaying and show the increase of spectral efficiency (in bps/Hz), especially at the low SNR regime, suggesting that opportunistic cooperation could be used for faster communication. Power efficiency is also discussed, given the fact that opportunistic relaying does not require all relays to listen and therefore does not have reception energy deficiencies, proportional to the number of relays. Additionally, important transmission energy savings can be realized, with gains quantified further in section 3.3. Finally, the required overhead for best relay selection, is shown in section 3.4 to be reasonably small, in slow fading environments.

3.1 Diversity-Multiplexing Tradeoff

We now consider the impact of opportunistic relaying on the cooperative diversity scenario. The main result of this section is that opportunistic relaying can be used to simplify a number of cooperative diversity protocols involving multiple relays. In particular we focus on the cooperative diversity protocol in [43] which requires the relays to use a space-time code while simultaneously transmitting towards the destination. We show that this protocol can be simplified considerably by simply selecting the best relay in the second stage. Perhaps surprisingly, this simplified protocol achieves the same diversity multiplexing tradeoff achieved in [43]. Furthermore, it does not matter whether the relay implements an amplify and forward or a decode and forward protocol in terms of the diversity-multiplexing tradeoff. We also note that opportunistic relaying can be used to simplify the non-orthogonal relaying protocols proposed in [2]. However the detailed performance analysis is left for future work.

3.1.1 Channel Model

We [1] consider an i.i.d slow Rayleigh fading channel model following [44]. A half duplex constraint is imposed across each relay node, i.e. it cannot transmit and listen simultaneously. We assume that the nodes (transmitter and relays) do not exploit the knowledge of the channel at the physical layer. Note that in the process of discovering the best relay described in the previous section the nodes do learn about their channel gains to the destination. However, we assume that this knowledge of channel gain is limited to the network layer protocol. The knowledge of channel gain is not exploited at the physical layer in order to adjust the code rate based on instantaneous channel measurements. In practice, the hardware at the physical layer could be quite constrained to allow for this flexibility to change the rate on the fly. It could also be that the transmitter is operating at the minimum transmission rate allowed by the radio hardware. Throughout this section, we assume that

[1] For clarity and completeness, we repeat the main assumptions presented in the introduction.

the channel knowledge is not exploited at the physical layer at either the transmitter or the relays.

If the discrete time received signal at the destination and the relay node are denoted by $Y[n]$ and $Y_1[n]$ respectively, then:

$$Y[n] = a_{sd}X[n] + Z[n],\ n = 1, 2 \ldots \frac{T}{2} \text{ (source transmits destination receives)} \quad (3.1)$$

$$Y[n] = a_{rd}X_1[n] + Z[n], n = \frac{T}{2}, \frac{T}{2} + 1 \ldots, T \text{(best relay transmits dest.receives)} \quad (3.2)$$

$$Y_1[n] = a_{sr}X[n] + Z_1[n]\ n = 1, 2 \ldots \frac{T}{2} \text{ (source transmits best relay receives)} \quad (3.3)$$

Here a_{sd}, a_{rd}, a_{sr} are the respective channel gains from the source to destination, best relay to destination and source to the best relay respectively. The channel gains between any two pair of nodes are i.i.d $\mathcal{N}(0,1)$[2]. The noise $Z[n]$ and $Z_1[n]$ at the destination and relay are both assumed to be i.i.d circularly symmetric complex Gaussian $\mathcal{N}(0, \sigma^2)$. $X[n]$ and $X_1[n]$ are the transmitted symbols at the transmitter and relay respectively. T denotes the duration of time-slots reserved for each message and we assume that the source and the relay each transmit orthogonally on half of the time-slots. We impose a power constraint at both the source and the relay: $E[|X[n]|^2] \leq P$ and $E[|X_1[n]|^2] \leq P$. For simplicity, we assume that both the source and the relay to have the same power constraint. We will define $\rho \triangleq P/\sigma^2$ to be the effective signal to noise ratio (SNR). This setting can be easily generalized when the power at the source and relays is different.

The following notation is necessary in the subsequent sections of the paper. This notation is along the lines of [2] and simplifies the exposition.

Definition 1 *A function $f(\rho)$ is said to be exponentially equal to b, denoted by $f(\rho) \doteq \rho^b$, if*

$$\lim_{\rho \to \infty} \frac{\log f(\rho)}{\log \rho} = b. \quad (3.4)$$

[2]The channel gains from the best relay to destination and source to best relay are not $\mathcal{N}(0,1)$. See Lemma 3 in the Appendix.

We can define the relation $\dot{\leq}$ *in a similar fashion.*

Definition 2 *The exponential order of a random variable* X *with a non-negative support is given by,*

$$V = -\lim_{\rho \to \infty} \frac{\log X}{\log \rho}. \tag{3.5}$$

The exponential order greatly simplifies the analysis of outage events while deriving the diversity multiplexing tradeoff. Some properties of the exponential order are derived in Appendix 6, lemma 2.

Definition 3 *(***Diversity-Multiplexing Tradeoff***) We use the definition given in [84]. Consider a family of codes* C_ρ *operating at SNR* $\equiv \rho$ *and having rates* $R(\rho)$ *bits per channel use. If* $P_e(R)$ *is the outage probability (see [75]) of the channel for rate* R*, then the multiplexing gain* r *and diversity order* d *are defined as*[3]

$$r \triangleq \lim_{\rho \to \infty} \frac{R(\rho)}{\log \rho} \qquad\qquad d \triangleq -\lim_{\rho \to \infty} \frac{\log P_e(R)}{\log \rho} \tag{3.6}$$

What remains to be specified is a policy for selecting the best relay. We essentially use the policy 1 (equation (2.1)) in the previous section.

Policy 1 *Among all the available relays, denote the relay with the largest value of* $\min\{|a_{sr}|^2, |a_{rd}|^2\}$ *as the best relay.*

To justify this choice, we note that relay selection under policy I is more efficient than policy II, as it will be shown in section 3.4. Furthermore, we will see in this section that this choice is optimum in that it enables opportunistic relaying to achieve the same diversity multiplexing tradeoff of more complex orthogonal relaying schemes in [43]. We next discuss the performance of the amplify and forward and decode and forward protocols.

[3] *We will assume that the block length of the code is large enough, so that the detection error is arbitrarily small and the main error event is due to outage.*

3.1.2 Digital Relaying - Decode and Forward Protocol

We will first study the case where the intermediate relays have the ability to decode the received signal, re-encode and transmit it to the destination. We will study the protocol proposed in [43] and show that it can be considerably simplified through opportunistic relaying.

The decode and forward algorithm considered in [43] is briefly summarized as follows. In the first half time-slots, the source transmits and all the relays and receiver nodes listen to this transmission. Thereafter, *all* the relays that are successful in decoding the message, re-encode the message using a distributed space-time protocol and collaboratively transmit it to the destination. The destination decodes the message at the end of the second time-slot. Note that the source does not transmit in the second half time-slots. The main result for the decode and forward protocol is given in the following theorem :

Theorem 1 ([43]) *The achievable diversity multiplexing tradeoff for the decode and forward strategy with M intermediate relay nodes is given by $d(r) = (M+1)(1-2r)$ for $r \in (0, 0.5)$.*

The following Theorem shows that opportunistic relaying achieves the same diversity-multiplexing tradeoff if the best relay selected according to policy 1.

Theorem 2 *Under opportunistic relaying, the decode and forward protocol with M intermediate relays achieves the same diversity multiplexing tradeoff stated in Theorem 1.*

Proof 1 *We follow along the lines of [43]. Let $\mathcal{E}$ denote the event that the relay is successful in decoding the message at the end of the first half of transmission and $\bar{\mathcal{E}}$ denote the event that the relay is not successful in decoding the message. Event $\bar{\mathcal{E}}$ happens when the mutual information between source and best relay drops below the code rate. Suppose that we select a code with rate $R = r \log \rho$ and let $I(X;Y)$ denote the mutual information between the*

source and the destination. The probability of outage is given by

$$\begin{aligned}
P_e &= \Pr(I(X;Y) \leq r\log\rho|\mathcal{E})\Pr(\mathcal{E}) + \Pr(I(X;Y) \leq r\log\rho|\bar{\mathcal{E}})\Pr(\bar{\mathcal{E}}) \\
&= \Pr\left(\frac{1}{2}\log(1+\rho(|a_{sd}|^2+|a_{rd}|^2)) \leq r\log\rho\right)\Pr(\mathcal{E}) + \\
&\quad \Pr\left(\frac{1}{2}\log(1+\rho|a_{sd}|^2) \leq r\log\rho\right)\Pr(\bar{\mathcal{E}}) \\
&\leq \Pr\left(\frac{1}{2}\log(1+\rho(|a_{sd}|^2+|a_{rd}|^2)) \leq r\log\rho\right) + \\
&\quad \Pr\left(\frac{1}{2}\log(1+\rho|a_{sd}|^2) \leq r\log\rho\right)\Pr\left(\frac{1}{2}\log(1+\rho|a_{sr}|^2) \leq r\log\rho\right) \\
&\leq \Pr\left(|a_{sd}|^2+|a_{rd}|^2 \leq \rho^{2r-1}\right) + \Pr\left(|a_{sd}|^2 \leq \rho^{2r-1}\right)\Pr\left(|a_{sr}|^2 \leq \rho^{2r-1}\right) \\
&\leq \Pr\left(|a_{sd}|^2 \leq \rho^{2r-1}\right)\Pr\left(|a_{rd}|^2 \leq \rho^{2r-1}\right) + \Pr\left(|a_{sd}|^2 \leq \rho^{2r-1}\right)\Pr\left(|a_{sr}|^2 \leq \rho^{2r-1}\right) \\
&\dot{\leq} \rho^{2r-1}\rho^{M(2r-1)} + \rho^{2r-1}\rho^{M(2r-1)} \doteq \rho^{(M+1)(2r-1)}
\end{aligned}$$

In the last step we have used claim 2 of Lemma 3 in the appendix with $m = M$.

We next study the performance under analog relaying and then mention several remarks.

3.1.3 Analog relaying - Basic Amplify and Forward

We[4] will now consider the case where the intermediate relays are not able to decode the message, but can only scale their received transmission (due to the power constraint) and send it to the destination.

The basic amplify and forward protocol was studied in [44] for the case of a single relay. The source broadcasts the message for first half time-slots. In the second half time-slots the relay simply amplifies the signals it received in the first half time-slots. Thus the destination receives two copies of each symbol. One directly from the source and the other via the relay. At the end of the transmission, the destination then combines the two copies

[4]I am grateful to my friend and colleague Ashish Khisti for his help in the derivation of this section. Without his help, the proof would be incomplete.

of each symbol through a matched filter. Assuming i.i.d Gaussian codebook, the mutual information between the source and the destination can be shown to be [44],

$$I(X;Y) = \frac{1}{2}\log\left(1+\rho|a_{sd}|^2 + f(\rho|a_{sr}|^2, \rho|a_{rd}|^2)\right) \quad (3.7)$$

$$f(a,b) = \frac{ab}{a+b+1} \quad (3.8)$$

The amplify and forward strategy does not generalize in the same manner as the decode and forward strategy for the case of multiple relays. We do not gain by having all the relay nodes amplify in the second half of the time-slot. This is because at the destination we do not receive a coherent summation of the channel gains from the different receivers. If γ_j is the scaling constant of receiver j, then the received signal will be given by $y[n] = \left(\sum_{j=1}^{M}\gamma_j a_{rd}^j\right)x[n] + z[n]$. Since this is simply a linear summation of Gaussian random variables, we do not see the diversity gain from the relays. A possible alternative is to have the M relays amplify in a round-robin fashion. Each relay transmits only one out of every M symbols in a round robin fashion. This strategy has been proposed in [43], but the achievable diversity-multiplexing tradeoff is not analyzed.

Opportunistic relaying on the other hand provides another possible solution to analog relaying. Only the best relay (according to policy 1) is selected for transmission. The following theorem shows that opportunistic relaying achieves the same diversity multiplexing tradeoff as that achieved by the (more complicated) decode and forward scheme.

Theorem 3 *Opportunistic amplify and forward achieves the same diversity multiplexing tradeoff stated in Theorem 1.*

Proof 2 *We begin with the expression for mutual information between the source and destination (3.7). An outage occurs if this mutual information is less than the code rate* $r\log\rho$.

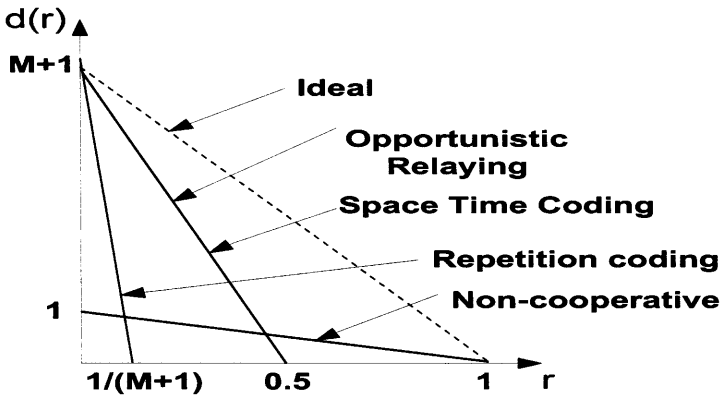

Figure 3-1: The diversity-multiplexing of opportunistic relaying is exactly the same with that of more complex space-time coded protocols.

Thus we have that

$$\begin{aligned}
P_e &= \Pr\left(I(X;Y) \leq r\log\rho\right) \\
&= \Pr\left(\log(1+\rho|a_{sd}|^2 + f(\rho|a_{sr}|^2, \rho|a_{rd}|^2) \leq 2r\log\rho\right) \\
&\leq \Pr\left(|a_{sd}|^2 \leq \rho^{2r-1}, f(\rho|a_{sr}|^2, \rho|a_{rd}|^2) \leq \rho^{2r}\right) \\
&\overset{(a)}{\leq} \Pr\left(|a_{sd}|^2 \leq \rho^{2r-1}, \min\left(|a_{sr}|^2, |a_{rd}|^2\right) \leq \rho^{2r-1} + \rho^{r-1}\sqrt{1+\rho^{2r}}\right) \\
&\overset{(b)}{\doteq} \rho^{2r-1}\rho^{M(2r-1)} = \rho^{(M+1)(2r-1)}
\end{aligned}$$

Here (a) follows from Lemma 4 and (b) follows from Lemma 3, claim 1 in appendix 6 and the fact that $\rho^{r-1}\sqrt{1+\rho^{2r}} \to \rho^{2r-1}$ *as* $\rho \to \infty$.

3.1.4 Discussion

Space-time Coding vs. Relaying Solutions

The (conventional) cooperative diversity setup (e.g. [43]) assumes that the cooperating relays use a distributed space-time code to achieve the diversity multiplexing tradeoff in Theorem 1. Development of practical space-time codes is an active area of research. Re-

cently there has been considerable progress towards developing practical codes that achieve the diversity multiplexing tradeoff over MIMO channels. In particular, it is known that random lattice based codes (LAST) can achieve the entire diversity multiplexing tradeoff over MIMO channels [21]. Moreover, it is noted in [58] that, under certain conditions, the analytical criterion such as rank and determinant criterion for MIMO links also carry over to cooperative diversity systems [5]. However some practical challenges will have to be addressed to use these codes in the distributed antenna setting: (a) The codes for MIMO channels assume a fixed number of transmit and receive antennas. In cooperative diversity, the number of antennas depends on which relays are successful in decoding and hence is a variable quantity. (b) The destination must be informed either explicitly or implicitly which relays are transmitting.

Opportunistic relaying provides an alternative solution to space time codes for cooperative diversity by using a clever relaying protocol. The result of Theorem 3 suggests that there is no loss in diversity multiplexing tradeoff[6] if a simple analog relaying based scheme is used in conjunction with opportunistic relaying. Even if the intermediate relays are digital, a very simple decode and forward scheme that eliminates the need for space-time codes can be implemented. The relay listens and decodes the message in the first half of the time-slots and repeats the source transmission in the second half of the time-slots when the source is not transmitting. The receiver simply does a maximal ratio combining of the source and relay transmissions and attempts to decode the message. Theorem 2 asserts once again, that the combination of this simple physical layer scheme and the smart choice of the relay is essentially optimal.

The diversity-multiplexing tradeoff is plotted in fig. 3-1. Even though a single terminal with the "best" end-to-end channel conditions relays the information, the diversity order in the high SNR regime is on the order of the number $M+1$ of all participating terminals. Moreover, the tradeoff is exactly the same as when space-time coding across M relays is used.

[5] However it is assumed that the destination knows the channel gain between source and relay for Amplify and Forward.

[6] compared to the orthogonal transmission protocols in [43]

Non-orthogonal Cooperative Diversity Schemes

The focus in this work is on the multiple relay cooperative diversity protocols proposed in [43], since they require that the transmitter and relay operate in orthogonal time-slots in addition to the half duplex constraints. The orthogonality assumption was amenable to practical implementation (section 2.3), since the decoder is extremely simple. More recently, a new class of protocols that relax the assumption that the transmitter and relay operate in orthogonal time-slots, (but still assume the half duplex constraint) have been proposed in [2]. These protocols have a superior performance compared to [43], albeit at the cost of higher complexity both at the decoder and network layer. Opportunistic relaying could be naturally used to simplify those protocols [7] and details of such simplifications and its performance are underway.

Impact of Topology

The analysis for diversity-multiplexing tradeoff was presented assuming that average channel gains between each pair of nodes is unity. In other words the impact of topology was not considered. We observe that the effect of topology can be included in the analysis using techniques used in [44]. In the high SNR regime, we expect fixed multiplicative factors of path loss to contribute little in affecting the diversity-multiplexing tradeoff. However topology is certainly important for finite SNR case as observed in [37].

3.2 Outage Capacity

Slow fading environments, where the channel remains the same for several transmission blocks, is the typical case for many wireless applications, including static transceivers as well as "slowly" mobile terminals (for example, walking users of cellular telephony is a typical case of slow fading). At the absence of CSI at the transmitter, there can be no guarantee

[7] An Alamouti[1] type code could be used if the relay and source are simultaneously transmitting.

for reliable communication and in information theoretic terms, the Shannon capacity is zero. That doesn't mean that wireless communication is impossible. It rather emphasizes that there is no rate of wireless transmission, for which, reception can be achieved with arbitrarily small probability of error. In such settings cooperative diversity can provide an attractive solution. On the other hand, in *fast* fading (also known as *ergodic*) environments, the Shannon capacity is non-zero and in that case the transmission/reception scheme ought to exploit the wireless channel fluctuations.

From the above, it is clear that slow fading is the most difficult case of wireless communication, since unreliable communication is due not only because of noise at the receiver (*error probability*) but also and more importantly, because of the wireless channel fluctuations that occur for extended periods of time and reduce the information rate that the wireless link can sustain. Such event is typically called *outage event* and for point-to-point communication can be mathematically described by the following relationship:

$$\begin{aligned} W\ log(1+|a_{sd}|^2\ P/N_o) &\leq R \Leftrightarrow \\ |a_{sd}|^2 &\leq (2^{R/W}-1)/(P/N_o) \Leftrightarrow \\ |a_{sd}|^2 &\leq (2^{\rho}-1)/SNR \Leftrightarrow \\ \gamma_{sd} &\leq \Theta \end{aligned} \tag{3.9}$$

In short, the wireless channel conditions as described by the magnitude of the channel coefficient $|a_{sd}|^2 \equiv \gamma_{sd}$, correspond to a received SNR that cannot sustain the desired rate R (in bps) and spectral efficiency $\rho = R/W$ (in bps/Hz). The probability of the outage event $Pr(\gamma_{sd} \leq \Theta)$ could be reduced by increasing the transmission power or decreasing the desired rate or spectral efficiency. In other words, if we use more power or slower communication, then probability of failure decreases.

In this section we show that opportunistic relaying is an efficient way to combat fading without sacrificing precious communication resources. We will show with a concrete example that opportunistic relaying for fixed transmission power and bandwidth can increase the outage rates (the spectral efficiency in bps/Hz for a given outage probability) by exploiting

several users as wireless channel sensors and selecting the most appropriate one for relaying purposes. The importance of opportunistic relaying will be emphasized even at cases where simple multihop communication could not provide for increased received SNRs.

As described in the previous chapter, opportunistic relaying selects the "best" relay b that maximizes a function of wireless channel conditions towards source ($\gamma_{si} \equiv |a_{si}|^2$) and destination ($\gamma_{id} \equiv |a_{id}|^2$):

$$b = \underbrace{arg}_{i} \; max\{min\{\gamma_{si}, \gamma_{id}\}\}, \quad i \in [1..M] \tag{3.10}$$

The communication through the "best" opportunistic relay fails due to outage when the following even happens:

$$P_r(\gamma_{sb} < \Theta_2 \bigcup \gamma_{bd} < \Theta_2) \tag{3.11}$$

Θ_2 is given in the following equation:

$$\Theta_2 = 2\ (2^{2\rho} - 1)/SNR \tag{3.12}$$

Notice that opportunistic relaying has been defined as a 2-step scheme: at the first step the source transmits and at the second step the best relay relays. In order to compare it in a fair way with direct communication, we need to fix the total transmission power P. We choose to allocate half of the power to direct communication (at the original source P_s) and half of the power to the best relay transmission (P_b): $P_s = P_b = P/2$. This is why Θ_2 has a factor of 2 at the beginning, compared to Θ in equation 3.9. Since communication happens in two steps using half-duplex, same frequency radios, the required spectral efficiency is now 2ρ, so that the communication application at the receiver receives information with end-to-end spectral efficiency ρ. This is why, Θ_2 has exponent 2ρ when compared with direct communication, in equation 3.9.

Equation 3.11 simply states that opportunistic relaying fails if either of the two hops (from source to best relay and from best relay to destination) fail. This probability can be ana-

lytically calculated for the case of Rayleigh fading:

$$\delta = P_r(\gamma_{sb} < \Theta_2 \bigcup \gamma_{bd} < \Theta_2) \tag{3.13}$$

$$\equiv P_r(min\{\gamma_{sb}, \gamma_{bd}\} < \Theta_2) \tag{3.14}$$

$$\overset{3.10}{=} P_r(\underbrace{max}_{i}\{\ min\{\gamma_{si}, \gamma_{id}\}\} < \Theta_2), \quad i \in [1..M] \tag{3.15}$$

$$\overset{(*)}{=} P_r(\underbrace{max}_{i}\{\gamma_{sid}\} < \Theta_2), \quad i \in [1..M] \tag{3.16}$$

$$= \prod_{i=1}^{M} P_r(\gamma_{sid} < \Theta_2) \tag{3.17}$$

$$= \prod_{i=1}^{M}(1 - exp(-\Theta_2/\overline{\gamma}_{sid})) \tag{3.18}$$

where we have exploited in (*) the fact that the minimum of two independent exponentials is again an exponential random variable, with parameter the sum of the two parameters:

$$\frac{1}{\overline{\gamma}_{sid}} = \frac{1}{\overline{\gamma}_{si}} + \frac{1}{\overline{\gamma}_{id}} \tag{3.19}$$

Equation 3.18 provides the outage probability of opportunistic relaying: relaying information through the "best" possible relay. Such calculation is pessimistic in the sense that it neglects the direct transmission between source and destination. Incorporating the direct transmission, further reduces the end-to-end outage probability:

$$P_r^{out} = (1 - exp(-\Theta_2/\overline{\gamma}_{sd})) \prod_{i=1}^{M}(1 - exp(-\Theta_2/\overline{\gamma}_{sid})) \tag{3.20}$$

3.2.1 Numerical Examples

From equation 3.18, we can calculate the spectral efficiency ρ for a given outage probability δ. We will study two cases:

- (a) Symmetric case where all relays are equidistant to source and destination, with distance equal to source-destination distance: $d_{sd} = d_{si} = d_{id}, \forall i \in [1..M]$. Therefore, there is no *multihop gain* by choosing to communicate to a nearby intermediate relay

node, towards the final destination.

- (b) Multihop case where all relays are half-way between source and destination: $d_{sd} = d_{si} + d_{id}, d_{si} = d_{id}, \forall i \in [1..M]$. In that case, choosing to communicate to nearby nodes towards the destination potentially could offer the advantage of increased average received SNR due to shorter distance (*multihop gain*). Two cases of path loss exponent are presented: $\overline{\gamma}_{ij} \propto 1/d_{ij}^{v}$ for $v = 3, 4$.

It is straightforward from equation 3.19 that $\overline{\gamma}_{sid} = \overline{\gamma}_{sd}/2$ for case (a) and $\overline{\gamma}_{sid} = 2^{v-1}\ \overline{\gamma}_{sd}$ for case (b). Normalizing $\overline{\gamma}_{sd} = 1$ and using equations 3.9 and 3.18, we can plot the spectral efficiency of opportunistic relaying, as a function of number of cooperating relays, for the two cases considered, having in mind that i) the total transmission power is held fixed (we do not input tx power to the system by inserting additional relays) and ii) the spectral efficiency plotted is slightly smaller than the actual, given the fact that equation 3.18 does not include the direct communication path, between source and destination. Fig. 3-2 plots the above scaling for outage probability $\delta = 1\%$ and SNR=10.

$$\rho_{opport} = \frac{1}{2}\ \log_2(1 - ln(1 - \delta^{1/M})\ \frac{SNR}{2}\ \overline{\gamma}_{sid}) \tag{3.21}$$

$$\rho_{direct} = \log_2(1 - \ ln(1 - \delta)\ SNR\ \overline{\gamma}_{sid}) \tag{3.22}$$

Fig. 3-2 shows that opportunistic relaying increases the outage capacity compared to direct communication, even for the symmetric case (a) of equidistant relays, where there is no multihop gain. We emphasize the fact that the comparison assumes same (fixed) total transmission power: by adding more relays into the network, we do not add power/energy into the system. The increased rates are due to the smart relay selection algorithm that facilitates intelligence at the network layer. Notice that a single relay does not increase the overall capacity for case (a) or for case (b), when $v \leq 3$. The latter result is in coherence with previous reported results, suggesting that a single relay (or a three terminal cooperative network), could not increase the capacity of wireless communication, when CSI is not exploited at the transmitters [44].

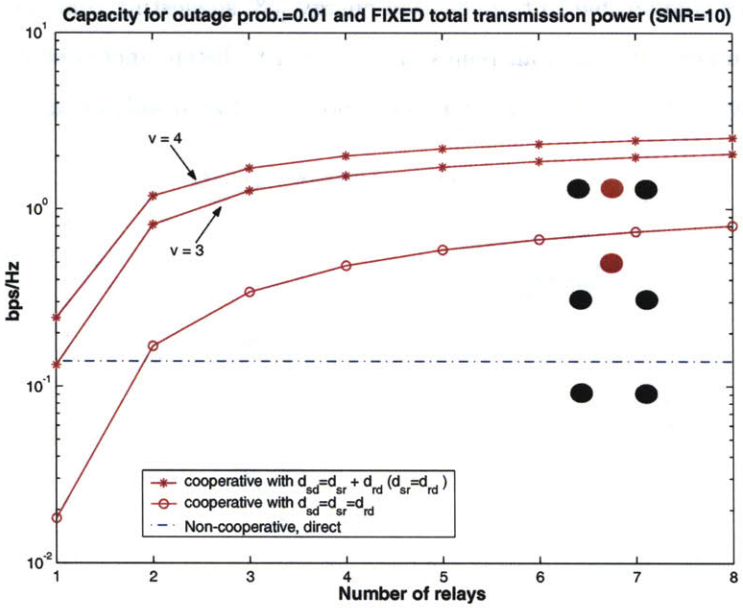

Figure 3-2: Under a total *tx* power constraint, the practical scheme of opportunistic relaying increases the outage capacity, compared to direct communication. Selecting the appropriate path at the RF level exploits users as an additional degree of freedom, apart from power and rate. Two topologies are used as an example: the first corresponds to the symmetric case of all relays equidistant to source and destination. The second topology corresponds to relays half distance between source and destination, for path loss exponent v=3, v=4.

For different values of SNR, the outage rates are plotted in fig. 3-3. Notice the surprising gains of opportunistic relaying at the low SNR regime, compared to direct communication. Those plots suggest that cooperation in the form of opportunistic relaying could be translated to substantial energy gains: reliability can be achieved with additional relays that participate in the relay selection but could "go to sleep mode" after the relay selection, since only the best relay participates in forwarding the information. Moreover, best relay selection can be performed in a small fraction of the coherence time of the channel, leaving the rest for information forwarding, as we will show in a subsequent section. This approach is in sharp contrast to existing proposals in the field that require all relays to remain in listening mode until all information is eventually transmitted. Reception energy is not negligible and especially in communication schemes where error correction is used, reception

energy becomes comparable to transmission energy [53] suggesting that the energy cost of reliable communication (when all relays are required to listen) increases linearly with the number of relays. Opportunistic relaying does not have this disadvantage.

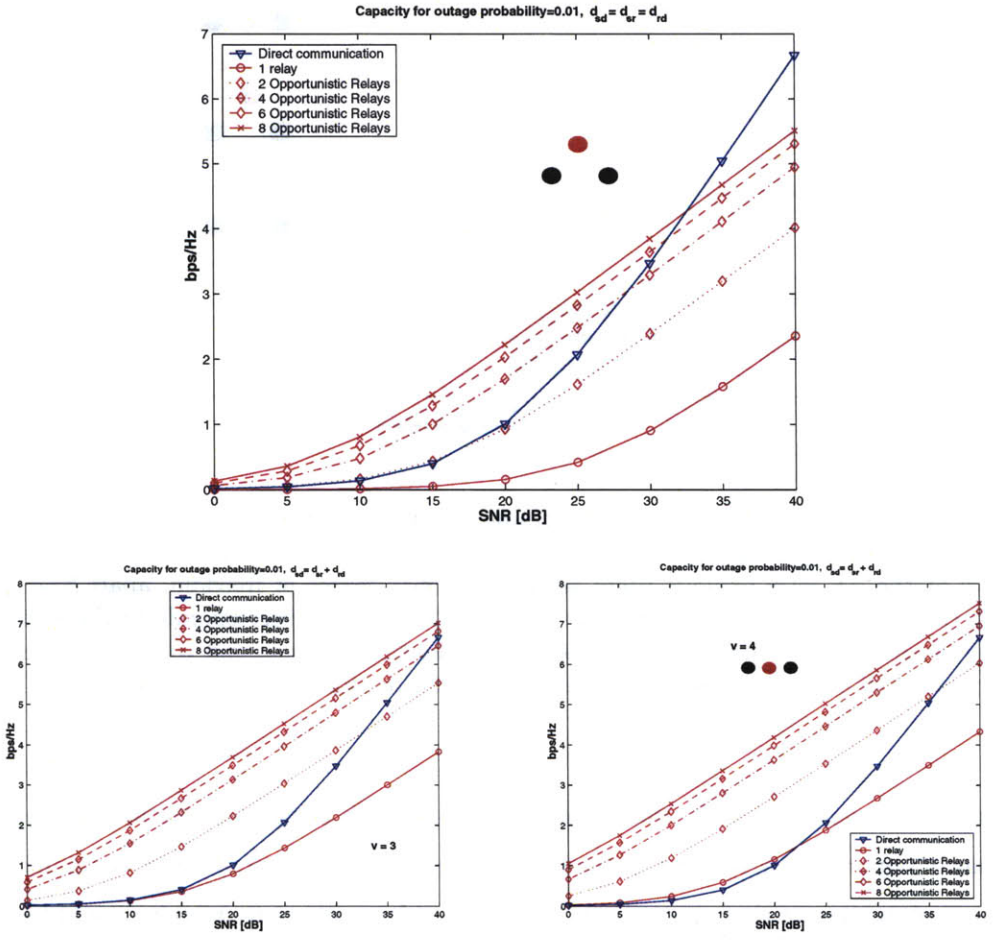

Figure 3-3: Outage rates for various SNRs in opportunistic relaying. Top: symmetric case. Bottom: asymmetric case for v=3 and v=4.

From a quick inspection of fig. 3-3, we can see that at transmitted SNR on the order of 20 dB, opportunistic cooperation of a small number of relays, on the order of 6, increases the spectral efficiency by a factor of 2. Additionally, for fixed spectral efficiency, a similar number of opportunistic relays could provide for transmission power gains on the order of 10 dB (a factor of 10) which is also significant. In the following section, we further attempt

to quantify the gains in transmission power of cooperative relaying.

3.3 Power Savings

We have already showed that a limited number of opportunistic relays can double the spectral efficiency (in bps/Hz) of wireless communication or lead to substantial energy gains, without any prior requirements of topology, among the participating nodes. In this section we will further study the three-node terminal from a practical perspective and show that cooperation can lead to substantial energy gains under certain conditions. We provide answers to previously reported research questions in the field [41] and emphasize the fact that previously reported solutions to the problem of cooperative diversity communication, require a priori knowledge of network geometry in order to be efficient. In contrast, opportunistic relaying requires no network topology, since the terminals find out the most appropriate path using distributed monitoring (*sensing*) of the wireless environment.

In the case of a single transmitter, single relay and receiver cooperative communication exploits the direct transmission, as well as the relayed transmission from a neighboring relay. The receiver combines direct and relayed transmission to detect information. In cases where direct transmission is not possible (for example, the receiver is out of range), multi-hop communication can be viewed as a special case of cooperative communication. Given the existence of a single relay, it is interesting to see what is the optimal signal processing strategy at the relay: it could either decode and re-encode (digital regeneration) or simply, amplify and forward the received information (plus its own noise) and leave the decision to the destination. Additional strategies could be used such as compressing the received information and forwarding which has been studied in [37] building upon the original work on the relay channel by El Gamal and Thomas.

We will concentrate on the two simple strategies of regenerate (decode) or amplify and forward, since the purpose of this section is to emphasize the importance of network intelligence in cooperative diversity schemes, so that power gains can emerge, regardless of the signal processing at each individual node. Both strategies can be found in the literature and have

different performance. In [41], it was reported that analog amplify and forward is better than digital regeneration for a relay located half distance between transmitter and receiver, when a Maximum Ratio Combining (MRC) receiver is used in Rayleigh fading. This is true, since we have showed that the areas of beneficial relay location are different for the above relaying techniques: uncoded analog relaying areas are symmetrical half-way distance between transmitter and receiver [7], while digital regeneration is beneficial only closer to the transmitter, since the scheme is limited by the probability of error in communication from transmitter to relay.

In fig. 3-4, we calculate the Symbol Error Probability (SEP) for 8-PSK modulation in Rayleigh fading with various propagation coefficients v, when MRC combining is used at the receiver and total transmission energy is split in half among the transmitter and the relay. Relay decodes and encodes (digital relay) and is placed closer to the transmitter, 1/4 the distance between source and destination. Performance is compared to direct (non-cooperative) transmission, when all the energy is used for direct, one-hop transmission.

For the digital case, we can calculate the end-to-end symbol error probability as one minus the probability of correct transmission which is basically the product of probability of correct reception between transmitter and relay and probability of correct reception of a MRC receiver when the two copies come from two different paths, one from the transmitter and one from the intermediate relay:

$$SEP = 1 - (1 - SEP_{1\to 2})(1 - SEP_{\substack{1\to 3\\2\to 3}}) \tag{3.23}$$

where the symbol error probabilities for M-PSK, are calculated by the following equations:

$$SEP_{1\to 2} = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi} \frac{sin^2(\theta)}{sin^2(\theta) + \sin^2(\pi/M)\ \overline{\gamma}_{1\to 2}}\, d\theta \tag{3.24}$$

$$SEP_{\substack{1\to 3\\2\to 3}} = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi} \frac{sin^2(\theta)}{sin^2(\theta) + \sin^2(\pi/M)\ \overline{\gamma}_{1\to 3}}\ \frac{sin^2(\theta)}{sin^2(\theta) + \sin^2(\pi/M)\ \overline{\gamma}_{2\to 3}}\, d\theta \tag{3.25}$$

$$\overline{\gamma}_{i\to j} = \mathcal{E}[\|a_{i\to j}\|^2]\,\frac{E_i}{N0}, \qquad \mathcal{E}[\|a_{i\to j}\|^2] \propto \frac{1}{d^v}, \tag{3.26}$$

with $a_{i\to j}$, the wireless channel between transmitter i and receiver j and E_i the symbol energy transmitted by node i.

We can see from figure 3-4 that the cooperative scheme is more reliable for the same transmission energy used, or it needs less transmission energy for the same performance. For SEP=1/1000, the plot is inverted and transmission energy savings are depicted in the form of ratios between transmission energy needed in the non-cooperative case vs the transmission energy needed in the cooperative case.

We can also observe improved performance, when a dense constellation is used, in combination with cooperation. For example, using a constellation of 3 bits per symbol (8-PSK) with cooperative transmission, performs more reliably than a constellation of 1 bit per symbol (2-PSK) of direct communication for rayleigh fading with $v \geq 3$ and digital relaying (we have omitted the plots due to space restrictions). Therefore, cooperation can increase throughput in uncoded systems by 50%, under certain conditions[8].

3.3.1 Areas of useful cooperation

In figure 3-5, the regions where digital relaying is beneficial compared to repetitive transmission are depicted, for the case of 8-PSK in Rayleigh fading and various signal-to-noise ratios (SNR) and two propagation coefficients v, normalized to point-2-point distance between transmitter and receiver. Specifically, we plot the space area where $SEP_{\substack{1\to 3\\1\to 3}}/SEP \geq 1$. We can see that provided that there is a relay close to transmitter, between transmitter and receiver, digital relaying (and consecutively cooperation) is beneficial at the low SNR regime, in highly attenuating propagation environments ($v \geq 3$).

Observe also that the regions are not symmetric, but they are "squeezed" toward the transmitter, since the probability of error is affected by the probability of correct transmission

[8] 3 bits per symbol, over two channel usages, one for direct and one for relayed transmission, result in 1.5 bits per channel usage versus 1 bit per channel usage for binary constellation and direct transmission.

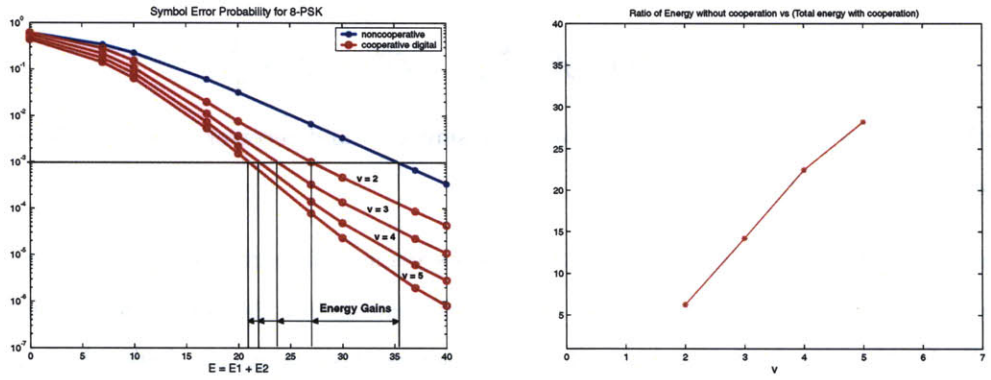

Figure 3-4: Performance of cooperative communication compared to non-cooperative communication in left figure (using 8-PSK and various propagation coefficients) and total transmission energy ratio for target Symbol Error Probability (SEP)=10^{-3} in right figure (using 8-PSK and v=4), in Rayleigh wireless channels. Relay decodes and encodes (digital relay) and it is placed closer to the transmitter, 1/4 the distance between source and destination. We can see that cooperative communication is more reliable compared to traditional point-to-point communication, leading to higher reliability or transmission energy savings. Left: SEP in 8-PSK for various environments and $E = E_1 + E_2$, $E_1 = E_2$. Right: corresponding ratio $E/(E_1 + E_2)$ for SEP=10^{-3}.

to the relay. Therefore, halfway the distance between transmitter and receiver, is NOT the optimal location to place a digital relay.

We have also studied analog amplify-and-forward in the context of uncoded M-PSK communication. The regions in that case are symmetric between transmitter and receiver, as opposed to the digital case. We have omitted the presentation of the plots due to space restrictions. More results for the analog case can be found in [7]. All the above findings explain why an analog amplify and forward relay outperforms a digital decode and forward relay both placed half-way between source and destination, as reported in [41] without thorough justification: the areas of useful cooperation are simply different for the two cases of signal processing at each relay.

Notice that the above improvements in energy gains are based on the assumption that there is a relay inside the appropriate area and *the transmitter knows that (i.e. the transmitter has decided that relaying is more beneficial than repetition).* Such decision could be based on knowledge of relay location, at the source! However, such knowledge is not trivial to acquire,

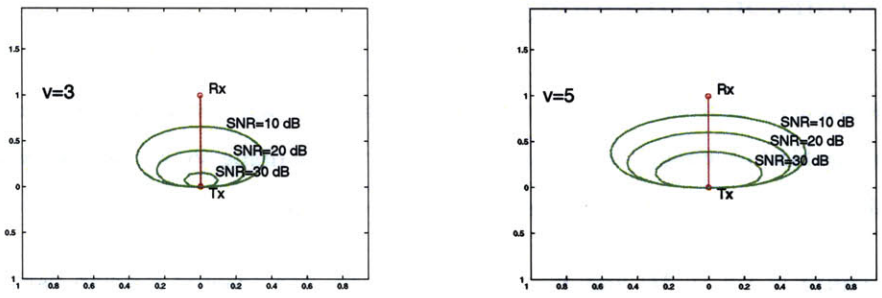

Figure 3-5: Left: v=3. Right: v=5. Regions of intermediate node location where it is advantageous to digitally relay to an intermediate node, instead of repetitively transmit. M=8 and the depicted ratio is the ratio of SEP of repetitive transmission vs SEP of user cooperative digital communication. The cooperative receiver optimally combines direct and relayed copy. Distances are normalized to the point-to-point distance between transmitter and receiver.

especially in the case of mobile nodes. It could be either estimated or provided by other external means (such as GPS). In such cases, relay location information should be provided to the source. In other words, transmission energy gains of traditional cooperative diversity schemes depend on network topology. More importantly, estimation of network topology might have significant overhead that could cancel the benefits of cooperative diversity and therefore, such overhead needs to be explicitly identified and quantified. Additionally, practical schemes for coordination and topology estimation need to be devised before the above simple three-node scheme could be implemented in practice. Attributing all the necessary overhead to an external service, such as GPS, might be one solution but is that a cost-effective solution? What happens when such services are not available (for example in indoor environments)?

On the other hand, opportunistic relaying not only scales cooperative diversity with larger than one number of relays, but also provides solutions for the required selection of appropriate relays, using distributed algorithms that require no topology estimation services (such as GPS). In the following section, we show that the network can react to the instantaneous channel conditions *fast* and with reasonably small overhead.

3.4 Collision Probability

In this section we provide an analytic way to calculate a close-form expression of equation (2.8) for any kind of wireless fading statistics. But before doing so, we can easily show that this probability can be made arbitrary small, close to zero.

If $T_b = \min\{T_j\}, j \in [1, M]$ and $Y_1 < Y_2 < \ldots < Y_M$ the ordered random variables $\{T_j\}$ with $T_b \equiv Y_1$, and Y_2 the second minimum timer, then:

$$Pr(any\ T_j < T_b + un \mid j \neq b) \equiv Pr(Y_2 < Y_1 + un) \tag{3.27}$$

From the last equation, we can see that this probability can be made arbitrarily small by decreasing the parameter un. For short range radios (on the order of 100 meters), this is primarily equivalent to selecting radios with small switch times (from receive to transmit mode) on the order of a few microseconds.

Given that $Y_j = \Lambda/h_{(j)}$, $Y_1 < Y_2 < \ldots < Y_M$ is equivalent to $1/h_{(1)} < 1/h_{(2)} < \ldots < 1/h_{(M)}$[9], equation (3.27) is equivalent to

$$Pr(Y_2 < Y_1 + un) = Pr(\frac{1}{h_{(2)}} < \frac{1}{h_{(1)}} + \frac{un}{\Lambda}) \tag{3.28}$$

and $Y_1 < Y_2 < \ldots < Y_M \Leftrightarrow h_{(1)} > h_{(2)} \ldots > h_{(M)}$ (h, Λ, un are positive numbers).

From the last equation (3.28), it is obvious that increasing Λ at each relay (in equation (2.3)), reduces the probability of collision to zero since equation (3.28) goes to zero with increasing Λ.

In practice, Λ can not be made arbitrarily large, since it also "regulates" the expected time, needed for the network to find out the "best" relay. From equation (2.3) and Jensen's inequality we can see that

$$E[T_j] = E[\Lambda/h_j] \geq \Lambda/E[h_j] \tag{3.29}$$

or in other words, the expected time needed for each relay to flag its presence, is lower bounded by Λ times a constant. Therefore, there is a tradeoff between probability of

[9]The parenthesized subscripts are due to ordering of the channel gains.

collision and speed of relay selection. We need to have Λ as big as possible to reduce collision probability and at the same time, as small as possible, to quickly select the best relay, before the channel changes again (i.e. within the coherence time of the channel). For example, for a mobility of $0-3$ km/h, the maximum Doppler shift is $f_m = 2.5\ Hz$ which is equivalent with a minimum coherence time on the order of 200 milliseconds. Any relay selection should occur well before that time interval with a reasonably small probability of error. From figure 3-6, we note that selecting $un/\Lambda \approx 1/200$ will result in a collision probability less than 0.6% for policy I. Typical switching times result in $un \approx 5\mu s$. This gives $\Lambda \approx 1ms$ which is two orders of magnitude less than the coherence interval. More sophisticated radios with $un \approx 1\mu s$ will result in $\Lambda \approx 200\mu s$, which is three orders of magnitude smaller than the coherence time [10].

3.4.1 Calculating $Pr(Y_2 < Y_1 + un)$

In order to calculate the collision probability from (3.27), we first need to calculate the joint probability distribution of the minimum and second minimum of a collection of M i.i.d[11] random variables, corresponding to the timer functions of the M relays. The following theorem provides this joint distribution:

Theorem 4 *The joint probability density function of the minimum and second minimum among $M \geq 2$ i.i.d. positive random variables $T_1,\ T_2, \ldots,\ T_M$, each with probability density function $f(t) \equiv \frac{dF(t)}{dt}$ and cumulative distribution function $F(t) \equiv Pr(T \leq t)$, is given by*

[10] Note that the expected value of the minimum of the set of random variables(timers) is smaller than the average of those random variables. So we expect the overhead to be much smaller than the one calculated above

[11] The choice of identically distributed timer functions implicitly assumes that the relays are distributed in the same geographical region and therefore have similar distances towards source and destination. In that case, randomization among the timers is provided only by *fading*. The cases where the relays are randomly positioned and have in general different distances, is a scenario where randomization is provided not only because of fading, but also because of different moments. In such asymmetric cases the collision probability is expected to decrease and a concrete example is provided.

the following equation:

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} M\ (M-1)\ f(y_1)\ f(y_2)\ [1-F(y_2)]^{M-2} & for\ 0 < y_1 < y_2 \\ 0 & elsewhere. \end{cases}$$

where $Y_1 < Y_2 < Y_3 \ldots < Y_M$ *are the* M *ordered random variables* $T_1,\ T_2, \ldots,\ T_M$.

Proof 3 *Please refer to appendix 6.*

Using Theorem 1, we can show the following lemma that gives a closed-form expression for the collision probability (equation 3.27):

Lemma 1 *Given* $M \geq 2$ *i.i.d. positive random variables* $T_1,\ T_2, \ldots,\ T_M$, *each with probability density function* $f(x)$ *and cumulative distribution function* $F(x)$, *and* $Y_1 < Y_2 < Y_3 \ldots < Y_M$ *are the* M *ordered random variables* $T_1,\ T_2, \ldots,\ T_M$, *then* $Pr(Y_2 < Y_1 + un)$, *where* $un > 0$, *is given by the following equations:*

$$Pr(Y_2 < Y_1 + un) = 1 - I_{un} \tag{3.30}$$

$$I_{un} = M\ (M-1)\ \int_{un}^{+\infty} f(y)\ [1-F(y)]^{M-2}\ F(y-un)\ dy \tag{3.31}$$

Proof 4 *Please refer to appendix 6.*

Notice that the statistics of each timer T_i and the statistics of the wireless channel are related according to equation (2.3). Therefore, the above formulation is applicable to any kind of wireless channel distribution.

3.4.2 Results

In order to exploit theorem 4 and lemma 1, we first need to calculate the probability distribution of T_i for $i \in [1, M]$. From equation (2.3) it is easy to see that the cdf $F(t)$

and pdf $f(t)$ of T_i are related to the respective distributions of h_i according to the following equations:

$$F(t) \equiv \ CDF_{T_i}(t) = Pr\{T_i \ \leq \ t\} = 1 - CDF_{h_i}(\frac{\Lambda}{t}) \tag{3.32}$$

$$f(t) \equiv \ pdf_{T_i}(t) = \frac{d}{dt}F(t) = \frac{\Lambda}{t^2} \ pdf_{h_i}(\frac{\Lambda}{t}) \tag{3.33}$$

After calculating equations (3.32), (3.33), and for a given un calculated from (2.6) or (2.7), and a specific Λ, we can calculate probability of collision using equation (3.30).

Before proceeding to special cases, we need to observe that for a given distribution of the wireless channel, collision performance depends on the ratio un/Λ, as can be seen from equation (3.28), discussed earlier.

Rayleigh Fading

Assuming $|a_{si}|$, $|a_{id}|$, for any $i \in [1, M]$, are independent (but not identically distributed) Rayleigh random variables, then $|a_{si}|^2$, $|a_{id}|^2$ are independent, exponential random variables, with parameters β_1, β_2 respectively ($E[|a_{si}|^2] = 1/\beta_1$, $E[|a_{id}|^2] = 1/\beta_2$).

Using the fact that the minimum of two independent exponential r.v.'s with parameters β_1, β_2, is again an exponential r.v with parameter $\beta_1 + \beta_2$, we can calculate the distributions for h_i under policy I (equation 2.1). For policy II (equation 2.2), the distributions of the harmonic mean, have been calculated analytically in [31]. Equations (3.32) and (3.33) become:

under policy I:

$$F(t) \ = \ e^{-\ (\beta_1+\beta_2)\ \Lambda/t} \tag{3.34}$$

$$f(t) \ = \ \frac{\Lambda\ (\beta_1 + \beta_2)}{t^2}\ e^{-(\beta_1+\beta_2)\ \Lambda/t} \tag{3.35}$$

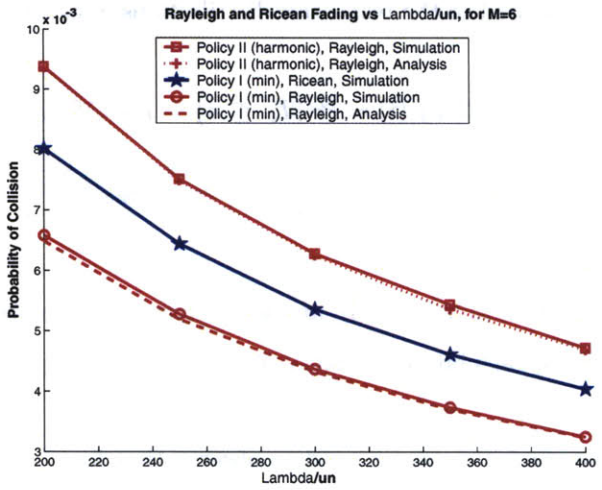

Figure 3-6: Performance in Rayleigh and Ricean fading, for policy I (min) and Policy II (harmonic mean), various values of ratio Λ/un and $M = 6$ relays, clustered at the same region. Notice that collision probability drops well below 1%.

under policy II:

$$F(t) = \frac{\Lambda\sqrt{\beta_1\,\beta_2}}{t}\, e^{-\Lambda\,(\beta_1+\beta_2)/(2t)}\, K_1\left(\frac{\Lambda\sqrt{\beta_1\beta_2}}{t}\right) \tag{3.36}$$

$$f(t) = \frac{\Lambda^2}{2\,t^3}\,\beta_1\beta_2\, e^{-\Lambda(\beta_1+\beta_2)/(2t)}\,\left[\frac{\beta_1+\beta_2}{\sqrt{\beta_1\,\beta_2}}K_1\left(\frac{\Lambda\sqrt{\beta_1\,\beta_2}}{t}\right)+2\,K_0\left(\frac{\Lambda\sqrt{\beta_1\,\beta_2}}{t}\right)\right] \tag{3.37}$$

where $K_i(x)$ is the modified Bessel function of the second kind and order i.

Equation (3.30) is calculated for the two policies, for the symmetric case ($\beta_1 = \beta_2 = E[|a_{si}|^2] = E[|a_{id}|^2] = 1$) of $M = 6$ relays. Monte-Carlo simulations are also performed under the same assumptions. Results are plotted in fig. 3-6, for various ratios Λ/un. We can see that Monte-carlo simulations match the results provided by numerical calculation of equation (3.30) with the help of equations (3.34)-(3.37).

Collision probability drops with increasing ratio of Λ/c as expected. Policy I ("the minimum"), performs significantly better than Policy II ("the harmonic mean") and that can be attributed to the fact that the harmonic mean smooths the two path SNRs (between source-relay and relay-destination) compared to the minimum function. Therefore, the ef-

fect of randomization due to fading among the relay timers, becomes less prominent under Policy II. The probability can be kept well below 1%, for ratio Λ/un above 200.

Ricean Fading

It was interesting to examine the performance of opportunistic relay selection, in the case of Ricean fading, when there is a dominating communication path between any two communicating points, in addition to many reflecting paths and compare it to Rayleigh fading, where there is a large number of equal power, independent paths.

Keeping the average value of any channel coefficient the same ($E[|a|^2] = 1$) and assuming a single dominating path and a sum of reflecting paths (both terms with equal total power), we plotted the performance of the scheme when policy I was used, using Monte-Carlo simulations (fig. 3-6). We can see that in the Ricean case, the collision probability slightly increases, since now, the realizations of the wireless paths along different relays are clustered around the dominating path and vary less, compared to Rayleigh fading. Policy II performs slightly worse, for the same reasons it performed slightly worse in the Rayleigh fading case and the results have been omitted.

In either cases of wireless fading (Rayleigh or Ricean), the scheme performs reasonably well.

Different topologies

For the case of all relays not being equidistant to source or destination, we expect the collision probability to drop, compared to the equidistant case, since the asymmetry between the two links (from source to relay and from relay to destination) or the asymmetry between the expected SNRs among the relays, will increase the variance of the timer function, compared to the equidistant case. To demonstrate that, we study three cases, where $M = 6$ relays are clustered half-way ($d/2$), closer to transmitter ($d/3$) or even closer to transmitter ($d/10$) (case 1,2,3 respectively in fig. 3-7 and d is the distance between source and destination) and

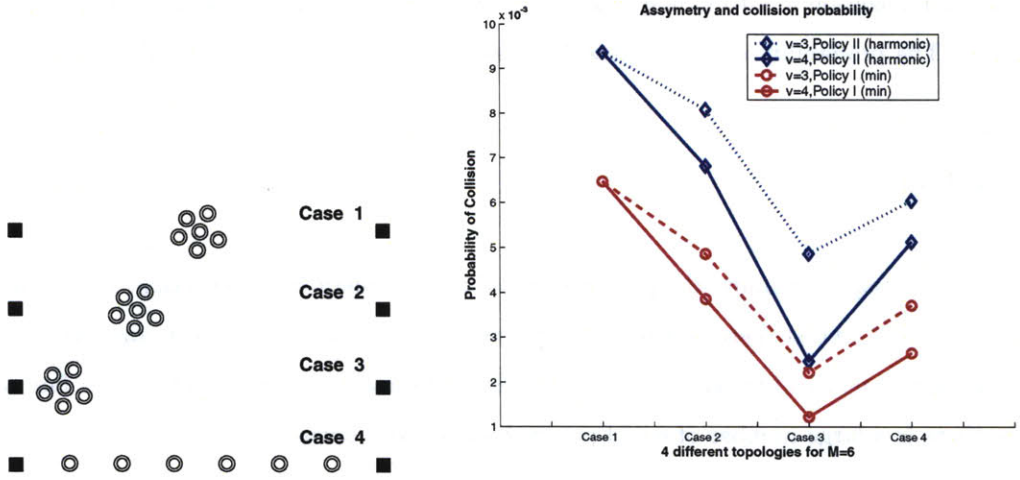

Figure 3-7: Unequal expected values (moments) among the two path SNRs or among the relays, reduce collision probability. M=6 and $un/\Lambda = 1/200$ for the four different topologies considered.

one case where the relays form an equidistant line network between source and destination (case 4 in fig. 3-7).

Assuming Rayleigh fading, $un/\Lambda = 1/200$ and expected path strength as a non-linear, decreasing function of distance ($E[|a_{ij}|^2] = 1/\beta_{ij} \propto (1/d_{ij})^v$), we calculate the collision probability for $M = 6$ relays, using expressions (3.34)-(3.37) into (3.30) for cases 1, 2, 3 while for case 4 we used Monte-Carlo simulation: in case 1, $\beta_1 = \beta_2 = 1$, in case 2, $\beta_1 = (2/3)^v, \beta_2 = (4/3)^v$ and in case 3, $\beta_1 = (1/5)^v, \beta_2 = (9/5)^v$. For case 4, $\beta_1 = (2/7)^v, \beta_2 = (12/7)^v$ for the closest terminal to source, $\beta_1 = (4/7)^v, \beta_2 = (10/7)^v$ for the second closest terminal to source, $\beta_1 = (6/7)^v, \beta_2 = (8/7)^v$ for the third closest to source terminal. Due to symmetry, the expected power and corresponding β factors of the paths, for the third closer to destination, second closer to destination and closest terminal to destination, are the same with the ones described before (third closer terminal to source, second closer terminal to source and closest to source terminal respectively), with β_1 and β_2 interchanged.

We can see in fig. 3-7 that the collision probability of asymmetric cases 2, 3 and 4 is strictly smaller compared to the symmetric case 1. Policy I performs better than Policy

II and collision probability decreases for increasing factor v ($v = 3, 4$ were tested). This observation agrees with intuition that suggests that different moments for the path strengths among the relays, increase the randomness of the expiration times among the relays and therefore decrease the probability of having two or more timers expire within the same time interval.

We note that the source can also participate in the process of deciding the best relay. In this special case, where the source can receive the CTS message, it could have its own timer start from a value depending upon the instantaneous $|a_{sd}|^2$. This will be important if the source is not aware whether there are any relays in the vicinity that could potentially cooperate.

The proposed method as described above, involving instantaneous SNRs as a starting point for each relay's timer and using *time* (corresponding to an assessment of how good is a particular path within the coherence time of the channel) to select *space* (the best available path towards destination) in a distributed fashion, is novel and has not been proposed before, to the best extent of our knowledge.

Chapter 4

Scaling and Extensions

4.1 To Relay or not to Relay?

One of the major findings in the previous chapter was that opportunistic relay selection and use of the single "best" relay, incurs no performance loss when compared to simultaneous transmissions of multiple relays which use complex space-time coding. The relevant analysis was performed with the high SNR tool of diversity-multiplexing gain tradeoff. Because of the high SNR nature of that tool, power allocation at the relays is meaningless and cannot be studied: SNR is increased towards infinity and at the same time the spectral efficiency is increased with $\log_2 SNR$, in order to calculate the corresponding tradeoff between diversity (reliability) and multiplexing gain (rate).

In this section, we attempt to study the problem of power allocation, in the case of multiple amplify-and-forward relays. We surprisingly discover that distributing the total transmission power to a set of simple[1] relay radios, is suboptimal when compared to opportunistic relaying and in fact, the penalty of performance loss (or inversely the gain of opportunistic relaying) increases logarithmicaly with the number of relays. This important result suggests

[1]Throughout this dissertation, we have excluded beamforming scenarios, since such hardware capability is difficult in practice, especially in the case of distributed single antenna radios.

that relay selection can provide for important gains in amplify-and-forward relays systems, when compared to "all relays-transmit" schemes, proposed in the literature [30], [14].

Total transmission power is an important network resource, especially for battery-operated applications and networks that seek to maximize network lifetime. Traditional studies of scalability, tend to examine the performance gains when multiple nodes (with their own tx power) enter the network [24]. Such studies investigate a communication performance measure, such as ergodic capacity or outage probability, as a function of number of relay nodes. Therefore, such studies implicitly assume that total transmission power increases with the number of participating nodes. In this section, we study scalability with a more careful treatment of total transmission power. Since we are interested in comparing different schemes for the same number of participating elements, we will explicitly fix the total transmission power.

We assume again a two step transmission scheme: during the first phase the source transmits and the relays and destination listen, while during the second step, the relays relay using a version of amplify-and-forward. For completeness, we allow the transmitter to transmit a different symbol, during the second step, even though we will relax this assumption in the subsequent analysis.

During the first slot, the destination receives $y_{D,1}$ while each relay Ri receives $y_{Ri,1}$. P_{SX} is the average normalized received power (or energy if multiplied with an appropriate scaling factor) between source and terminal X and includes the transmitted power, as well as other propagation phenomena, like shadowing. a_{SX} is the unit-power fading coefficient, which for the numerical results of the subsequent section will be assumed complex, circularly symmetric, Gaussian random variable, ($a = h_1 + jh_2$, where h_1, h_2 are i.i.d normal r.v's $\mathcal{N}(0, 1/2)$), corresponding to Rayleigh fading. Similarly, n_x is additive white complex gaussian noise, with power $N_0/2$ per dimension ($n = n_a + jn_b$, where n_a, n_b are i.i.d normal r.v's $\mathcal{N}(0, N_0/2)$). x_1 is the unit power symbol, sent from the source during the first slot. We will further assume independence among the noise and channel terms among different relays.

1st Slot:

$$y_{D,1} = \sqrt{P_{SD}}\, a_{SD}\, x_1 + n_{D,1} \tag{4.1}$$

$$y_{Ri,1} = \sqrt{P_{SRi}}\, a_{SRi}\, x_1 + n_{Ri,1},\ \forall\ i\ \in\ [1, M] \tag{4.2}$$

Notice that the expected power of each symbol received at each relay Ri can be easily calculated, taking into account the assumptions above: $\mathcal{E}[|y_{Ri,1}|^2] = P_{SRi} + N_0$. Each relay normalizes its received signal with its average power and transmits $\frac{y_{Ri,1}}{\sqrt{\mathcal{E}[|y_{Ri,1}|^2]}}$. This is a normalization followed in the three terminal analysis (one source, one destination and one relay) presented in [58]. Here, we can easily generalize it to the case of multiple relays, during the second slot:

2nd Slot:

$$y_{D,2} = \sqrt{P_{SD}}\, a_{SD}\, x_2 + \sum_{i=1}^{M} \sqrt{P_{SRi}}\, a_{SRi}\, \frac{y_{Ri,1}}{\sqrt{\mathcal{E}[|y_{Ri,1}|^2]}} + n_{D,2} \Leftrightarrow \tag{4.3}$$

$$y_{D,2} = \sqrt{P_{SD}}\, a_{SD}\, x_2 + \sum_{i=1}^{M} \frac{\sqrt{P_{SRi}}\,\sqrt{P_{RiD}}}{\sqrt{P_{SRi} + N_0}}\, a_{SRi}\, a_{RiD}\, x_1 + \underbrace{+ n_{D,2} + \sum_{i=1}^{M} \frac{\sqrt{P_{RiD}}}{\sqrt{P_{SRi} + N_0}}\, a_{RiD}\, n_{Ri,1}}_{\widetilde{n}_{D,2}}$$

$$y_{D,2} = \sqrt{P_{SD}}\, a_{SD}\, x_2 + \sum_{i=1}^{M} \frac{\sqrt{P_{SRi}}\,\sqrt{P_{RiD}}}{\sqrt{P_{SRi} + N_0}}\, a_{SRi}\, a_{RiD}\, x_1 + \widetilde{n}_{D,2} \tag{4.4}$$

Again, here P_{XD} is the average normalized received power (or energy if multiplied with an appropriate scaling factor) between terminal X and destination and includes the transmitted power, as well as other propagation phenomena, like shadowing. a_{XD} is the unit power fading coefficient, which will be assumed complex, circularly symmetric Gaussian random variable (corresponding to Rayleigh fading), for the numerical results of the subsequent section. x_2 is the unit power symbol, sent from the source during the second slot.

From the last equation, we can see that the received signal at the destination, can be

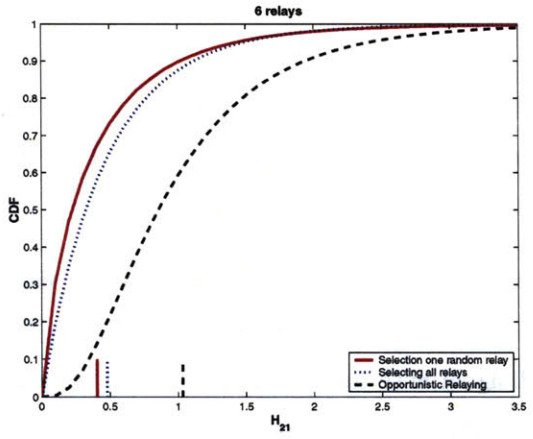

Figure 4-1: Cumulative Distribution Function (CDF) of $\tilde{H}_{12}$ (eq. 4.12, 4.13, 4.14), for the three cases examined (one, all, "best" relay(s) transmit). The expected value is also depicted, at the bottom of the plot.

written as the sum of two terms, corresponding to the two transmitted information symbols plus one noise term. Assuming that the destination has knowledge of the wireless channel conditions between the relays and itself (for example, the receiver can estimate the channel using preamble information), the noise term in equation 4.4 becomes complex Gaussian with power easily calculated[2]:

$$\mathcal{E}[\tilde{n}_{D,2}\ \tilde{n}^*_{D,2}\ |H_{R\to D}] = N_0\ \underbrace{(1+\sum_{i=1}^{M}\frac{P_{RiD}\ |a_{Rid}|^2}{P_{SRi}+N_0})}_{\omega^2} = \omega^2\ N_0 \tag{4.5}$$

Therefore, the system of the above equations can be easily written in matrix notation:

$$\begin{bmatrix} y_{D,1} \\ \\ \frac{y_{D,2}}{\omega} \end{bmatrix} = \left[\begin{array}{c|c} \sqrt{P_{SD}}\ a_{SD} & 0 \\ \hline \frac{1}{\omega}\ \sum_{i=1}^{M}\frac{\sqrt{P_{SRi}}\ \sqrt{P_{RiD}}}{\sqrt{P_{SRi}+N_0}}\ a_{SRi}\ a_{RiD} & \frac{1}{\omega}\ \sqrt{P_{SD}}\ a_{SD} \end{array}\right] \begin{bmatrix} x_1 \\ \\ x_2 \end{bmatrix} + \begin{bmatrix} n_{D,1} \\ \\ \frac{\tilde{n}_{D,2}}{\omega} \end{bmatrix}$$

[2]Notice that we do not need knowledge of the wireless channels conditions at the receiver between source and relays, for the above assumption to hold

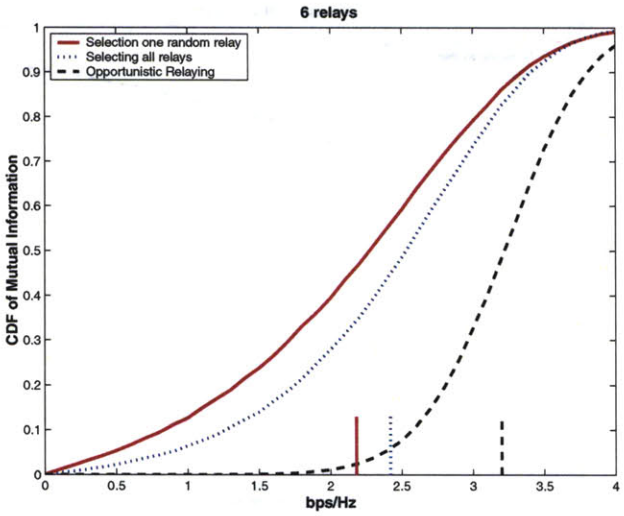

Figure 4-2: Cumulative Distribution Function (CDF) of mutual information (eq. 4.11), for SNR=20 dB. Notice that the CDF function provides for the values of outage probability.

The above notation can be summarized as:

$$\mathbf{y} = \begin{bmatrix} \sqrt{P_{SD}}\, a_{SD} & 0 \\ H_{21} & \frac{1}{\omega}\sqrt{P_{SD}}\, a_{SD} \end{bmatrix} \mathbf{x} + \mathbf{n} \tag{4.6}$$

$$\mathbf{y} = \mathbf{H}\,\mathbf{x} + \mathbf{n} \tag{4.7}$$

The noise term, under the above assumptions, has covariance matrix given below,[3] where $\mathbf{I_2}$ is the 2x2 unity matrix:

$$\omega^2 = (1 + \sum_{i=1}^{M} \frac{P_{RiD}\,|a_{Rid}|^2}{P_{SRi} + N_0} \tag{4.8}$$

$$\mathbf{E}[\mathbf{n}\,\mathbf{n}^{\mathbf{T}}\,|H_{R\to D}] = N_0\,\mathbf{I_2} \tag{4.9}$$

For the subsequent section, we will further simplify the cooperation scheme and will not allow the transmission of a new symbol x_2 during the second slot, in coherence with the

[3]The symbols $*, \mathbf{T}$ correspond to complex-conjugate and conjugate-transpose respectively

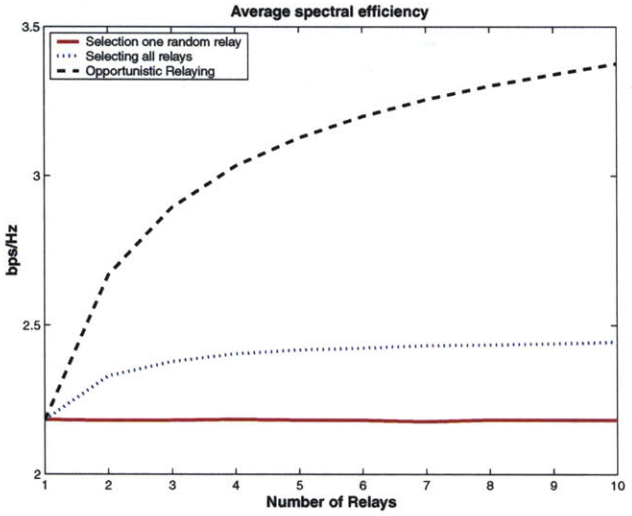

Figure 4-3: Expected value of mutual information (eq. 4.11), corresponding to the ergodic capacity, as a function of number of relays. Notice that using all relays incurs a penalty that increases with number of relays, compared to opportunistic relaying.

communication scheme studied in the previous chapter. In that way, the second column of the matrix $\mathbf{H}$ is zero and $\mathbf{H}$ becomes a column vector (the first column of $\mathbf{H}$ above).

The mutual information for the above assumptions can be easily calculated for the above linear system, using the result from Telatar's work [75]:

$$I_{AF} = \frac{1}{2}\ \log_2(1 + \frac{P_{SD}}{N_0}\ |a_{SD}|^2 + \frac{|H_{21}|^2}{N_0}) \tag{4.10}$$

Since we are interested in the power allocation of total transmission power P_R at the relays, we further dismiss the direct connection term between source and destination. Practically, that corresponds to the case when source and destination are not within communication range or simply, destination does not exploit that connection. P_{SD} corresponds to the power the source spends during the transmission at the first slot:

$$I_{AF} = \frac{1}{2}\ \log_2(1 + \frac{P_{SD}}{N_0}\ |\widetilde{H}_{21}|^2) \tag{4.11}$$

We further assume that all relays are *equivalent*: all relays have the same *average* power

terms P_{RiD} which practically means that $P_{RiD} = const$ for the M relays. We will test three different cases: a) all power P_R is used at one random relay, b) power is distributed at all relays $P_{RiD} = P_R/M$ and c) all power P_R is used at the best, opportunistic relay:

$$|\widetilde{H}_{21}|^2_{one} = \frac{1}{\frac{P_{SD}+N_0}{P_R} + |a_{RiD}|^2} |a_{SRi}\, a_{RiD}|^2 \tag{4.12}$$

$$|\widetilde{H}_{21}|^2_{all} = \frac{1}{\frac{P_{SD}+N_0}{P_R/M} + \sum_{i=1}^{M} |a_{RiD}|^2} |\sum_{i=1}^{M} a_{SRi}\, a_{RiD}|^2 \tag{4.13}$$

$$|\widetilde{H}_{21}|^2_{opp} = \frac{1}{\frac{P_{SD}+N_0}{P_R} + |a_{RbD}|^2} |a_{SRb}\, a_{RbD}|^2 \tag{4.14}$$

where $\min\{|a_{SRb}|^2,\ |a_{RbD}|^2\} \geq \min\{|a_{SRi}|^2,\ |a_{RiD}|^2\}, \forall i \in [1, M]$

The first term in equations 4.12, 4.14 is greater than the first term in equation 4.13. The second term in 4.13 corresponds to the magnitude of the sum of complex numbers with random phases. Therefore, the addition of an increasing number of those terms does not necessarily results in a proportional increase of the magnitude: that would be possible, only under equal phases (beamforming).

For $P_R = P_{SD} \gg N_0$ then $\frac{P_{SD}+N_0}{P_R} \to 1$, the Cumulative Distribution Function ($CDF(x) = Pr(\widetilde{H}_{21} \leq x)$ for the three above cases, is depicted in figure 4-1, for the case of Rayleigh fading for all coefficients a_{SRi}, a_{RiD}.

In fig. 4-1 it is shown that $Pr(|\widetilde{H}_{21}|^2_{one} \leq x) \geq Pr(|\widetilde{H}_{21}|^2_{all} \leq x) \geq Pr(|\widetilde{H}_{21}|^2_{opp} \leq x)$ which means that, in general, $|\widetilde{H}_{21}|^2_{one} \leq |\widetilde{H}_{21}|^2_{all} \leq |\widetilde{H}_{21}|^2_{opp}$.

Consecutively, the mutual information statistics are depicted in figure 4-2, in the form of CDF function (corresponding to the outage probability) and in figure 4-3, in the form of expected values (corresponding to ergodic capacity). Both plots show the superiority of opportunistic relaying, compared to the case of having all relays transmit. They also show, that choosing a random relay is a suboptimal technique, compared to the all relays case.

We can see in fig. 4-3 that selecting a single, "best" relay, provides performance gains that increase with the number of relays, compared to the "all-relays transmit" case, under a

sum power constraint. This is an important result, given the popularity of the "all relays-transmit" approach in the literature. This result clearly suggests that, the advantages of multiple nodes in a relay network, do not arise because of complex reception techniques, as the "all relays transmit" approach requires, but rather emerge because of the fact that multiple possible paths exist between source, the participating relays and the destination. Opportunistic relaying, simply exploits the best available path.

In that sense, opportunistic relaying can be viewed as a "smart" scheduling algorithm of RF energy, coming from another node (the source) and destined for another user (the destination). Through the method of distributed timers presented and analyzed in the previous chapters, the network schedules the transmission over the most appropriate relay path (relaying as scheduling), via a decentralized way.

In the following section we show that opportunistic relaying can be easily viewed as RF scheduling, in more involved settings.

4.2 Extensions: Scheduling Multiple Streams

In the previous sections, we described opportunistic relaying as a distributed way to select the relay b, that maximizes a function of the instantaneous channel conditions between source/relay and relay/destination. As we saw, the minimum function was a viable solution and the best relay is the one, according to $\min\{|a_{SRb}|^2,\ |a_{RbD}|^2\} \geq \min\{|a_{SRi}|^2,\ |a_{RiD}|^2\}$, $\forall i \in [1, M]$, for the M relays. Assuming similar radio hardware, we can safely further assume that the thermal noise at all relays has the same average power. Therefore, we can extend the opportunistic relaying rule to incorporate instantaneous SNR conditions at each relay, rather than just instantaneous channel conditions. The timer functions use SNR values and the two rules are essentially equivalent, both from a conceptual perspective, as well as from a practical (implementation) point of view :

$$b = \underbrace{arg}_{i}\ max\{min\{SNR_{si}, SNR_{id}\}\} = max\{SNR_{sid}\}, \quad i \in [1..M] \tag{4.15}$$

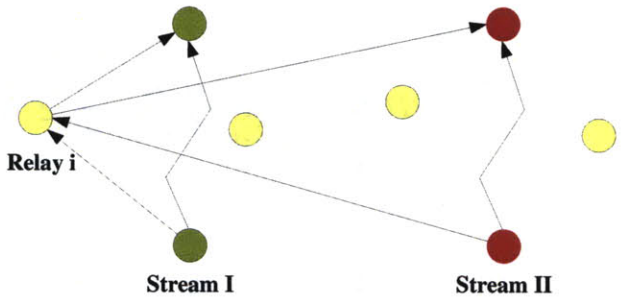

Figure 4-4: Relaying as scheduling for multiple streams.

Each relay, willing to assist the transmission stream I (fig. 4-4), for which it has gathered information overhearing the pilot signals RTS/CTS transmitted initially by the communicating source and destination, is affected by the simultaneous transmission from stream II (fig. 4-4). Stream II simultaneous transmission affects stream I effective path Signal-to-Interference-and-Noise Ratio (SINR) and therefore, the path (relay) selection rule caqn be changed from a notation point of view:

$$b = \underbrace{arg}_{i} \; max\{min\{SINR_{si}, SINR_{id}\}\} = max\{SINR_{sid}\}, \quad i \in [1..M] \tag{4.16}$$

From a practical point of view, taking into account other concurrent streams does not affect implementations: relays assisting stream I, need not know anything regarding stream II, since its influence automatically appears in the SINR term.

Following the notation of [3], based on work in [20], we assume N streams and denote $G_{ij} \equiv |a_{(i)(j)}|^2$, the square magnitude of the channel condition between the source of stream j and destination of stream i. Stream i is successfully transmitted if its corresponding SINR is above a threshold θ_i. Assuming P_j, the transmission power of transmitter in stream j, the system of equations describing successful communication of the N streams is summarized

as follows:

$$SINR_i = \frac{G_{ii}P_i}{\sum_{j\neq i} G_{ij}P_j + n_i} \geq \theta_i \tag{4.17}$$

$$(\mathbf{I} - \mathbf{F})\ \mathbf{P} \geq \theta \tag{4.18}$$

where

$$\mathbf{P} = (P_1 \ \ P_2 \ \ \ldots P_N)^T \tag{4.19}$$

$$\theta = (\frac{\theta_1\ n_1}{G_{11}} \ \ \frac{\theta_2\ n_2}{G_{22}} \ldots \ \frac{\theta_N\ n_N}{G_{NN}})^T \tag{4.20}$$

$$F_{ij} = 0, \ \ i = j \tag{4.21}$$

$$F_{ij} = \frac{\theta_i\ G_{ij}}{G_{ii}}, \ \ i \neq j \tag{4.22}$$

if such $\mathbf{P}$ *exists, then :*

$$\mathbf{P} \geq \mathbf{P}^* = (\mathbf{I} - \mathbf{F})^{-\mathbf{1}}\ \theta \tag{4.23}$$

If the requirement of $SINR_i \geq \theta_i, \ \forall i \in [1, N]$ streams, the transmitted power vector should satisfy equation 4.23 which shows the minimum required transmitted power.

The above notation provides a compact way to evaluate performance if *interference* is treated explicitly from a network point-of-view, rather than treating interference and thermal noise as a single quantity. In terms of best relay selection, in the presence of multiple streams the algorithm still "works", both from a conceptual as well as from a practical standpoint.

Additional extensions would require coordination among all participating wireless terminals and they are left for future work. Specific network coordination techniques, based on network time keeping are presented in the following chapter.

Chapter 5

Relevant Time Keeping Technologies

In chapter 2 we described a method of "best" relay selection, based on distributed timers, that exploit functions of the instantaneous channel conditions. We quantified the performance of such technique in section 3.4 and explained why an explicit time synchronization protocol is not required.

However, the notion of network time keeping is of primal importance in distributed environments and specifically, in scalable wireless networks. Accurately synchronized clocks enable services and provide the basis for efficient communications. Autonomous sensor array operation is facilitated by accurate time stamps [63], [62]. Global Positioning System, as well as proposed Ultra-Wide Band urban and intra-building location systems [60] rely on precise timing measurements. Internet performance can be evaluated from accurate measurement of the delay between various nodes in the network. Various important Internet Protocols such as TCP could benefit from accurate time keeping [64].

Additionally, a common time reference is important for many applications of distributed sensing, especially when the individual sensor nodes span a large geographical area and communicate over wireless. Time synchronization among the nodes becomes non-trivial when

all the individual nodes are several hops away and therefore a single broadcast signal from a particular node (a "server") is not sufficient, as it cannot reach all nodes. Energy constraints of individual sensor nodes prohibit extensive communication among them, complicating further the problem of time synchronization. Sensor Networks ought to self-configure and work unattended and therefore, any synchronization scheme should have minimal complexity: at the network level, any time keeping should require minimal communication and coordination among the nodes, while at the individual sensor node level, it should require simple digital signal processing (as measured in floating point operations per second and required memory size).

In this chapter, we present two novel approaches to the problem of Network Time Keeping.

In the first approach, we follow the methodology of one of the oldest Internet protocols, the Network Time Protocol (NTP) [52], where a client node tries to steer its local clock parameters, using time messages exchanged with a remote time server, over a "noisy" and uncontrollable network connection. We propose an adaptive filtering technique, based on Kalman filtering and contrast it to other techniques in the field, for two cases of noise (additive Gaussian noise and Self-similar (chaotic) noise). One of the interesting findings was that our proposed technique can reduce the estimation error, faster than $\sqrt{N}$, where N is the number of messages exchanged (bandwidth), outperforming other techniques based on a simple averaging (where error decrease on the order of $1/\sqrt{N}$ is expected) or more involved techniques, based on linear programming.

The second proposed approach to the problem of Network Time Keeping, is based on a completely decentralized technique: no servers are used and time keeping is performed using schemes inspired by natural phenomena of synchronization: the way fireflies blink in unison, even though they interact only locally or the way cardiac neurons fire in sync. A simple demonstration was constructed and discussed to illustrate the principles and measurements as well as theoretical analysis were performed. One of the interesting findings was that synchronization error does not necessarily increase with diameter of the network: by adding nodes into the system, the network establishes a common time reference without additional overhead and might have smaller synchronization error, depending on the individual clock

characteristics of the participating wireless nodes.

After presenting basic definitions of clocks and time synchronization, we present the two approaches based on centralized and decentralized time keeping.

5.1 Clock Basics

Using the representation $C(t)$ for a clock reading and $T(t) = t$ for true time, the following definitions are presented:

- **time offset:** the difference between the time reported from a clock and the "true" time: $C(t) - T(t) = C(t) - t$. In this work we will refer to the time offset calculated for $t = 0$ as θ and for $t \neq 0$ as x.
- **frequency offset** (also referred as *skew*): the difference in frequencies between a clock and the "true" time: $C'(t) - T'(t) = C'(t) - 1$. In this work, we will refer to frequency offset as $\phi - 1$.
- **drift:** the long-term frequency change of a clock. Drift is caused by changes in the components of the oscillator and its environment.

Typical quartz oscillators (without any type of temperature compensation) exhibit frequency offsets on the order of a few parts per million (PPM). For example a 10 PPM oscillator will introduce an uncertainty (i.e. error) of 36 msec in one hour. Cesium beam atomic clocks on the other hand, exploiting the stabilities of the quantum world perform better with uncertainties close or smaller than 1 nsec in 24 hours.

Modeling a clock as a piecewise linear function of time is a reasonable step since any function can be approximated in a similar manner. The client should estimate only two parameters, namely the time and frequency offset θ, $\phi - 1$ respectively, compared to the source of true time $T(t)$ as depicted in fig. (5-1-LEFT), since only two parameters are needed to define a line.

However, for this model to be realistic, it is important to keep the duration of the measurement process as small as possible, before ϕ and θ at the client clock are modified. The parameters ϕ, θ change with a rate related to the clock drift and it has been found that for most free running oscillators used in current computer systems, this change happens at intervals on the order of 1-2 hours or more [47]. That is reasonable to expect since macroscopic factors that heavily influence crystal oscillators, such as temperature change no faster than that rate.

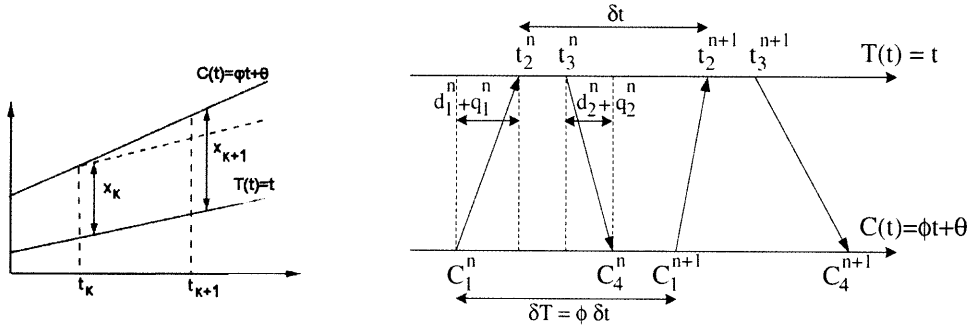

Figure 5-1: LEFT: Frequency offset $\phi - 1$ and time offset θ of $C(t)$, compared with the source of "true" time $T(t)$. RIGHT: Exchanging timestamps between client and time server. Notice that a time difference of δt according to server clock is translated to $\phi \delta t$ according to client clock.

A statistical tool that provides a stationary measure of the stochastic behavior regarding time deviation residuals and their associated frequency fluctuation estimates, is the Allan variance [4]. Allan variance associates frequency fluctuation estimates with specific observation duration and therefore could be used to quantify how often the above clock parameters change. For an excellent review of oscillators, Allan variance, time and frequency metrology, the interested reader could refer to [48].

5.2 Centralized Network Time Keeping

5.2.1 Problem Formulation

After describing the clock nomenclature followed in this work, we are ready to formulate the problem. The client clock $C(t)$ is *synchronized* to a time source $T(t) = t$ when both frequency offset ϕ and time offset θ are estimated.

The client timestamps ($C(t_1)$) a UDP packet according to each own clock $C(t)$ and sends the message to a time source server which timestamps the packet upon reception and retransmission (t_2, t_3 respectively) back to the originating client (fig. 5-1). The client timestamps again the message upon reception and therefore acquires a set of 4 timestamps: $(C(t_1), t_2, t_3, C(t_4))$. For convenience, we will notate $C(t_1)$ as C_1 and $C(t_2)$ as C_2 from now on. The same process can be repeated for a set of N consecutive messages. Therefore we should answer the following questions:

- What is the optimal processing of N messages

  $(C_1^1, t_2^1, t_3^1, C_4^1), (C_1^2, t_2^2, t_3^2, C_4^2), \ldots, (C_1^N, t_2^N, t_3^N, C_4^N)$ so as to obtain unbiased estimates with minimum error?

- What is the cost of obtaining estimates of ϕ and θ in terms of bandwidth spent (number N,inter-departure time between packets)?

- Do the algorithms employed in the estimation of ϕ, θ impose special restrictions in the operation of client (or server) operating system (i.e. are there any major non-algorithmic modifications in the operation of existing client/time server daemons)?

The number of packets N exchanged between client and server (fig.(5-1-RIGHT)) is a crucial parameter of any algorithm eventually adopted, considering the heavy load of current Internet time servers, on the order of $1000 - 1200$ requests per second and increasing every year [49]. Moreover, the inter-departure intervals of the NTP-like messages should not be

very large since closely spaced packets ensure that the clock parameters are not changing during the measurements from N packets.

Finally, we need to emphasize that the queuing delay q_1 across the *forward* path (from client to server) is never constant and generally different from the queuing delay q_2 across the *reverse* path (from server to client) (fig. (5-1-RIGHT)). Moreover, since the messages are carried through UDP packets, the forward and reverse routes could be physically different and therefore the propagations delays [1] d_1, d_2 could be unequal across the forward and reverse paths.

$$d_1 + q_1 \neq d_2 + q_2 \tag{5.1}$$

5.2.2 Prior Art on Centralized Client-Server Schemes

NTP estimates the time offset using the 4 timestamps of a message, according to the following equation:

$$\hat{x}_n = \frac{C_1^n - t_2^n - t_3^n + C_4^n}{2} \tag{5.2}$$

Since the round-trip time (rtt) is on the order of a few msecs, the contribution of the frequency offset on the error for a single measurement is negligible (e.g. a 10 ppm oscillator for a 10 msec rtt exhibits 0.1 μsec which is on the order of "noise" due to the operating system) and therefore excluded from Eq.(5.2). The frequency offset can be estimated using several measurements of x.

From a closer look on Eq. (5.2), NTP estimates are erroneous by a quantity proportional to half the difference between forward and reverse path delays (*assymetry*).

$$\hat{x}_n = x_n + \frac{d_2^n + q_2^n - d_1^n - q_1^n}{2} \Rightarrow \tag{5.3}$$

$$\hat{x}_n = x_n + w_n \tag{5.4}$$

That is why the NTP error is upper bounded by half the round-trip time. If we make the

[1] Time needed for the first bit to arrive at the destination as opposed to transmission delay which is related to the speed of the link.

assumption that the assymetry, depicted as "noise" w_n in Eq. (5.4) for the $n-th$ NTP message, is an Additive White Gaussian, zero-mean random variable, then the estimate of Eq. (5.2) is the Maximum Likelihood estimate, equivalent to the efficient[2] minimum variance, unbiased estimator for this particular case, according to the Gauss-Markov theorem. However, the assymetry is not always Gaussian, as we discuss in the following sections.

Line fitting techniques, based on the median slope calculated from averaged one way delay measurements [65] or linear programming [56] are alternative proposals for frequency offset estimation. The linear programming technique proposed in [56] is revisited with a slightly different derivation which provides not only for frequency offset $(\phi - 1)$ estimation but also for time offset estimation (θ).

In the Gaussian case, averaging N measurements from Eq. (5.2) can improve the estimates (decreasing the standard deviation of the estimate) by a factor of $\sqrt{N}$. This is an idea exploited in the client-server synchronization schemes deployed by the National Institute of Standards and Technology (NIST) using dedicated phone lines [46] or the Internet [47]. A variant of this method is discussed in this work. A similar approach based on averaging is also investigated in [78].

Finally, Kalman filtering is an attractive alternative for clock parameter estimation [6], since Kalman filters are the optimal linear estimators for the Gaussian case i.e. the linear estimators that minimize the Mean Square Error (MSE) [25]. As we see in the next section, the problem can be formalized using the Kalman filtering notation and due to the optimality property (at least for the Gaussian case) excels over a range of recursive estimators like phased lock loops [6].

The optimality and the appealing recursive nature of Kalman filtering, the intuitive structure (as explained below) of the linear programming technique and the simplicity of the averaging technique (referred as *Averaged Time Differences (ATD)*) as well as its wide deployment, were the reasons behind the selection of the above algorithms for comparative performance evaluation.

[2]The efficient estimator when exists achieves the minimum variance of the estimate, equal to the Cramer-Rao bound.

5.2.3 The Algorithms

Kalman Filtering

The motivation behind the adoption of Kalman filtering stems from a simple observation: a time interval δt according to "true" time is translated to $\phi\,\delta t$ according to client clock. Therefore, it is sufficient for the client to send messages at constant intervals δT measured according to its local clock and estimate the inter-arrival intervals at the server, using the timestamps $\{t_2^n\}$ which correspond to "true" time. Variation of forward and reverse one-way delays are interpreted as *noise* in the estimation process.

With the above, the formulation of the problem using Kalman filtering becomes clear: the client sends the NTP packets at constant intervals δT and estimates the inter-arrival interval $s = \frac{\delta T}{\phi}$ in the presence of network delay variations v, exploiting the measured inter-arrival intervals $y_n = t_2^{n+1} - t_2^n$ for $n \in [1..N]$.

The measurement and state model of the Kalman filter easily follow (fig. (5-1-RIGHT)):

$$
\begin{aligned}
y_n &= t_2^{n+1} - t_2^n && (5.5)\\
&= t_1^{n+1} + d_1^{n+1} + q_1^{n+1} - (t_1^n + d_1^n + q_1^n) && (5.6)\\
&= \underbrace{t_1^{n+1} - t_1^n}_{\delta t} + \underbrace{(d_1^{n+1} + q_1^{n+1})}_{e^{n+1}} - \underbrace{(d_1^n + q_1^n)}_{e^n} \Rightarrow \\
y_n &= \delta t + e^{n+1} - e^n = \delta t + v_n, && (5.7)\\
s_n &\equiv \delta t = \frac{\delta T}{\phi}, \quad \text{n} \in [1..\text{N}] \Rightarrow && (5.8)\\
y_n &= s_n + v_n, \text{ measurement model} && (5.9)\\
s_{n+1} &= s_n + w_n, \text{ state model} && (5.10)
\end{aligned}
$$

The measurement noise v_n accounts for the variation of travel time, when the NTP message is transmitted from client to server and it is assumed a zero mean process throughout this work. This is the type of noise that depends on the network path between client and server. Its power can be minimized only if the client selects a shortest path route toward the server. The state model noise w_n accounts for the fact that inter-departure times between

consecutive packets from the client could not be constant, possibly due to operating system delay variations. The power of this noise process is fully controlled by the client and could be estimated by client's own timestamps $\{C_1^n\}$. Alternatively, we can treat that noise as additional measurement noise (v_n) and simply ignore it ($w_n = 0$). That is the approach followed in this work.

Assuming v_n a zero mean process and e^n (from Eq. (5.7)) a stationary, non-zero mean process with uncorrelated consecutive samples, the following equation is derived:

$$\begin{aligned} E[v_i\ v_j] &= \begin{cases} R & \text{i} = \text{j} \\ -R/2 & \text{i} = \text{j} + 1 \\ 0 & \text{otherwise} \end{cases} \\ R &= variance(y_n),\ \text{n} \in 1..\text{N} \end{aligned} \tag{5.11}$$

Under the above assumptions and using vector notation, the measurement and state model equations become:

$$\begin{bmatrix} y_n \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} \delta t \\ \delta t \end{bmatrix} + \begin{bmatrix} v_n \\ v_{n-1} \end{bmatrix} \Rightarrow \tag{5.12}$$

$$\overline{\mathbf{y}}_\mathbf{n} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \delta t + \overline{\mathbf{v}}_\mathbf{n}, \tag{5.13}$$

$$s_{n+1} = s_n = \delta t,\ \text{n} \in 1..\text{N}. \tag{5.14}$$

The Kalman filter *"predict"* and *"update"* equations are omitted and could be found in a relevant textbook [61]. The Kalman filtering technique is a recursive scheme, therefore the estimate s_n converges to the correct value of δt after a number of messages $(C_1^n, t_2^n, t_3^n, C_4^n)$. The initial predicted value $s_{0|-1}$ was set to δT while the associated error variance was set to R. After the Nth packet, the frequency of the client clock is obtained by the output $\hat{s} = s_{N|N}$ of the kalman filter:

$$\hat{\phi} = \frac{\delta T}{\hat{s}} = \frac{\delta T}{\hat{\delta t}} \tag{5.15}$$

From Eq. (5.7), averaging N measurements results in the following equation:

$$\frac{1}{N}\sum_{n=1}^{N} y_n = \delta t + \frac{1}{N}(\underbrace{e^2 - e^1}_{v_1} + \underbrace{e^3 - e^2}_{v_2} \ldots + \underbrace{e^{N+1} - e^N}_{v_N}) \tag{5.16}$$

$$\frac{1}{N}\sum_{n=1}^{N} y_n = \delta t + \frac{1}{N}(e^{N+1} - e^1) \tag{5.17}$$

The average value of N measurements could be used as a "naive" estimator of δt (and consecutively of clock rate via Eq. (5.15)). The variance of this estimate, under the same assumptions for the noise process v_n, drops with N^2, since $var(e^{N+1} - e^1) = 2\ var(e_n) = var(v_n)$. Despite its attractive simplicity, this estimator provides large errors, compared to all the other approached presented in this work, especially when small number (N) of messages are used, as we will see in the following sections.

For the estimation of time offset θ we could use Eq. (5.2). However for a large number N of packets used, the duration of the experiment multiplied by the frequency skew could contribute to a significant synchronization error (e.g. 100 packets spaced 1 sec from each other correspond to an additional time offset of 4 msec for a 40 ppm clock). Therefore the estimate of the frequency offset should be exploited in the time offset calculation.

From fig. (5-1-RIGHT) we have the following relationships:

$$C_1^n - \phi\, t_2^n = \theta - \phi\,(d_1 + q_1)^n \tag{5.18}$$

$$C_4^n - \phi\, t_3^n = \theta + \phi\,(d_2 + q_2)^n \tag{5.19}$$

$$\Downarrow$$

$$C_1^n - \phi\, t_2^n \leq \theta - \phi\, d_1 \tag{5.20}$$

$$C_4^n - \phi\, t_3^n \geq \theta + \phi\, d_2 \tag{5.21}$$

Therefore, an estimate of θ is obtained by the following relationship:

$$\hat{\theta} = \frac{max(C_1^i - \hat{\phi}\, t_2^i) + min(C_4^j - \hat{\phi}\, t_3^j)}{2} \tag{5.22}$$

Alternatively, Kalman filtering could be used again for the estimation of time offset θ. The estimate of clock rate ϕ from the above technique could be exploited to "adjust" the timestamps $C_1^n \leftarrow C_1^n/\hat{\phi}$, $C_4^n \leftarrow C_4^n/\hat{\phi}$ at the client side. Then measurements of time offset θ according to Eq. (5.2), could be filtered using standard, one dimensional Kalman equations, with measurement model given by Eq. (5.4). The output estimate of θ after Kalman filtering of N measurements is also reported in the experimental results section.

Linear Programming

This line fitting technique exploits both the forward and reverse path timestamps, by estimating a "clock line" that minimizes the distance between the line and the data, leaving all the data points below the line on a (t_2, C_1) plane or above the line on a (t_3, C_4) plane. The following equations describe the problem and it's solution:

.....forward path (from client to server)

$$\begin{aligned} \alpha_1 &= \phi \\ \beta_1 &= \theta - \phi\, d_1 \end{aligned} \tag{5.23}$$

$$\text{Eq. (5.20)} \Rightarrow \alpha_1\, t_2^n + \beta_1 - C_1^n \geq 0,\ \forall\, n \in [1..N] \tag{5.24}$$

Find α_1, β_1 that minimize

$$f(\alpha_1, \beta_1) = \sum_{n=1}^{N} (\alpha_1\, t_2^n + \beta_1 - C_1^n)$$

under the constraint of Eq. (5.24)

.....reverse path (from server to client)

$$\begin{aligned} \alpha_2 &= \phi \\ \beta_2 &= \theta + \phi\, d_2 \end{aligned} \tag{5.25}$$

$$\text{Eq. (5.21)} \Rightarrow C_4^n - \alpha_2\, t_3^n - \beta_2 \geq 0,\ \forall\, n \in [1..N] \tag{5.26}$$

Find α_2, β_2 that minimize

$$f(\alpha_2, \beta_2) = \sum_{n=1}^{N} (C_4^n - \alpha_2 \, t_3^n - \beta_2)$$

under the constraint of Eq. (5.26)

$$\hat{\phi} = \frac{\alpha_1 + \alpha_2}{2} \tag{5.27}$$

$$\hat{\theta} = \frac{\beta_1 + \beta_2}{2} \tag{5.28}$$

The simple and intuitive derivation above sets this technique as a strong candidate for clock parameter estimation.

Averaged Time Differences

This method can be best described by fig. 5-1. The time offset x_n is computed according to the NTP formula (Eq. 5.2) and therefore this method has all the limitations discussed at the NTP section above. Differences of the time offset estimates provide estimates for the frequency offset. Particularly, clusters of 25-50 closely spaced messages are used, time offsets are computed and the results are averaged to a single data point for the time offset. Then that data point is used in the following formula for frequency offset estimation.

$$\hat{f}(t_{n+1}) = \frac{x_{n+1} - x_n}{\tau} \tag{5.29}$$

$$\hat{f} \equiv \hat{\phi} - 1$$

$$y(t_{n+1}) = \frac{y(t_n) + \alpha \hat{f}(t_{n+1})}{1 + \alpha} \tag{5.30}$$

The value of τ nominally should be equal to $t_{n+1} - t_n$ however this quantity cannot be measured by the client's own clock. Nevertheless for small values of the frequency offset this can be set to $C(t_{n+1}) - C(t_n)$, since that is what the client can measure. The estimated

frequency offset is averaged again using an exponential filter with a time constant α that depends on the stability of the local oscillator. Then the filtered frequency offset is used in the following formula, which is also depicted in fig. (5-1-LEFT).

$$\hat{x}(t_{n+1}) = \hat{x}(t_n) + y(t_n)(\tau) \tag{5.31}$$

A variant of this method is used in this work. Frequency offsets are calculated using Eq. (5.29) and then filtered using the above exponential filter with $\alpha = 0.5$. The final frequency offset estimation is the mean of all the N exponentially filtered frequency offsets calculated at each epoch.

The power of this method is its simplicity. For the Gaussian case where consecutive measurements are independent from each other, an increase of samples averaged by a factor of N reduces the variance of the estimate by a factor of N. Therefore, there is a trade-off between accuracy achieved and cost of realizing it.

5.2.4 Performance

In this section we evaluate the performance of the three algorithms in two separate cases:

- The "Gaussian case" where the queuing delay difference between two consecutive NTP messages is a Gaussian random variable. Consequently, the dispersion of the packets at the server is also a Gaussian random variable. In this experiment, consecutive measurements are independent.
- The "Self-Similar case" where multiple pareto connections aggregate and form cross-traffic with long-range dependence.

The estimate, the variance of the estimate and the number N of packets used at each epoch are reported. In both cases the true clock frequency offset $\phi - 1$ was $+40$ ppm and the time offset θ was 20 msec. The NTP messages were transmitted at intervals of 1000 msec. Each experiment was run 300 times.

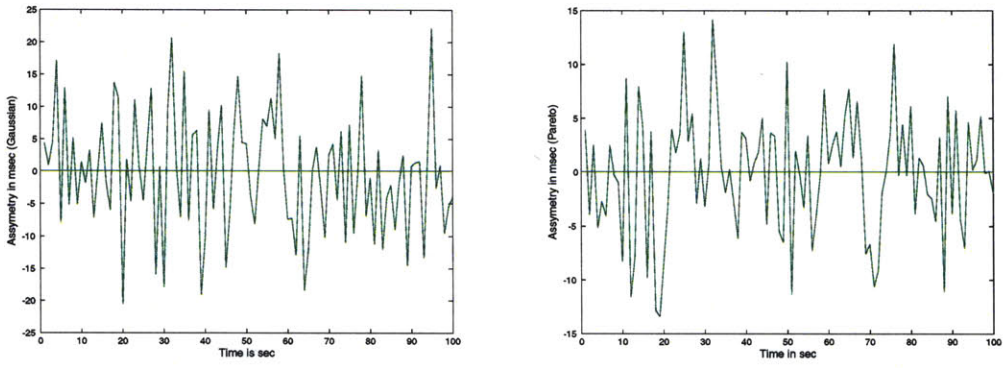

Figure 5-2: Asymmetry of delays between forward (to server) and reverse (to client) path. LEFT: Gaussian case. RIGHT: Self-similar case.

The Gaussian Case

In fig. (5-2-LEFT) we present the asymmetry between forward (to server) and reverse (to client) path, from a sample run. The average round-trip time was on the order of 40 msecs and consecutive measurements were independent and identically distributed. In figure (5-3) and figure (5-4) we present the estimate and the standard deviation of the estimate for the frequency offset $\phi - 1$ and time offset θ respectively, as a function of number N of packets used.

The Kalman filter performed better when the number of packets N was above the minimum number of samples needed for convergence (on the order of 25-30 packets). This experimental finding is validated by the fact that the Kalman filter (at steady-state) is the optimal linear estimator in the presence of Gaussian noise. The LP technique performed better than both averaging techniques (ATD and "Naive" estimator), which performed well only if large number of messages were used.

Frequency offset estimate variance was decreased with number N of packets used. From fig. (5-3) it is shown that the standard deviation of the estimate drops slightly faster than linearly with N (variance drops with N^2) for Kalman filtering, while it drops linearly with N for the "Naive" estimator, as expected, while the variance as well as the error is smaller

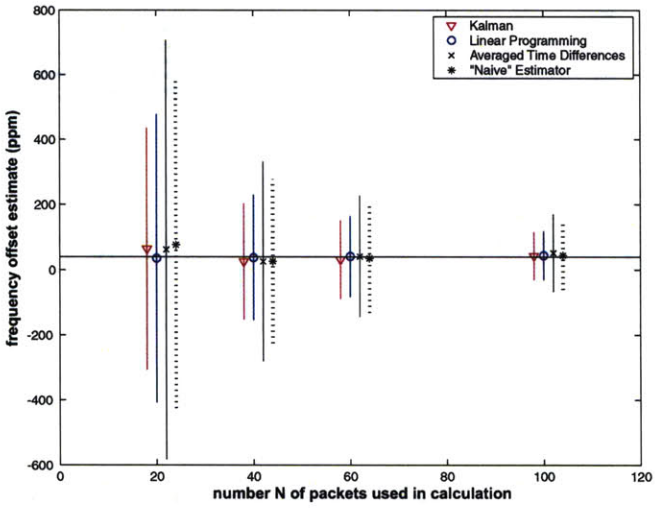

Figure 5-3: Gaussian case. Frequency offset estimate and standard deviation as a function of N (number of packets used).

for the Kalman algorithm. Time offset estimates were close to the real value, regardless of N. This can be justified by the fact that the algorithms presented here focus on the accurate calculation of frequency offset which was set at 40 ppm in this experiment. Error in the calculation of a 40 ppm quantity over a duration of 100 sec (1 packet every 1000 msec) is negligible in the calculation of time offset (using the algorithms described above)[3] and of course not visible at the time scales of fig. (5-4).

The Self-Similar Case

In this section, we are investigating the performance of the three algorithms in the presence of bursty traffic. It has been shown that the aggregation of many on/off sources could form a self-similar source, exhibiting long range dependence [74]. The fact that Local Area

[3]Time offset θ was estimated using the same algorithm for Kalman, ATD and "Naive", described in the Kalman filtering section. Kalman filtering for both time and frequency offset, is depicted as Kalman/Kalman.

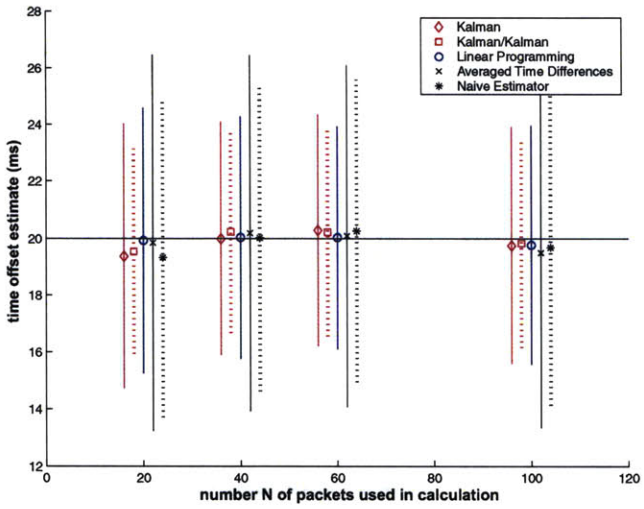

Figure 5-4: Gaussian case. Time offset estimate and standard deviation as a function of N (number of packets used).

Network traffic demonstrates chaotic (self-similar) behavior [45] motivates the test of the three algorithms in a self-similar environment which is fundamentally different from the Gaussian case for which Kalman filtering seems appropriate.

Fig. (5-5) displays the simulation setup in Network Simulator 2 (ns-2) [87]. The utilization of the links was 90%, the average round-trip time on the order of 40 msecs and the assymetry between the forward and reverse path is depicted in fig. (5-2-RIGHT). The inter-departure time of NTP packets remains 1000 msecs.

Fig. (5-6) shows in a sample run how well the Kalman filter "locks" onto the correct inter-arrival time δt and frequency offset value $(\phi - 1)$. Fig. (5-7-RIGHT) shows how well the ATD technique (with the exponential filter) "locks" onto the frequency offset value $(\phi - 1)$. The internal line is the filtered waveform through a low pass filter. Fig. (5-7-LEFT) displays the one-way delay across the reverse path as a function of time. The trend of the plot is coherent with the following derivation. The "clock line" $\hat{\phi}\, t_3^n + \hat{\theta}$ with parameters estimated

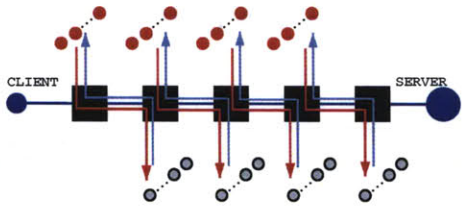

Figure 5-5: Simulation in ns-2 with pareto cross traffic. 14 connections per link per direction.

by the LP technique is also depicted.

$$
\begin{aligned}
Eq.\ (5.21) \quad &\Rightarrow \qquad\qquad (5.32)\\
C_4^n - t_3^n \quad &= \quad (\phi - 1)\, t_3^n + \phi\, (d_2^n + q_2^n) + \theta \Rightarrow\\
C_4^n - t_3^n \quad &\geq \quad (\phi - 1)\, t_3^n + \phi\, d_2^n + \theta
\end{aligned}
$$

Fig. (5-8) shows the histogram of frequency offset estimates for the self-similar case, for $N = 100$ and fig. (5-9-TOP) shows the performance of the three algorithms in the estimation of frequency offset, for various number N of packets used in the calculation. Time offset estimation $\hat{\theta}$ resulted in significant errors due to assymetry between forward and reverse path, as expected (fig. (5-9)-BOTTOM)).

From the above diagrams, it is deduced that the Kalman filtering technique no longer produces the best estimates with the smallest variance. The noise is no longer Gaussian so Kalman filtering is not optimal and LP performs better in the presence of bursty traffic both in terms of estimation error (accuracy) and its variance (precision). For the same reason (burstiness and asymptotically long range dependence as opposed to the Gaussian distribution around the mean), ATD and "Naive" estimator perform inferiorly than the LP technique. All algorithms for frequency offset estimation reduce the standard deviation (and therefore variance) of the estimate with increased number N of packets used, and the

relation between that improvement and N seems faster than linear for the case of Kalman filtering and Linear Programming or linear for the case of averaging (therefore variance drops with N^2), as can be seen in fig. (5-9-TOP).

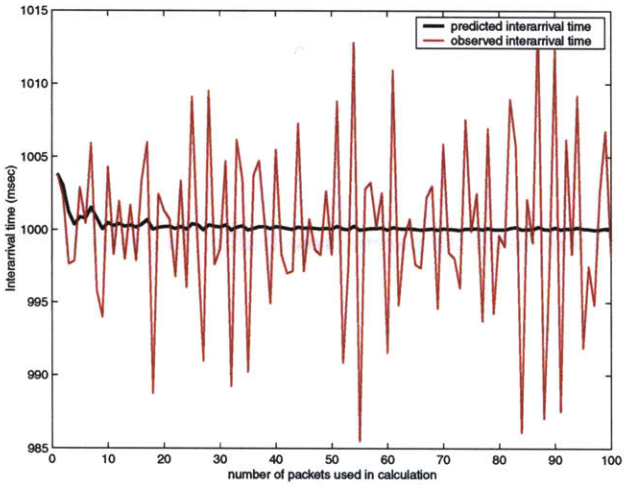

Figure 5-6: Predicted inter-arrival and measured inter-arrival interval using the Kalman filter for self-similar cross traffic.

5.2.5 Measurements

In order to emphasize the end-to-end character of the algorithms evaluated (especially for the case of Kalman filtering and LP), we modified NTP client daemon and exchanged 100 packets at intervals of 1 sec with a stratum-0 server (connected to GPS). The time server was geographically located at Palo Alto CA, 3100 miles away from our client machine, with average round-trip time 85 msec, 18 hops away. We then processed the packets according to the algorithms evaluated above and the frequency offset estimation results are presented at Table 5.1.

An interesting idea could be averaging the two estimates calculated according to Kalman and LP since the former performed better at the Gaussian case and the latter at the Self-similar one.

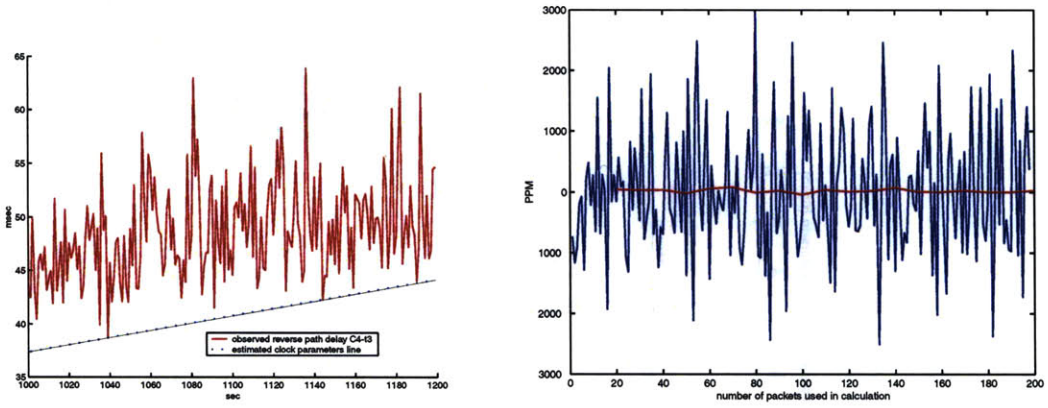

Figure 5-7: LEFT: Delay $C_4^n - t_3^n$ from the reverse path and clock line estimation using LP for self-similar cross traffic. RIGHT: Estimation of frequency offset $\phi - 1$ using the ATD technique. Low pass filtering of data is also plotted.

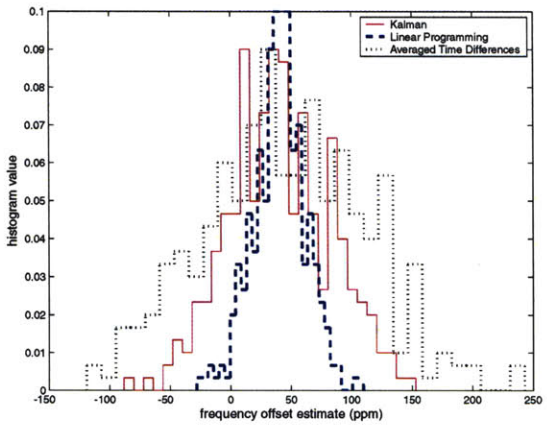

Figure 5-8: Histogram of the frequency offset estimates for self-similar cross traffic.

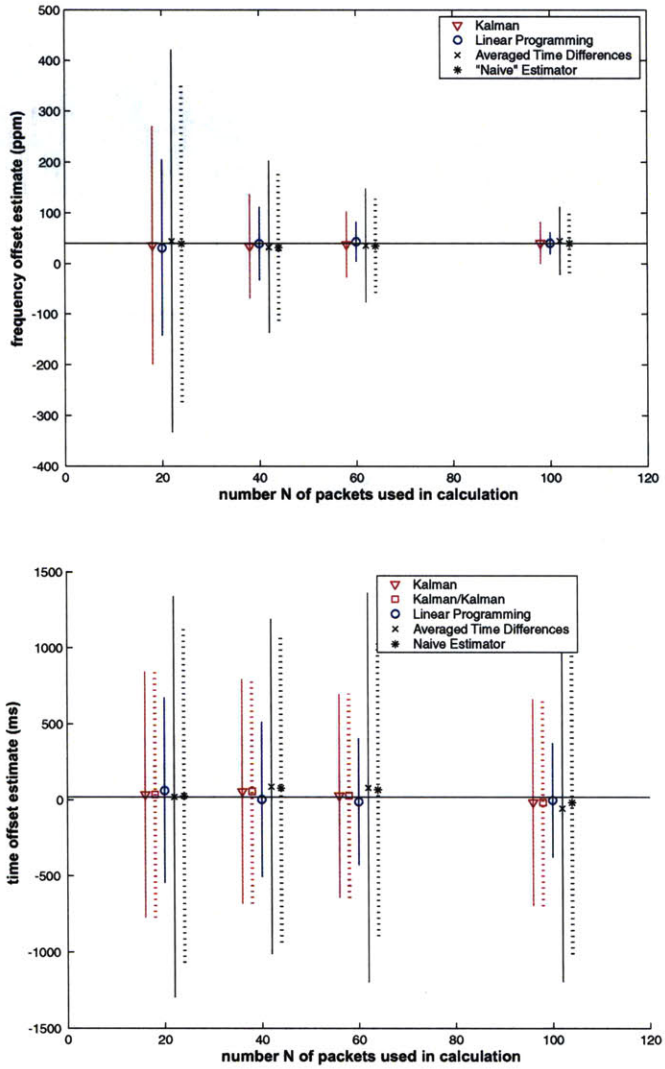

Figure 5-9: Self-similar case. TOP: Frequency offset estimate and standard deviation as a function of number N of packets used in calculation. BOTTOM: Time offset estimate and standard deviation as a function of number N of packets used in calculation.

Table 5.1: Frequency offset estimation using an existing NTP/GPS server.

	Kalman	LP	ATD
$\widehat{\phi} - 1$ (PPM)	101.6	54.7	122.7

5.2.6 Discussion

The Kalman filtering technique, optimal for the Gaussian case, needed a considerable number of packets in order to converge (on the order of 20-30 packets for the formulation adopted and the experimental setup). Nevertheless, the technique performed well at the Self-similar case as well, with improved performance in terms of error and variance of the estimate when the number of packets N increased. The algorithm estimates the variance of network delay (jitter) and uses that estimate to calculate the frequency and time offset model variables. The algorithm can be applied without major operation requirements in the NTP-client daemon and requires no modifications in the NTP-server daemon. It could be benefitted by scheduled transmission from the client system that ensure minimum delay variance due to the operating system.

The Linear Prediction technique surpassed all the other techniques at the case of bursty traffic approximating real-word long-range dependence (chaotic) conditions even though it had inferior performance when measurements where completely independent. Its intuitive structure makes it attractive for straightforward implementation.

Finally, averaging as exploited and implemented in the Averaging Time Differences technique *(where equal intervals between measurements were used and therefore averaging differences of time was equivalent to averaging frequency offset estimates)* performed inferiorly to the LP and Kalman filtering techniques at the Self-similar case where measurements are not independent. However, its simplicity makes it attractive, especially at the cases where a small number of measurements are available or a trade-off between accuracy and cost of realizing it cannot be avoided.

All three techniques showed improvement in terms of frequency offset estimation error and

variance of the estimate with increasing number of packets N. For Kalman filtering, the relationship between improvement and N seems slightly faster than linear (for standard deviation of frequency offset estimate) or quadratic (for variance of frequency offset estimate) and therefore increased accuracy is expensive in terms of number of packets used (communication bandwidth). This work tried to quantify that cost, by comparatively evaluating a number of diverse techniques.

5.3 Decentralized Network Time Keeping

In this work we implement a time synchronization technique for multi-hop, energy, communication, computing constrained sensor networks which is completely beacon (or server) free. Moreover, it requires no global coordination since all nodes in the network communicate with nearest neighbors for time-synchronization purposes. Therefore the scheme has no centralized point of control (or failure) and, it has no network routing overhead and it is appropriate for ad-hoc sensor networks where the topology might change (often due to mobility) or might be unknown.

Our initial goal was an experimental evaluation of time synchronization in multi-hop networks, in a real-world setup. For that cause, we implemented a distributed orchestra, where each node could have a speaker to output a song, while at the edges of the network, two nodes were equipped with LED displays (figure 5-10). At the same time, we wanted to quantify in practice, the observed accuracy and precision of the algorithm against its required communication and computation overhead using our embedded wireless network. Evaluation of the scheme through implementation in a real-world embedded network reveals the important limitations on computation, communication and complexity sensor networks encompass.

Evaluations of synchronization schemes only through simulations usually underestimate the limited resources in terms of memory, computation and communication of each node and also assume worst case scenarios that might not reflect reality. Even though experimental study of time synchronization has been reported before in *single-hop* embedded wireless networks,

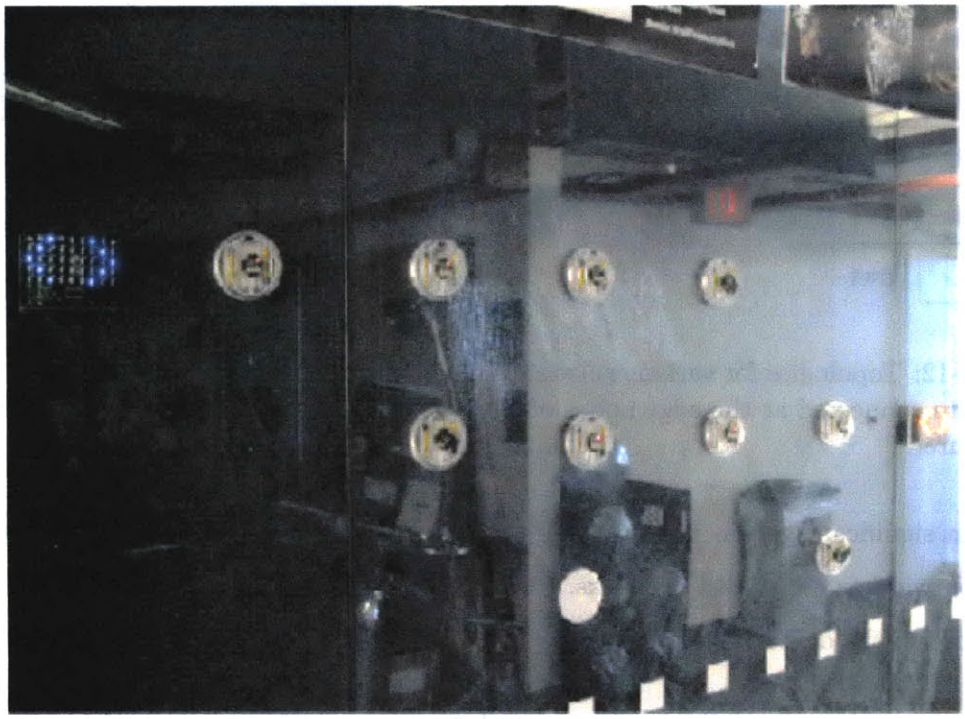

Figure 5-10: Demo on a glass wall: each node can communicate with at most 4 immediate neighbors. The network manages to synchronize all nodes so that they can "output" through speakers the same music. At the edges of the network, the nodes are equipped with LED displays instead of speakers, to provide for visual proof of synchrony. All nodes are communicating with immediate neighbors only and there is no point of central control.

Figure 5-11: The individual nodes used in this work. Speakers and displays provided for audio-visual output. LEFT: 4-IR Pushpin without speaker. The four IR transceivers provide directional communication only along the horizontal and vertical axis. CENTER: 4-IR Pushpin with speaker. RIGHT: 45-LED display. A 4-IR Pushpin is connected behind the LED grid.

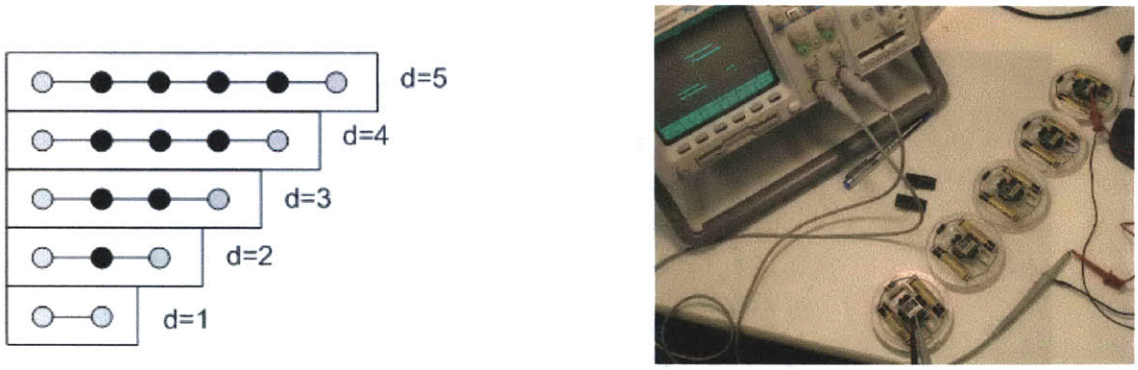

Figure 5-12: Topologies for various network diameters d used in this work. The oscilloscope probes are connected at the edge nodes of the network. The case for $d = 4$ is shown in the right figure.

there is a significant gap in measurements of time synchronization error in realizations of *multi-hop* wireless embedded networks. To our knowledge, this work is the first to fill this gap. Video of the demonstration could be found at [88].

The proposed scheme and the implemented demo were inspired by natural phenomena of synchronization: the way fireflies blink in unison, even though they interact only locally or the way cardiac neurons fire in sync.

5.3.1 Experimental Setup

The goal in this work is to demonstrate a time synchronization scheme that would be:

a) *transparent* to the sensing or actuating tasks of any node in the network. Each node should communicate only locally with its immediate neighbors and avoid explicit connections to remote servers of "true time" one or more hops away.

b) *self-calibrating* with no coordination requirements upon deployment or during operation. The multi-hop network should *spontaneously* converge to a common time reference without centralized control.

To make matters more realistic, we choose to evaluate the transparent and self-calibrating (as defined above) character of the scheme at the extremes: we evaluated the scheme at the

edges of the network, when connectivity is established only through intermediate nodes. RF communication range could be on the order of hundreds of meters, therefore it would be more appropriate to utilize short-range and directive communication links in order to demonstrate multi-hop performance. We used 8051-based micro-controllers (8-bit, 2 Kbytes of RAM and 32 Kbytes of program space) connected to short-range, 4-way infrared transceivers. Those are the pushpin nodes [50], [90], that we packaged in round battery holders as shown in figure 5-11. Pushpins practically allowed evaluation of the synchronization scheme at the edges of the network, for several values of network diameter d as shown in fig. (5-12-LEFT). The experimental setup for $d = 4$ is shown in fig. (5-12-RIGHT).

The goal was to demonstrate network multi-bit clock synchronization among all nodes in a distributed fashion, not just synchronization to a reference signal coming from a specialized server [52], [10] or beacon [18]. No prior knowledge of network topology was assumed and all nodes would be loaded with the same code. All nodes could be equipped with small speakers (fig. 5-11) and as a proof of synchrony they would play the same piece of music at the same time. According to ([15] p.95), the smallest perceivable time difference from humans is on the order of 30-50 milliseconds, therefore clock synchronization error above that limit could be perceived.

Apart from the oscilloscope measurements at the edge networks and the audio outputs at many intermediate nodes, visual patterns at the edge nodes could provide for visual proof of synchrony. Displays from the rf-Badges [39], [89] were connected to 4-IR pushpins and used in this work (fig. 5-11).

5.3.2 The Algorithm and its Implementation in our Embedded Network

Lamport, in his 1978 seminal work in the context of computer clocks and processes synchronization [40], described a simple algorithm, based on the fact that *time* is a strictly monotonically-increasing quantity. Therefore events happening in subsequent times should have timestamps ordered accordingly, otherwise a correction in the clocks should be made. Although Lamport's work has been extensively referenced in the area of sensor networks

time synchronization, there has been no validation and testing in embedded networks so far (at least to the extent of authors knowledge). Since time is viewed as a non-decreasing quantity in Lamport's algorithm, its implementation has probably been considered problematic in memory-restricted and communication-constrained sensor networks.

- **Broadcast:** node i transmits its clock value $C_i(t)$ at regular intervals. Time-stamping occurs just before transmission and the MAC protocol has been modified accordingly.
- **Receive and Compare:** upon reception from node j of a clock value $C_i(t)$ from node i, node j compares and keeps the highest value: if $C_i(t) > C_j(t)$ then $C_j(t) \leftarrow C_i(t)$ else ignore.

In this work, we modify Lamport's algorithm to fit the memory and communication constraints of sensor networks and through implementation in a multi-hop, embedded network, we prove that the new algorithm can sufficiently synchronize the whole network, in a distributed, *transparent* and *self-calibrating* way, satisfying many real-world scenarios.

The first modification in Lamport's algorithm is that time is no longer considered a monotonically increasing quantity: clock $C_j(t)$ in every network node j is bounded above and upon reaching that bound, time is reset. Therefore, clock function $C_j(t)$ follows a "saw-type" periodic waveform and its period should be set according to the natural phenomenon which is sensed by the sensor network. In this work, since the goal was distributed synchronized play of music, the period T of each clock was set to 13 seconds approximately.

The first reason behind upper bounding time, was the fact that timestamps are communicated among neighboring nodes and therefore their size in bits should be kept minimal, because of memory, bandwidth and energy constraints. In this work, clock value $C_j(t)$ of node j is represented by an unsigned 16-bit variable, incremented each time another 16-bit *counter* resets. This reset occurs every 5.9 msecs approximately, limiting the resolution of each clock variable $C_j(t)$ in the millisecond regime. The counter is interrupt-driven and since it controls time increments, it is assigned the highest priority interrupt.

The second reason behind upper bounding time, was our desire to explicitly study *self-calibration* capabilities of the algorithm and show in practice that even though clocks reset periodically (in this case, every 13 seconds), the network as a whole re-synchronizes quickly and unattended (*spontaneously*), while it performs its sensing and actuating tasks.

Note that in this realization, we have time $C_j(t)$ of node j to be represented as 16-bit integer, with resolution set by another 16-bit counter. However, only the first 16-bit value is communicated to nearest neighbors. The length of the clock value $C_j(t)$ (in bits) and its resolution both depend on the physical phenomenon that needs to be sensed. For example, for environmental sensing of moisture, a 16-bit clock incremented every 1.3 seconds would need 24 hours approximately, to reset. Therefore, the same "saw-type" definition of time would suffice, the information communicated over the network would be the same as in the example of this paper and the only modification would be in the clock resolution in every node of the network. The network would reset in synchrony every 24 hours instead of 13 seconds. The slower period in this work (and resolution on the order of milliseconds) allowed us to quickly validate the fact that the network *re-calibrates* after every clock variable $C_j(t)$ expiration, without unwanted periods of instability.

In other words, length and resolution of the clock variable $C_j(t)$ depend on the physical phenomenon to be sensed and the introduced algorithm can be used with success in many different contexts and applications such as environmental sensing.

The second modification in Lamport's algorithm is the fact that broadcasting of time-stamped information is controlled by an independent timer and not by the clock of each node. The reason behind such implementation decision was that we wanted to decouple the two stages of the algorithm (broadcast and receive), simplify design and avoid *bootstrapping* problems that might occur, if we had used the same timer to control both *when* as well as *what* to transmit. Time-stamping during the broadcast phase occurred just before transmission, therefore the Medium Access Control (MAC) protocol in every node had been modified accordingly.

Table 5.2 lists the clock period T and the resolution of each node's clock, the time needed for each node to transmit timing packet information to its neighbors and how often every node

Table 5.2: Period and resolution of each clock, transmission delay and bandwidth used for timing packets (in packets per second).

	T of $C(t)$	res of $C(t)$	tx delay	bw
r1	13.2 sec	5.9 msec	1.24 msec	0.3 pps
r2	13.2 sec	5.9 msec	1.24 msec	3 pps

broadcasts its clock value in packets per second (pps), for two scenarios ($r1$, $r2$) evaluated in this work.

It is important to note that the packet each node transmits at regular intervals $\frac{1}{bw}$, contains only the 16-bit time variable, a protocol header byte and one additional byte with Cyclic Redundancy Check (CRC) information. In other words, the 4-byte packet transmitted contains no information about node source id, destination id or any other kind of routing information, since communication is happening with nearest neighbors. Therefore, the synchronization scheme is *transparent* (as defined previously) to the sensing or actuating task of each node.

The algorithm as modified, customized and implemented in this work is aimed to provide for distributed, unattended and spontaneous synchronization and will be evaluated in practice in the following section. Naturally, we called the new scheme "Spontaneous Synchronization". Video of the demonstration could be found at [88].

5.3.3 Results

We run experiments with duration 500 seconds each and measured the absolute synchronization error $|C_i(t) - C_j(t)|$, where nodes i, j are the edge nodes of the network as shown in fig. 5-12. To do so, each node output a pulse, when its clock variable reached a specific value ($C_i(t) = max/2$)[4]. We have already described that time is represented by an unsigned 16-bit integer (reaching its maximum value and then resetting every T seconds), incremented from the overflow of a 16-bit counter (controlled by the crystal oscillator of

[4]note that max need not be $2^{16} - 1 = 65535$ but it could be set to a smaller value: $max = \frac{T}{res}$.

each node and overflowing every *res* milliseconds, from table 5.2). Therefore, we measured the absolute synchronization error at the edges of the network every T seconds and for $T \simeq 13$ *sec*, the 500 seconds experiment corresponded to 37 measurements per experiment.

The network managed to synchronize all individual nodes so they could play the same piece of music repeatedly, as long as the nodes were switched on. That provided a quick proof of synchronization error smaller than 30 milliseconds, since that is the smallest time difference perceived by humans ([15], p.95). Moreover, that demonstration assured that time resetting at each individual node didn't cause instabilities but on the contrary, the network managed to re-calibrate and converge to a common time reference, continuously and unattended.

The oscilloscope measurements helped us quantify the performance of the synchronization scheme (fig. (5-12-RIGHT)). Average absolute error $|\widehat{\epsilon(t)}|$ and its standard deviation for different network diameters are shown in fig. 5-13. All experiments were run twice since apart from network diameter (d), we wanted to study performance against different bandwidth (bw from table 5.2) used for broadcasting time (broadcast phase of the algorithm).

From fig. 5-13 we can see that the absolute synchronization error $|\epsilon|$ is on the order of a few milliseconds. This is not a surprising result since the clock resolution of each network node is on the order of milliseconds (table 5.2). Moreover, as we will see below, the synchronization error depends on the transmission delay which is, again, on the order of milliseconds. We discuss ways to reduce the error, because of those two factors, down to the μsecond regime in section 5.3.4.

What is surprising about these measurement results, is the fact that synchronization error does NOT increase linearly with the diameter of the network as it has been reported previously in simulation setups. A simple analysis follows to justify the above findings: we could model the timer $C_i(t)$ of each node i as a linear function. Time increases with a rate ϕ_i that depends on the crystal oscillator of each node. The difference $\phi - 1$ is called frequency skew and for the crystals used in our nodes, it is on the order of $\pm$ 50 parts per million (ppm).

We ignore for now the fact that time resets at each node and we assume that node i transmits its timestamp at time t_0. The packet will be received and processed by neighboring node j

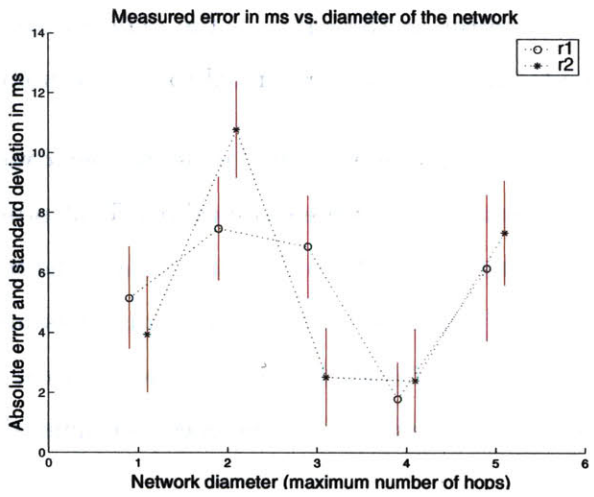

Figure 5-13: Measured average time synchronization absolute error and its standard deviation in milliseconds, as a function of network diameter. Clock resolution and transmit time is on the order of milliseconds, limiting the error in the millisecond regime, as expected. Notice that error is not increased linearly with number of hops, since error depends on the sign of clock drift differences between neighboring nodes (equation 5.37). $r1$, $r2$ correspond to different communication speeds (table 5.2)

at time $t_0 + x$.

$$C_i(t_0) = \phi_i \, t_0 + \theta_i \tag{5.33}$$

$$C_j(t_0 + x) = \phi_j \, (t_0 + x) + \theta_j \tag{5.34}$$

$$\begin{aligned} x = \; & \textit{propagation delay} \; + \\ & + \textit{transmission delay} \; + \\ & + \textit{operating system delay} \end{aligned} \tag{5.35}$$

Time duration x includes the propagation time of the signal which is basically the time needed for the first bit to arrive at the destination (distance/speed of light), the transmission time which is the time needed for the transmitter electronics to transmit the waveform (tx delay at table 5.2 in section 5.3.2) and finally the time needed at the operating system at the receiver to process the received packet. Propagation delay is negligible, on the order of a couple of μseconds for short range transceivers, therefore x is dominated by tx delay and

os delay. In our system, tx delay is 1.24 msec (since we are using slow transceivers) while operating system delay has been kept one order of magnitude smaller, given the fact that we are using pipelined, RISC micro-controllers driven by 22.11084 MHz crystals. Medium Access Control has been modified in order to avoid adding delays in the transmission of timing packets. If $C_i(t_0) > C_j(t_0 + x)$ then $C_j(t_0 + x) \leftarrow C_i(t_0)$ and the absolute error $\mid \epsilon \mid$ at time $t_0 + x$ becomes:

$$\begin{aligned} \mid \epsilon(t_0 + x) \mid &= \mid C_i(t_0 + x) - C_j(t_0 + x) \mid \\ &= \mid C_i(t_0 + x) - C_i(t_0) \mid \Rightarrow \\ \mid \epsilon(t_0 + x) \mid &= \phi_i \, x \end{aligned} \tag{5.36}$$

Therefore, the error at time $t_0 + x$ is on the order of $(1 \pm 50 \ 10^{-6}) \ x \approx tx \ delay = 1.24 \ msec$. Thereinafter, the error might increase or decrease depending on the frequency skew differences of node i, j clocks, since it is not difficult to see that according to this linear representation of time in equation 5.33, the error at time $t_c > t_0$ becomes[5]:

$$\begin{aligned} \epsilon(t_c) &= C_i(t_c) - C_j(t_c) = \\ &= \epsilon(t_0 + x) + (\phi_i - \phi_j) \ \Delta t \end{aligned} \tag{5.37}$$

$$\Delta t = t_c - (t_0 + x) \tag{5.38}$$

We can see that the error at time t_c might decrease if $\phi_i - \phi_j < 0$ or increase if $\phi_i - \phi_j > 0$. The amount of increase or decrease is on the order of $(50 \ 10^{-6} - (-50 \ 10^{-6})) \ T/2 \approx 650 \ \mu sec$, since we have at least one packet transmission per T seconds. From the above, it is straightforward to understand that the measured absolute error might decrease below *tx time* and there were occasions when the absolute error could drop at the *μsecond* regime.

The fact that time resets at each node does not affect the above analysis: reseting changes θ at each clock, not ϕ (which depends on the crystal oscillator on-board) and time differences

[5]provided that there is no time modification during the receive-and-compare phase of the algorithm at node j

using our algorithm depend on frequency skew differences $\Delta\phi$ (equation (5.37)), therefore changes of θ due to reseting, don't matter.

Even in the case of a node's clock resetting and then receiving a clock value from another clock which is close to reset, it can be seen that there are no instabilities in the overall system: both clocks eventually reset and the synchronization error between them starts from ϕ s and increases or decreases, depending on the sign of their frequency skew difference.

From fig. 5-13 we can see that increasing the broadcasting rate from 0.3 packets per second (r1) to 3 packets per second (r2), doesn't dramatically affect the overall error, since that increase of rate just decreases Δt in equation (5.37) but it doesn't affect x which is the dominating factor in the error. Increasing the broadcast rate (or decreasing Δt) allows for finer increase or decrease of the error (on the order of 650 $\mu sec/10 = 65$ μsec for r2 compared to 650 μsec for r1). Increasing the broadcast rate would make more sense for oscillators with higher frequency skew, than those used in this work ($\pm$ 50 ppm).

From the above analysis, it is now obvious why the average absolute error is not increasing monotonically with the diameter of the network. That is because the error as we saw, depends on the sign of the frequency skew among the clocks (equation 5.37), therefore by inserting additional nodes in a chain topology (fig. (5-12-LEFT)), the sign might be negative, leading to smaller synchronization errors. Analysis that shows that error increases linearly with the diameter of the network [40] assumes worst case scenarios i.e. the sign of $\Delta\phi$ in equation (5.37) is always positive, therefore the error builds up with the number of hops. This interesting behavior as depicted in fig. 5-13 would not have been observed if we hadn't implemented our algorithm in a real-world embedded network.

5.3.4 Further Improvements

The synchronization error could be further reduced by minimizing x. That can be achieved if the packet transmission time (which is deterministic and known) is incorporated in the transmitted timestamp during the broadcast phase of the algorithm. That basically means that each node broadcasts at time t, $C(t)+tx\ time$ instead of $C(t)$. Moreover, the operating

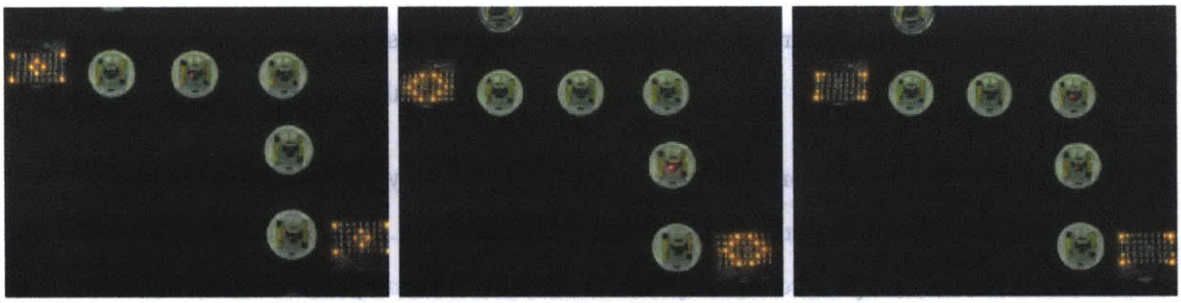

Figure 5-14: Visual proof of synchrony. A "heartbeat" pattern is synchronized over the network and displayed at the edges. The distributed, server-free approach for network synchronization resembles the decentralized coordination of colonies of fireflies and inspired this work.

system delays could be minimized or anticipated (and therefore incorporated as well in the transmitted timestamp). It is also useful to reduce uncertainties due to the channel access scheme in the MAC layer (allowing for time-stamping at the MAC layer could be one solution).

We implemented the above modifications in a RF, embedded, single-hop network and the synchronization error was reduced down to the μsecond regime. The interested reader could refer to [8] for additional information regarding the RF, single-hop case.

5.3.5 Spontaneous Order and its Connection to Biological Synchronization

What we have seen so far, is that coupling between neighboring oscillators with similar (but not exactly the same) frequency skew and periodic (due to reset) time waveforms, is able to globally provide network synchrony.

This global phenomenon of sync emerged as a consequence of local interactions between homogeneous elements and resembles similar phenomena found in nature: the way fireflies manage to globally blink in unison, even though they interact locally or the way millions of cardiac neurons fire in sync to produce the cardiac pulse. Those phenomena depend on coupling between oscillators, they have nothing to do with averaging of similar quantities (like timestamps for example) and they are canonical examples of *entrainment* ([73], p.72).

In the above examples of entrainment, including our work, synchrony is not controlled by any centralized authority but it is the natural emergent result of local interactions.

Inspired by the fireflies phenomenon, we attached two display-equipped nodes (fig. 5-11) at the edges of the network in order to visualize synchrony (fig. 5-14). The displays output a "heartbeat waveform" synchronized by the distributed scheme presented in this work. The difference with fireflies is that fireflies need only 1-bit synchronization as opposed to the 16-bit synchronization presented in this work.

5.3.6 Relevant Work on Distributed Sync and Discussion

We were particularly interested in multi-hop time synchronization algorithms. Work using simulations and reported in [72], [22] and [27] fall into this category. The basic idea is that each node exchanges two-way timing information with its closest neighbors, its neighbors with their neighbors and so on, up until the reference node. In [72], the same clock model of equation 5.33 is used, the time offset between two nodes is linear (as we also saw in equation 5.37) and bounds on the time offset between two nodes are derived. Those bounds are used to outcast redundant information and estimate in a sub-optimal, low complexity scheme the offset between any two nodes. In [22], the same clock model is used and two-way timing information is used to estimate time offset between two neighboring nodes using the NTP algorithm [52]. The major problem with such approach is the fact that delays between any two nodes are not symmetric in general. In [27] two versions of pair-wise synchronization are used: one based on the implementation of a spanning tree starting from the reference node down to the edges and one distributed implementation using node-2-node measurements from a node that needs synchronization up to the reference node.

In all the above implementations there are two distinct differences compared to our Spontaneous approach: a) the above need two-way measurements, meaning that each node needs to send a timing packet and receive its response back, while in our case we need just one way transmissions and therefore, the overhead is smaller and more importantly b) all the above approaches need the maintenance of a hierarchy from the edge nodes up to the reference

node in the form of an NTP-like hierarchy ([22], [72] or Spanning tree version in [27]) or find appropriate communication paths toward the reference path (distributed version in [27]). This is significant overhead compared to our approach where routing is not needed, since the edges of the network are coupled through nearest neighbor communication and no relay of timing packets is needed.

In [67] it was suggested that sensor network nodes need to be able to provide information about when a specific event happened, according to their own clock instead of trying to synchronize all nodes to the same reference. Ordering of events matters, according to the same reasoning and each node should be able to provide an interval, according to each own clock, on when a specific event happened as an approximation of global time. Bounds provided by that scheme are a function of network diameter and that is an important contribution. However, approaches like that would make distributed actuation scenarios (like in our work) difficult to implement in practice.

Reference signals could be used to trigger time-stamping in different receivers, reducing the variability of a wireless transmission time due to channel access delay and propagation time of a signal [18]. Then the two receiving nodes would exchange information to find out their time offset. Such approach is quadratic in the number of participating nodes and therefore more expensive than the approaches discussed so far. It also needs special handling in the case of several beacons and nodes in the vicinity of more than one beacon. Also it is not an infrastructure-free approach compared with our Spontaneous scheme.

It has been argued that time synchronization might be viewed as an offline problem, depending on the application [19]. *Post-facto* synchronization in [19] basically suggests that clocks should be left "free running" and time offsets should be calculated offline, after data have been time-stamped and gathered. This is true for a large range of potential wireless sensor network applications.

Finally, we should mention the cooperative construction of a propagating signal waveform across several hops, that could be used for synchronization [32]. It remains to be implemented and evaluated in practice.

5.3.7 Discussion

We presented and practically implemented a simple time synchronization algorithm for wireless multi-hop sensor networks, which requires no beacons/servers in the network and no global coordination but only local communication among the nodes. We quantified its error as a function of the diameter of the network and theoretically analyzed the observations. Within certain limitations, the algorithm could practically provide for 16-bit synchronization in realistic embedded sensor networks, with msec down to μsec synchronization error without extensive communication or computing overhead requirements. We experimentally showed that the synchronization error is not necessarily increased with the number of hops and provided a simple analytical explanation. Video of the demonstration could be found at [88].

Chapter 6

Conclusion

We proposed a practical cooperative diversity scheme that employees multiple relay radios, distributed in space. Cooperative diversity has been previously proposed as a solution to improve wireless communication, especially in slow fading environments. Our proposed scheme is based on *intelligent* antenna sharing, where a single relay is used among a collection of participating nodes.

It was surprising to find out that using a single relay, chosen according to the opportunistic rule, is as efficient as schemes which require simultaneous transmissions of several relays, from a diversity-multiplexing gain point-of-view. This finding suggests that the virtual antenna array formed by opportunistic relays, *eliminates* the need for complex space-time coding which is required when all, or many relays transmit (relay) simultaneously.

There is considerable amount of effort in the research community to invent practical space-time coding for multiple relay systems and to this end, there is no authoritative solution. The relay channel is fundamentally different from the Multiple Input Multiple Output (MIMO) channel where antennas belong to the same transmitter, and therefore, space-time coding invented for MIMO links, is not directly applicable to relay systems. Additionally, simultaneous transmissions (at the same time and frequency bands) require receiver architectures that deviate from existing RF-front ends, which are based on a single transmission/reception principle. Opportunistic relaying eliminates simultaneous transmissions

and required space-time coding at the physical layer, by exploiting relays as sensors at the network layer. Therefore it is a practical and cost-effective cooperative diversity scheme, for existing radio hardware.

From a diversity-multiplexing gain tradeoff point of view, the surprising performance of opportunistic relaying is justified by the fact that there is a large number of relays and a smart selection is performed, optimizing end-to-end performance. The gains do not arise from a smart processing at the best relay node and that was effectively demonstrated by the fact that, as we proved, a simple amplify-and-forward at the best relay, has the same tradeoff performance as more involved decode-and-forward schemes. Therefore, this thesis suggests that multiple relay systems should be studied not only as a space-time coding problem at the physical/link layers, but also as a best relay selection problem, at the link/routing layers.

At the low SNR regime, we demonstrated two important findings: a) opportunistic relaying increases the outage rates compared to direct, non-cooperative communication with increasing number of relays and b) opportunistic relaying increases the outage and ergodic rates compared to schemes where all relays transmit (with amplify and forward), with increasing number of relays, under a total transmission power constraint (in both cases). Those findings suggest that the number of cooperating relays can be treated as an additional degree of freedom, apart from transmission power and bandwidth, under opportunistic relaying.

We also demonstrated that opportunistic relaying can provide performance gains even at cases when a simple two-hop, single relay, scheme fails to do so. That is the case when all nodes have similar average received SNRs. Performance gains, in terms of spectral efficiency (bps/Hz) under a transmission power constraint, can also be translated to transmission power savings for a given spectral efficiency requirement, suggesting that the practical, antenna array formed by opportunistic relays could be used in applications, where tx power savings are required.

Furthermore, the fact that a single, "best" node is used for relaying allows for the remaining, participating relays to enter "sleep" mode and therefore, suggests that opportunistic relaying

allows for network *reception* power savings. Reception power, is an important parameter in the energy budget of any communication receiver and it has been shown that it can become comparable to transmission power when involved error correction is used. Opportunistic relaying benefits from multiple relays, without requiring all of them to transmit. In fact, we can say that opportunistic relaying benefits even from the relays that do not transmit.

Smart selection is performed in a distributed fashion, that minimizes the required communication overhead among the relays. Specifically, the transmission of maximum two packets is required, independently of the number of participating relays. We quantified how well that selection can be performed, treating the selection scheme as an access scheme and showed that the selection can be performed *fast*, requiring orders of magnitude smaller time than the channel coherence time. The relay selection is based on a novel method that exploits time as function of instantaneous received SNRs at each relay, that characterize the overall end-to-end path source-relay-destination. We kept the relay selection performance analysis separate from the communication performance analysis, since in the future, different single relay selection schemes might be proposed.

The simplicity of the technique allowed us to implement a cooperative diversity demonstration, with simple radio hardware. To the best extent of out knowledge, this demonstration is one of the first of its kind. Since diversity benefits come from a smart relaying protocol, without simultaneous transmissions, we used low-cost, existing radio modules. Opportunistic relaying requires diversity reception at the physical layer, has characteristics of an access scheme at the link layer and involves several nodes at the routing layer. Therefore, it is inherently a cross layer approach and in order to realize it, we implemented from scratch all relevant layers. This might be viewed as a disadvantage, since the whole wireless industry has been structured according to layering that accommodates, mostly, point-to-point, non-cooperative communication links. This thesis suggests that layering needs to be modified in order to benefit from multiple, cooperative relays and the specific implementation presented, can serve as an example. Moreover, this thesis quantifies the important benefits of practical cooperative diversity schemes, further motivating and justifying modifications in the existing Open Systems Interconnect (OSI) 7-layers architecture.

Time keeping was viewed as a network service that could simplify coordination, in a distributed wireless network. Two approaches based on client/server or fully decentralized setups, were presented, analyzed and implemented: the first was based on Kalman filtering and it was contrasted to Linear Programming and simple Averaging, for two cases of network noise. The second approach on decentralized time keeping, exploited oscillators' coupling and required no routing overhead, since it employed local communication to provide global network synchronization. It was inspired by biological phenomena of synchronization, as found in colonies of fireflies or cardiac neurons and was demonstrated in simple hardware. One of the interesting finding was that by adding local oscillators into the network, global synchronization error does not necessarily increase but it could decrease, depending on the individual oscillators characteristics.

Future work on cooperative systems could include application specific studies. For example, sensing of a physical quantity from multiple nodes, and communicating elements of that sensed information to the edges of the network, using relay nodes that have already sensed the same information, provides redundancy at the signal level at each node, that should be exploited at the communication layer[1]. In general, large scale sensing, monitoring and signal processing, that require coverage of large geographical areas, is an application domain for cooperative relay networks and a fruitful area for future research. Multiple antennas at each relay, is another interesting option, studied in the recent literature. Such studies usually focus on the special modulation and coding techniques required for MIMO links. However, more work is needed for the invention of (really) distributed solutions: practical relay schemes that provide benefits without requiring global knowledge about the state of all participating elements, is an art and a field for exciting future reseach endeavors.

Attempts to bridge the gap between information theoretic approaches and practical implementations is usually difficult, given the fact that real world restrictions are in general different than the initial theoretical assumptions. Therefore, when it comes to implementation, the initial problem has considerably changed and more research work is needed to solve

[1]The source-channel separation theorem does not hold for multiple node systems and one relevant study can be found in [24].

it. Consider for example, the case of amplify-and-forward relay networks, where it is not uncommon in the research community, to assume that the final destination has knowledge of the wireless channel conditions between the initial source and all participating relays. Such assumption might simplify analysis but it is highly unpractical, since the destination has no way to estimate those parameters and the only valid solution is to have all relays estimate and communicate such information. Regardless of the elegance or sophistication of the underlying coding algorithm, the above scheme might require significant communication overhead that cancels the benefits of cooperative diversity, even for the case of slow fading environments. In this work, we followed a holistic approach to the problem of cooperative diversity, trying to analyze in theory and implement in practice our solution. Hopefully, the research methodology used might set an example for future research in wireless relay networks.

Cooperative diversity and opportunistic relaying create virtual antenna arrays that improve the efficiency of local wireless communication across several dimensions, including spectral efficiency (in bps/Hz) and transmission/reception power management. These improvements could be viewed as a first step towards the greater vision: scalable, wide-area wireless communication and networking that alleviate the 4th of July problem.

Appendix A

Theorem 4 The joint probability density function of the minimum and second minimum among M i.i.d. positive random variables $X_1,\ X_2, \ldots,\ X_M$, each with probability density function $f(x)$ and cumulative distribution function $F(x)$, is given by the following equation:

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} M\ (M-1)\ f(y_1)\ f(y_2)\ [1-F(y_2)]^{M-2} & for\ 0 < y_1 < y_2 \\ 0 & elsewhere. \end{cases}$$

where $Y_1 < Y_2 < Y_3 \ldots < Y_M$ are the M ordered random variables $X_1,\ X_2, \ldots,\ X_M$.

Proof 5 $f_{Y_1,Y_2}(y_1,y_2)\ dy_1\ dy_2 = Pr(Y_1\ \in\ dy_1,\ Y_2\ \in dy_2) =$
$Pr(one\ X_i\ in\ dy_1,\ one\ X_j\ in\ dy_2\ (with\ y_2 > y_1\ and\ i \neq j),\ and\ all\ the\ rest\ X_i's\ greater\ than\ y_2) =$

$= 2\ \binom{M}{2}\ Pr(\ X_1 \in dy_1,\ X_2 \in dy_2\ (y_2 > y_1),\ X_i > y_2,\ i \in\ [3,M])\ =$
$= 2\ \binom{M}{2} f(y_1)\ dy_1\ f(y_2)\ dy_2\ [1-F(y_2)]^{M-2} =$
$= M\ (M-1)\ f(y_1)\ f(y_2)\ [1-F(y_2)]^{M-2}\ dy_1\ dy_2,\ for\ 0 < y_1 < y_2.$

The third equality is true since there are $\binom{M}{2}$ *pairs in a set of M i.i.d. random variables. The factor 2 comes from the fact that ordering in each pair matters, hence we have a total number of* $2\ \binom{M}{2}$ *cases, with the same probability, assuming identically distributed random variables. That concludes the proof.*

Using Theorem 1, we can prove the following lemma:

Lemma 1 Given M i.i.d. positive random variables $X_1,\ X_2, \ldots,\ X_M$, each with probability density function $f(x)$ and cumulative distribution function $F(x)$, and $Y_1 < Y_2 < Y_3 \ldots < Y_M$ the M ordered random variables $X_1,\ X_2, \ldots,\ X_M$, then $Pr(Y_2 < Y_1 + un)$, where $un > 0$, is given by the following equations:

$$Pr(Y_2 < Y_1 + un) = 1 - I_{un} \tag{1}$$

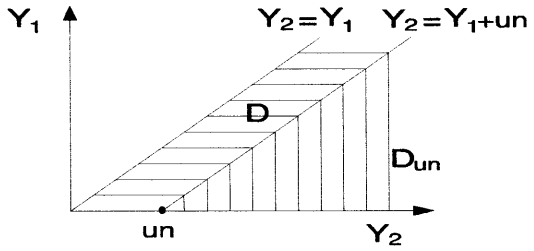

Figure -1: Regions of integration of $f_{Y_1,Y_2}(y_1, y_2)$, for $Y_1 < Y_2$ needed in Lemma I for calculation of $Pr(Y_2 < Y_1 + un)$, $un > 0$.

$$I_{un} = M\ (M-1) \int_{un}^{+\infty} f(y)\ [1 - F(y)]^{M-2}\ F(y - un)\ dy \tag{2}$$

Proof 6 *The joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ integrates to 1 in the region $D \cup D_{un}$, as it can be seen in fig. -1. Therefore:*

$$\begin{aligned} Pr(\ Y_2 < Y_1 + un) &= \int\int_D f_{Y_1,Y_2}(y_1, y_2)\ dy_1\ dy_2 \\ &= 1 - \int\int_{D_{un}} f_{Y_1,Y_2}(y_1, y_2)\ dy_1\ dy_2 \\ &= 1 - I_{un} \end{aligned}$$

Again from fig. -1, I_{un} can easily be calculated:

$$I_{un} = {}_{M(M-1)} \int_{y_2=un}^{+\infty} f(y_2)\ [1 - F(y_2)]^{M-2} \int_0^{y_2-un} f(y_1)\ dy_1\ dy_2$$

$$= {}_{M(M-1)} \int_{y_2=un}^{+\infty} f(y_2)\ [1 - F(y_2)]^{M-2}\ F(y_2 - un)\ dy_2 \tag{3}$$

The last equation concludes the proof.

Appendix B

We repeat *Definition 1* and *Definition 2* in this section for completeness. The relevant lemmas follow.

Definition 1: A function $f(\rho)$ is said to be exponentially equal to b, denoted by $f(\rho) \doteq \rho^b$. if

$$\lim_{\rho \to \infty} \frac{\log f(\rho)}{\log \rho} = b. \tag{4}$$

We can define the relation $\dot{\leq}$ in a similar fashion. *Definition 2:* The exponential order of a random variable X with a non-negative support is given by,

$$V = -\lim_{\rho \to \infty} \frac{\log X}{\log \rho}. \tag{5}$$

Lemma 2 *Suppose $X_1, X_2, \ldots, X_m$ are m i.i.d exponential random variables with parameter λ (mean $1/\lambda$), and $X = \max\{X_1, X_2, \ldots X_m\}$. If V is the exponential order of X then the density function of V is given by*

$$f_V(v) \doteq \begin{cases} \rho^{-mv} & v \geq 0 \\ 0 & v < 0 \end{cases} \tag{6}$$

and

$$\Pr(X \leq \rho^{-v}) \doteq \rho^{-mv} \tag{7}$$

Proof 7 *Define,*

$$V_\rho = -\frac{\log X}{\log \rho}.$$

Thus V_ρ is obtained from definition 2, without the limit of $\rho \to \infty$.

$$\begin{aligned}
\Pr(V_\rho \geq v) &= \Pr(X \leq \rho^{-v}) \\
&= \Pr(X_1 \leq \rho^{-v}, X_2 \leq \rho^{-v}, \ldots X_m \leq \rho^{-v}) \\
&= \prod_{i=1}^{m} \Pr(X_i \leq \rho^{-v}) \\
&= \left(1 - \exp(-\lambda \rho^{-v})\right)^m \\
&= \left(\lambda \rho^{-v} + \sum_{j=2}^{\infty} \frac{(-\lambda)^j}{j!}\ \rho^{-jv}\right)^m
\end{aligned}$$

Note that $\Pr(V_\rho \geq v) \approx \rho^{-mv}$. Differentiating with respect to v and then taking the limit $\rho \to \infty$, we recover (6).

From the above it can be seen that for the simple case of a single exponential random variable ($m = 1$), $\Pr(X \leq \rho^{-v}) = \Pr(V_\rho \geq v) \doteq \rho^{-v}$.

Lemma 3 *For relays, $j = 1, 2, \ldots, m$, let a_{sj} and a_{jd} denote the channel gains from source to relay j and relay j to destination. Suppose that a_{sr} and a_{rd} denote the channel gain of the source to the best relay and the best relay to the destination, where the relay is chosen according to rule 1. i.e.*

$$\min(|a_{sr}|^2, |a_{rd}|^2) = \max\{\min(|a_{s1}|^2, |a_{1d}|^2), \ldots, \min(|a_{sm}|^2, |a_{md}|^2)\}$$

Then,

1. $\min(|a_{sr}|^2, |a_{rd}|^2)$ *has an exponential order given by (6).*

2. $$\Pr(|a_{sr}|^2 \leq \rho^{-v}) = \Pr(|a_{rd}|^2 \leq \rho^{-v}) \dot{\leq} \begin{cases} \rho^{-mv} & v \geq 0 \\ 1 & \textit{otherwise} \end{cases}$$

Proof 8 *Let us denote* $X^{(j)} \triangleq \min(|a_{sj}|^2, |a_{jd}|^2)$. *Since each of the* $X^{(j)}$ *are exponential random variables with parameter 2, claim 1 follows from Lemma 2. Also since* $|a_{sd}|^2$ *and* $|a_{rd}|^2$ *cannot be less than* $\min(|a_{sd}|^2, |a_{rd}|^2)$ *claim 2 follows immediately from claim 1.*

Lemma 4 *With* $f(\cdot,\cdot)$ *defined by relation (3.8), we have that*

$$\Pr\left(f(\rho a, \rho b) \leq \rho^{2r}\right) \leq \Pr\left(\min(a,b) \leq \rho^{2r-1} + \rho^{r-1}\sqrt{1+\rho^{2r}}\right).$$

Proof 9 *Without loss in generality, assume that* $a \geq b$.

$$\begin{aligned} f(\rho a, \rho b) &= \rho \frac{ab}{a+b+\frac{1}{\rho}} \\ &= \rho b \left(\frac{a}{a+b+\frac{1}{\rho}}\right) \\ &\overset{(a)}{\geq} \rho b \left(\frac{b}{2b+\frac{1}{\rho}}\right) \end{aligned}$$

Here (a) follows since $\frac{a}{a+K}$ *is an increasing function in* a, *for* $K > 0$ *and* $a \geq b$.

Now we have that

$$\begin{aligned} \Pr\left(f(\rho a, \rho b) \leq \rho^{2r}\right) &\leq \Pr\left(\frac{b^2}{2b+\frac{1}{\rho}} \leq \rho^{2r-1}\right) \\ &= \Pr\left(b^2 \leq 2\rho^{2r-1}b + \rho^{2r-2}\right) \\ &= \Pr\left((b-\rho^{2r-1})2 \leq \rho^{4r-2} + \rho^{2r-2}\right) \\ &\overset{(a)}{=} \Pr\left(b \leq \rho^{2r-1} + \rho^{r-1}\sqrt{1+\rho^{2r}}\right) \end{aligned}$$

Where (a) follows since $b \geq 0$ *so that* $\Pr(b < 0) = 0$.

Bibliography

[1] S. M. Alamouti, "A Simple Transmitter Diversity Scheme for Wireless Communications", IEEE J. Select. Areas Commun., vol. 1, pp. 1451-58, October 1998.

[2] K. Azarian, H. E. Gamal, and P. Schniter, "On the Achievable Diversity-vs-multiplexing Tradeoff in Cooperative Channels", *IEEE Trans. Information Theory*, submitted, 2004.

[3] N. Bambos, S. C. Chen, G. J. Pottie, "Channel Access Algorithms with Active Link Protection for Wireless Communication Networks with Power Control", IEEE/ACM Transactions on Networking, vol. 8, no. 5, October 2000.

[4] J.A. Barnes, et. al. "Characterization of Frequency Stability", IEEE Transactions on Instrumentation and Measurement, Vol. IM-20, No. 2, pp. 105-120, May 1971.

[5] S. Z. Biswas, *Opportunistic Routing in Multi-Hop Wireless Networks*, M.S. Thesis, MIT, February 2005.

[6] A. Bletsas, *T*ime Keeping in Myriad Networks: Theories, Solutions and Applications, M.S. thesis, MIT, June 2001.

[7] A. Bletsas and A. Lippman, "Efficient Collaborative (Viral) Communication in OFDM Based WLANs", Proceedings of IEEE/ITS International Symposium on Advanced Radio Technologies (ISART 2003), Institute of Standards and Technology, Boulder Colorado, March 4-7, 2003.

[8] A. Bletsas, A. Lippman, "Natural Spontaneous Order in Wireless Sensor Networks: Time Synchronization Based on Entrainment", Technical Report, MIT Media Lab, December 2003.

[9] A. Bletsas, "Evaluation of Kalman Filtering for Network Time Keeping", Proceedings of IEEE International Conference on Pervasive Computing and Communications (PERCOM), Fort-Worth Texas, March 2003.

[10] A. Bletsas, "Evaluation of Kalman Filtering for Network Time Keeping", IEEE Transactions on Ultrasonics, Ferromagnetics and Frequency Control, to appear. Available at http://web.media.mit.edu/ aggelos/papers/tuffc_nov04.pdf

[11] A. Bletsas, A. Lippman, "Spontaneous Synchronization in Multi-hop Embedded Sensor Networks: Demonstration of a Server-free Approach", Proceedings of IEEE 2nd European Workshop on Sensor Networks 2005.

[12] A. Bletsas, A. Lippman, D.P. Reed, "A Simple Distributed Method for Relay Selection in Cooperative Diversity Wireless Networks, based on Reciprocity and Channel Measurements", accepted for publication, IEEE 61st Semiannual Vehicular Technology Conference May 30 - June 1 2005, Stockholm, Sweden.

[13] A. Bletsas, A. Khisti, D.P. Reed, A. Lippman, "A Simple Cooperative Diversity Method based on Network Path Selection", submitted for publication, IEEE Journal on Selected Areas of Communication, Special Issue on 4G, January 2005.

[14] D. Cassioli, M. Win, A. Molisch, "The Ultra-Wide Bandwidth Indoor Channel: From Statistical Model to Simulations", IEEE Journal On Selected Areas of Communications, vol. 20, no. 6, August 2001.

[15] P. R. Cook, *Music, Cognition and Computerized Sound. An Introduction to Psychoacoustics*, MIT Press, Cambridge Massachusetts, 1999

[16] G.M. Crippen, T.F. Havel, "Distance Geometry and Molecular Conformation", John Wiley & Sons, New York 1988.

[17] Q. Dong, Z. Wu, "A Geometric Build-Up Algorithm for Solving the Molecular Distance Geometry Problem with Sparse Distance Data", Journal of Global Optimization, no. 26, pp. 321-333, Kluwer Academic Publishers 2003.

[18] J. Elson, L. Girod, D. Estrin, "Fine-Grained Network Time Synchronization using Reference Broadcasts", Proceedings of the Fifth Symposium on Operating Systems Design and Implementation (OSDI), Boston MA, 2002.

[19] J. Elson, K. Römer, "Wireless Sensor Netowkrs: A New Regime for Time Synchronization", Proceedings of the First Workshop on Hot Topics in Networks, October 2002.

[20] G. J. Foschini, Z. Miljanic, "A Simple Distributed Autonomous Power Control Algorithm and its Convergence", IEEE Transactions on Vehicular Technology, vol. 42, no. 4, November 1993.

[21] H.E. Gamal, G. Caire, M.O. Damen, "Lattice Coding and Decoding Achieve the Optimal Diversity-Multiplexing Tradeoff of MIMO Channels," IEEE Transactions pp. 968-985, June 2004.

[22] S. Ganeriwal, R. Kumar and M. Srivastava, "Timing Sync Protocol for Sensor Networks", Proceedings of the First International Conference on Embedded Networked Sensor Systems, Los Angeles CA, 2003.

[23] D. Ganesan, R. Govindan, S. Shenker and D. Estrin, "Highly Resilient, Energy Efficient Multipath Routing in Wireless Sensor Networks", Mobile Computing and Communications Review (MC2R), Vol 1., No. 2. 2002.

[24] M. Gastpar and M. Vetterli, "On the Capacity of Large Gaussian Relay Networks". IEEE Transactions on Information Theory, 51(3):765-779, March 2005.

[25] N. Gershenfeld, The Nature of Mathematical Modeling, Cambridge University Press, 2000.

[26] P. Goud Jr, C. Schlegel, W.A. Krzymien, R. Hang, "Multiple Antenna Communication Systems - An Emerging Technology", Canadian Journal of

Electrical and Computer Engineering, accepted for publication. Also found in http://www.ece.ualberta.ca/ hcdc/publications.html

[27] J. V. Greunen, J. Rabaey, *"Lightweight Time Synchronization for Sensor Networks"*, Proceedings of the 2nd ACM International Workshop on Wireless Sensor Networks and Applications, San Diego, CA, September 2003.

[28] M. Grossglauser, D. Tse, "Mobility increases the capacity of wireless networks", Proceedings of IEEE Infocom 2001, Anchorage 2001.

[29] P. Gupta, P.R. Kumar, "The Capacity of Wireless Networks", Transactions on Information Theory, vol. 46, no. 2, March 2000.

[30] I. Hammerstroem, M. Kuhn, and A. Wittneben, "Impact of Relay Gain Allocation on the Performance of Cooperative Diversity Networks", IEEE Vehicular Technology Conference, VTC Fall 2004, Los Angeles, Sept. 2004.

[31] M. O. Hasna, M.S. Alouini, "End-to-End Performance of Transmission Systems With Relays Over Rayleigh-Fading Channels", IEEE Transactions on Wireless Communications, vol. 2, no. 6, November 2003.

[32] A. Hu, S. D. Servetto, "Asymptotically Optimal Time Synchronization in Dense Sensor Networks", Proceedings of the 2nd ACM International Workshop on Wireless Sensor Networks and Applications, San Diego, CA, September 2003.

[33] T. Hunter and A. Nosratinia, "Cooperation Diversity Through Coding", in Proc. IEEE Int. Symp. Information Theory, Lausanne, Switzerland, June 2002, p. 220

[34] M. Janani, A. Hedayat, T.E. Hunter, A. Nostratinia, "Coded Cooperation in Wireless Communications: Space-Time Transmission and Iterative Decoding", IEEE Transactions on Signal Processing, vol. 52, no. 2, February 2004.

[35] Y. Jing, B. Hassibi, "Distributed Space-time Coding in Wireless Relay Networks-Part I: Basic Diversity Results", Submitted to IEEE Trans. On Wireless Communications, July 2004. Available at http://www.cds.caltech.edu/ yindi/publications.html

[36] A. Khandani, J. Abounadi, E. Modiano, L. Zheng, "Cooperative Routing in Wireless Networks", Allerton Conference on Communications, Control and Computing, October 2003.

[37] G. Kramer, M. Gastpar and P. Gupta. "Cooperative Strategies and Capacity Theorems for Relay Networks". Submitted to IEEE Transactions on Information Theory, February 2004. Available at http://www.eecs.berkeley.edu/ gastpar/relaynetsIT04.pdf

[38] S. R. Kulkarni, P. Viswanath, "A Deterministic Approach to Throughput Scaling in Wireless Networks", Transactions on Information Theory, vol. 50, no. 6, pp. 1041-1049, June 2004.

[39] M. Laibowitz, J. Paradiso, "Wearable Wireless Transceivers", Circuit Cellar, no.163, February 2004.

[40] L. Lamport, "Time, Clocks, and the Ordering of Events in a Distributed System", Communications of the ACM, vol. 21, no.7, July 1978.

[41] J.N. Laneman, G.W. Wornell, "Energy-Efficient Antenna Sharing and Relaying for Wireless Networks", Proc. IEEE Wireless Comm. and Networking Conf. (WCNC), Chicago, IL, Sept. 2000.

[42] J.N. Laneman, *Cooperative Diversity in Wireless Networks: Algorithms and Architectures*, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, Aug. 2002.

[43] J. N. Laneman and G. W. Wornell, "Distributed Space-Time Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks", *IEEE Trans. Inform. Theory*, vol. 59, pp. 2415–2525, October 2003.

[44] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior", *IEEE Trans. Inform. Theory*, Accepted for publication, June 2004.

[45] W. E. Leland, M. S. Taqqu, D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic", IEEE/ACM Transactions on Networking, vol. 2, February 1994.

[46] J. Levine, "An Algorithm to Synchronize the Time of a Computer to Universal Time", IEEE/ACM Transactions on Networking, vol. 3, no. 1, February 1995.

[47] J. Levine, "Time Synchronization Using the Internet", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, vol. 45, no. 2, March 1998.

[48] J. Levine, "Introduction to Time and Frequency Metrology", Review of Scientific Instruments, vol. 70, no. 6, June 1996.

[49] J. Levine, personal communication, NIST, March 2002.

[50] J. Lifton, D. Seetharam, M. Broxton, J. Paradiso, "Pushpin Computing System Overview: a Platform for Distributed, Embedded, Ubiquitous Sensor Networks", Pervasive 2002, Proceedings of the Pervasive Computing Conference, Zurich Switzerland, August 2002.

[51] J. Luo, R. S. Blum, L. J. Cimini, Jr., L. J. Greenstein, and A. M. Haimovich, "New Approaches for Cooperative Use of Multiple Antennas in Ad hoc Wireless Networks", IEEE Vehicular Technology Conference (VTC) Fall 2004.

[52] D. L. Mills, *N*etwork Time Protocol (Version 3) Specification, Implementation and Analysis, RFC 1305, University of Delaware, March 1992.

[53] R. Min, M. Bhardwaj, Seong-Hwan Cho, N. Ickes, Eugene Shih, Amit Sinha, Alice Wang, Anantha Chandrakasan, "Energy-Centric Enabling Technologies for Wireless Sensor Networks", IEEE Wireless Communications (formerly IEEE Personal Communications), vol. 9, no. 4, August 2002, pp. 28-39.

[54] R. Min, Personal Communication based on his 2003 Ph.D. work.

[55] A. F. Molisch and M. Z. Win, "MIMO Systems with Antenna Selection", IEEE Microwave Magazine, pp. 46-56, March 2004.

[56] S. B. Moon, P. Skelly, D. Towsley, "Estimation and Removal of Clock Skew from Network Delay Measurements", Proceedings of 18th Annual Joint Conf. of the IEEE Computer and Communications Societies (INFOCOM), 1999.

[57] R. U. Nabar, O. Oyman, H. Bolcskei, A. J. Paulraj, "Capacity Scaling Laws in MIMO Wireless Networks", Allerton Conference on Communication, Control, and Computing, Monticello, IL, pp. 378-389, Oct. 2003.

[58] R. U. Nabar, H. Blcskei, and F. W. Kneubhler, "Fading Relay Channels: Performance Limits and Space-time Signal Design", IEEE Journal on Selected Areas in Communications, June 2004, to appear, available from http://www.nari.ee.ethz.ch/commth/pubs/p/jsac03

[59] M. Neely and E. Modiano, "Capacity and Delay Tradeoffs in Mobile Ad Hoc Networks," (Invited Paper) IEEE BroadNets 2004, San Jose, CA, October, 2004.

[60] G. Opshaug, P. Enge, "GPS and UWB for Indoor Navigation", Institute of Navigation GPS Conference, Salt Lake City, September 2001.

[61] S. J. Orfanidis, *O*ptimum Signal Processing, An Introduction, 2^{nd} Edition, McGraw-Hill, 1988.

[62] J. Paradiso, K.Hsiao, J. Strickon, P. Rice, "New Sensor and Music Systems for Large Interactive Surfaces", Proceedings of the International Computer Music Conference (ICMC), August 2000.

[63] J. Paradiso, K. Hsiao, J. Strickon, J. Lifton, and A. Adler, "Sensor Systems for Interactive Surfaces", IBM Systems Journal, Volume 39, Nos. 3 & 4, October 2000, pp. 892-914.

[64] C. Parsa, J.J. Garcia-Luna-Aceves, "Improving TCP Congestion Control Over Internets with Heterogeneous Transmission Media", Proceedings of IEEE, ICNP 99, October 1999.

[65] V. Paxson, *M*easurement and Analysis of End-to-End Internet Dynamics, Ph.D. thesis, University of California at Berkeley, April 1997.

[66] T.S Rappaport, *Wireless communications: principles and practice*, Upper Saddle River, NJ: Prentice Hall, 1996.

[67] K. Römer, "Time Synchronization in Ad Hoc Networks", MobiHOC 2001, Long Beach CA.

[68] A. Scaglione, Yao-Win Hong, "Opportunistic Large Arrays: Cooperative Transmission in Wireless Multihop Ad Hoc Networks to Reach Far Distances", IEEE Transactions on Signal Processing, vol. 51, no. 8, August 2003.

[69] A. Sendonaris, E. Erkip and B. Aazhang. "User cooperation diversity-Part I: System description". IEEE Transactions on Communications, vol. 51, no. 11, pp. 1927-1938, November 2003.

[70] A. Sendonaris, E. Erkip and B. Aazhang, "User cooperation diversity. Part I: System description, IEEE Trans. on Communications, vol. 51, no. 11, pp.1927-1938, November 2003.

[71] T.J. Shepard, "A Channel Cccess Scheme for Large Dense Packet Radio Networks", SIGCOMM 1996, Palo Alto California.

[72] M.L. Sichitiu, C. Veerarittiphan, "Simple, Accurate Time Synchronization for Wireless Sensor Networks", IEEE WCNC 2003.

[73] S. Strogatz, *Sync, The Emerging Science of Spontaneous Order*, 1st ed. New York: Theia, 2003.

[74] M. S. Taqqu, W. Willinger, R. Sherman, "Proof of a Fundamental Result in Self-Similar Traffic Modeling", ACM Computer Communications Review, April 1997.

[75] E. Teletar, "Capacity of Multi-Antenna Gaussian Channels", *European Transac. on Telecom. (ETT)*, vol. 10, pp. 585–596, November/December 1999.

[76] S. Toumpis, A. Goldsmith, "Capacity regions for wireless ad hoc networks", IEEE International Conference on Communications, 28 April-2 May 2002.

[77] S. Toumpis and A. Goldsmith, "Large Wireless Networks under Fading, Mobility, and Delay Constraints", IEEE Infocom 2004.

[78] G. D. Troxel, *T*ime Surveying: Clock Synchronization over Packet Networks, Ph.D. thesis, MIT, May 1994.

[79] D. N. C. Tse and P. Viswanath, *F*undamentals of Wireless Communications. Working Draft, 2003.

[80] P. Viswanath, D.N.C. Tse, R. Laroia, "Opportunistic Beamforming Using Dumb Antennas", IEEE Transactions on Information Theory, vol. 48, no. 6, June 2002.

[81] A. Wittneben, B. Rankov, "Impact of Cooperative Relays on the Capacity of Rank-Deficient MIMO Channels", Proceedings of the 12th IST Summit on Mobile and Wireless Communications, Aveiro, Portugal, pp. 421-425, June 2003.

[82] A. Wittneben, I. Hammerstroem, and M. Kuhn, "Joint Cooperative Diversity and Scheduling in Low Mobility Wireless Networks", IEEE Global Telecommunications Conference, Globecom 2004, Nov. 2004.

[83] B. Zhao and M.C. Valenti, "Practical relay networks: A generalization of hybrid-ARQ", IEEE Journal on Selected Areas in Communications (Special Issue on Wireless Ad Hoc Networks), vol. 23, no. 1, pp. 7-18, Jan. 2005.

[84] L. Zheng and D. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple Antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–96, May, 2003.

[85] M. Zorzi, R.R. Rao, "Geographic Random Forwarding (GeRaF) for Ad hoc and Sensor Networks: Multihop Performance", IEEE Trans. on Mobile Computing, vol. 2, Oct.-Dec. 2003.

[86] M. Zorzi, R.R. Rao, "Geographic Random Forwarding (GeRaF) for Ad hoc and Sensor Networks: Energy and Latency Performance", IEEE Trans. on Mobile Computing, vol. 2, Oct.-Dec. 2003.

[87] ns-2, *"The Network Simulator-ns-2"*, http://www.isi.edu/nsnam/ns/

[88] Demo website,
http://web.media.mit.edu/ aggelos/demos/demos.html

[89] Badge website, http://www.media.mit.edu/resenv/badge

[90] Pushpin website, http://web.media.mit.edu/ lifton/PushPin/