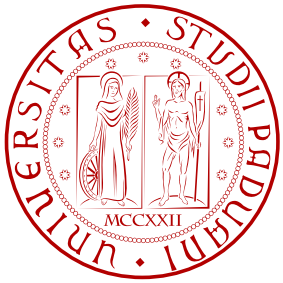


Università degli Studi di Padova

Dipartimento di Elettronica e Informatica

Scuola di Dottorato in Ingegneria dell'Informazione

Indirizzo in Scienze e Tecnologie dell'Informazione

XXII Ciclo

Cooperative Techniques for Wireless Ad Hoc and Cellular Networks

Tesi di: *Federico Librino*

Supervisore:
Chiar.mo Prof. Michele Zorzi

Direttore della Scuola:
Chiar.mo Prof. Matteo Bertocco

Anno Accademico 2009/2010

"It is difficult to say what is impossible,
for the dream of yesterday
is the hope of today
and the reality of tomorrow "

Robert H. Goddard

A mio padre,
che apprezza i miei girovagare per il mondo,
e a mia madre,
che in fondo riesce a comprenderne le ragioni

Contents

Abstract

The impact of fading and other impairments in wireless channels can be counteracted by leveraging communication diversity and introducing cooperative paradigms, where third-party nodes contribute to assist the communication. Cooperation can be implemented in a variety of schemes, and with a number of different purposes. In this work, we investigate how different cooperative techniques can be practically applied in different environments.

In the first part, we study the performance improvement offered by Coded Cooperation in wireless Ad Hoc networks. Our analysis aims at clarifying the impact of cooperation from a network-wide perspective, and a cooperative protocol is designed to illustrate a practical implementation. Cooperation is shown to be effective both locally and globally. On the one side, it can strengthen a single link, when one or more cooperators help the source in delivering the packet to its destination; on the flip side, cooperators may be exploited to find out alternative routes to deliver a packet to a far destination, thus avoiding bad links or crowded areas of the network. A cooperative scheme based on Coded Cooperation is applied to a network whose nodes are equipped with multiple antennas. MIMO technology in fact makes it easier to coordinate the nodes involved in each communication, since multiple concurrent transmissions can be decoded. In our scheme, a distributed form of HARQ is implemented, which is able to react to interference variations and channel fading by exploiting both coding gain and spatial diversity.

The second part discusses the use of relays in wireless CDMA cellular networks for uplink transmissions, with particular attention put to the constraints and the standard parameters that may affect the effectiveness of cooperation in a realistic scenario. Simple cooperative techniques like Amplify & Forward and Decode & Forward are tested. The goal is to identify some design directions that should be used when applying these cooperative techniques: cooperators positions and capabilities, in fact, have an impact on the obtained performance. Moreover, carrier assignment is important to keep interference under control.

Our investigation shows that the coverage area can be increased through use of cooperation, or equivalently, that power saving is possible when a mobile terminal is helped by a relay.

In order to derive a theoretical limit to the capacity of these cooperative systems, which requires combinatorial calculations and the derivation of the areas covered by one or more Stations, an iterative algorithm has been developed and is also presented in the last part of this work.

Sommario

Nei canali wireless, l'impatto del fading può essere ridotto utilizzando tecniche e paradigmi cooperativi che traggano vantaggio dalla diversità in trasmissione, laddove nodi non impegnati in trasmissioni assistano le comunicazioni dei propri vicini. La cooperazione può essere implementata attraverso diverse strategie, e con differenti scopi. Il lavoro presentato in questa tesi è perciò incentrato sullo studio delle modalità di applicazione concreta di tecniche cooperative diverse in ambienti differenti.

La prima parte del lavoro mira ad identificare quale miglioramento delle prestazioni possa essere garantito attraverso meccanismi di Coded Cooperation in reti wireless Ad Hoc: obiettivo dell'analisi è quindi identificare l'impatto della cooperazione a livello di rete. Inoltre, viene presentato e descritto un protocollo cross-layer allo scopo di esemplificare una possibile implementazione pratica di tale schema cooperativo. I risultati mostrano come la cooperazione sia efficace sia localmente che globalmente. Da un lato infatti essa può irrobustire il singolo collegamento tra due nodi della rete, allorchè uno o più cooperatori utilizzino le proprie risorse per aiutare la sorgente nella consegna del pacchetto alla relativa destinazione. Dall'altro lato, è possibile sfruttare la presenza dei cooperatori per identificare percorsi alternativi nella consegna del pacchetto a destinazioni più distanti, aggirando aree della rete ad elevata congestione ed evitando link soggetti a cattive condizioni di canale. Uno schema cooperativo basato sulla Coded Cooperation è stato applicato ad una rete i cui nodi dispongono di antenne multiple. Infatti, la tecnologia MIMO rende più semplice la coordinazione dei nodi coinvolti nella trasmissione cooperativa, poiché rende possibile la decodifica di più trasmissioni simultanee. Nello schema proposto viene implementata una forma distribuita di HARQ, in grado di reagire a variazioni di interferenza e di fading sfruttando sia il guadagno di codice che la diversità spaziale.

Nella seconda parte del lavoro l'analisi si incentra sui benefici derivanti dall'utilizzo di relay in reti cellulari di tipo CDMA relativamente al canale di uplink: particolare atten-

zione viene posta alla necessità di adattare l'implementazione dei meccanismi cooperativi alla presenza di vincoli dettati dagli standard vigenti; tali vincoli possono infatti limitare la flessibilità degli schemi adottati e, di conseguenza, l'efficacia della cooperazione. Sono state analizzate pertanto tecniche cooperative facilmente implementabili, quali Amplify & Forward e Decode & Forward. Obiettivo del lavoro è stata l'identificazione di linee guida che possano essere tenute in considerazione in fase di progettazione di protocolli cooperativi per questo tipo di reti: infatti, sia la posizione che le capacità del cooperatore hanno un notevole impatto sulle prestazioni. Inoltre, l'assegnazione delle frequenze risulta fondamentale per contenere l'interferenza. Lo studio qui presentato mostra come l'uso della cooperazione garantisca un incremento del raggio di copertura, o, equivalentemente, una riduzione della potenza richiesta al terminale mobile in presenza di relay.

Allo scopo di individuare un limite teorico alla capacità del sistema cooperativo, che richiede calcoli combinatori nonché la determinazione delle aree coperte da una o più stazioni radio o relay, è stato infine sviluppato un algoritmo iterativo, presentato nell'ultima parte del lavoro.

Chapter 1

Introduction

The number of wireless networks and mobile devices have increased exponentially in recent years, due to their ease of deployment. On the one hand, mobile phones are spread in almost each country, where the widespread diffusion of cellular networks guarantees coverage in most locations. On the other hand, wireless LANs are becoming more and more reliable, leading to a huge number of wireless connections to the Internet, today commonly accessible from airports, shops, stadiums and many other public places. In addition, sensor networks also rely on wireless transmissions in order to collect the data measured by a number of nodes. This information is transmitted through wireless links to be then processed at the sink. A different paradigm of wireless networks is given by Ad Hoc networks: here multiple nodes can independently access (and leave) an infrastructureless wireless network, and multiple communications can be kept active towards other nodes without the need of a centralized control.

It is clear that wireless networks are desirable for a number of reasons. First of all, they can be easily deployed: no expensive cables are required, but only a central node, or an access points, for some kind of wireless networks. This also makes it possible to set up networks in hardly reachable environment, such as mountains, forests, or war zones. Secondly, wireless connections grant the advantage of mobility, since wireless terminals no longer need to be located close to wired accesses to the network.

On the flip side, however, wireless communications are known to be strongly influenced by multipath fading and interference problems. Node mobility, as well as the multipath propagation of the signal, cause the channel to be extremely time-variant, especially when line of sight is not available. Doppler effect is also to be considered, thus requiring an ad-

ditional complexity at the receiver to equalize the channel and decode the transmission. A second important aspect is represented by interference. The wireless medium is intrinsically a broadcast channel: the signal transmitted by each source is received by all the terminals within a certain range, thus potentially creating a huge amount of interference. Without a proper control mechanism, this may cause collisions among packets transmitted by different sources, leading to a significant throughput loss.

While error control and adaptation to channel conditions appear to be fundamental issues to cope with the wireless channel, different techniques have been designed to combat the effects of interference.

In Ad Hoc networks, a smart protocol design can achieve this goal. One of the most commonly used strategy is to use carrier sensing: when a node is ready to transmit, it listens to the channel to ensure the absence of other communications with which its one would probably interfere. A similar result is obtained by collision avoidance protocols, which make use of short signaling packets to block all the possible interfering communications. Clearly, both this schemes, which try to avoid the occurrence of packet collisions, are conservative, and do not take advantage of the spatial diversity offered by the wireless channel. Moreover, collisions are still possible in particular situations (hidden terminal and exposed terminal problem).

A different solution is based on resource allocation strategies, and is commonly adopted in infrastuctured networks, where a centralized control is possible, for example in cellular networks. In this case, concurrent transmissions can take place, given that they are assigned different (preferably orthogonal) resources. The main examples are TDMA, where communications are separated in time, FDMA, where different frequencies are assigned to each transmission, or CDMA, where codes are used to highly reduce the interference.

A third approach aims at improving the receiver capabilities in order to make the decoding of multiple incoming transmissions possible. Efforts in this direction have led to the concept of interference cancellation: when interfering signals are present, it may be possible to decode them, and successively subtract the interference contribution from the received signal. Such a technique is even favoured by the use of MIMO technology. Nodes equipped with multiple antennas can handle interference even more easily, thus letting multiple links to be connected within the same area.

Recently, however, a completely different strategy has been investigated to improve the performance in wireless networks. Although the presence of other nodes in the network

have always been regarded as a problem, being the source of potential interference, in a completely different perspective it may be considered, on the contrary, a precious resource. As a matter of fact, idle nodes are able to receive and decode the transmissions among neighboring nodes, and can then behave as helpers for those nodes. This is the concept of cooperation. In other words, the main idea is that nodes which are not currently transmitting can offer their own resources to improve the performance of their transmitting neighbors. In doing this, they are consuming part of their energy to help other nodes, but an overall gain is likely to be achieved, since transmissions can be completed in a shorter time, and interference may be lowered as well.

A number of cooperative strategies have been designed and studied, since cooperation can be applied at different layers, and with different purposes. At the physical layer, a form of cooperation is given by Space-Time Codes: a distributed code is applied by two or more terminals to strenghten their transmission towards the same destination, using a technique which was originally applied to MIMO communications, but requiring a high precision in synchronization. At the MAC layer, a wide research effort has been put to the use of relays, nodes which usually have no own traffic, and can hence dedicate their resources in helping other nodes. The two main schemes are Amplify and Forward (A&F), where the cooperator acts as a repeater, amplifying and retransmitting the received signal, and Decode and Forward (D&F), where the relay tries to decode the signal and to send it to the destination. While the former solution is easier to be implemented, and introduces a lower delay, the latter avoids noise propagation, and allows a selection among the incoming signals, thus limiting interference from unnecessarily amplified transmissions. Practically, it is also possible for any node to act as relay, thus exploiting the idle time between two subsequent transmissions.

Finally, if Coded Cooperation is used, the cooperator is required to decode the signal coming from the source, and generate some kind of redundancy to be transmitted to the destination. A codeword is hence received, taking advantage also of the spatial diversity due to the different positions of source and cooperator. Generally, multiple cooperators can be active at the same time, although issues like cooperator selection and nodes coordination are to be specifically addressed.

In this thesis, we want to investigate how well known cooperative schemes can be practically applied to wireless networks, which kind of benefits can be achieved, whether constraints are necessary, and which tradeoffs must be considered. In particular, we will focus

on both Ad Hoc networks and cellular networks. Though sharing some common features, these two kinds of wireless networks differ in several design aspects, as well as in the kind of imposed constraints.

As regards Ad Hoc networks, we will mainly focus on Coded Cooperation. The originality of our work is twofold. On the one side, we want to test cooperation for different purposes. From a local point of view, cooperation can improve the reliability of a single link through the help of a third (or more) node. This is also the general three-node scenario commonly studied in literature. From a network-wide perspective, cooperation may be useful in increasing routing effectiveness. Relays can offer the choice of different paths to a given destination, maybe avoiding congested areas, or in response to varying channel conditions. In doing this, we will also propose a practical scheme, and a complete protocol, designed from the physical to the network layer, thus taking into account all the coordination and overhead necessary for the cooperation to work.

On the other side, we want to explore cooperation from a wider perspective. The use of cooperators, which is likely to locally improve performance, also implies a higher number of transmissions, which may have a detrimental effect on the other nodes of the same network. An overall analysis is presented, which shows the global effects of cooperation on the performance of the whole network.

In cellular networks the existing standards impose a limit on the design of cooperative schemes. In this work, we focus on Uplink performance, and we try to determine the effectiveness of simple techniques like A&F and D&F in a CDMA network. In this part, we investigate which kind of benefits can be obtained through cooperation, and we give some design directions regarding the cooperators capabilities and locations that should be used to achieve the highest performance. Practical limitations, such as power constraints or half-duplex capability, are taken into account, to make the results closer to what can be achieved with a practical implementation. Moreover, an implementation in a fully detailed simulator has been also included. In this simulation, all the approximations have been removed, and the practical advantages of using relays are discussed.

From the presented results, cooperation effectiveness can be evaluated in a wide range of environments. We want to point out that this work paves the way for additional studies. In Ad Hoc networks, it may be arguable whether cooperation is likely to occur, since it requires a non selfish behavior of the nodes, or if it should be promoted with proper rewards. In cellular networks, other more refined cooperative techniques may be applied,

and integrated with mechanisms like soft handover and power control, to further boost the advantages offered by cooperation.

This thesis is structured as follows: in Chapter 2 cooperation in wireless ad hoc networks is investigated from a network-wide perspective, with a full protocol design and simulation testing. Chapter 3 focuses on the application of relaying techniques in CDMA cellular networks. The analysis spans from a simple three node scenario, where outage performances are compared, and a whole multi-cell scenario, where attention is put on the gains in terms of coverage range, throughput, and power saving. Moreover, four cooperative power control algorithms are designed and compared through preliminary results. Chapter 4 concludes the thesis. Finally, in Appendix A, we show a side investigation on distributed estimation in wireless sensors networks, where nodes cooperate to achieve a common goal: a practical protocol is designed and its performance is compared to the theoretical predictions.

Chapter 2

Cooperation in Ad Hoc Wireless Networks

Interference control is a fundamental issue in wireless networking, and in fact it is the focal element in the design of access protocols for wireless networks. In ad hoc networks, where nodes manage access and maintain connectivity in a distributed fashion, the need for protecting the active receivers from interference has led to approaches based on the concept of *collision*. The collision rationale assumes that multiple transmissions in a certain area of the network, often coinciding with the definition of neighborhood, always result in failed packet decoding at the receiving nodes.

This rationale led to the design of carrier-sensing and collision avoidance protocols [1], whose goal is to avoid the occurrence of multiple transmissions. This is clearly a conservative approach, that in some cases, and for some topologies, achieves poor performance in terms of spatial reuse of the channel resource. Moreover, carrier sense-based access control is subject to the well known *exposed* and *hidden* terminal problems [2].

The significant advancements achieved in the recent years in the design of multiple-input multiple-output (MIMO) [3] architectures are leading to new networking approaches based on MIMO technology, where the transmission and reception of multiple information streams have been shown to considerably increase the capacity of point-to-point links.[1] Nevertheless, a flourishing research effort is being devoted to the design of wireless networks that exploit the potential of MIMO to let multiple links be activated at the same time and in the same area. Roughly speaking, MIMO processing may be exploited to reduce the sensi-

[1] By links we mean communications between neighbors of the network.

tivity of the receiver to incoming interference, thus making the coexistence of multiple links more profitable, without the need for multiplexing them in time/frequency.

The control of these networks requires a big shift of perspective in the design of the networking protocols. A collision based approach is no longer suitable, as the goal is to boost the parallelism of the transmissions, rather than to avoid it. Due to simultaneous access of multiple nodes, efficiency, adaptation to channel conditions and error control are key issues for the performance of the network. In fact, the time required to deliver a packet, including possible retransmissions, determines the interference load as a function of the offered traffic. Thus, an efficient communication scheme may shift the congestion point of the network, thus increasing the achieved maximum aggregate throughput. It is important to observe that interference represents a further source of channel variation, as interfering transmissions may start and end during a communication. Thus, transmission and error control protocols must be able to efficiently cope with rapid and unpredictable variations of the channel conditions.

This is especially true when looking at packet forwarding in the network. In fact, as channel conditions are extremely variable due to interference, static routing may be highly inefficient. Consider for instance an area of the network where multiple transmitters are active, that is, an area of the network potentially generating high interference. Thus, if the forwarding path of a packet passes through that area it may suffer an increased delivery failure rate (and increase that perceived at the other active receivers), or anyway require a greater time to be delivered, due to a smaller feasible transmission rate, as well as the need for multiple retransmissions.

Hence, in the presence of unpredictable and quickly changing interference conditions, we need a routing protocol able to promptly react and to opportunistically adapt to changing interference conditions in the network.

To improve this situation, and achieve reliability at the upper layers when relying on a wireless physical layer, many researchers have proposed to adopt cooperative paradigms [4–7]. This means that one or more intermediate nodes intervene in the communication between a transmitter and a receiver, so that either the communication is re-routed over a better path, or the original link is kept in use but its quality is strengthened thanks to diversity provided by these cooperators. The former approach is known as cooperative (or opportunistic) routing [4, 8]. In the simplest version, one intermediate node simply acts as a relay between the transmitter and the receiver. However, if the communication between

transmitter and receiver is part of a multihop transmission, it may be re-routed over an entirely new path which no longer involves this receiver. To this end, a distance metric is needed to verify that the selection of a given intermediate node as next hop still sends the message toward the end-to-end destination and not further away from it. Additionally, a negotiation phase is also required where intermediate nodes can volunteer as next hops whenever the link to the intended receiver does not guarantee sufficiently high quality. The latter technique utilizes instead cooperative diversity based on coded transmissions. For this reason, we will refer to it as coded cooperation [9, 10]. The rationale of this approach is that the original communication link is still kept, but intermediate nodes try to overhear the transmission and reinforce its quality during a retransmission phase. This means that if the message is not correctly received and thus a retransmission is dictated, not only does the original transmitter re-send it, which would correspond to a plain Automatic Repeat reQuest (ARQ) approach, but also intermediate nodes which were able to overhear the packet do so. Moreover, they properly encode the information so that the multiple transmissions can be combined at the receiver's side to improve the link quality. Therefore, cooperative intermediate nodes bring two separate benefits, namely, they allow to exploit alternative links with different quality, i.e., space diversity, and also introduce coding gain, which gives to this approach the same advantage that Hybrid ARQ (HARQ) has over traditional ARQ techniques [10, 11]. Inspired by theoretical findings showing the need for cooperation in wireless networks [6, 7, 12], both these approaches have recently attracted a great deal of attention in the research community. However, we believe that many studies tend to evaluate them in local scenarios with small topologies and simplified assumptions about Medium Access Control (MAC) or networking issues. In this work, we aim at providing a description of these techniques from a network-wide perspective, in order to determine whether the advantages brought by cooperation can be exploited in wireless networks, which new problems or challenges they introduce, and how they can be addressed and solved. To gain some visual insight of the new dimension we add to the problem, consider Fig. 2.1. Standard cooperative approaches focus on scenarios like the left-hand part of this figure, where a single-hop transmission is locally aided by some casual cooperators, whereas we consider a scenario where all nodes apply cooperative paradigms, as shown in the right-hand part. This implies that the effects of cooperative actions are not only local; for example, cooperative relays assisting a communications can in turn find new cooperators to carry on the packet toward the final destination. The originality of our investigation is twofold. We focus

at first on implementation details at the medium access and network layers, describing possible strategies to practically achieve cooperation (including opportunistic routing, coded cooperation, or both). Moreover, we evaluate the performance of different cooperative approaches from the network viewpoint. For example, node density has an impact on which of the two aforementioned solutions is preferable. Coded cooperation may suffer problems in crowded networks, as multiple transmissions locally improve link quality but cause an interference increase for neighboring connections. Conversely, opportunistic routing may be difficult in sparse networks, as it might be difficult to find how to re-route a packet. This leads us to the idea of combining these approaches, to exploit the cooperation paradigm in a comprehensive manner. Actually, we show that even a simple joint approach which tries to use opportunistic routing first, and employs coded cooperation when the former is not viable, achieves a great improvement with respect to the two individual approaches. Such a result opens up the possibility of defining advanced techniques, which merge together cooperative routing and coded cooperation, so as to take full advantage of the cooperation paradigm. Finally, we observe that the study of these issues can lead to further considerations, which appear as interesting directions for future research. In many related papers, cooperation is assuredly granted, as it was assumed to come from goodwill of wireless users. Indeed, one may doubt if cooperation among wireless terminals is likely to occur, unless it is properly encouraged and rewarded. This is supported by many studies which apply, for example, game theory to wireless networks [13]. From a selfish standpoint, cooperative nodes have nothing to gain and everything to lose; the benefits of cooperation are evident only at a global level, not locally. Surely, evidence of the increased welfare in a cooperative network can be convincing for some of the wireless users, and mechanisms to promote cooperation among them can be used. However, in order to analyze these behaviors it is in our opinion mandatory to include medium access and network details, as done in the present work, to determine where and how cooperation can be introduced.

2.1 Related Work

MIMO systems have been widely investigated in the technical literature since the pioneering work by Foschini [3, 14]. Nevertheless, their implementation in ad hoc networks is relatively recent, and mostly refers to directional antennas and spatial bemforming [15–20]. The design of protocols for wireless networks with omnidirectional transmissions, where

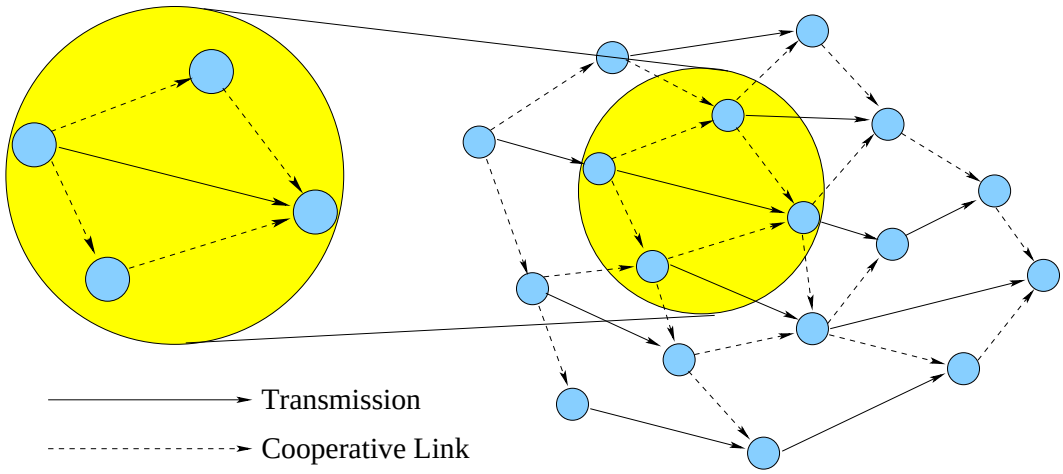


Figure 2.1. *Cooperation from a network-wide perspective.*

the superimposed signals are separated through signal processing at the receiver, has been considered in [21–24]. When speaking of MIMO wireless networks we refer to the latter class of works mentioned above.

Cooperation, by letting nodes transmit information belonging to other nodes' communication, improves throughput and reliability of wireless links by means of channel diversity [6,7,25]. There exists a wide literature addressing cooperation in wireless networks, and a variety of techniques have been proposed. These works mainly analyze cooperation at the physical layer, and over a single link, and often ignore the rest of the surrounding network. In [25] the transmission is based on a TDMA system, which lets the cooperator retransmit what it received from the source if necessary, and the importance of feedback is highlighted, whereas in [6,7] two nodes cooperate in delivering their data by simultaneously using both their channels towards the same destination. A similar technique is proposed also in [26], where the authors add the possibility of combining the data of two sources via Space Time Block Codes. The capacity and achievable rates of the basic cooperative relaying scheme, where the cooperator only retransmits the packet received from the source, were investigated in [12,27,28], and protocols based on this technique have been designed and studied in [10,29–31].

In coded cooperation, the cooperating nodes transmit different parts of the source codeword, which are then recombined at the destination in order to improve the decoding probability. This idea was first explained in [32], and the frame error probability is calculated in [33], where block fading and orthogonal channels are assumed. The performance of different coding techniques, applied to coded cooperation, has been calculated in [33–36], and

nodes equipped with multiple antennas were considered in [36]. Cooperation in location–based sensor networks was exploited to improve the reliability of multi–hop paths in [37], where the cooperator uses Decode&Forward to help when the next link of the path suffers harsh channel conditions. In [11] a cooperation scheme based on hybrid automatic retransmission request (HARQ) has been proposed for an ad hoc network scenario similar to those considered in this paper. Cooperation, together with the great adaptability of HARQ error control, has been shown to increase the performance of the network. A few works have recently focused on various cooperative routing techniques [4, 38–45]. The simplest and most studied scheme consists of using idle nodes as relays, either in place of the original source or in addition to it. In the former case a new route is opportunistically created while forwarding, taking advantage of the links of good quality; in the latter, multiple relays try to combine their signals at the next hop of the route, either superimposing them coherently or transmitting with different delays. These techniques are based on the Decode&Forward cooperative scheme, since only nodes which correctly decoded the entire packet can help as cooperators. In [38], cooperative relaying is analyzed for UWB networks, where routing is based on predefined routes, but other nodes which overhear a transmission can become additional relays. The main purpose is to exploit spatial diversity, by dividing the power between the source and the cooperator. The choice of the best relay is investigated in [39], based on the conditions of both channels, from the cooperator to the source and to the destination. Similar conclusions are drawn in [40]: in order to avoid collisions, the best relay is the first which retransmits the packet received, once a link quality metric is defined. Multiple relays instead are considered in [41], where they create an effective virtual antenna array. The authors here introduce two different protocols, where the link cost is based respectively on the power consumption and on the physical distance. However, cooperators are supposed to adjust their transmitted power and phase in order to superimpose their transmissions at the destination.

In most of these schemes, the assumption of a fixed routing table, where all paths are recorded, is removed, thus allowing a dynamic route selection and cooperative packet forwarding. In [42] the forwarding of the packet is based on a relaying structure from a source to a destination. All nodes are assigned a decoding order, depending on the number of hops between them and the destination, such that all the nodes at the same hop distance cooperatively forward the packet. However, a large amount of information is needed to avoid interference problems due to simultaneous transmissions. The same idea is analyzed also

in [43], where the number of relays for each hop is limited to 2, which cooperate by using a Space Time Block Code, and in [4, 44, 45], where the problem of finding the route which minimizes the total energy consumption is treated under some assumptions on the physical layer structure.

Unlike the previously mentioned work on cooperative routing, the scheme presented in this work proposes the integration of advanced cooperative error control techniques and opportunistic routing.

2.2 Network Description

We consider a multi-hop wireless ad hoc network where nodes are equipped with an array of antennas. Transmissions are omnidirectional and a MIMO architecture is implemented at the transmitter and the receiver. The proposed protocol does not require a specific MIMO system, and we keep the discussion general until Section 2.4, where we provide a specific case study for concreteness. However, our working assumption is that the multiuser capabilities of the underlying MIMO system implemented at the nodes of the network enables the superposition of multiple concurrent transmissions. We remark that this does not mean that we are assuming an idealized physical layer model. In fact, we explicitly include error control in our protocol in order to cope with decoding errors due to interference and fading.

As mentioned before, carrier sense and collision avoidance are not suitable protocols, as the MIMO physical layer reduces the impact of incoming interference and therefore multiple links can be established simultaneously. In the following we describe the basic operations of the network.

All the nodes of the network have a finite-length buffer, in which they store packets to be sent to a destination randomly chosen among the other nodes.[2] As an initial setting, we assume that packets are forwarded to their destination according to a predefined shortest-path routing table. We adopt the following nomenclature:

- *source*: node that generated the packet;
- *destination*: final recipient of the packet;

[2]We consider unicast communications. A destination is independently selected for each new packet in the buffer.

- *transmitter*: node currently in charge of forwarding the packet;
- *receiver*: node currently receiving the packet;
- *next hop relay*: node selected by the current transmitter of a packet for packet forwarding according to its predefined routing table;
- *opportunistic relay*: node opportunistically selected for packet forwarding according to instantaneous channel and network conditions;
- *HARQ relay*: node that cooperates with the transmitter by sending HARQ packets to the receiver.

Before transmitting data, transmitter-receiver pairs perform a handshake phase, during which they exchange short control messages in order to check the receiver's availability and to estimate the quality of the channel. Those control messages are sent at a low transmission rate, in order to have a high probability of being correctly received.[3] Thus, the transmitter sends a request packet (RP), and the intended receiver, if it is not involved in another communication and the signal-to-interference-plus-noise ratio (SINR) at the output of the receiver is above a certain threshold Γ_t, replies with a confirmation message (CP). If the destination is available, but the perceived SINR is below the threshold Γ_t, then a negative CP (nCP) is sent. We underscore that, unlike in CSMA protocols, the control messages are sent by the transmitter and the receiver without a preliminary channel sensing, and that the neighbors who decode those messages are not forced to be silent. The threshold Γ_t avoids the transmission of data packets over channels experiencing fading dips or high interference, in order not to increase the interference load in congested areas of the network with transmission of packets with a small probability of being successfully received.

If the handshake succeeds, then the transmission of a data packet starts, otherwise the transmission of a further request message is scheduled after a random backoff interval.[4] The transmission of a data packet is performed following a hybrid ARQ rationale. Packets are encoded with a low rate error correction code. After the handshake, the transmitter sends a fragment of the obtained codeword, called HARQ packet in the following, which, if received correctly, allows the receiver to recover the whole information content. If this is the case, the receiver sends an acknowledgment packet (ACK). If the decoding fails, then a negative acknowledgment (nACK) is sent, and the source transmits a different fragment of the

[3] In some MIMO architectures this means that they also generate less interference.

[4] Note that if the receiver is non-available, then the CP packet is not sent. In this case, the transmission ends.

codeword with a second HARQ packet, which will result in a further ACK/nACK, and so on. A codeword fragment plus acknowledgment transmission is referred to as *transmission phase* in the following. In each transmission phase, the receiver collects a new fragment of the codeword and combines it with those previously received in order to have a lower rate codeword (and thus to achieve a smaller error probability). The amount of redundancy sent in a phase may be, in general, a function of the channel quality and of the amount of redundancy sent in the previous phases. The transmitter-receiver pair continues to perform transmission phases until an ACK is eventually received or the maximum number of phases M_{ph} is reached. We call *delivery attempt* the transmissions performed by the source and the receiver from the handshake to the first received ACK or to the M_{ph}–th nACK.[5]

Upon the reception of the M_{ph}–th nACK, the current delivery attempt is aborted and a new one is scheduled after a random backoff interval. After M_{tx} unsuccessful delivery attempts, including failed handshakes, the packet is returned to the upper layers.

In case of a successfully received packet, the receiver checks the final destination of the packet. If the current receiver is the destination, then the packet is sent to the upper layers, otherwise it will be scheduled for transmission to the next hop node.

We do not force a synchronous time structure for delivery attempts. Therefore, a node may start a new delivery attempt even if other delivery attempts are currently ongoing. Due to this asynchronous start and end of delivery attempts, the interference load during a delivery attempt may suddenly change. The incremental redundancy HARQ error protocol described above allows the transmitter to adapt the coding rate to the instantaneous fading and interference conditions. This is of fundamental importance, since, besides increasing the delivery success rate, the adaptation capabilities of the HARQ protocol improve the transmission efficiency and reduce the interference in the network.

The SINR threshold, the random backoff and the HARQ error control scheme can control the interference load and preserve the reliability of the communications. Nevertheless, it is clear that a packet may incur a long delay if its predefined route has as the next hop a node in a congested area, as the node currently in charge of forwarding the packet may enter backoff due to the reception of a nCP. Moreover, the additional delay also impacts the other packets in the buffer of the node. Note that if the congestion conditions persist, a packet may be dropped when the node fails M_{tx} times to deliver it. In the following section we present

[5] We refer to delivery attempt also in case of failed handshake.

the integrated cooperative packet forwarding and error control protocol that enables a more flexible and effective delivery of the packets to distant destinations.

2.3 Integrated Cooperative Packet Forwarding and HARQ Scheme

As observed in the previous section, in the presence of highly variable instantaneous interference conditions standard routing approaches that use fixed or slowly-changing routing tables may fail to provide satisfactory performance in terms of delivery success rate and delay.

In traditional ad hoc networks, failure events are mostly connected with transient fast fading impairments, and the routing tables may be able to follow the relatively slow variations of the average channel gain of the link to the next hop relays. Conversely, the evolution of the local interference conditions is a bursty stochastic process, where there may be substantial variations of the average local interference even during the transmission of a single packet. Therefore, a link may be declared as interrupted due to short-term high interference load conditions, that may be entirely different by the time a new route to the destination is discovered.

In order to cope with these issues, we propose a cooperative protocol, referred to as Cooperative HARQ and Opportunistic Routing (CHOR) protocol in the following, that adaptively manages packet forwarding by adaptively and opportunistically selecting an efficient route to the destination. We underscore again that an inefficient route, where each hop requires long delivery attempts, *i.e.*, attempts with many transmission phases, may increase the interference load of the network and damage other communications by increasing both their failure rate and their waiting time before packet transmission.

The CHOR protocol adds two main features to the basic communication protocol:

- it enables other nodes that overheard the transmission of a packet to send fragments of the codeword obtained by re-encoding it.
- when needed, it opportunistically switches to a different next hop relay in order to minimize the amount of redundancy sent and to guarantee a geographical advancement to each packet.

In the following, we assume that all nodes know the hop distance with respect to the other nodes of the network. This topological knowledge is required in order to guaran-

tee that the opportunistic path moves toward the destination. In fact, we need a metric $\mathcal{H}(\mathrm{N}_1, \mathrm{N}_2)$ to indicate how "far" from each other the two nodes N_1 and N_2 are. This metric has to satisfy obvious relationships of being non-negative, zero if and only if $\mathrm{N}_1 = \mathrm{N}_2$, and subject to triangular inequality. The trivial choice is hence to choose $\mathcal{H}(\mathrm{N}_1, \mathrm{N}_2)$ as the number of hops between N_1 and N_2, which is indicated by the routing table. We also define $\mathcal{S}(\mathrm{N}_1, \mathrm{N}_2)$ as the instantaneous post-processing SINR perceived by N_2 when decoding the signal from N_1.

Let us explain step-by-step the CHOR protocol following the diagram in Fig. 2.2. Consider a packet in the buffer of node T that is to be delivered to node D, and let node R be the next hop relay. The preliminary handshake is analogous to that described in the previous section.

If node T receives a positive CP, *i.e.*, $\mathcal{S}(\mathrm{T}, \mathrm{R}) > \Gamma_t$, then it transmits the first HARQ packet. If R correctly decodes the message, then it transmits an ACK and the delivery attempt successfully ends. If R fails to decode the packet, than a nACK is sent and a further transmission phase starts. Differently from the basic communication protocol, the transmitter initially keeps idle in order to collect special control messages, namely cooperative ACK (cACK) from some of its neighbors. In particular, if a node N belonging to both T's and R's neighborhoods[6] correctly decoded the packet sent by T and its number of hops to the destination of the packet is equal to or smaller than that of R, that is $\mathcal{H}(\mathrm{R}, \mathrm{D}) \geq \mathcal{H}(\mathrm{N}, \mathrm{D})$, then N transmits a cACK packet with a special flag active, namely $\mathrm{cACK_f}$. If T receives a $\mathrm{cACK_f}$, then the delivery attempt ends and N stores the packet in its buffer and is in charge of forwarding it to D. In fact, a $\mathrm{cACK_f}$ means that the packet has been decoded by a node that guarantees a routing and geographical advancement toward the destination D, thus the current relay opportunistically decides not to continue with the delivery attempt, thus being able to start other communications and reducing the amount of interference generated to forward the packet. Thus, N becomes an opportunistic relay.

If the node N failed to decode the packet, or $\mathcal{H}(\mathrm{R}, \mathrm{D}) < \mathcal{H}(\mathrm{N}, \mathrm{D})$, but the following condition holds

$$\mathcal{S}(\mathrm{R}, \mathrm{N}) \geq \Gamma_t, \tag{2.1}$$

then it sends a cACK, with the aforementioned flag not active. $\mathcal{S}(\mathrm{R}, \mathrm{N})$ is the post processing

[6] As will be clarified in the next section, the neighborhood of a node N, corresponding to the set of nodes that can directly communicate with N, is defined by a threshold on the minimum average signal-to-noise ratio.

SINR that N measured when receiving the ACK sent by R. The goal is to select the neighbors of both T and R that have good channel conditions to R. Note that the information N acquires by decoding the ACK is a biased version of $\mathcal{S}(\mathrm{N},\mathrm{R})$. In fact, although the fading coefficient of the channel between N and R is the same at both nodes, they are in a different position, and thus the perceived interference may lead to different post-processing SINRs.

Upon the reception of one or more cACK, T keeps idle, while all the nodes who sent the cACKs transmit HARQ packets. Those nodes are referred to as HARQ relays according to the previous definitions. The specific content of those HARQ packets depends on the specific implementation, the architecture of the receiver and the code used. For instance, if a node decoded the packet, but does not meet the condition on the hop count, then it can re-encode the packet and send randomly chosen redundancy fragment. If it failed the decoding, than it can send an amplified version of the received signal. In case of hard decoding and block codes, the HARQ relay may be able to send a fragment of redundancy obtained from a reduced version of the code. We will explain this case in detail in the next section.

The HARQ relays then continue to transmit HARQ packets until either an ACK is sent by R or the maximum number of phases is reached. In the former case, node T considers the packet as delivered, while in the latter case it schedules a new attempt after a random backoff interval (or drops the packet if the number of attempts already performed is equal to M_{tx}).

Thus, if the packet cannot be released to a node closer to the destination, the HARQ relays help the current transmitter to forward the packet to the predefined next hop relay.

If the transmitter receives a negative CP (nCP), then instead of quitting the delivery attempt as in the basic protocol, it keeps idle for a time sufficient to detect the transmission of cooperative CP (cCP) messages. In fact, a neighbor of N that overheard the handshake, and that is not yet involved in other communications, may declare its availability for cooperation. If the following conditions are met:

$$\begin{aligned} \mathcal{H}(\mathrm{R},\mathrm{D}) &\geq \mathcal{H}(\mathrm{N},\mathrm{D}); \\ \mathcal{S}(\mathrm{T},\mathrm{N}) &\geq \Gamma_t \end{aligned} \tag{2.2}$$

then N sends a cCP message with a flag active $\mathrm{cCP_f}$, that indicates that it is a good alternative to R, as it is experiencing good channel conditions to the current relay and provides a positive advancement toward the destination.

If T receives a $\mathrm{cCP_f}$, then it starts the first transmission phase, but sending the packet to

the sender of the cCP_f packet instead of sending it to R, who returns idle. The transmission is then performed as described in the previous case, potentially admitting other nodes cooperation upon the reception of a nACK. If multiple cCP_fs are received, then T randomly selects one of the senders as the next hop relay.

If one or more cCPs are received by T, but no cCP_fs have been sent, then T performs the first transmission phase with intended receiver R. In fact, in this case there is no proper alternative for releasing the packet, but the conditions on the SINR ensure that there are potential HARQ cooperators that will help the delivery of the packet to the next hop relay.

If no cCPs are received, T quits the attempt.

A graphical representation of a packet delivery on a two-hop path is depicted in Fig. 2.3. The initial transmitter changes the next hop relay due to fading/interference impairments. Then two HARQ relays help the node currently in charge of forwarding the packet in delivering it to its destination.

As a final remark, we do not assume that control messages are error-free. The incorrect decoding of CP, nCP, ACK, nACK messages are interpreted by the receiving nodes as negative CP and acknowledgments, while if T fails to decode a cCP and cACK message then it disregards the existence of the message. Note that the failed decoding of some control messages may result in the de-synchronization of the operations of the nodes involved in a communication. Moreover, if the destination is unavailable, then the cooperative packet forwarding cannot start. However, the nodes are, in some cases, able to resynchronize by decoding the subsequent control packets sent over the channel.

2.4 Example of Implementation

In the previous Section we provided a general description of the proposed cooperative framework. In this Section, as a concrete example, we describe in detail a specific instantiation of this concept in a protocol stack based on the powerful Bell Laboratories Layered Space-Time (BLAST) architecture and on linear erasure codes (LEC) packet encoding.

2.4.1 Physical Layer

Our work focuses on ad hoc networks where nodes are equipped with multiple antennas and use an MMSE version of the LASTMUD scheme described in [46]. This scheme is a particular implementation of the BLAST architecture, but is not necessary for our proposed

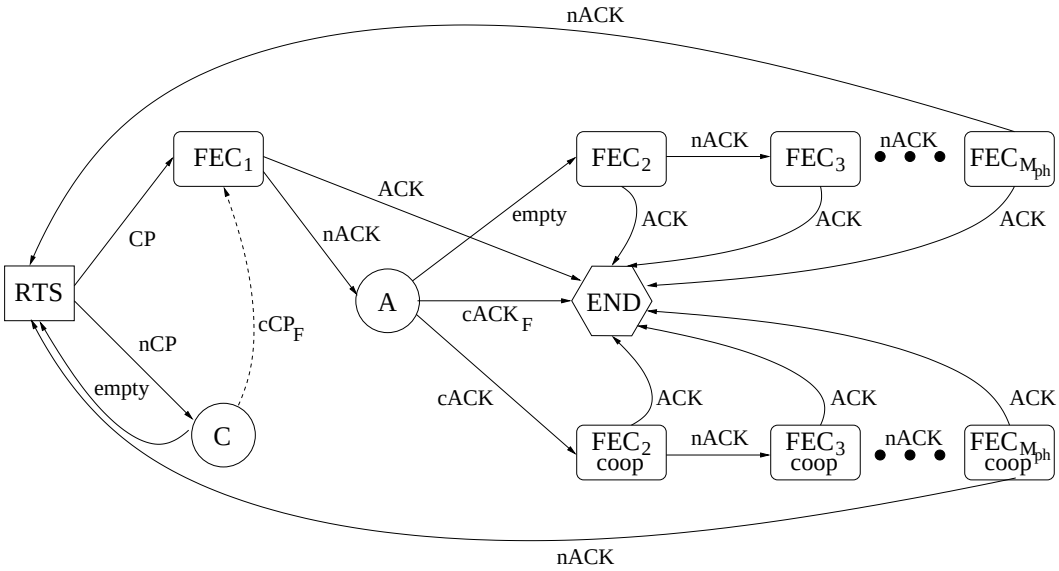


Figure 2.2. *Graphical representation of the proposed communication scheme.*

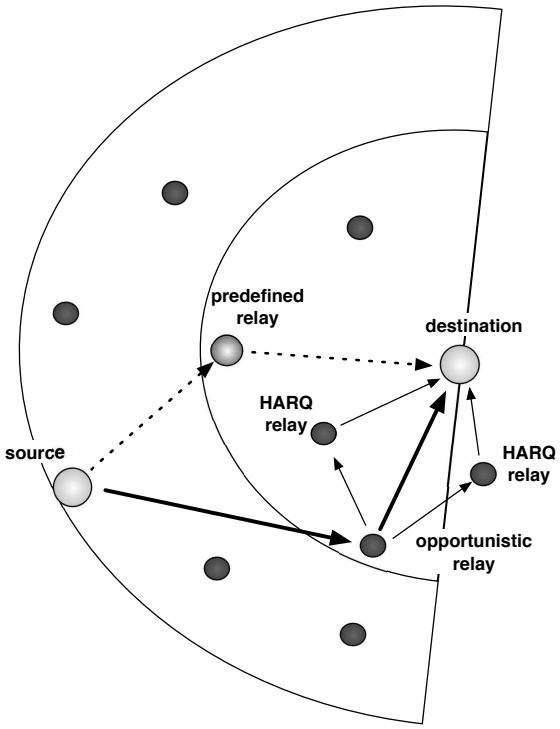


Figure 2.3. *Example of packet forwarding with the CHOR scheme.*

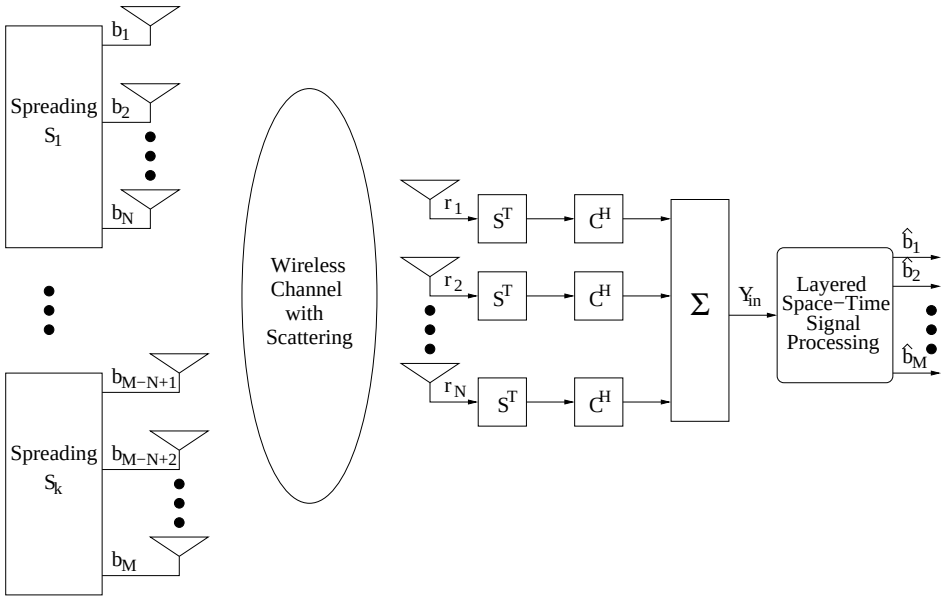


Figure 2.4. *Example of MIMO scenario. M nodes equipped with N antennas trasmit to the same destination, also equipped with N antennas. The superimposed signals are decoded through use of a BLAST decoder.*

protocol to work, and could be replaced by any other scheme which allows the existence of multiple simultaneous transmissions. A general picture is depicted in Fig. 2.4.

We assume that each node, equipped with N antennas, is assigned a specific spreading code of length N_s, thus creating a CDMA system. We also assume symbol synchronization, although this is not strictly necessary. In fact, asynchronous transmissions may be handled by the receivers at the cost of an increased complexity. The symbols to be transmitted are equally divided among the antennas, thus achieving the benefits of Spatial Multiplexing, spread with the source node spreading sequence, and finally BPSK modulated. The power P_{TOT} used by each node is fixed, and is equally divided among the antennas used.

The channel between antennas i and j, belonging to two different nodes n and m, is modeled as flat Rayleigh fading, and is considered constant during the transmission of a block of bits. It can be expressed as $h_{i,j}^{(n,m)} = \sqrt{\alpha^{(n,m)}}\phi_{i,j}^{(n,m)}$, where $\alpha^{(n,m)}$ is the path loss between the source n and the destination m, while $\phi_{i,j}^{(n,m)}$ is a complex Gaussian random variable with zero mean and unit variance, accounting for fading. Moreover, the channel is time correlated with correlation factor ρ, hence the fading coefficient at time k is computed from the one at time $k-1$ as $\phi_{i,j}^{(n,m)}(k) = \rho\phi_{i,j}^{(n,m)}(k-1) + \sqrt{1-\rho^2}\xi(k)$, where $\xi(k)$ is an i.i.d. complex Gaussian random process with zero mean and unit variance.

We give now a brief description of the decoding algorithm. If nodes t=$1, 2, \cdots, T$ trans-

mit in a given time slot, the signal received at node m can be evaluated as follows. We consider the $N_s \times T$ matrix $\boldsymbol{S}$, whose columns are the spreading sequences of the T transmitting nodes, and the vector $\boldsymbol{H}_j^{(t,m)}=[h_{1,j}^{(t,m)}, \cdots, h_{N,j}^{(t,m)}]$, which collects the complex channel gains between all the antennas of node t and the j–th antenna of node m. Finally, we denote as $\boldsymbol{d}$ the column vector containing the bits transmitted by all the antennas of all the T transmitting nodes. The N_s chips received at the j–th antenna of node m can then be written as:

$$\boldsymbol{r}_j^{(m)}=\boldsymbol{S}'\boldsymbol{C}_j^{(m)}\boldsymbol{d}+\boldsymbol{\omega}_j^{(m)}=\boldsymbol{R}_j^{(m)}\boldsymbol{d}+\boldsymbol{\omega}_j^{(m)}, \tag{2.3}$$

In the above equation, $\boldsymbol{C}_j^{(m)}$ is a diagonal $TN \times TN$ matrix, whose diagonal is given by the elements of all the vectors $\boldsymbol{H}_j^{(t,m)}$ for $1 \leq t \leq T$, i.e., by the channel coefficients between all the transmitting antennas and the j–th antenna of m. The matrix $\boldsymbol{S}'$ is obtained as $\boldsymbol{S} \otimes \boldsymbol{1}_N$, where $\otimes$ is the Kronecker product and $\boldsymbol{1}_N$ is a row vector of length N whose elements are all equal to 1, whereas $\boldsymbol{\omega}_j^{(m)}$ is a column vector containing N_s i.i.d. complex Gaussian random variables with zero mean and power σ^2.

The output of the matched filter is now given by

$$\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{R}}^{(m)}\boldsymbol{d} + \boldsymbol{n}^{(m)}, \tag{2.4}$$

where we used

$$\tilde{\boldsymbol{R}}^{(m)} = \sum_{j=1}^{N} \boldsymbol{R}_j^{(m)H}\boldsymbol{R}_j^{(m)}, \qquad \boldsymbol{n}^{(m)} = \sum_{j=1}^{N} \boldsymbol{R}_j^{(m)H}\boldsymbol{\omega}_j^{(m)} \tag{2.5}$$

and ^{H} denotes the Hermitian operator.

The algorithm then sorts the received streams in descending order of SNR, and starts decoding the one with the highest SNR. Each time a stream is decoded, it is then subtracted from the total received signal, until all streams are decoded. For a more exhaustive description of the algorithm, the interested readers are referred to [46]. In [46], the receivers use zero-forcing weights, corresponding to the Moore-Penrose transform of the correlation matrix $\boldsymbol{R}$, to reduce the impact of still-to-decode signals on the signal currently decoded. In our study, we use instead minimum mean square error weights, i.e., the Moore-Penrose transform of $\boldsymbol{R}^{(m)}+\boldsymbol{I}\sigma^2$, where $\boldsymbol{I}$ is a $T \times T$ identity matrix, in order to improve the performance [47].

We assume that only the neighbors of a receiver are included in the decoding algorithm, while farther away interfering signals are treated as an additional noise term. In order to

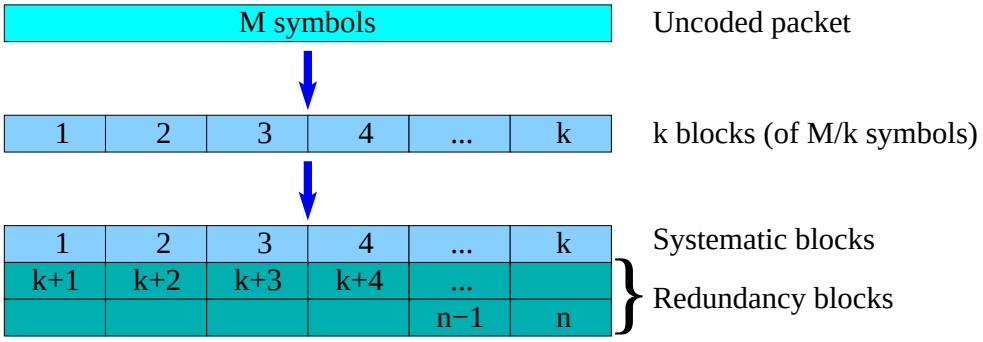


Figure 2.5. *Example of LEC applied to a packet of M symbols. The packet is first fragmented in k blocks, which are then encoded to generate n blocks. The packet can be decoded upon reception of any k out of the n encoded blocks.*

speed up network simulations, the output SINRs of the algorithm are computed with the Gaussian approximation presented in [48].

2.4.2 Packet Encoding

We implement the incremental redundancy error control protocol using linear erasure codes (LEC) [49] These codes are well suited since they also allow the encoding of fractions of the packet and are particularly effective for transmissions affected by bursts of errors [50], but other solutions can be investigated as well. The proposed encoding is depicted in Figure 2.5.

We now briefly introduce the coding technique used. All packets are encoded through a LEC, which is described by the triple $(n, k, \boldsymbol{G})$, where $n{>}k$ and $\boldsymbol{G}$ is a full rank $n \times k$ matrix whose elements are taken from the Galois field GF(2^r). The i–th coded symbol $\boldsymbol{p}_i$ is generated from the original k data symbols $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k$ as

$$\boldsymbol{p}_i = [\boldsymbol{G}]_{i,1}\boldsymbol{s}_1 \oplus [\boldsymbol{G}]_{i,2}\boldsymbol{s}_2 \oplus \ldots \oplus [\boldsymbol{G}]_{i,k}\boldsymbol{s}_k \,, \tag{2.6}$$

where $\oplus$ is the element-wise sum of vectors in GF(2^r). We consider a systematic LEC, such that $\boldsymbol{p}_i = \boldsymbol{s}_i,\ i = 1, 2, \ldots, k$.

In order to retrieve the data symbols, let $\mathcal{D}{=}\{\kappa_1, \kappa_2, \ldots, \kappa_k\}$ be the set of the indices of the first k correctly decoded symbols, and call $\bar{\boldsymbol{G}}_{\mathcal{D}}$ the matrix containing the rows of $\boldsymbol{G}$ whose indices belong to $\mathcal{D}$. Since $\boldsymbol{G}$ is full rank, also $\bar{\boldsymbol{G}}_{\mathcal{D}}$ is full rank, and can be inverted in GF(2^r). Hence, the original symbols can be computed as follows:

$$\boldsymbol{s}_q = [\bar{\boldsymbol{G}}_{\mathcal{D}}^{-1}]_{q,1}\boldsymbol{p}_{\kappa_1} \oplus [\bar{\boldsymbol{G}}_{\mathcal{D}}^{-1}]_{q,2}\boldsymbol{p}_{\kappa_2} \oplus \ldots \oplus [\bar{\boldsymbol{G}}_{\mathcal{D}}^{-1}]_{q,k}\boldsymbol{p}_{\kappa_k} \tag{2.7}$$

for q=$1, 2, \ldots, k$. Since the only requirement for decoding, if $\boldsymbol{G}$ is full rank, is that $\mathcal{D}$ contains at least k elements, no matter which ones, it can be inferred that decoding is possible if any k different coded symbols are received, which is a peculiar property of the Linear Erasure Codes.

In our system, a packet of L_p bits is first split into F blocks $\mathcal{S}$=$\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_F\}$, each of L_b=L_p/F bits. These blocks are the systematic set, which is then encoded with a code $(V, F, \boldsymbol{G})$, where each block can be seen as a code symbol belonging to GF(2^{L_b}). The coded set $\mathcal{C}$=$\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_V\}$ is obtained. Before transmission, each block is finally added a CRC code, such that at the destination the number of missing blocks after each transmission is known. Once the destination receives at least F different blocks, it can retrieve the entire original packet.

2.4.3 Networking

In this section, we show how the networking protocols proposed in this paper are implemented based on the physical layer and the packet encoding described before.

We assume that time is divided into slots, and transmissions start only at the beginning of each slot, although their duration may involve several consecutive slots. Note that this does not mean that the start of all delivery attempts is synchronous, as they can start in slots when other delivery attempts are already active. The channel between each pair of antennas is considered constant during a single slot, and the use of short training sequences at the beginning of each slot lets the receiving nodes estimate the channel and perform LASTMUD decoding.

Thus, packets are encoded with the LEC described before, and we assume that the transmission of a block of symbols generated by the encoder matches the duration of a slot.[7]

We also assume that control messages are encoded with a convolutional code of rate $1/2$, they are transmitted simultaneously with all the antennas (if only one antenna is available, other techniques are possible, such as a repetition code applied to the whole signaling packet) and their transmission fits the duration of one slot.

The error control protocol requires that different parts of the codeword be sent in the various transmission phases. We specify now how this process is carried out based on the use of LEC. As explained in Section 2.4.2, the packet is divided into blocks, whose set is

[7]We observe that thus the duration of a slot is inversely proportional to the number of antennas. In fact, the transmitter splits the bits to be sent among the available antennas.

called $\mathcal{S}$. The encoding of these blocks generates the coded set $\mathcal{C}$. Since a systematic code is used, $\mathcal{S}$ is also a subset of $\mathcal{C}$. In the first phase, the transmitter transmits all the F blocks of the systematic part $\mathcal{S}$, one for each time slot, and with all the N antennas, as previously explained. If the receiver failed to decode the packet, then it sends a nACK specifying the number of blocks N_{BL} still missing, and also the set $\mathcal{D}_1 \subset \mathcal{C}$ of correctly decoded blocks. The source consequently starts the transmission of other N_{BL} blocks, randomly chosen in $\mathcal{C}\backslash\mathcal{D}_1$, followed by a new ACK by the destination. The possible further transmission phases are carried out in an analogous way, where at the $f+1$–th phase the set of blocks in which the transmitter selects is $\mathcal{C}\backslash\mathcal{D}_f$, and $\mathcal{D}_f$ is the set of blocks correctly received in $1,\ldots,f$.

The threshold Γ_t is numerically computed through Montecarlo simulations such that the failure probability of the subsequently sent packet is below a fixed threshold (0.9 in the numerical results). These simulations are carried out in the absence of interference from other transmissions, while the mutual interference of the signals sent by the various antennas of the transmitter is taken into account.[8] We have chosen to numerically evaluate Γ_t due to the high complexity of the receiver and the correlated statistics of the channel coefficients. Moreover, we cannot set an a priori statistics of the interference, since this is a function of the protocol implemented as well as of the distribution of the nodes in the network.

In the CHOR algorithm, HARQ relays that decoded all the blocks sent in the first transmission phase randomly select the blocks to be sent in each phase following the same procedure described before. In case the HARQ relay has available only a subset of the blocks sent in the first phase, then it uses a reduced code to obtain a different coded set. In particular, a node can be a HARQ relay only if it decoded all the blocks missing at the receiver, and in that case it randomly picks up the blocks to be sent in the next transmission phase in the coded set obtained by encoding the blocks in $\mathcal{S}\backslash\mathcal{D}_1$.

As a final detail, the backoff periods are uniformly chosen in the exponentially increasing window $[1, 2^p W]$, where p is the number of failed delivery attempts for a given packet.

[8] This choice is motivated by the impossibility to have an a priori estimation of the interference in the network. The hop distance is thus computed in the absence of interference with a margin for the error probability sufficient to preserve the reliability of the link when the entire network is considered. We remark that this approximation is only used to evaluate the threshold Γ_t, whereas in the simulation results of Section 2.5 we accurately account for the total interference in the network.

2.4.4 Routing

The predefined routes are computed by means of a shortest path algorithm, once a maximum hop length d_{hop} has been fixed. Similarly to Γ_t, the value of d_{hop} is numerically computed as the maximum distance between a transmitter and a receiver providing a failure probability below a certain threshold. Again, this quantity is evaluated in the absence of interference. It must also be highlighted that the metric used by the chosen shortest path algorithm is the number of hops independent of the instantaneous channel conditions. As observed before, due to the quick variations of the interference conditions, this may lead to significant delay and throughput loss. It is worth to observe that routing protocols based on periodically computed tables or computing the whole route before the source starts to transmit a packet (proactive and reactive routing, respectively) may fail to provide satisfactory performance due to the quick variations of the channel conditions. However, any other routing protocol can be used as a base for the proposed integrated error control and opportunistic scheme. In fact, the protocol only needs for a reference selection for the next hop relay and a distance measure to the destination.

2.5 Numerical Results

In this section, numerical results comparing the performance of CHOR with that achieved by the basic communication protocol are shown. We also plot the performance associated with the protocol that implements opportunistic routing only, *i.e.*, it does not enable the transmission of HARQ packets by other nodes. We refer to the latter protocol as opportunistic routing (OR) protocol. The results are obtained through extensive MATLAB simulations, since an exhaustive theoretical analysis of the entire network is infeasible, also due to the complex interference interactions between the communications of the various nodes of the network.

We developed a simulator which describes the network from the physical to the routing level. Performance is averaged over random topologies, where nodes are uniformly placed in a fixed square area of side s=500m. Packet arrivals in the buffer of each node are generated according to a Poisson process of intensity λ_g packets per second per node, and the destination of each packet is randomly chosen among all the other nodes in the network. We assume that a fixed fraction of the queue is left for packets received from other nodes which must be forwarded to their final destinations. A timeout is also set, and packets which

Table 2.1. *System Parameters.*

Phy Layer and Environment	
Modulation	BPSK
Number of array elements	2
P_{TOT}	0.25 W
Spreading factor N_s	16
Bit-rate per antenna	$7.5/N_s$Mb/s
Control pkt FEC polynomial (rate 1/2)	$133_8, 171_8$
Noise power σ_w^2	-170 dBm
Path loss factor β	4
Channel correlation ρ	0.9
Packet generation parameter λ_g	5 pkt/s per node
MAC and HARQ	
Number of blocks per packet F	8
Block length	512 bits
Erasure code parameters (n,k)	(8, 24)
Maximum number of phases M_{ph}	5
Maximum number of attempt M_{tx}	8
Initial windows W	4
N_{fails}^{MAX}	5
P_G	0.99
Threshold SNR_{TH}	9.44 dB
Queue length	64 pkts
Relaying queue length	32 pkts
Routing	
Maximum hop length	163 m

have not yet been delivered when it expires are dropped. Finally, each node, as previously explained, has a routing table, generated via Shortest Path algorithm, in which the routes towards every possible destination are stored.

The slow fading channel is modeled as described in Section 2.4.1, with fixed correlation factor ρ, and the path loss coefficient is chosen such that $\alpha^{(m,n)} \propto [d^{(m,n)}]^{-\beta}$, where $d^{(m,n)}$ is the distance between node m and node n. As stated before, the performance of the physical layer was modeled as in [48]. The parameters used in the simulations are collected in Table 2.1.

Fig. 2.6 depicts the end–to–end throughput for the basic, OR and CHOR protocols as a

function of the number of nodes in the network, each equipped with only one antenna. As can be seen, the throughput grows for all the protocols as the number of nodes increases, until it reaches a floor. In fact, as the number of nodes is increased, the network is initially able to support the greater traffic generated by the transmissions, whereas beyond a certain value the level of parallelism that the network can sustain saturates. This is due to the effect of interference, that does not allow other communications to start (SINR threshold) and results in an increased failure rate of the delivery attempts.

The OR protocol achieves an improved throughput with respect to the basic protocol. In fact, although a slightly larger signaling overhead is required, the capability of choosing opportunistically the path improves the probability that the packet is eventually delivered to its destination. Moreover, the avoidance of bad links enables more efficient communications, and thus the interference load decreases and each packet is delivered in a shorter time. The CHOR protocol outperforms both the OR protocol and the basic communication scheme. The possibility for the neighbors with a good channel to the next hop relay to help the HARQ process by sending HARQ packets on behalf of the source speeds up the delivery process and reduces the probability that nodes enter backoff. However, cooperative HARQ may result in bursty transmissions if the number of HARQ relays is large, an event that may happen in densely populated networks. The burst of cooperative transmissions may generate short-term interference congestion conditions, that may be harmful to the other ongoing communications. The effectiveness at the intended receiver depends on the tradeoff between the amount of collected redundancy and the quality of the channel experienced by each HARQ packet. This effect can be observed in Fig 2.6 when the number of nodes is larger than 120 and the throughput of the CHOR protocol slightly decreases.

Fig. 2.7 depicts the success probability on paths with different path lengths. Both the OR and CHOR protocols grant an improved throughput when delivering packets to far-away destination. For one–hop paths, the improvement granted by CHOR is due to the effectiveness of distributed HARQ. The OR protocol shows the same performance as the basic protocol, since they behave identically for single hop communications. The help of HARQ relays in the CHOR protocol increases the success probability of each hop of the path, and thus the overall benefit for long paths is more pronounced than that of the OR protocol. For instance, for four-hop paths CHOR shows a success probability that is more than doubled compared to the basic communication scheme. OR also provides better performance than the basic protocol, thanks to the more efficient paths.

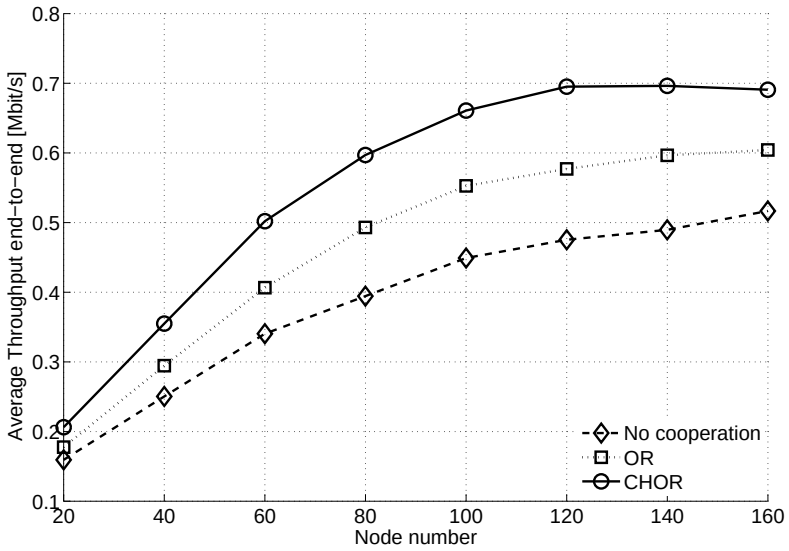


Figure 2.6. *End–to–end throughput as a function of the number of nodes in the network. The area is kept constant, and the number of antennas for each node is 1.*

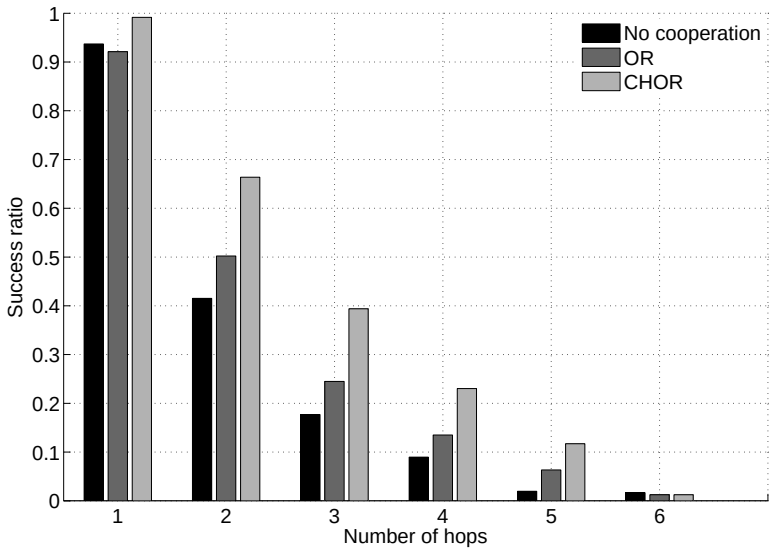


Figure 2.7. *End–to–end success delivery ratio, on paths with different numbers of hops. Here the number of array elements is 1, and the total number of nodes is 40.*

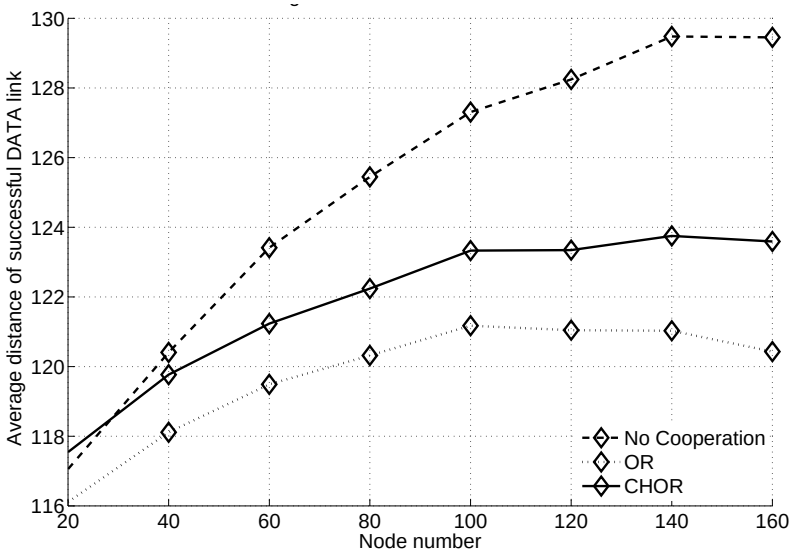


Figure 2.8. *Average distance of successfull DATA links for basic, OR and CHOR protocols as a function of the number of nodes in the network. The area is kept constant, and the number of antennas for each node is 1.*

We plot in Fig. 2.8 the average distance between the transmitter and the receiver of successful delivery attempts. The basic protocol offers the largest advancement, since the routes are calculated such that the next hop is the node closest to the final destination among those within the transmission range of the source. As observed in Fig. 2.7, however, this choice may lead to a smaller success probability for a larger interference level, and consequently to a worse throughput. OR on the contrary selects the shortest links. In fact, in long range transmissions the transmitter is more likely to receive a nCP, and thus a next hop relay at a short distance may be chosen. Moreover, when multiple relays are available, the one with the best channel towards the source is chosen, which is likely to be the closest to the source satisfying the condition on the hop distance. The CHOR protocol lies between the basic and the OR protocol. In fact, it implements the next hop opportunistic switching, that on average results in shorter links, but the distributed HARQ increases the probability that medium and long range delivery attempts are successfully completed.

The effect of the number of antennas used at each node is investigated in Fig. 2.9, where the throughput achieved by the basic and the CHOR protocol is shown for one and two antennas. The use of multiple antennas lets nodes increase the spatial multiplexing level, thus increasing the raw transmission rate. The throughput improves when nodes use two antennas for both the basic and the CHOR protocol. However, while in low-density topologies

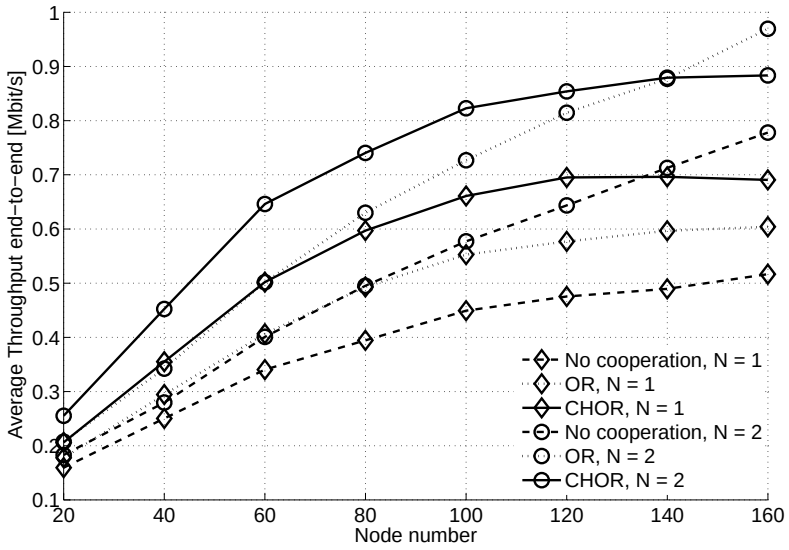


Figure 2.9. *End–to–end throughput for basic, OR and CHOR protocol as a function of the number of nodes, with the number N of antennas equal to 1 and 2. The spreading factor is $N_s = 16$.*

CHOR outperforms the basic protocol in both cases, as the number of nodes becomes larger the difference between the performance of the two protocols gets closer in the two antenna case. This is due to the balance between the increased multiuser capabilities at the receivers and the larger number of information flows due to the presence of two antennas. The CHOR protocol is more sensitive to interference issues, due to the possibility of interference bursts due to cooperative transmissions. This effects reduces the benefit granted by cooperation to the overall network performance.

Finally, Fig. 2.10 depicts the average end–to–end throughput as a function of the spreading factor. It can be observed that the CHOR protocol provides a throughput gain even when the spreading factor is small, and thus when the system is more sensitive to interference. This means that the advantage granted by the reception of multiple HARQ packets overcomes the effect of bursty interference. The OR and the basic protocol perform similarly when the spreading factor is small, since in this case also the interference perceived by the neighbors is likely to be such that the SINR condition is not met. It can be also observed that when the spreading factor is large the OR and the CHOR curves get closer, as the network incurs fewer failures and then the transmission by the HARQ relays is not often needed.

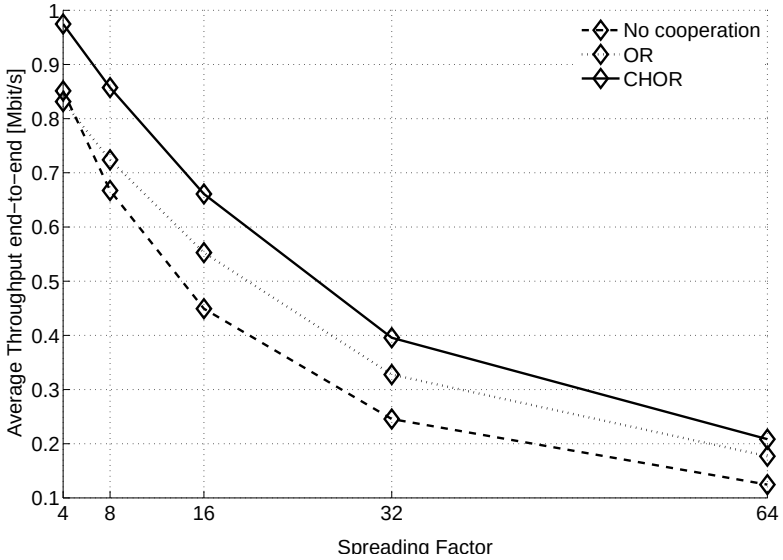


Figure 2.10. *End–to–end throughput for basic, OR and CHOR protocol as a function of spreading factor. The number of nodes is* 100*.*

2.6 Discussion

In this part of the work, we presented a novel cooperative framework for MIMO multi-hop ad hoc networks that integrates opportunistic routing and distributed error control. We focused on MIMO ad hoc networks where the potential of the physical layer is exploited to increase the parallelism of the communications in the network. Our goal is to increase the efficiency and the performance in terms of delay, delivery success rate and interference load per packet achieved by the network. We provided a general description of the proposed algorithm as well as a specific case study where nodes implement a BLAST transmitter-receiver architecture and packets are encoded by means of a low-rate linear erasure code. We presented numerical results showing that the proposed approach, by combining cooperation, error control and routing, grants a significant performance gain to the network.

Chapter 3

Cooperation in CDMA Cellular Networks

The main contribution of this part of the work is to investigate how different cooperative techniques can be practically applied to CDMA cellular networks. In doing this, it is our goal to understand which kinds of benefits may be achieved, which constraints are to be taken into account and how the system should be designed in order to include cooperative nodes.

It is clear that the use of cooperation requires modifications to the current standards, which are likely to be accepted by operators only if the achievable gains are high enough. A second observation is that the benefits obtained through the use of cooperative nodes may have impact on different metrics. For instance, an increased coverage area due to the presence of relays can be instead turned into a power saving gain, if the range is kept constant. Finally, we want to highlight the importance of an analysis which takes into account a whole cell, hence considering also the interference effects and the possibility that a single cooperator may serve more than one source.

Two well known cooperative schemes have been investigated in details, namely Amplify and Forward (A&F) and Decode and Forward (D&F), which appear to be the most practical in cellular networks. Coded cooperation, in fact, may require excessive modifications in the protocol stack.

We focused on Uplink performance, implying that a number of transmission sources, the Mobile Terminals, share the same destination, corresponding to the Base Station. It is well know that, in the Uplink, the coverage range is limited by the maximum transmission power allowed for a Mobile Terminal and by the interference, which, although reduced by

the use of spreading codes, is still proportional to the number of transmitting terminals. If the extension of the coverage range is the aim, then cooperation appears to be a good solution: not only can the presence of a cooperator significantly reduce the equivalent path loss to the Base Station, but, with a proper carrier assignment, it is also possible to reduce the overall interference at the destination, as will be explained in this Chapter.

The Chapter is divided in three parts.

The first part focuses on the analysis of a simple three-node scenario. The aim of this part is to compare different cooperative schemes in terms of outage probability, in order to determine which one is likely to offer the best performance. Moreover, an investigation on how the improvement depends on the position of the cooperator and on its capabilities is also carried out. The results obtained from this preliminary study were useful to figure out which kinds of benefits may be offered by cooperation, using the transmission constraints and the parameters commonly adopted in the study of cellular networks.

In the second part, we exploit the previous results to develop a deeper analysis of the A&F cooperation scheme applied in a wider scenario, where an entire cell is considered. The additional interference generated by the cooperating nodes is taken into account, as well as the cooperator selection among a set of available nodes. The gain in terms of transmission power is derived for the helped nodes, and a new coverage area is defined accordingly. An interesting result on the relationship between the coverage area and the association of the mobile terminals to the different relays leads to important observations regarding the system design, and lays the foundations for an investigation on algorithms to determine the capacity of the cooperative system.

In the third part of the work, we tested cooperation in a full system, multi-cell scenario. When considering multiple cells, and modeling all the layers, from physical to MAC, a theoretical analyis becomes excessively involved. Moreover, the aim of this third part is to remove all the simplifying assumptions commonly accepted in theroretical work (therefore, including power control, traffic model, HSPA mechanisms...) to observe if the gain shown in the analysis can be really achieved in a real cellular system. The cooperative scheme, as well as the cooperator capabilities, were chosen based on the previously shown results, whereas some assumptions and parameters have been removed or modified according to what is stated in the latest standards.

3.1 Related Work

Considerable research efforts have been focused on cooperation in wireless networks due to the promise for increased performance and robustness against channel impairments. Several cooperative techniques have been developed in recent years. However, the fundamental rationale is to let nodes help a communication between a given source-destination pair by transmitting over the channel information belonging to the source transmission, thanks to the broadcasting nature of the wireless channel, that allows the neighbors of a source to overhear its transmission. Cooperation is often represented as a distributed antenna array system, where the array is formed by the different nodes participating in the communications.

Several theoretical studies have appeared in a number of articles. In [27] a four–node network is investigated: the advantages of both transmitter cooperation and receiver cooperation are compared, and capacity bounds are derived. In [51] power allocation is also considered. The outage behaviour of various schemes is investigated in [25] in high SNR regime, where the importance of channel estimation and of limited feedback from the destination are analyzed. Kramer *et al.*, in [12] focus on the best position of the cooperator when various cooperative schemes are used, and generalize the problem to small networks with more than one source or one destination. Cooperation applied to cellular systems is studied in [52], where multiple nodes, grouped in clusters, cooperate in sending their packets to a common base station. Throughput optimization of the whole network, based on the rate of the cooperative links, is analyzed in [53], while a cooperative scheme for nodes with full duplex capability was designed by Hunter and Nosratinia in [26, 54], and extended to wideband channels in [55]. Here, two nodes listen to each other in a first phase and then cooperatively transmitting in a second phase. Such a scheme is less effective in fast fading environments, but a feasible solution for cellular networks, where different channels are available, was then developed in [56], which takes advantage of space-time block codes such as the Alamouti code [27]. A cooperative scheme based on CDMA was presented by Sendonaris, Erkip and Aazhang in [6, 7], where two nodes combine their transmissions towards the same Base Station by exploiting RCPC codes and full duplex capability at the source, and Multi User Detection at the destination. Other cooperative techniques applied to wireless LANs and cellular networks can be found in [57] and in [58] respectively.

Although most prior work focused on ad hoc network scenarios, and generally consid-

ered simple three-node topologies, increased attention has been recently put to the impact of cooperation on the higher layers. In [38], a routing algorithm for UWB wireless networks has been investigated, where cooperating nodes can help the packet forwarding along a predetermined route. A more general algorithm, not based over a static non cooperative route, is instead presented in [45]. Its performance are significantly better than other non cooperative routing algorithms, but interference from other cooperative transmissions in the network is not taken into account. This aspect is well analyzed in [59], where cooperation in a multi–source, multi–destination, multi–hop wireless network is considered, taking into account the performance reduction due to interference among two or more symultaneous cooperative transmissions. A more general cross–layer framework, which aims at guaranteeing fairness and efficiency, while optimizing the node performances, like throughput or outage, can be found in [60].

3.2 Single Link Analysis

This part of the work investigates and discusses some issues and the potential benefits resulting from the use of simple and practical cooperative techniques, such as Decode&Forward (D&F) and Amplify&Forward (A&F) in UMTS cellular networks. We include cooperation in a realistic setting that takes into account the features of UMTS systems and we compute the performance gains in terms of coverage range and success probability. We include in our discussion considerations on how the characterization of the cooperative node influences performance. In particular, mobile nodes and base stations have different antenna gains. If we assume the cooperating node to be a mobile itself we can not take advantage of an increased gain, while if it is a sort of relay station (RS) we can assume it has greater antenna gain then a regular mobile. Note that in the D&F case the cooperating node decodes the source message. This also raises some privacy and security issues that must be considered when designing the system. We show that different schemes and cooperating node characterizations lead to different design directions in terms of optimal cooperating node position to get the maximum coverage range gain.

3.2.1 System Description

In this section, we describe the UMTS–like environment where we want to apply a cooperative scheme. More specifically, we focus on the transmission from a mobile terminal,

which will be labeled as S, towards the Base Station D. In addition, we call C a third terminal, which helps the source in delivering its packets when cooperation is used. This third terminal may be another Mobile Terminal (MT) or a Relay Station (RS). In the latter case we assume that the cooperating node has an increased antenna gain with respect to a MT. This may have an effect on the performance of a cooperative scheme, as will be shown in Section 3.2.2.

We call d_{xy} the distance between terminal x and terminal y. Each MT can choose its transmission power $P_t \leq P_m$, where P_m is the maximum transmission power. We denote the antenna gain at the Base Station with B_g, and with M_g the antenna gain of a MT. In this work we consider a fixed transmission rate R, and a fixed CDMA spreading factor. The use of CDMA grants an additional gain C_g to the received SNR. We consider a Rayleigh channel with flat fading, with path loss between x and y proportional to $d_{xy}^{-\alpha}$. The fading coefficient δ_{xy} is given by a complex Gaussian random variable with zero mean and unit variance, and is considered constant during the transmission of a single packet. Finally, we denote with N_0 the noise power and we assume a SNR threshold model for decoding, where the destination successfully decodes the incoming packet if the perceived SNR is greater than η. All the considered values for the parameters are listed in Table 3.1. As regards the channel, we consider a first-order correlation model, where if at time t the complex channel gain is $\delta_{sd}^{(t)}$, at time $t+1$ it can be written as $\delta_{sd}^{(t+1)} = \rho\delta_{sd}^{(t)} + \sqrt{1-\rho^2}\xi$, where ξ is another complex Gaussian random variable with zero mean and unit variance. The outage probability can be written as

$$\zeta = \mathcal{P}\left[\Gamma < \eta\right], \tag{3.1}$$

where $\mathcal{P}[\cdot]$ denotes probability and Γ is the perceived SNR at the BS. We define the coverage range as the maximum distance at which transmission is possible with outage probability lower than a target value $\bar{\zeta}$. Four communication schemes are investigated. The first is the traditional direct transmission towards the BS; the second is the Double Direct Transmission (DDT) scheme, where the source performs in the subsequent time slot a second transmission. The receiver uses Maximum Ratio Combining (MRC), meaning that the global SNR can be obtained as the sum of the two SNRs. The third scheme is based on the A&F cooperative technique. We assume that the packet is transmitted by S on an uplink channel. A third node C which is listening to the transmission from S also retransmits the received signal, amplified by a factor β, to the base station D. Different solutions are possible: C may transmit on another uplink channel, thus requiring a full–duplex connection, or in the sub-

Table 3.1. *System Parameters.*

Maximum TX power P_m	0.125 W
Thermal noise density	-174 dBm/Hz
Receiver noise figure	5 dB
Receiver noise power N_0	-103.2 dBm
Required SNR η	5 dB
Processing gain C_g	25 dB
Base Station antenna gain B_g	18 dB
Rate R	12.2 kbps
Correlation factor ρ	0.9

sequent time slot, as for DDT. The same variants hold for the fourth transmission scheme, based on D&F, where C retransmits only if it decodes the packet from S.

3.2.2 Derivation of Outage Probability

In this section we compute the outage probability for the considered communication schemes as a function of the distances between the various nodes involved in the communication and their antenna gains. This will be instrumental to evaluate the coverage range gain.

Direct Transmission

If a single transmission is performed by the source, the SNR at the destination is given by:

$$\Gamma_{sd} = \frac{P_t}{GN_0} d_{sd}^{-\alpha} |\delta_{sd}|^2 \tag{3.2}$$

where $G = A/(B_g C_g)$ and A is a path loss factor. Since δ_{sd} is a complex Gaussian random variable, $|\delta_{sd}|^2$ is distributed as an exponential random variable of unit mean. Therefore, we can write:

$$\zeta = \mathcal{P}\left[\frac{P_t}{GN_0} d_{sd}^{-\alpha} |\delta_{sd}|^2 < \eta\right] = 1 - e^{-\frac{\eta G N_0}{P_t} d_{sd}^{\alpha}} \tag{3.3}$$

Double Direct Transmission

In this scheme, two subsequent transmissions of the same packet are performed by the source S, in order to exploit time diversity and increase the SNR at the destination thanks to MRC.

It can be shown that once $\delta_{sd}^{(t)}$ is given, $\delta_{sd}^{(t+1)}$ is equal to $V_R + jV_I$, where V_R and V_I are two Gaussian random variables whose means depends on the real and imaginary part of $\delta_{sd}^{(t)}$ respectively, whereas their variances are equal, and depend only on the correlation factor ρ. Hence, the conditional probability distribution of $|\delta_{sd}^{(t+1)}|^2$ can be derived from the pdf of a chi-square random variable. Analytically, the expression of the conditional pdf is:

$$f_{X|Y}(x|y) = \frac{1}{1-\rho^2} e^{-\frac{\rho^2 y}{1-\rho^2}} e^{-\frac{x}{1-\rho^2}} I_0\left(\frac{2\rho}{1-\rho^2}\sqrt{xy}\right) \tag{3.4}$$

where we call, for ease of notation, $Y = |\delta_{sd}^{(t)}|^2$ and $X = |\delta_{sd}^{(t+1)}|^2$, and $I_0(x)$ is the modified Bessel function of the first kind and order 0. As can be observed from this distribution, $|\delta_{sd}^{(t+1)}|^2$ depends on the magnitude of $\delta_{sd}^{(t)}$ but not on its phase. The outage probability can hence be computed as

$$\begin{aligned} \zeta_{DT} &= \mathcal{P}\left[\left|\delta_{sd}^{(t)}\right|^2 + \left|\delta_{sd}^{(t+1)}\right|^2 < \frac{\eta G N_0}{P_m} d_{sd}^{\alpha}\right] \\ &= \int_0^l \int_0^{l-x} f_{XY}(x,y)\,dy\,dx \end{aligned} \tag{3.5}$$

Here, $l = \eta G N_0 d_{sd}^{\alpha}/P_m$, and $f_{XY}(x,y)$ is the joint probability distribution of X and Y; since Y is exponential, we can write:

$$f_{XY}(x,y) = \frac{1}{1-\rho^2} e^{-\frac{x+y}{1-\rho^2}} I_0\left(\frac{2\rho}{1-\rho^2}\sqrt{xy}\right) \tag{3.6}$$

The outage probability is again a monotonically increasing function of d_{sd}; more precisely, the integrand function remains the same, but as d_{sd} increases the area over which $f_{XY}(x,y)$ is integrated becomes larger. Since $f_{XY}(x,y)$ is a pdf, the value of the integral also increases.

Amplify&Forward

By this communication scheme node C retransmits the received signal from the source in the subsequent slot. We consider also in this case MRC at the destination, where the total SNR hence has a component due to the direct transmission and another one due to the relayed transmission. The powers used by S and C are called P_s and P_c respectively, whereas β is the power amplification factor at the cooperator. As said in Section 3.2.1, M_g is the antenna gain at the cooperator, which may be different from B_g.

The outage probability is given by:

$$\zeta_{AF} = \mathcal{P}\left[\Gamma_{sd} + \Gamma_{cd} < \eta\right] \tag{3.7}$$

where Γ_{sd} is the term due to the direct transmission, and Γ_{cd} the one due to the relayed transmission. The former is expressed by (3.2), with $P_t = P_s$. We now focus on the latter. The signal power received at the cooperator is

$$r_c = \frac{P_s M_g}{A} d_{sc}^{-\alpha} |\delta_{sc}|^2 \tag{3.8}$$

This power is then amplified by the factor β, and sent to the destination D. Considering the path loss and the complex gain of the second channel, the useful power received at the destination is

$$r_d = \frac{P_s M_g \beta}{AG} d_{sc}^{-\alpha} d_{cd}^{-\alpha} |\delta_{sc}|^2 |\delta_{cd}|^2. \tag{3.9}$$

It should be noticed that the spreading gain C_g is applied only at the destination D, where the packet is decoded. The SNR is now given by the ratio between r_d and the total received noise power. This power can be derived as follows. The noise power introduced at the cooperator is N_0. This power is also amplified by the factor β, and then sent to the destination D, where an additional noise power, also equal to N_0, is added. Thus, the total noise power received at the destination is

$$r_n = N_0 \left(1 + \frac{\beta B_g}{A} d_{cd}^{-\alpha} |\delta_{cd}|^2\right) \tag{3.10}$$

Please note that the spreading gain C_g is never applied to the noise. The ratio r_d/r_n gives the term Γ_{cd} in (3.7), namely

$$\Gamma_{cd} = \frac{P_c M_g \beta d_{sc}^{-\alpha} d_{cd}^{-\alpha}}{AGN_0 \left(1 + \frac{\beta B_g}{A} d_{cd}^{-\alpha} |\delta_{cd}|^2\right)} |\delta_{sc}|^2 |\delta_{cd}|^2 \tag{3.11}$$

As intuition suggests, the outage probability depends on the three distances d_{sd}, d_{sc} and d_{cd}, on the source power P_s and on the amplification factor β. This factor, however, can be rewritten as a function of P_c, since $\beta = P_c/(r_c + N_0)$. If we replace β, after some manipulations we get

$$\zeta_{AF} = \mathcal{P}\left[\psi_{sd}|\delta_{sd}|^2 + \frac{\psi_{sc}|\delta_{sc}|^2 \psi_{cd}|\delta_{cd}|^2}{1 + \psi_{sc}|\delta_{sc}|^2 + \psi_{cd}|\delta_{cd}|^2} < \frac{\eta}{C_g}\right] \tag{3.12}$$

with

$$\psi_{sd} = \frac{P_s B_g}{AN_0} d_{sd}^{-\alpha} \quad \psi_{sc} = \frac{P_s M_g}{AN_0} d_{sc}^{-\alpha} \quad \psi_{cd} = \frac{P_c B_g}{AN_0} d_{cd}^{-\alpha} \tag{3.13}$$

The terms $\psi_{\cdot}$ can be seen as the average received powers on the three channels, except for the spreading gain C_g, which affects only the channels towards the destination D. In order to complete the derivation, the three gains $|\delta_{sd}|^2$, $|\delta_{sc}|^2$, $|\delta_{cd}|^2$ can be considered as three

independent exponential random variables of unit mean X, Y and Z. Two of them are combined in one random variable Q, whose form is

$$Q = \frac{YZ}{1+aY+bZ} \tag{3.14}$$

The cdf of Q can be analytically derived as

$$F_Q(t)=1-e^{-(a+b)t}2\sqrt{t+abt^2}K_1\left(2\sqrt{t+abt^2}\right) \tag{3.15}$$

where $K_1(x)$ is the modified Bessel function of the second kind and order 1. The expression of the outage probability can be consequently expressed in terms of the two random variables X and Q:

$$\begin{aligned}\zeta_{AF} &= \mathcal{P}\left[cX+kQ<l\right]=\mathcal{P}\left[Q<\frac{l}{k}-\frac{c}{k}X\right] \\ &= \int_0^{\frac{l}{c}} F_Q\left(\frac{l}{k}-\frac{c}{k}h\right)f_X(h)dh\end{aligned} \tag{3.16}$$

In the last equation, the parameters for our problem are $l=\eta/C_g$, $c=\psi_{sd}$ and $k=\psi_{sc}\psi_{cd}$, whereas $a=\psi_{sc}$ and $b=\psi_{cd}$. Now, since X is exponential, as said above, its pdf is simply $f_X(x)=e^{-x}$. The integral in (3.16) can be further manipulated, giving the following result:

$$\zeta_{AF}=1-e^{-\frac{n}{m}}-e^{-\frac{n}{m}(t+s)}\int_0^{\frac{n}{m}} e^{(t+s-1)h}g(h)K_1(g(h))dh$$

where we put

$$g(h)=2\sqrt{n-mh+st\left(\frac{n}{m}-h\right)^2} \tag{3.17}$$

$$m=\frac{1}{\psi_{sc}}\left(\frac{d_{cd}}{d_{sd}}\right)^{\alpha} \qquad n=\frac{\eta}{C_g}\frac{1}{\psi_{sc}\psi_{cd}} \tag{3.18}$$

$$s=\left(\frac{d_{cd}}{d_{sd}}\right)^{\alpha} \qquad t=\frac{B_g}{M_g}\left(\frac{d_{sc}}{d_{sd}}\right)^{\alpha} \tag{3.19}$$

Decode&Forward

In the D&F scheme the third node C decodes the source's packet and retransmits it, thereby avoiding the noise propagation that affects A&F. Thus, cooperator's transmission is conditioned on the correct decoding of the packet at its first transmission. We assume, as in the A&F case, that the cooperator C retransmits the packet in the time slot following the one during which S performed its first transmission. MRC is used at the destination, resulting in a global SNR given by the sum of the two terms Γ_{sd} and Γ_{cd}: the former is due

to the transmission from the source, the latter to that relayed by C, just as in the A&F case previously described.

The second term, Γ_{cd}, is added to Γ_{sd} only if C succeeds in decoding the transmission from S. The outage probability can be expressed as:

$$\begin{aligned}\zeta_{DF}=&\mathcal{P}\left[\psi_{sd}|\delta_{sd}|^2<\frac{\eta}{C_g}\right]\mathcal{P}\left[\psi_{sc}|\delta_{sc}|^2<\frac{\eta}{C_g}\right]+\\&+\mathcal{P}\left[\psi_{sd}|\delta_{sd}|^2+\psi_{cd}|\delta_{cd}|^2<\frac{\eta}{C_g}\right]\mathcal{P}\left[\psi_{sc}|\delta_{sc}|^2\geq\frac{\eta}{C_g}\right]\end{aligned}\tag{3.20}$$

In this scheme, also the cooperator takes advantage of the spreading gain C_g, since it performs decoding before retransmitting to the destination. The probabilities in (3.20) can be calculated as follows. Given the distributions of the $|\delta|^2$ we have

$$P\left[\psi|\delta|^2<\frac{\eta}{C_g}\right]=1-e^{-\frac{\eta}{C_g}\frac{1}{\psi}}\tag{3.21}$$

The term with the contribution of both S and C is in the form $\mathcal{P}[\psi_{sd}X+\psi_{cd}Y<\eta/C_g]$, with X and Y independent exponential random variables. We can write

$$\mathcal{P}[\psi_{sd}X+\psi_{cd}Y<l]=1-\frac{\psi_{cd}e^{-\frac{l}{\psi_{cd}}}-\psi_{sd}e^{-\frac{l}{\psi_{sd}}}}{\psi_{cd}-\psi_{sd}}\tag{3.22}$$

where we put $l=\eta/C_g$. The total outage probability can hence be rewritten as

$$\zeta_{DF}=1-e^{-\frac{l}{\psi_{sd}}}\left(1-e^{-\frac{l}{\psi_{sc}}}\right)-e^{-\frac{l}{\psi_{sc}}}\frac{\left(\psi_{cd}e^{-\frac{l}{\psi_{cd}}}-\psi_{sd}e^{-\frac{l}{\psi_{sd}}}\right)}{\psi_{cd}-\psi_{sd}}.\tag{3.23}$$

The outage probability depends on all three distances among the nodes, and also on the two powers P_s and P_c used by S and C respectively. A deeper insight in such topics will be provided in the next Section.

3.2.3 Coverage Range of Cooperative Schemes

We define the coverage range as the maximum distance between the source and the destination at which the target outage probability $\bar{\zeta}$ is achieved. In the case of non–cooperative schemes, such as Direct Transmission and DDT, the outage probability depends only on the distance d_{sd} and on the power used. Therefore, once the power is fixed to the maximum allowed value P_m, the coverage range can be found, either analytically or by numerical calculation. In the single transmission case it is simply

$$\Omega_T=\left(-\frac{\ln\left(1-\bar{\zeta}\right)P_m}{\eta G N_0}\right)^{\frac{1}{\alpha}}.\tag{3.24}$$

The value Ω_{DT} for which ζ_{DT}, computed as in Eq. (3.5), is equal to $\bar{\zeta}$ is the coverage range of the double transmission scheme.

In cooperative systems, however, the outage probability depends on three distances. This means that even when the powers of S and C are chosen, and the distance d_{sd} is fixed, different outage probabilities can be reached, based on the position of the cooperator.

In the following, we analyze how the position of node C and its capabilities affect the coverage range.

Amplify & Forward

In the Amplify and Forward scheme the main problem is given by the noise amplification at the cooperator. A cooperator placed near the source makes the scheme similar to DDT, but its transmission suffers a high attenuation on the link towards the destination. On the contrary, if C is located near D, it receives a signal with low SNR from the source, so that C amplifies and retransmits a signal corrupted by noise.

Since the success probability $\phi_{AF} = 1-\zeta_{AF}$ increases with both P_s and P_c, once the nodes are placed the minimum outage probability is reached for $P_s = P_c = P_m$. Moreover, it can be seen that ϕ_{AF} decreases as the distances among the nodes become larger. This means that, once d_{sd} is fixed, the best choice for the cooperator is to place it aligned with S and D, such that $d_{sc} = d_{sd} - d_{cd}$.

We numerically evaluated the success probability ϕ_{AF} for a given distance d_{sd} as a function of d_{sc}, considering the three nodes aligned, and plotted it in Figure 3.1 for different values of the antenna gain M_g. We observe that when M_g is equal to 1, as in the case of another mobile terminal serving as a cooperator, the best position is quite close to the source. This is due to the strong asymmetry between the two channels $S - C$ and $C - D$; the SNR at the cooperator is much lower than the one which would be obtained at a base station placed at the same distance.

As M_g increases, the maximum value of ϕ_{AF} also becomes higher. Moreover, a second local maximum is reached when the cooperator is near the destination, and when $M_g = B_g$ this second maximum has the same value as the first one. Looking at equations (3.17)-(3.19), it can be noticed that the values of m and n are the same for $d_{sc} = d_1, d_{cd} = d_2$ and for $d_{sc} = d_2, d_{cd} = d_1$. This is not true for s and t, but their values, if $M_g = B_g$, appear to be reversed in the two cases. Since however they always appear in (3.17) added or multiplied

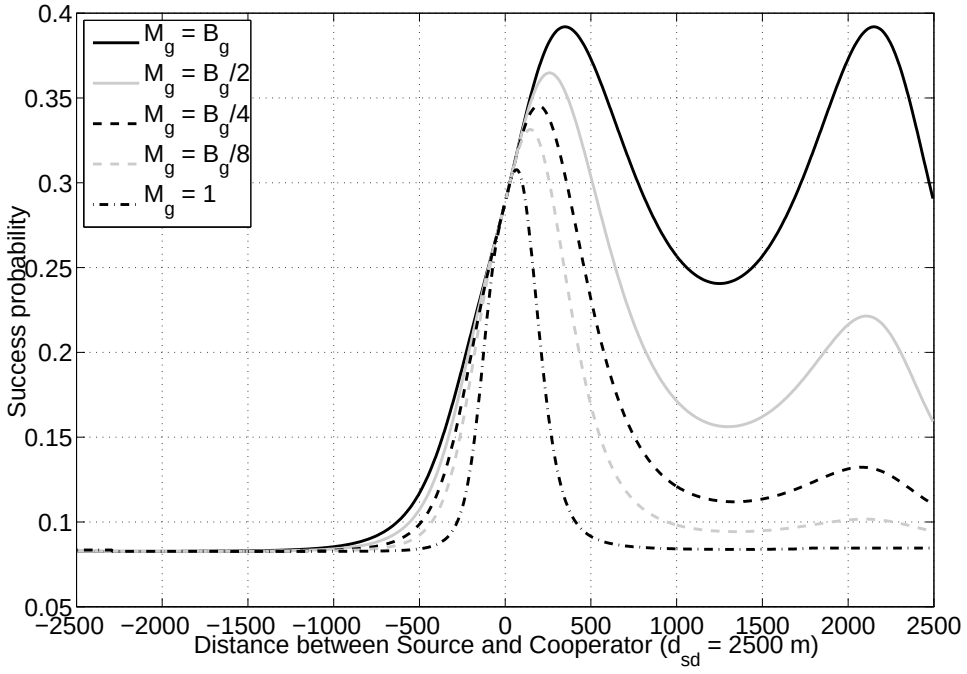


Figure 3.1. *Success probability for A&F as a function of d_{sc}, with $d_{sd} = 2500m$, for different values of M_g.*

together, it follows that ζ_{AF} is a symmetric function of d_{sc}, centered in $d_{sd}/2$, when the antenna gain at the cooperator is equal to that at the destination.

In order to better understand the reason of such symmetry, we analyze the composition of the average powers received at the cooperator and at the destination, once d_{sd} is fixed, and $P_s = P_c = P_m$.

The average useful power received at C is given by

$$\bar{r}_c = \frac{P_m M_g}{A} d_{sc}^{-\alpha} \tag{3.25}$$

whereas the noise power is N_0, and the total received power is $\bar{r}_{tot} = \bar{r}_c + N_0$. Both the useful and the noise power are amplified before retransmission. Hence, the signal and noise components of the average retransmitted power are respectively

$$\lambda_s = P_m \frac{\bar{r}_c}{\bar{r}_{tot}} \qquad \lambda_n = P_m \frac{N_0}{\bar{r}_{tot}} \tag{3.26}$$

The average power received at the destination can be found by multiplying both the components by the path loss of the channel $C - D$. At the destination, however, a third component is added, due to the noise introduced at the BS receiver. In addition, the term of useful power is increased by the factor C_g, which cannot be applied to the other two noise components.

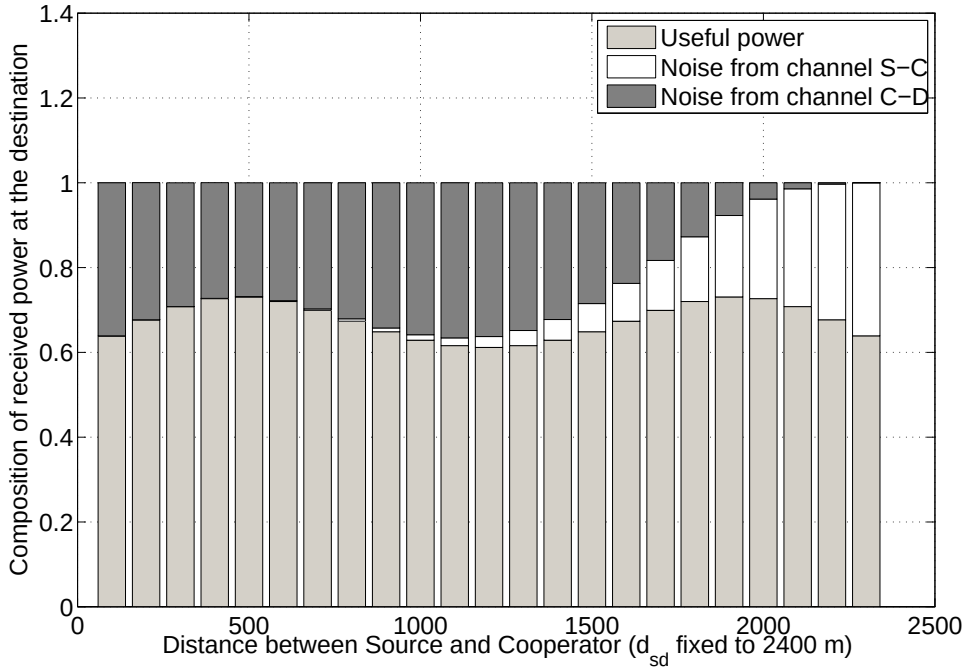


Figure 3.2. *Composition of the power received at destination D as a function of d_{sc}, for $d_{sd} = 2400\ m$.*

The three components are then written as

$$\lambda_s^d = \lambda_s \frac{B_g C_g}{A} d_{cd}^{-\alpha}, \quad \lambda_{n_1}^d = \lambda_n \frac{B_g}{A} d_{cd}^{-\alpha}, \quad \lambda_{n_2}^d = N_0. \tag{3.27}$$

The values of the three components are plotted in Figure 3.2, normalized over their sum, for a given value of d_{sd}. The graph shows that the noise introduced in the second channel has a stronger impact if compared to the noise component due to the first transmission, except when the cooperator is quite close to the destination. Nevertheless, the sum of the two noise components has a symmetric shape, centered in $d_{sd}/2$, and shows two peaks, as predicted by equation (3.17). We recall that the SNR due to the direct transmission must be added, but since this SNR does not depend on the position of the cooperator nor on the value of M_g, the global SNR as a function of d_{sc} keeps the same shape as shown in Figure 3.2. The best position of the cooperator, as a function of d_{sd}, can be found by numerically evaluating (3.17). We can then define the outage probability at distance d_{sd} as the minimum outage probability that can be reached at that distance, by placing a cooperator in the best position. We plot this function in Figure 3.3. The coverage range d_{AF} is hence derived as the distance at which the outage probability is equal to the target value $\bar{\zeta}$. Since the outage probability depends on M_g, the same holds for the coverage range, which increases as the antenna gain at the cooperator becomes higher. Finally, once the target outage probability $\bar{\zeta}$ is fixed, and limiting our analysis to only aligned configurations of S, C and D, we can

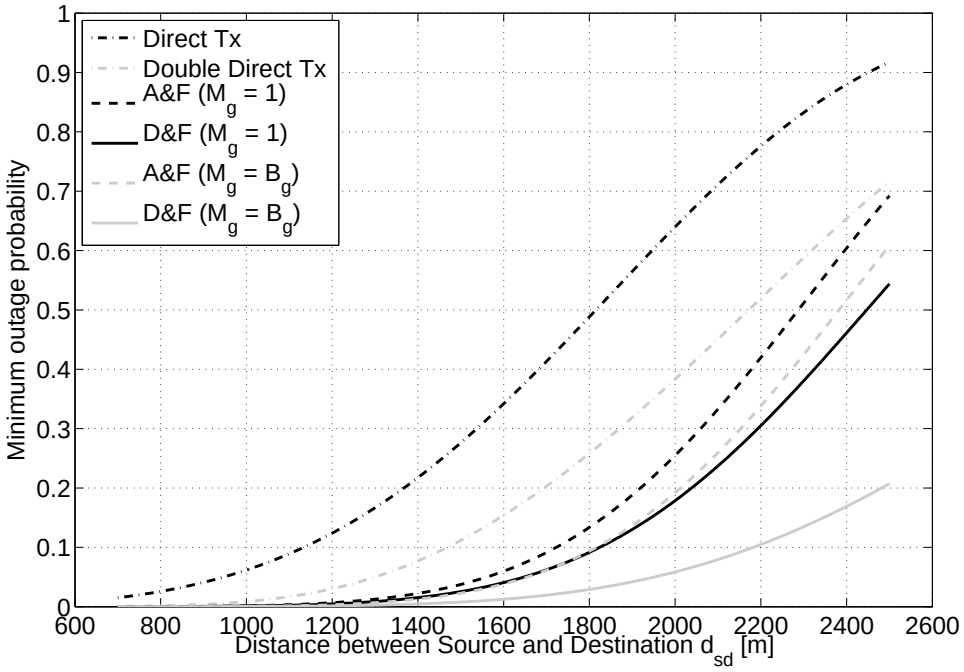


Figure 3.3. *Outage probability for the four analyzed schemes, as a function of d_{sd}.*

identify the region of all the positions of the cooperator which make possible a transmission with outage probability not greater than $\bar{\zeta}$. The region of feasible positions can be found by numerically evaluating the outage probability ζ_{AF} with $P_s = P_c = P_m$.

Decode and Forward

In D&F, the position of the cooperator still heavily influences the reliability. A cooperator too close to the source results in a scheme similar to DDT with increased channel diversity. A cooperator far away from S is not likely to decode the packet. As in the A&F case, the success probability $\phi_{DF}=1-\zeta_{DF}$ increases with both P_s and P_c, and is then maximized when using the maximum transmission power P_m at both nodes. Furthermore, ϕ_{DF} decreases when the distances among the three nodes are larger; this means that also for D&F, for given d_{sd} the best position of the cooperator lies on the line between S and D.

In Figure 3.4 the success probability is shown as a function of d_{sc} for a given value of d_{sd}, and for different values of the antenna gain at the cooperator M_g. If a simple mobile terminal is used as cooperator, meaning $M_g = 1$, the best position is closer to the source, due to the asymmetry between the two channels $S - C$ and $C - D$, as already explained for A&F. When M_g is greater, however, the maximum value of ϕ_{DF} increases, and the best position for a cooperator moves towards the middle point. For $M_g = B_g$ the best choice

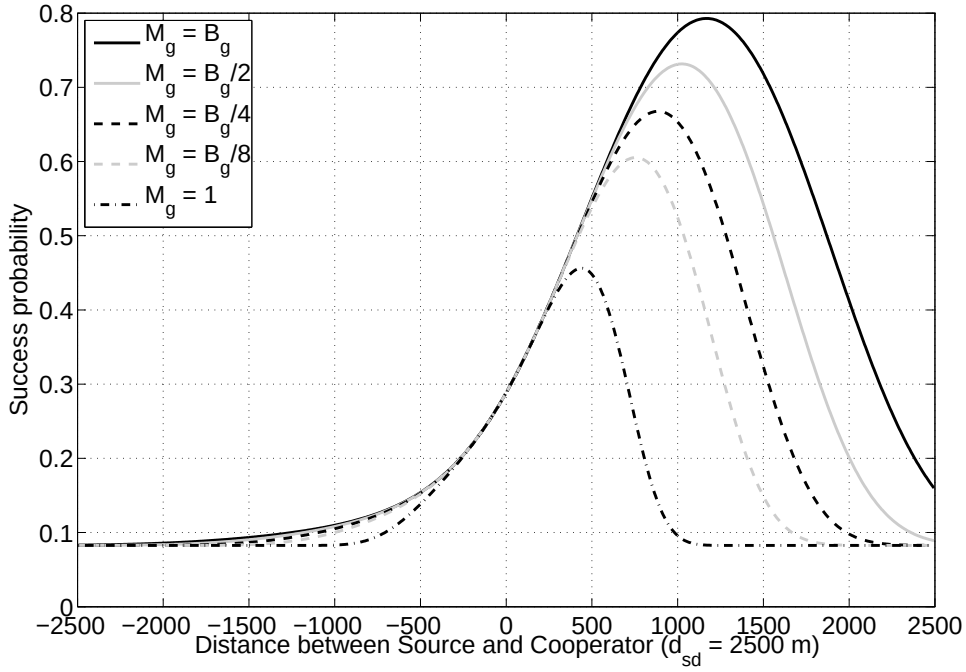


Figure 3.4. *Success probability for D&F as a function of d_{sc}, with $d_{sd} = 2500m$, for different values of M_g.*

is d_{sc} only slightly less $d_{sd}/2$. Comparing Figure 3.1 and Figure 3.4, it appears that for the same value of d_{sd} D&F achieves a much higher success probability. This in turn affects the coverage range, which is itself higher than that achieved by A&F, once defined the outage probability at distance d_{sd} as done for the A&F. The outage probability as a function of d_{sd}, obtained by numerically evaluating Eq. (3.23), is reported in Figure 3.3. The coverage range d_{DF} is again calculated as the distance at which the outage probability is equal to the target value $\bar{\zeta}$.

3.2.4 Design Directions

In the previous section we derived the outage probability and the coverage range of different cooperation techniques as a function of the distances between the nodes involved in the communications and the antenna gain of the cooperator. In this section we discuss how these result can offer design directions for the setting of cooperative system in UMTS cellular networks.

As a first observation, different cooperative techniques have different optimal placement of the cooperator. Moreover, whether C is a MT or a RS its position influences the achieved performance. In the D&F case we achieve the best performance when C is approximately

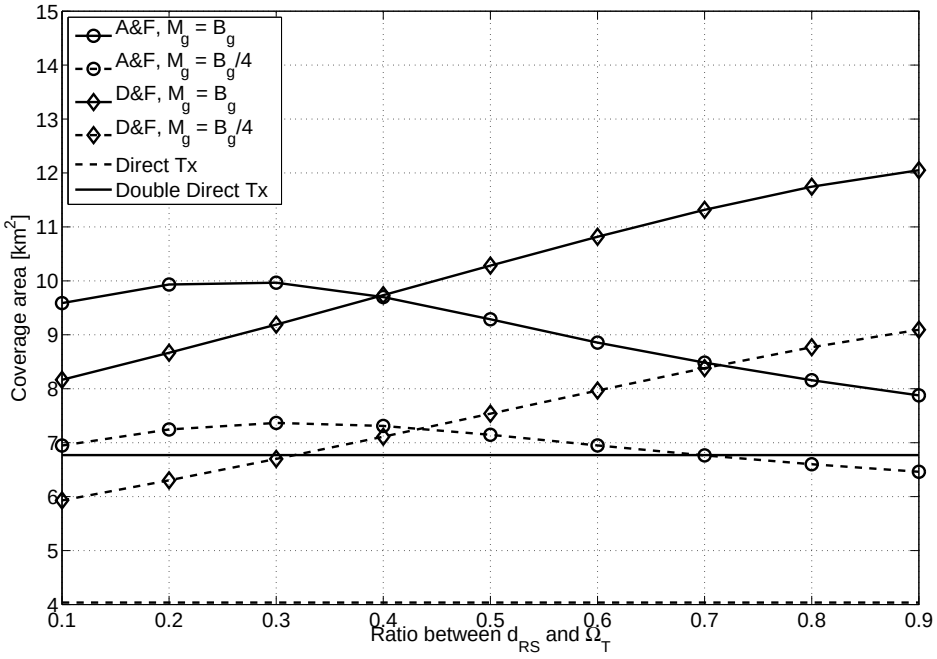


Figure 3.5. *Coverage area for A&F and D&F, as a function of the ratio d_{RS}/Ω_T, where d_{RS} is the distance of the RSs from the BS. Here $N_B = 3$.*

in the middle between S and D and as M_g is decreased it shifts toward S. In the A&F case we have a wide area in the middle between S and D in which we get poor performance and as M_g is decreased we have that C can be only very close to S. If C is a MT, we can choose dynamically the cooperator in order to achieve a performance improvement, but it is clear that when the antenna gain of C is that of a MT we have a small area in which this happens. If C is a RS the feasibility region is larger and a cooperator placed in a given position is likely to be helpful to many MTs. However, in this case the cooperator has a fixed position and therefore we have to carefully place it in order to get the maximum improvement. In Figure 3.5 we plot the coverage area of a cell with a given number N_B of RSs placed at a given range with uniform angle. As could be expected, the area is wider for higher values of N_B, and when using D&F rather than A&F. Moreover, the maximum area for D&F is reached when the RSs are located far from the center. This is right, since it was shown that the best position for the cooperator in D&F is nearly in the middle point between S and D; hence, a RS which can decode the signal from the source at distance $d_{sc} = d_1$ can also reliably retransmit it to a destination at distance almost equal to d_1. With A&F, on the contrary, the maximum area improvement is for the RSs quite close to the BS, according to Figure 3.1.

3.2.5 Discussion

In this part we have investigated the benefits of cooperation in cellular UMTS networks from a simple three-node scenario perspective. Although this preliminary analysis does not consider interference issues, which may strongly affect the effectiveness of cooperation, it still offers some suggestions about how cooperation may be applied. In particular, it was observed that:

- D&F appears to grant higher benefits than A&F, which on the contrary suffers from noise amplification and retransmission; moreover, an A&F cooperator cannot distinguish among the incoming signals, thus potentially retransmitting also interference (intra-cell and inter-cell), and cannot apply more sophisticated receivers (for instance, with interference cancellation mechanisms). On the contrary, A&F requires a much simpler practical implementation, and introduces much lower delays, hence granting the advantages coming from soft handover.

- Higher improvement is achieved when the cooperator can exploit higher antenna gains. This suggests that Mobile Terminals could bring little advantage as cooperators. Relay Stations, whose antennas are comparable to that of the Base Station, appear to be a much better solution, although they must be deployed in fixed positions. Moreover, their number is likely to be much lower than that of the mobiles, thus leading to a scenario where a single Relay Station has to serve a number of transmission sources.

- As was pointed out before, the position of the cooperator affects the performance. If fixed Relay Stations are used, their positions should be properly chosen, given the statistics of the spatial distribution of the Mobiles. When uniform distributions are considered, as in most of the theoretical papers, the positions of the cooperators can be determined to maximize the overall gain. If, on the contrary, clustered distributions are investigated, or obstacles are assumed to increase the path loss of given mobiles, the cooperators should be deployed accordingly, thus granting even higher benefits.

3.3 Algorithms for Cooperative Power Control in CDMA networks

The aim of this Section is to investigate different power control schemes which could be applied to a cooperative uplink transmission in a UMTS like environment. We have already analyzed the outage probability which can be achieved by Amplify and Forward and by

Decode and Forward; moreover, we put our attention to the relationship between these probabilities, the position of the cooperator and its capabilities. Results showed that while for A&F the best position is quite close to the source S or to the destination D, when D&F is used the midpoint between S and D becomes the most advantageous location.

In the analytical derivation of all the previous results, however, the effects of fast fading on the channel gain were averaged, in order to find, for example, the maximum distance at which the average SNR is equal to a certain threshold or the outage probability becomes higher than a fixed value.

In practice, however, even when the distance between the source and the destination is set to a constant value, the instantaneous SNR varies, being affected by both shadowing and fast fading. Power control mechanisms are then used, such that the transmission power is frequently changed and can compensate for the channel variations.

In the following, we first explain the channel model we used to test some power control algorithms; then, we briefly recall the principles of the traditional power control mechanism adopted in cellular systems. Finally, we describe some power control algorithms which could be applied to a cooperative transmission relying on Amplify and Forward.

3.3.1 Rayleigh channel model

In order to verify the effectiveness of the power control algorithms, we adopt a more refined channel model. Since power control is meant to counteract mainly the effects of fast fading, we do not consider shadowing. Hence, the channel gain is given by the contribution of two terms: the path loss, which depends on the distance between the source S and the destination D, and the fast fading term, which instead takes into account the multipath propagation and the mobile speed.

The path loss term is usually derived by the Okumura-Hata model [61] for macro urban cells:

$$L_p(dB) = 69.55 + 26.16 \log_{10} f_c - 13.82 \log_{10} h_b - a(h_m) + (44.9 - 6.55 \log_{10} h_b) \log_{10} R \quad (3.28)$$

where $L_p(dB)$ is the path loss expressed in dB, f_c is the carrier frequency, h_b is the height of the Base Station, h_m is the height of the Mobile Terminal and R is the distance between the BS and the MT, expressed in kilometers. Finally, $a(h_m)$ is a correction factor, which is usually set to 0 for $h_m = 1.5$ m.

Hence, by assuming $h_m = 1.5$ m, $h_b = 30$ m, and $f_c = 1950$ MHz, we can rewrite:

$$L_p(dB) = 137.4 + 35.2 \log_{10} R \tag{3.29}$$

In the previous sections, we used a similar model, where the path loss term is given instead by $d_{sd}^{-\alpha}/A$, where $A = 1000$ and d_{sd} is the distance between transmitter and receiver. However, by selecting the path loss exponent α equal to 3.5, and expressing d_{sd} in kilometers, our propagation model can also be rewritten in dB as

$$L_q(dB) = 135 + 35 \log_{10} d_{sd} \tag{3.30}$$

As regards multipath fading, if a Rayleigh flat fading channel is considered, then the channel gain due to fast fading can be expressed by a single complex gaussian random variable of zero mean and unit variance h_{sd}. If we also add the processing gain C_g and the base station antenna gain B_g, the instantaneous SNR at time t at the base station is given by

$$SNR_{sd}(t) = \frac{P_t B_g C_g}{A N_0} d_{sd}^{-\alpha} \left| h_{sd}(t) \right|^2 \tag{3.31}$$

where N_0 is the noise power. This result has already been derived in the previous sections. However, also the channel time autocorrelation must be properly designed. For a Rayleigh fading channel, it is known that the autocorrelation is linked to the maximum Doppler frequency f_D:

$$R(\tau) = E\left[h_{sd}(t) h_{sd}^*(t-\tau)\right] = J_0(2\pi f_D \tau) \tag{3.32}$$

where $J_0(x)$ is the zero-order Bessel function of the first kind. As a consequence, the power spectral density of the channel process is

$$S_c(f) = \begin{cases} \dfrac{1}{\pi f_D} \dfrac{1}{\sqrt{1 - \left(\frac{f}{f_D}\right)^2}} & |f| < f_D \\ 0 & \text{otherwise} \end{cases} \tag{3.33}$$

Various techniques have been proposed in literature to obtain a random process with the intended power spectral density. We used the following model: a complex gaussian noise is generated and fed to an IIR filter $F(z)$, designed such that its squared magnitude response is approximately equal to the Doppler spectrum reported in (3.33) for a fixed value of the Doppler rate (defined as $\rho = f_D T$, where $1/T$ is the sampling rate). The output of the filter is then properly amplified and interpolated through a polyphase interpolator, in order to get a process with the intended Doppler rate without excessive computational complexity.

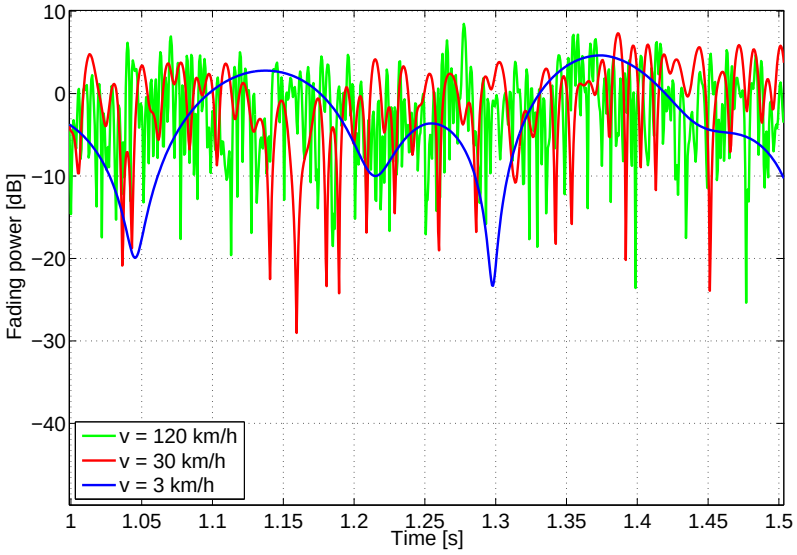


Figure 3.6. *Fading power for a flat Rayleigh fading channel, with different Mobile Terminal speeds.*

The maximum Doppler frequency f_D is linked to the mobile terminal speed v, since $f_D = v/\lambda$, where λ is the wavelength used.

We report in Figure 3.6 some channel fading simulations for different values of the mobile terminal speed, with carrier frequency $f_c = 1980$ MHz.

The simulated channel, as said before, is a flat fading channel, where a single set of unresolvable paths between the source and the destination is considered, resulting in a single channel coefficient. Nevertheless, a more refined model should take into account also the possibility of multiple resolvable paths, which instead results in a frequency selective channel. In this model, the channel impulse response is not given anymore by a single complex value, but by K_t taps, corresponding to different paths. If a rake receiver is implemented at the destination, the power carried by some of these paths can be collected, too. Under this assumption, the total received power is given by the sum of the powers received through the various paths.

In the ITU channel models, the different paths are generally considered independent, each described by a random process with the power spectral density reported in (3.33). However, the average amplitudes of the random processes are different, as listed in Table 3.2.

As can be seen, Pedestrian A channel model is similar to the single path channel model, whereas Vehicular A gives significant multipath diversity. We report in Figure 3.7 some channel realizations with $v = 3$ km/h, comparing the single path model, ITU Pedestrian A

Table 3.2. *ITU channel models, taps in chip resolution*

	Pedestrian A	Vehicular A
Tap 1	0.0 dB	0.0 dB
Tap 2	-12.5 dB	-2.4 dB
Tap 3	-	-6.5 dB
Tap 4	-	-9.4 dB
Tap 5	-	-12.7 dB

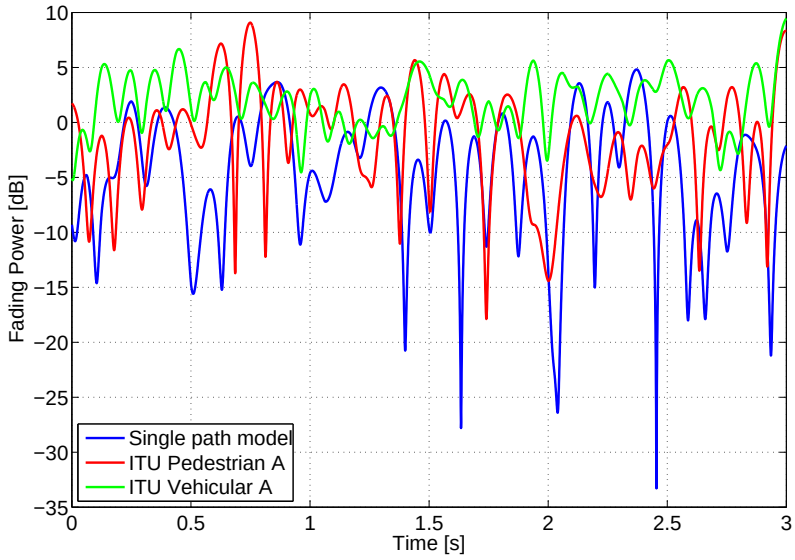


Figure 3.7. *Fading power for various channel models. For Pedestrian A and Vehicular A, rake receiver is assumed at the destination, meaning that the total power is given by the sum of the powers of each tap. The Mobile Terminal speed is 3 km/h.*

and ITU vehicular A.

As can be observed, a rich multipath channel has a higher average fading power if compared to single path or poor multipath channels. In the following, we will use mostly ITU Vehicular A channel model.

3.3.2 Power control in cellular systems

As was pointed out in the previous section, channel variations due to multipath fading can cause the SNR to greatly vary during a single transmission. Since generally a given SNR Γ_t is required at the destination, an effective way to grant reliability is to use power control. Focusing on the uplink channel, during the transmission the Base Station can measure the

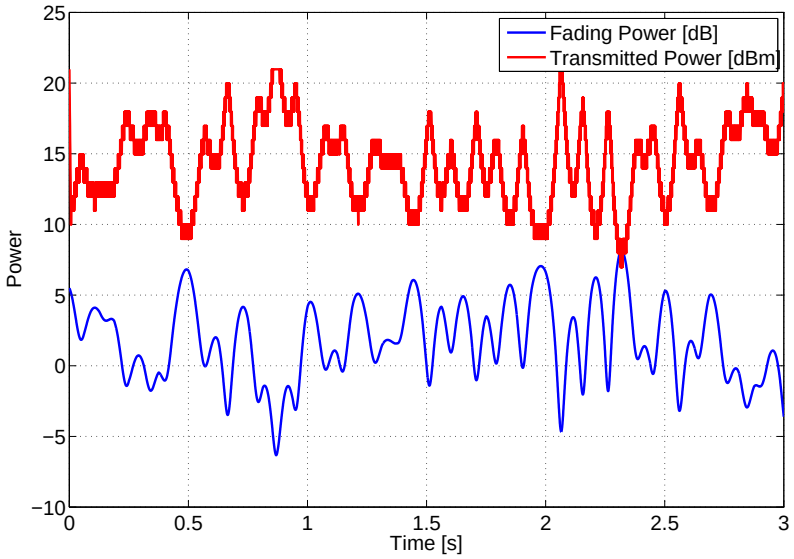


Figure 3.8. *Example of power control for a Vehicular A channel, with* $v = 3$ *km/h, and* $d_{sd} = 1500\ m$.

instantaneous SNR of the signal received by the mobile terminal; hence, it can use the downlink control channel to trasmit the power control commands, asking the mobile terminal to either increase its transmission power, if the SNR is above threshold, or decrease it, in case the SNR is below threshold. The basic algorithm is reported as Algorithm 1.

Algorithm 1 Basic algorithm for non cooperative transmission.

Measure the received Signal to Noise Ratio Γ_{sd};
Compare Γ_{sd} with the required SNR Γ_t;
if $\Gamma_{sd} > \Gamma_t$ **then**
  lower the transmission power at the MT
else
  increase the transmission power at the MT
end if

We consider that power control commands are sent by the Base Station at each time slot, and that power at the MT can be increased or decreased with steps of 1 dB. An example of power control is shown in Figure 3.8, where an ITU Vehicular A channel model is used, with $d_{sd} = 1500$ m.

3.3.3 Cooperative Power Control for A&F

We now investigate how to apply an effective power control scheme to a cooperative transmission based on Amplify and Forward. As a starting point, we make the assumption that each time slot is divided into two mini slots of half length. In the first mini slot, the source S transmits with power P_s to both the destination D and the cooperator C; in the second one the cooperator retransmits with power P_c an amplified version of the received signal to D. This scheme can be seen in fact as a mini TDMA between the source and the cooperator, which has the advantage that the Base Station can send the power control commands with the same delay as in the non cooperative scheme. A more effective scheme is based instead on using two different uplink channels, in order to let the cooperator retransmit a packet while the source is transmitting the following one. However, it requires full duplex capabilities at the cooperator, and introduces an additional delay to the power control commands, which we expect to be negligible, in terms of performance, for rich multipath channels and low to moderate MT speeds.

Average SNR and best position

In Section 3.2, the expression of the instantaneous SNR at the destination Γ_{cd}, if Maximal Ratio Combining is used, was derived as:

$$\Gamma_{cd} = C_g \left(\psi_{sd} |\delta_{sd}|^2 + \frac{\psi_{sc} |\delta_{sc}|^2 \psi_{cd} |\delta_{cd}|^2}{1 + \psi_{sc} |\delta_{sc}|^2 + \psi_{cd} |\delta_{cd}|^2} \right) \tag{3.34}$$

where

$$\psi_{sd} = \frac{P_s B_g}{A N_0} d_{sd}^{-\alpha} \quad \psi_{sc} = \frac{P_s M_g}{A N_0} d_{sc}^{-\alpha} \quad \psi_{cd} = \frac{P_c B_g}{A N_0} d_{cd}^{-\alpha} \tag{3.35}$$

and M_g is the antenna gain at the cooperator. The distances between the three terminals are given by d_{sd}, d_{sc} and d_{cd}. The average SNR can be obtained by taking the expectation on all the three fading gains; since $|\delta_{sd}|^2$, $|\delta_{sc}|^2$ and $|\delta_{cd}|^2$ can be seen as three independent exponential random variables of unit mean, we get:

$$\begin{aligned} \bar{\Gamma}_{cd} &= C_g \psi_{sd} + \frac{C_g \psi_{sc} \psi_{cd}}{(\psi_{sc} - \psi_{cd})^3} \left(\psi_{sc}^2 - \psi_{cd}^2 + (\psi_{sc} - \psi_{cd} + 2\psi_{sc}\psi_{cd})\, e^{1/\psi_{sc}} Ei\left(-\frac{1}{\psi_{sc}}\right) + \right. \\ &+ \left. (\psi_{sc} - \psi_{cd} - 2\psi_{sc}\psi_{cd})\, e^{1/\psi_{cd}} Ei\left(-\frac{1}{\psi_{cd}}\right) \right) \end{aligned} \tag{3.36}$$

for $\psi_{sc} \neq \psi_{cd}$, whereas

$$\bar{\Gamma}_{cd} = C_g \psi_{sd} + \frac{C_g}{6\psi_{sc}^2} \left(2\psi_{sc}^3 - \psi_{sc}^2 + \psi_{sc} + e^{1/\psi_{sc}} Ei\left(-\frac{1}{\psi_{sc}}\right) \right) \tag{3.37}$$

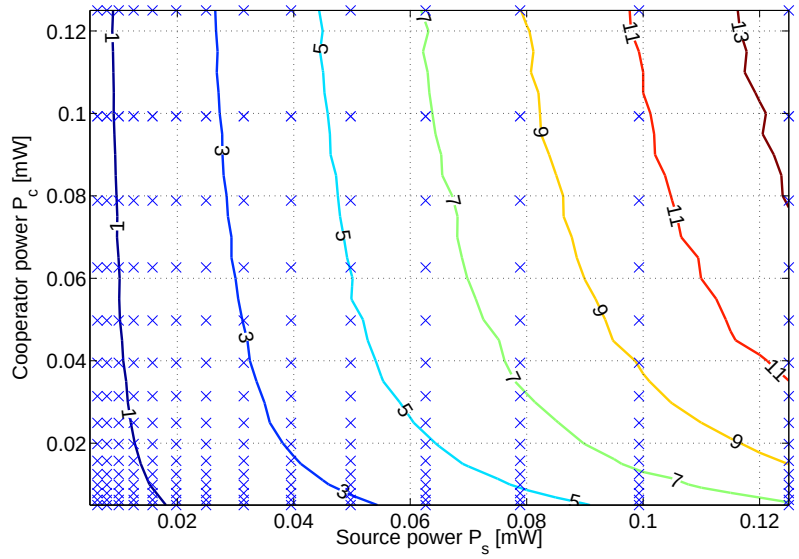


Figure 3.9. *Average SNR as a function of P_s and P_c, for $d_{sd} = 1800\ m$, $d_{sc} = 1400\ m$ and $d_{cd} = 400\ m$, calculated via Monte Carlo simulations. The possible power levels are also reported as 'x'.*

if $\psi_{sc} = \psi_{cd}$. In the above equations we used the exponential integral function $Ei(x)$, defined as:

$$Ei(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} dt \tag{3.38}$$

As can be observed, the average SNR for given positions of the three nodes depends on both P_s and P_c. This means that different choices of the two powers can be made in order to reach the required SNR Γ_t. Moreover, it is not generally true that using the same power at the source and at the destination is the best choice, in terms of power savings. As an example, we report in Figure 3.9 the average SNR as a function of both P_s and P_c for given d_{sd}, d_{sc} and d_{cd}. It is clear that there exists an optimal pair $(\bar{P}_s, \bar{P}_c)$ which minimizes the total power used while still guaranteeing the required average SNR.

This minimization problem can be solved numerically, but it is quite computationally expensive, given the complex expression of the average SNR (3.36). In addition, power control has to counteract the variations of the total SNR, which in these scheme now depends on all the three involved channels. In the following, we design four possible power control schemes, which rely on different assumptions. These algorithms are now being tested, through simulations, over a wide range of terminal positions, in order to verify their effectiveness.

Cooperative power control algorithms

In the non cooperative power control algorithms, power is generally increased or lowered in fixed steps. We assume that both the source and the cooperator use steps of 1 dB, and have the same set of power levels $\mathcal{P}$, with maximum power P_M and minimum power P_m. In a linear scale, the power levels belonging to $\mathcal{P}$ are logarithmically spaced, as depicted in Figure 3.9, where $P_M = 0.125$ W. This means that if P_s and P_c in a given time slot are set to two different values, it is not the same to increase the former or the latter, in terms of total used power. This will be clear after the algorithm description.

Algorithms 2 and 3 are based on the assumption that only the global SNR Γ_{AF} is known at the receiver; in algorithm 4 also the positions of S and C must be known, at least approximately. Finally, in algorithm 5 the knowledge of the SNRs on the three channels is instead required.

Algorithm 2

In this algorithm, we make the assumption that the receiver only knows the total SNR Γ_{AF}. The main idea is to lower the power of both the terminals until Γ_{AF} falls below the required value Γ_t. When this occurs, one of the two powers is raised, whereas the other one is lowered. If the SNR in the next slot is still below threshold, then both the powers are raised, until $\Gamma_{AF} > \Gamma_t$. The choice of which power should be raised when the SNR becomes too low is indicated by a flag, randomly set only the first time. Successively, the flag is changed each time both the powers must be raised to reach Γ_t.

In the $P_s - P_c$ plane, we report an example of the power walk in Figure 3.10, where the average SNR is considered, and $\Gamma_t = 5$. It can be observed that the walk converges to a position not too far from the optimal one. The choice of raising only one power when $\Gamma_{AF} < \Gamma_t$ lets the algorithm move towards the region with the minimum total power. However, when such region is reached, the same mechanism keeps the average SNR slightly less than the required value. Referring to Figure 3.10, the algorithm finally moves from A to D and viceversa. Since Γ_{AF} at B and C is below threshold, this means that the average SNR is also slightly less than Γ_t. The same effect was observed when the algorithm was applied to a fading channel. A possible solution is to set a threshold slightly higher than the one really required; this additional margin, and its dependence on the positions of the nodes, is still to be studied.

We plot in Figure 3.11 an example of cooperative power control algorithm 2, applied to a

Algorithm 2 Power control for cooperative transmission.

$k_c \leftarrow 0$;

$k_u \leftarrow 0$;

Measure the received Signal to Noise Ratio Γ_{AF};

Compare Γ_{AF} with the required SNR Γ_t;

if $\Gamma_{AF} > \Gamma_t$ **then**

  lower both the transmission powers P_s and P_c;

else

  if $k_c = 0$ **then**

    if $k_u = 1$ **then**

      raise P_s and lower P_c;

    else

      raise P_c and lower P_s;

    end if

    $k_c \leftarrow 1$;

  end if

else

  raise both the transmission powers P_s and P_c;

  $k_c \leftarrow 0$;

  $k_u \leftarrow 1 - k_u$

end if

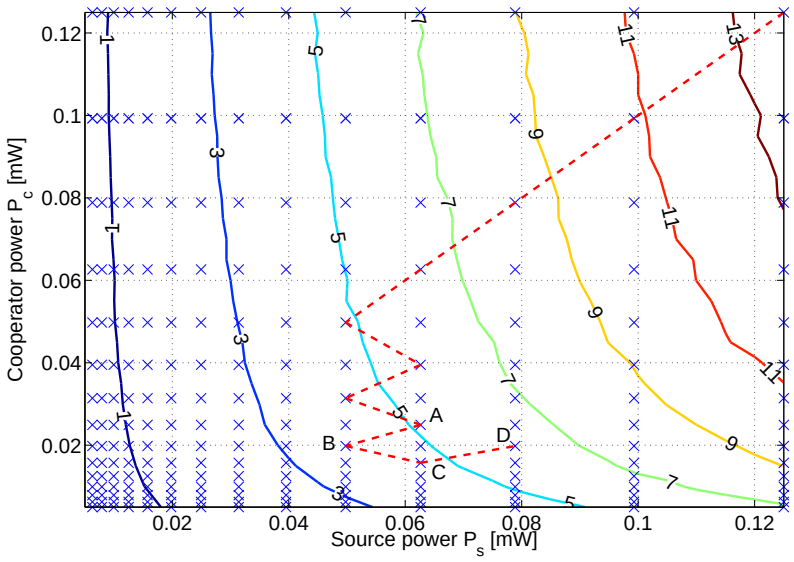


Figure 3.10. *Power walk of algorithm 2, starting from the point* (P_M, P_M).

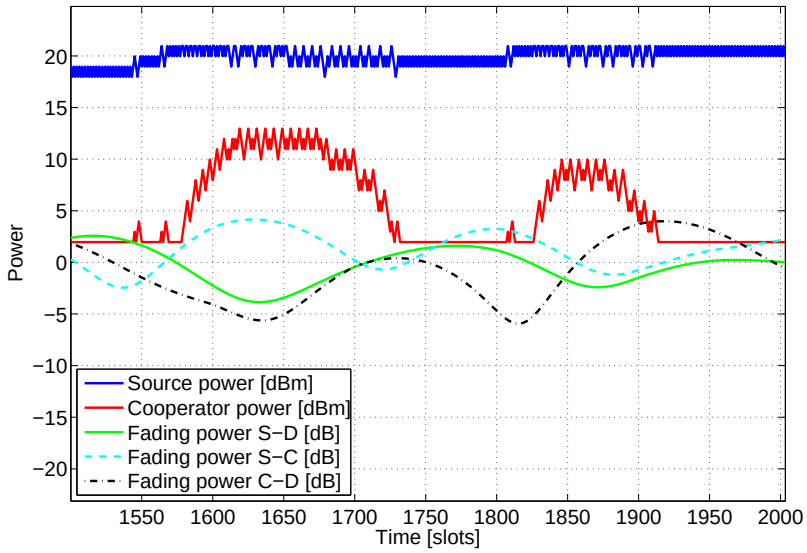


Figure 3.11. *Power control algorithm 2, applied to the fading channel. Here* $d_{sd} = 2000$ *m,* $d_{sc} = 1500$ *m and* $d_{cd} = 500$ *m.*

Vehicular A channel with speed $v = 3$ km/h. As can be seen, the cooperator usually sets its power P_c to the minimum value P_m, and P_c is raised only when the channel $S - D$ is bad. A higher power is spent at the cooperator if also one of the two cooperative links shows harsh conditions. As a result, most of the global power is spent by the source. This effect is due to the logarithmic scale of the admissible power levels of both P_s and P_c. In fact, when one of the two powers is much higher than the other one, to raise the higher power (by 1 dB) is much more effective in increasing the total SNR. As a consequence, the power walk in the $P_s - P_c$ plane shows a drift towards the regions far from the bisector. Likewise, when one of the power is small, it takes several slots to raise it, if only steps of 1 dB are possible.

Algorithm 3

As the previous one, also this algorithm requires only the knowledge of the global SNR Γ_{AF} at the destination. The main difference is the behaviour of the algorithm when the SNR falls below the threshold Γ_t. In algorithm 2, when $\Gamma_{AF} < \Gamma_t$, at first only one power is raised, assuming that this may be sufficient; if the SNR remains below threshold, then both powers are increased. In this algorithm, instead, when the SNR is too low, both P_s and P_c are immediately raised until $\Gamma_{AF} > \Gamma_t$; successively, one of them is lowered, whereas the other one is increased again. If the SNR remains high enough, then both powers are lowered again. The choice of which power must be decreased is determined by the flag k_u, as in algorithm 2: the flag is kept constant until a failure is encountered after having decreased

the selected power and raised the other one. It can be noticed that this algorithm too is likely to be affected by the drift effect described for algorithm 2. If, for example, according to the flag k_u the source power P_s must be raised when P_c is lowered, P_s early becomes much higher than P_c. This in turn means that increasing it will substantially increase also the global SNR, no matter what happens to P_c, and the algorithm will drift towards the region with P_s very close to the maximum value P_M, and P_c to the minimum value P_m, which is not in general the optimal solution. In order to limit this effect, we added an additional control flag, k_m; assuming again that P_s was chosen as the power to be increased, if after P_s is raised and P_c lowered $\Gamma_{AF} > \Gamma_t$ then, as said above, both are decreased; if the SNR then falls below threshold, and the two powers must be raised again, the value of k_u is also changed, and it is set up such that when it is possible again to raise only one power, the lower one will be increased.

In order to better understand the behaviour of the algorithm, an example of the power walk is reported in Figure 3.12. We report the walk with and without the control flag k_m. As in the case of algorithm 2 here again the possibility of increasing one power while lowering the other one is introduced in order to converge near the optimal choice. In this particular case, once point D is reached, the algorithm follows the path $D \rightarrow C \rightarrow A \rightarrow B \rightarrow A \rightarrow E \rightarrow F \rightarrow G \rightarrow A \rightarrow B \rightarrow A \rightarrow C$, and then comes again to D. Please note that in this cycle of twelve steps, the SNR is above threshold eight times. As a matter of fact, simulations show that the average SNR is slightly higher than the target one Γ_t.

An example of the behaviour of the algorithm in a simulated Rayleigh channel (Vehicular A, $v = 3$ km/h) is reported in Figure 3.13. Again, the cooperator uses less power than the source, depending on its position, closer to the destination. Its power is increased when the direct channel is bad. With respect to algorithm 2, however, the power distribution is more balanced.

Algorithm 4

Both algorithms 2 and 3 suffer from the abovementioned drift effect: when one of the two powers becomes much higher than the other one, it is the only one which can compensate fast fading, thus often resulting in an unbalanced power allocation, although the relationship between this effect and the positions of the nodes is still to be investigated. In algorithm 3 the control flag k_m is meant to avoid this situation.

In Algorithm 4, the main idea is that if the position of the nodes is at least approximately known, then the best power allocation can be numerically found. In the $P_s - P_c$ plane this

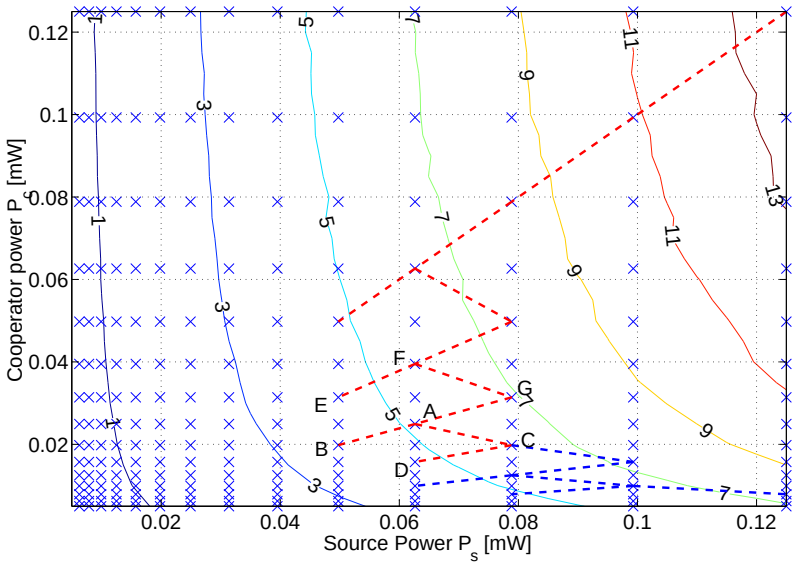


Figure 3.12. *Power walk of algorithm 2, starting from the point* (P_M, P_M)*. The blue path is obtained without the control flag* k_m*.*

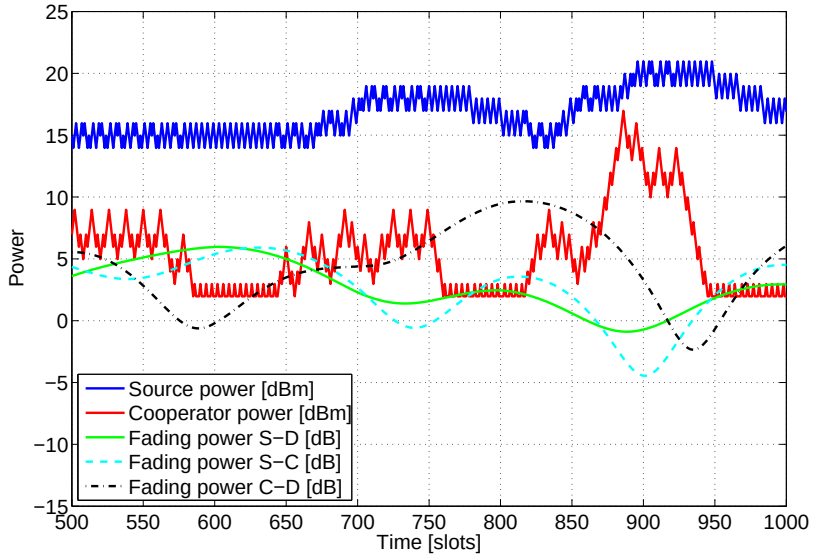


Figure 3.13. *Power control algorithm 3, applied to the fading channel. Here* $d_{sd} = 2000\ m$*,* $d_{sc} = 1500$ *m and* $d_{cd} = 500\ m$*.*

Algorithm 3 Power control for cooperative transmission.

$m \leftarrow 0$, $k_m \leftarrow 0$, $k_u \leftarrow 0$;
Compare the received Signal to Noise Ratio Γ_{AF} with the required SNR Γ_t;
if $\Gamma_{AF} < \Gamma_t$ **then**
  raise both the transmission powers P_s and P_c;
  if $m = 2$ or $m = 4$ **then**
    $k_u \leftarrow 1 - k_u$
  end if
  if $m = 1$ **then**
    if $k_m = 1$ **then**
      $k_u \leftarrow sign(P_s - P_c)/2 + 1$
    else
      $k_m \leftarrow 1$
    end if
  end if
  $m \leftarrow 3$
else
  if $m = 3$ **then**
    if $k_u = 1$ **then**
      raise P_c and lower P_s;
      $m \leftarrow 4$
    else
      raise P_s and lower P_c;
      $m \leftarrow 2$
    end if
  else
    if $m = 1$ **then**
      $k_m \leftarrow 0$
    end if
    lower both the transmission powers P_s and P_c;
    $m \leftarrow 1$;
  end if
end if

Algorithm 4 Power control for cooperative transmission (assuming $\tilde{P}_s$ and $\tilde{P}_c$ known).

Measure the received Signal to Noise Ratio Γ_{AF};
Compare Γ_{AF} with the required SNR Γ_t;
if $\Gamma_{AF} < \Gamma_t$ **then**
  raise both the transmission powers P_s and P_c;
else
  lower both the transmission powers P_s and P_c;
end if

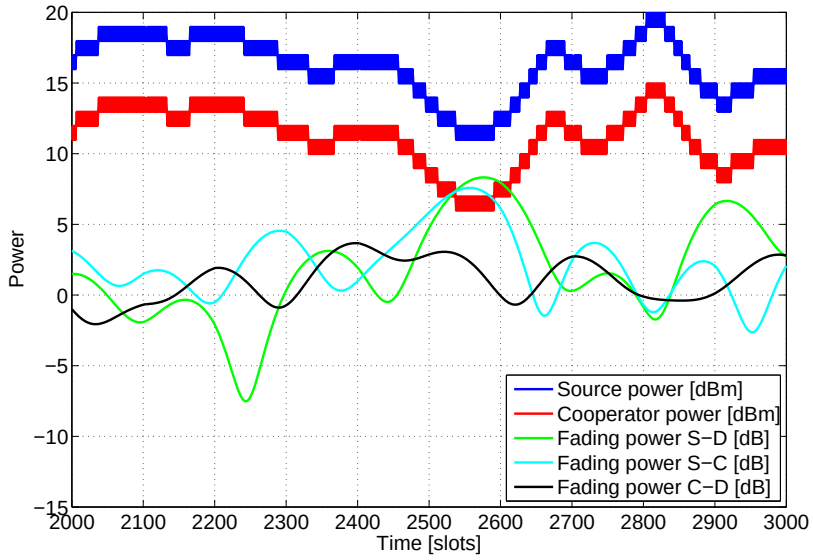


Figure 3.14. *Power control algorithm 4, applied to the fading channel. Here* $d_{sd} = 2000$ *m,* $d_{sc} = 1500$ *m and* $d_{cd} = 500$ *m.*

corresponds to finding the point $(\bar{P}_s, \bar{P}_c)$ which satisfies the condition $\bar{\Gamma}_{AF} > \Gamma_t$ while minimizing the sum $P_s + P_c$. Given the discreteness of the admissible power levels, an approximate solution belonging to the grid shown in Figure 3.9 should be found. If we assume to know this approximate solution $(\tilde{P}_s, \tilde{P}_c)$, it is possible to initially set $P_s = \tilde{P}_s$, $P_c = \tilde{P}_c$; in the following time slots, the algorithm becomes very similar to the non cooperative one: when $\Gamma_{AF} < \Gamma_t$, both powers are increased, and if instead $\Gamma_{AF} > \Gamma_t$ they are both lowered. The result is that the time evolution of the two powers is the same, but differently scaled. This algorithm is much simpler, and can achieve the target average SNR while granting higher power savings than algorithm 2 and 3; however, it requires the knowledge of the positions of the nodes, which should be updated, although with lower frequency.

We report in Figure 3.14 an example of the power allocation evolution in a Vehicular

A Rayleigh fading channel, with $v = 3$ km/h. In the example $d_{sd} = 2000$ m, $d_{sc} = 1500$ m and $d_{cd} = 500$ m; for these positions, the optimal power allocation that minimizes the total power when average SNR is considered was numerically calculated, and is given by $\bar{P}_s = 72.5574$ mW and $\bar{P}_c = 20.4643$ mW.

Algorithm 5

Since the global average SNR, as expressed in equation (3.34), depends on the SNRs of all the three channels, a better power control is likely to be achieved if they are known at the destination. Therefore, in this algorithm we make the assumption that the destination has knowledge about all the three channels. As a matter of fact, the destination can easily get information about channel S-D from the direct transmission; it also knows the second term of the sum on the right in equation (3.34), which we now call w, defined as:

$$w(t) = \frac{\psi_{sc} \left|\delta_{sc}(t)\right|^2 \psi_{cd} \left|\delta_{cd}(t)\right|^2}{1 + \psi_{sc} \left|\delta_{sc}(t)\right|^2 + \psi_{cd} \left|\delta_{cd}(t)\right|^2} \tag{3.39}$$

since it can get it from the relayed transmission. Hence, if it knows also $\psi_{cd} \left|\delta_{cd}\right|^2$, which may be inferred from the control channel on the link C-D, also the SNR between the source and the cooperator can be calculated. In the following, we assume that the Base Station knows all the three SNRs. Moreover, we consider that the BS keeps track of the powers used by both the source and the cooperator. Under these assumptions, if the channel can be considered slowly varying between two successive time slots, i.e., $\left|\delta_{sc}(t+1)\right|^2 \simeq \left|\delta_{sc}(t)\right|^2$ and $\left|\delta_{cd}(t+1)\right|^2 \simeq \left|\delta_{cd}(t)\right|^2$ the base station is able to predict the effect given by a power increase at either the source or the cooperator. Mathematically, we have that if the cooperator power P_c is increased, the estimated SNR in the following time slot is given by

$$\Gamma_{AF}(t+1) = C_g \left(\psi_{sd} \left|\delta_{sd}(t)\right|^2 + k_{cd} w(t)\right) \tag{3.40}$$

with

$$k_{cd} = \kappa \frac{1 + \psi_{sc} \left|\delta_{sc}(t)\right|^2 + \psi_{cd} \left|\delta_{cd}(t)\right|^2}{1 + \psi_{sc} \left|\delta_{sc}(t)\right|^2 + \kappa\psi_{cd} \left|\delta_{cd}(t)\right|^2} \tag{3.41}$$

and $\kappa = 10^{1/10}$, corresponding to 1 dB. On the contrary, if the source power is raised, the global SNR of the next time slot can be predicted as:

$$\Gamma_{AF}(t+1) = C_g \left(\kappa\psi_{sd} \left|\delta_{sd}(t)\right|^2 + k_{sc} w(t)\right) \tag{3.42}$$

where now

$$k_{sc} = \kappa \frac{1 + \psi_{sc} \left|\delta_{sc}(t)\right|^2 + \psi_{cd} \left|\delta_{cd}(t)\right|^2}{1 + \kappa\psi_{sc} \left|\delta_{sc}(t)\right|^2 + \psi_{cd} \left|\delta_{cd}(t)\right|^2} \tag{3.43}$$

Algorithm 5 Power control for cooperative transmission (assuming $\tilde{P}_s$ and $\tilde{P}_c$ known).

Compare the received Signal to Noise Ratio Γ_{AF} with the required SNR Γ_t;
if $\Gamma_{AF} < \Gamma_t$ **then**
  calculate predicted SNR Γ_1 after raising P_s and the predicted SNR Γ_2 after raising P_c;
  if $\Gamma_1 < \Gamma_t$ and $\Gamma_2 < \Gamma_t$ **then**
    raise both P_s and P_c
  else
    calculate $\Gamma_m = \min(\Gamma_1, \Gamma_2)$
    if $\Gamma_1 = \Gamma_m$ **then**
      if $\Gamma_1 < \Gamma_t$ **then**
        if $d < d_p$ **then**
          raise P_c; $d \leftarrow d + 1$;
        else
          raise both the powers P_s and P_c
        end if
      else
        raise P_s; $d \leftarrow d - 1$;
      end if
    else
      if $\Gamma_2 < \Gamma_t$ **then**
        if $d < d_p$ **then**
          raise P_s; $d \leftarrow d + 1$;
        else
          raise both the powers P_s and P_c
        end if
      else
        raise P_c; $d \leftarrow d - 1$;
      end if
    end if
  end if
else
  lower both the transmission powers P_s and P_c
end if

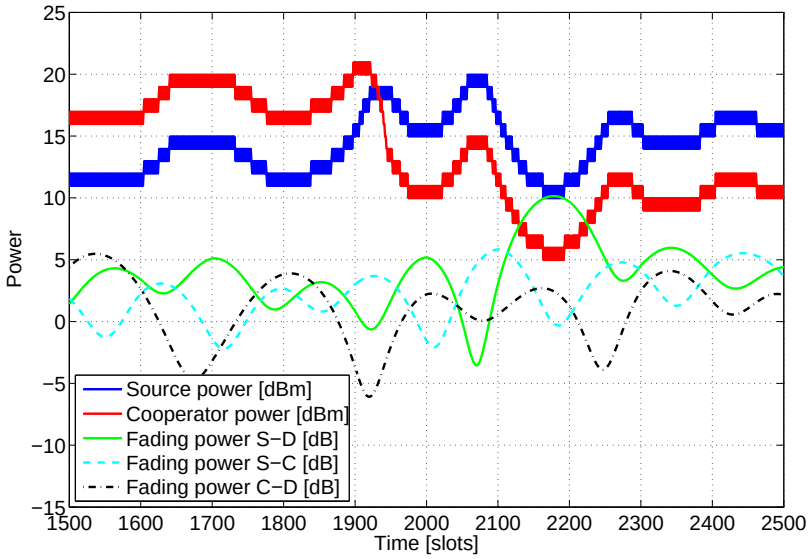


Figure 3.15. *Power control algorithm 5, applied to the fading channel. Here $d_{sd} = 2000$ m, $d_{sc} = 1500$ m and $d_{cd} = 500$ m, whereas the maximum allowed difference between the two powers is $d_p = 5$ dB.*

The algorithm then works as follows: if the SNR is above threshold, then both the powers are lowered, as in algorithm 2; on the contrary, if $\Gamma_{AF} < \Gamma_t$, the Base Station calculates the SNR that would be obtained by increasing P_s, P_c or both, and chooses the solution which grants the minimum SNR above threshold; if no acceptable solution is available, both the powers are raised. Once again, in order to avoid the drift effect observed in the previous explained algorithms, a maximum difference d_p in dB is fixed between P_s and P_c. If the optimal power allocation required the absolute value of the difference between the two powers (expressed in dB) to be greater than d_p, than both powers would be increased instead. Please note that in this algorithm it is not required that each power must be changed at each time slot, since it is sufficient that at least one is lowered or increased. In Figure 3.15 we show an example of the application of algorithm 5 to a Vehicular A Rayleigh fading channel, with speed $v = 3$ km/h. The two powers show a very similar evolution, meaning that their difference is often equal to the maximum allowed value d_p. Nonetheless, sometimes their roles are reversed. The effect of the positions of the three nodes on the frequency of such inversions is now under investigation.

In order to make a preliminary comparison, we report in Table 3.3 the average SNR and the average powers used at the source and at the destination when $d_{sd} = 2000$ m, $d_{sc} = 1500$ m and $d_{cd} = 500$ m; the behaviours of the four cooperative power control algorithms in this particular situation were also shown in Figures 3.11, 3.13, 3.14 and 3.15. Of course, the

Table 3.3. *Performance of the four cooperative power control algorithms.*

Algorithm	Average SNR	Average P_s	Average P_c	Average total power
Algorithm 2	4.91 dB	0.06702 mW	0.00226 mW	0.06928 mW
Algorithm 3	5.41 dB	0.05238 mW	0.01178 mW	0.06416 mW
Algorithm 4	5.07 dB	0.04183 mW	0.01325 mW	0.05508 mW
Algorithm 5	5.04 dB	0.04479 mW	0.02463 mW	0.06942 mW

performance of the algorithms may vary significantly for different positions of the nodes; we are now investigating this aspect through extensive simulations over a wide range of possible positions.

Discussion

In this part of the work we focused on practical power control algorithms which may be applied in cooperative uplink transmissions. The main challenge to be addressed is how to counteract the channel variations, due to fast fading processes, in three different channels, by acting on the transmission powers of source and cooperator. We assumed that A&F is used by the cooperator, in order to reduce processing delays, and that only steps of 1 dB are allowed for both the transmission powers. We proposed four different algorithms, which in general require different information regarding either the SNRs on the three channels or the positions of source and cooperator. Most of these algorithm show a particular behavior, such that after a transient phase the entire power control process relies on the power of only one of the two transmitting nodes. This behavior is not optimal, and can be avoided by means of minor modifications of the algorithms. All the algorithm, after a preliminary study, show good performance, although they are still to be tested in more complex scenarios. On the other side, if Maximum Ratio Combining is not applied, simplifications are likely to be applicable, since only two channels must be controlled.

3.4 Single Cell Analysis

In this part of the work, we address the design and the main issues arising when implementing cooperation in CDMA cellular networks (for concreteness, we refer to the UMTS standard [62], but the general ideas and results are equally applicable to CDMA2000 networks) and derive reasonable design directions. We want to underscore that a feasible ap-

plication of cooperation in this kind of network must comply with its structure, features and requirements. Moreover, even if it is very difficult to include a cooperative behavior without any modification to the standard, a high level of integration and compatibility is desirable, most of all from the point of view of mobile terminal operations.

An important distinction of this work with respect to prior work is the inclusion of inter-user interference in our discussion. In fact, in a CDMA cellular scenario we can not assume the protection from interference granted by the carrier sensing and collision avoidance protocols usually implemented in ad hoc networks.[1] Conversely, in a CDMA scenario, where multiple users simultaneously transmit at the same time and frequency, interference becomes a key issue. Thus, after having focused on the performance gain provided by cooperation to a single link, we now investigate the overall capacity of the whole cell in terms of allocation capabilities.

We focus on a scenario where fixed Relay Stations (RSs), placed at a given distance from the Base Station (BS) and equal angular separation, are in charge of retransmitting the signals of a group of Mobile Terminals (MTs) in order to increase their probability of successfully delivering packets to the BS. As for the cooperation technique, we consider A&F with a Time Division Multiple Access structure where first the various MTs of the group simultaneously transmit their signals to the RS in the same way as they usually transmit them to the BS and then, in the following slot, the RS transmits an amplified version of the overall received signal to the BS. We will discuss in Section 3.4.1 the motivations behind these choices, based on practical considerations.

We present an analytical framework for the coverage range of the various RSs and the BS as a function of the number of MTs already allocated to the RSs and of those directly transmitting to the BS. Based on this result, we derive the probability that a MT randomly placed within the cell can be allocated. Moreover, we derive the probability that in order to allocate a randomly placed MT we have to deallocate at least one other user. These results give us an idea of the allocation capabilities of the cell in the non-cooperative and cooperative cases and makes it possible to compare.

[1] Carrier sensing and collision avoidance protocols prevent neighbors from transmitting while a given communication is active. This somewhat justifies the assumption that interference is negligible and that nodes in a certain portion of the area are available for cooperation.

3.4.1 Network Design

As introduced before, a cellular CDMA scenario poses some limitations, dictated by practical considerations on operators' constraints, Quality of Service (QoS) requirements and compliance with the standard.

A first important observation, leading to a fundamental design direction, is that operators are not likely to leave part of the control of the network to MTs acting as cooperating nodes, as required by an effective implementation of a cooperative scheme. For instance, cooperating nodes should control the power transmitted by the MTs, in order to have a perceived SINR sufficient to retransmit to the BS a signal with an adequate component of useful power. This results in a MT (the cooperating node) transmitting on a control channel, that is probably not desirable for the operators. Furthermore, users are likely not interested in consuming power and channel resource to help other users' connectivity and improve fairness in the cell under heavy load. Due to the previous observations, we consider fixed Relay Stations (RSs) placed in the cell and controlled by the operator as cooperators. We want to underscore that RSs usually have an antenna gain and maximum transmission power greater than that of a simple MT. Thus, a cooperating RS grants considerably better performance than a cooperating MT, as was clearly shown in the previous part of the analysis. As a drawback, the number of RSs deployed in a cell is probably smaller than that of the MTs, and therefore a MT has typically a smaller set of cooperating nodes to choose from. For the same reason, we have to design a mechanism allowing each RS to serve multiple MTs.

Another common assumption is that cooperating nodes can simultaneously receive and transmit. In our scenario, we can not assume that cooperating nodes simultaneously receive the signal from a given source and transmit over the channel previously received packets or packets from their own queue. In fact, both the original and the cooperative transmission occur on an uplink channel, and the spacing between the carriers of the various uplink channels is not sufficient to provide the separation needed to avoid unsustainable self-interference. Therefore, we have to design a Time Division Multiple Access scheme for cooperative transmissions, that evidently diminishes the bandwidth of sources that rely on cooperation to connect to the base station. Preliminary results, not reported here for lack of space, show that the improved performance in terms of coverage range and dropping probability grants an overall throughput comparable to that of the non cooperative case, though with enhanced fairness.

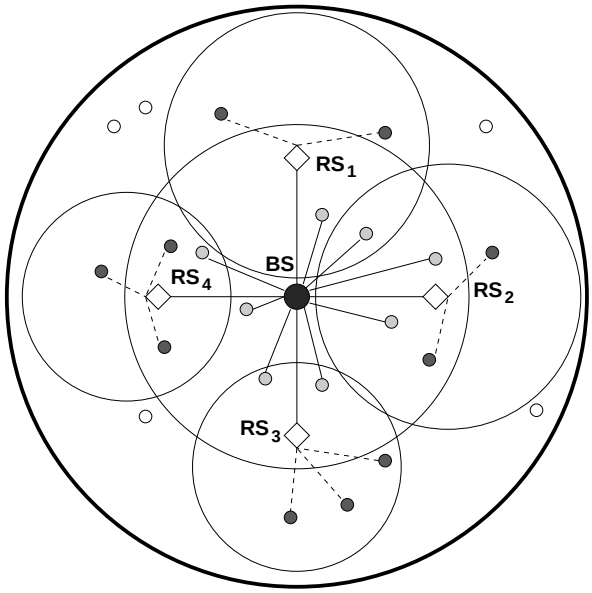


Figure 3.16. *Small white, dark grey and light grey circles represent MTs not covered by the cell, using cooperation and transmitting directly to the BS, respectively. The outer thick circle is the wanted coverage area, the other big circles are the coverage area of the BS and the various RSs.*

We also have to take into account the limitations imposed by the control mechanisms, such as power and rate control. It is important to observe that cooperating nodes can retransmit a signal only after its reception and processing. Thus, cooperation introduces a delay that is a function of the complexity of the signal processing required by the scheme and depends also on the TDMA structure of the communications. A delayed transmission from the cooperating nodes not only affects the latency of the packets, but also potentially diminishes the effectiveness of power control on the link. In fact, unless multiple control channels are allocated to the same communication, messages for the control of the MT and RS transmit power have to share the same channel, and thus the time between the received power estimation at the BS and the reception of the control packet with the associated power adjustment is increased. Moreover, additional latency has a detrimental effect on the quality of the voice traffic. It is possible, if different modulation, rate or codes are used, that there exists a direct link between the MT and the BS on the control channel, which would allow a faster control signalling. Nonetheless, such a scheme is feasible only if there is a radio connection between the MT and the BS. Henceforth, although in the previous Section D&F showed better performance, we focus instead on A&F scheme, that incurs negligible processing delay.

To summarize, we consider a scenario where a certain number of fixed RSs, placed at a

given distance from the BS and at uniform angles, employ A&F to help connectivity of MTs with bad channel conditions. All the MTs assigned to a given RS transmit in the same slot, on the same channel the BS assigned them. In the following slot, RSs transmit an amplified version of the overall received signal on the same uplink channels used by the MTs[2]. As for the MTs assigned to a RS, we assume that only the signal received by the RS is considered for packet decoding at the BS, while MT's transmission in the previous slot is taken as interference. Other choices can be made. For instance, Maximal Ratio Combining (MRC), where the BS combines the signals received in the two transmissions, potentially grants considerable SINR gain and robustness to fading. It was used in the previous Section, since it can grant additional benefit for in-range cooperation. However, in our case, where we deploy cooperation in order to extend the coverage range, the MTs assigned to the RSs are likely to have a channel to the BS with a low average received power. This holds especially when considering an allocation that looks at the throughput of the cell, and thus tries to assign a small number of MTs to the RSs and most likely the most distant ones (due to the bandwidth reduction required by cooperative transmission). In addition, as shown before, the use of MRC would require an increased complexity in the power control mechanism, which should operate on three different channels. Hence, MRC does not provide an improvement that justifies the complexity increase due to the storage and the processing of a significant number of signals at the BS.

Once defined the system, our task is to design the algorithm allocating resource (uplink channels, codes and time slots, but also RSs' cooperation) with the aim of maximizing the number of served MTs per cell.[3]

Adaptive modulation or rate adaptation may probably enhance the performance of our cooperative scheme: the relayed channels consist of two shorter links, where a higher modulation or rate could be reliably adopted, which could turn into a throughput increase. This analysis, however, is beyond the scope of this work.

We define the coverage range as the maximum distance from the BS (or a RS) that allows a MT to achieve the desired average SINR at the BS without violating a constraint on the

[2]RSs use the same channels/codes as those assigned to their MTs in order not to occupy further resource, potentially decreasing the number of channels available to MTs not using cooperation.

[3]As the bandwidth of MTs using cooperation is cut by half, the overall throughput of the cell potentially decreases. We will discuss this point later. However, we want to observe that successfully allocating MTs incurring high outage probability without cooperation improves the fairness in the cell. Therefore, cooperation in this scenario provides a way to trade off throughput for fairness.

maximum transmit power. It is clear that interference plays a key role in the determination of the coverage range. In fact, MTs transmitting on the same channel interfere with each other, and thus, as we will show in the following, the per-MT average transmitted power needed to have a certain SINR at the receiver increases.

We explain the problem with the help of Fig. 3.16, where we represented a configuration with four RSs. Assume we have only one channel. When we allocate a user to the BS or to one of the RSs, the radius of the coverage area of the others decreases, due to the additional interference. It is worth observing that RSs are receiving the signals of the MTs only every other slot and that the power control of the various MTs is performed based on the coefficient of their uplink channels. Thus, if a MT assigned to the BS is close to a RS, it potentially generates a very high interference to the MTs assigned to that RS. As a consequence, the coverage area of the RS shrinks, and it becomes essentially useless. Therefore, it appears as a good choice to have the MTs divided at least into two groups using different channels, one composed of the MTs assigned to the BS and the other of the MTs assigned to the various RSs. We can further divide MTs into other groups, each using a different channel. Even if it would be interesting to see how to form the various groups, we focus on a scenario with two channels (in the non-cooperative case both channels are assigned to MTs directly connected to the BS), investigating in depth the issues arising in this case and leaving further refinements as future work.

The choice of the channel and BS or RS to be assigned to a MT depends on the interaction between the coverage ranges (the big circles in the figure). As the power received by each assigned MTs is the same, the radius of the circles is a function only of the number of users allocated to the various RSs and the BS. It is important to observe that if another RS allocates a further MT, the coverage area of the other decreases, and potentially the MTs of the other RSs need too high a power to achieve the desired SINR.

3.4.2 Transmitted Power and Coverage Range

In this Section, we derive analytical expressions of the needed trasmission powers of both cooperating and non cooperating nodes, given their positions and the number of MTs in the cell. We consider a Rayleigh block fading channel, with pathloss described according to Hata model, as in the previous sections. Therefore, in a given time slot the power P_b

received at the BS from a MT at distance d_{sd} and transmitting with power P_t is

$$P_b = P_t \frac{B_g C_g}{A} \gamma d_{sd}^{-\alpha} |\delta_{sd}|^2 \tag{3.44}$$

where again B_g is the antenna gain of the BS, C_g is the processing gain due to the spreading code, A is a pathloss parameter depending on the adopted channel model, α is the pathloss exponent and δ_{sd} is the fading gain, modeled as a complex Gaussian random variable with zero mean and unit variance. Moreover, γ is an additional factor which takes into account shadowing. In a cellular network, however, power control is used to keep the received SNR around a target value, thus counteracting multipath fading. Assuming that the channel variations are slow enough, we can consider the target value as the effectively perceived SNR. If N_0 is the noise power, the target SNR $\bar{S}$ is given by

$$\bar{S} = P_t \frac{B_g C_g}{A N_0} \gamma d_{sd}^{-\alpha} \tag{3.45}$$

Here P_t represents the value that must be reached by the product between the effective transmitted power and the instantaneous fading gain, in order to have the SNR equal to $\bar{S}$.

Equation (3.45) can be used to determine the SNR of a transmission from a MT and the BS, once the distance and the transmission power are given. In case of cooperative transmission, where the MT first transmits towards a RS, which then amplifies and retransmits the same packet to the BS, the global SNR perceived at the BS can be easily calculated. If d_{sc} is the distance between the MT and the RS, and d_R the distance between the RS and the BS, the useful power received at the RS is simply $P_c = P_t G_r d_{sc}^{-\alpha}$. Here G_r is defined as $M_g \gamma / A$ (note that now the antenna gain at the RS M_g must be used), and the processing gain C_g is not applied since the signal is not decoded at the RS, as the A&F scheme is used.[4] The noise power instead is equal to N_0. Both these components are amplified by a factor β and sent over the channel between the RS and the BS, where eventually another noise component of power N_0 is added before decoding. Henceforth, the total average SNR at the destination can be written as

$$SNR_{AF} = C_g \frac{P_t G_r G_b M_g \beta d_{sc}^{-\alpha} d_R^{-\alpha}}{N_0 \left(1 + \beta M_g G_b d_R^{-\alpha}\right)} \tag{3.46}$$

In the equation above $G_b = B_g \gamma / A$. Please note that the antenna gain of the RS M_g is applied also when the packet is retransmitted towards the BS.

Equations (3.45) and (3.46) are useful to determine the power of a single MT and the amplification factor of a RS required to achieve a target SNR at the BS. Nonetheless, they do

[4] Different values of γ for the different channels could be used as well, without any significant change in our analysis.

$$SINR_i^j = \frac{C_g P_t^{(i)} \beta_j \sigma d_i^{-\alpha} d_R^{-\alpha}}{\sum\limits_{\substack{l \in \mathcal{C}_j \\ l \neq i}} P_t^{(l)} \beta_j \sigma d_l^{-\alpha} d_R^{-\alpha} + \sum\limits_{\substack{h=1 \\ h \neq j}}^{Q} \sum\limits_{m \in \mathcal{C}_h} P_t^{(m)} \beta_h \sigma d_m^{-\alpha} d_R^{-\alpha} + \sum\limits_{q \in \mathcal{D}} P_t^{(q)} G_b d_q^{-\alpha} + N_0 \left(1 + \beta G_b d_R^{-\alpha}\right)} \tag{3.47}$$

not take into account the presence of multiple interfering MTs, all transmitting towards the same destination.

Consider now the more general scenario previously introduced, and depicted in Figure 3.16, where Q RSs are deployed at distance d_R from the BS with uniform angle, and where a SINR target equal to Γ is required at the BS for all the transmissions.

Assume that there are both cooperating and non cooperating MTs (i.e.: both MTs transmitting to a RS and MTs transmitting directly to the BS), and that the generic i–th MT is placed at distance d_i from the destination it is assigned to, either the BS or one of the RSs. We now derive the expression of the transmission power $P_t^{(i)}$ of the i–th MT. We define $\mathcal{D}$ as the set of the MTs which transmit directly to the BS, and $\mathcal{C}_j$ as the set of the MTs which are assigned to the j–th RS. Moreover, we call $P_r^{(j)}$ and β_j the power and the amplification factor used by the j–th RS respectively, whereas $P_r = \sum_{j=1}^{Q} P_r^{(j)}$ and $\beta = \sum_{j=1}^{Q} \beta_j$. Finally, we put $n_j = |\mathcal{C}_j|$, $n = \sum_{j=1}^{Q} n_j$ and $k = |\mathcal{D}|$. If the MTs allocation is such that each MT generates a small interference to the RSs/BS it is not assigned to, the SINR at the BS from the i–th MT, with $i \in \mathcal{C}_j$, can be written as in (3.47), where we called $\sigma = G_r M_g G_b$ to simplify the notation. The first three terms at the denominator account respectively for the interference due to the MTs linked to the same RS, to the ones linked to the other RSs and to the non cooperating MTs. The last term is the noise power, comprehensive of the noise amplified by all the RSs.

Equation (3.47) can be significantly simplified if we highlight the following argument: in order to have the same SINR for all the users at the BS it is necessary that all the powers received at the BS are equal. This in turn implies also that the various signals received at a given RS must all have equal power, since they are all amplified by the same factor and sent over the same channel. Therefore, we can rewrite the SINR using the auxiliary variables ϕ_j, with $1 \leq j \leq Q$, one for each RS, and ϕ_0, accounting for the non cooperating MTs. They are defined as $\phi_j = P_t^{(i)} d_i^{-\alpha}$, $\forall i \in \mathcal{C}_j$, and $\phi_0 = P_t^{(i)} d_i^{-\alpha}$, $\forall i \in \mathcal{D}$. The total number of variables is now equal to $Q + 1$, and $P_t^{(i)}$, with $i \in \mathcal{C}_j$, can be immediately found as $\phi_j d_i^{\alpha}$ (and as $\phi_0 d_i^{\alpha}$ if

$i \in \mathcal{D}$). Consequently, the equation of the SINR can be rewritten as:

$$SINR^j = \frac{C_g \phi_j \beta_j \sigma d_R^{-\alpha}}{I_j + N_0 \left(1 + \beta G_b d_R^{-\alpha}\right)} \tag{3.48}$$

which holds for all the MTs assigned to the j–th RS, and where now the interference term I_j is simply given by

$$I_j = (n_j - 1)\, \beta_j \sigma d_R^{-\alpha} + \sum_{h=1, h \neq j}^{Q} n_h \phi_h \beta_h \sigma d_R^{-\alpha} + k \phi_0 G_b \tag{3.49}$$

Since the same equation can be applied to each RS, and a similar equation can be written for the non cooperative MTs, we obtain a linear system that can be easily solved to find the values of all the ϕ_k, and consequently of all the values of the transmission powers. We report the system solutions: in order to have a target SINR equal to Γ, the power required for a non cooperating MT i is

$$P_t^{(i)} = \frac{\Gamma}{C_g - (n + k - 1)\Gamma} \frac{N_0 d_i^{\alpha}}{G_b} \left(1 + \beta M_g G_b d_R^{-\alpha}\right) \tag{3.50}$$

whereas, if $i \in \mathcal{C}_j$, is equal to

$$P_t^{(i)} = \frac{\Gamma}{C_g - (n + k - 1)\Gamma} \frac{N_0 d_i^{\alpha}}{G_r} \left(\frac{\beta}{\beta_j} + \frac{1}{\beta_j M_g G_b d_R^{-\alpha}}\right) \tag{3.51}$$

The obtained results are very general, but can be applied only under the aforementioned interference assumptions. As was pointed out in Section 3.4.1, however, the non cooperating MTs are likely to create a large amount of interference at the RSs, thus justifying the choice of assigning them to a different channel. In this case, the expression of the required power for a cooperating node is still given by (3.51), with $k = 0$, whereas for the non cooperating nodes the traditional non cooperative scheme equations can be used, but as if only k MTs were deployed within the cell.

It is now possible to derive the expression of the coverage range, defined as the maximum distance at which the desired SINR can be reached without violating a constraint on the transmission power at the MTs. Assume now that each MT has a maximum value for the transmission power equal to P_m. For the non cooperative scheme, it can be easily shown that the coverage range d_T is a function of the number k of active users, and can be expressed as:

$$d_T(k) = \left(P_m \frac{G_b}{N_0} \left(\frac{C_g}{\Gamma} - (k - 1)\right)\right)^{\frac{1}{\alpha}} \tag{3.52}$$

The evaluation of the same quantity in the cooperative scenario is less straightforward. The SINR at the BS depends on the powers used both by the MT and by the RS. Therefore,

it is necessary to define a power allocation policy at the RSs. In this work, we assume that each RS uses a total transmission power $P_r^{(j)} = \Psi n_j$ proportional to the number of MTs n_j assigned to it (up to a maximum value). With this choice, the MTs assigned to a certain RS do not have to change their transmission power each time a new MT begins a transmission using the same RS as cooperator. Other choices are possible as well. The amplification factor β_j can then be rewritten as

$$\beta_j = \frac{P_r^{(j)}}{\sum_{i=1}^{n_j} P_t^{(i)} G_r d_i^{-\alpha} + N_0} = \frac{P_r^{(j)}}{n_j G_r \phi_j + N_0} \tag{3.53}$$

The value of $\beta = \sum_{i=1}^{Q} \beta_i$ can also be easily computed; by replacing them in (3.51), and setting $P_r = \Psi n$ we get

$$P_t^{(i)} = \frac{N_0}{G_r d_i^{-\alpha}} \frac{1 + \dfrac{N_0}{M_g G_b \Psi n d_R^{-\alpha}}}{\dfrac{n_j}{n}\left(\dfrac{C_g}{\Gamma}+1\right) - n_j\left(1+\dfrac{N_0}{M_g G_b \Psi n d_R^{-\alpha}}\right)} \tag{3.54}$$

Assuming P_m as the maximum allowed transmission power, and with $\Psi = P_m$, simple algebraic manipulations give the expression of the maximum distance at which a MT can be deployed in order to match the condition on the SINR at the BS. This distance, however, is from the RS rather than from the BS, meaning that the coverage area is here represented by Q circles, centered at the Q RSs. The radius of the circle around the j–th RS can be written as

$$d_j = \left[\frac{P_m G_r}{N_0}\left(\frac{n_j P_m}{n P_m + \frac{N_0}{M_g G_b d_R^{-\alpha}}}\left(\frac{C_g}{\Gamma}+1\right) - n_j\right)\right]^{\frac{1}{\alpha}} \tag{3.55}$$

and depends not only on the number of users assigned to the j–th RS, but also on the number of total users exploiting cooperation.

3.4.3 Resource Allocation and System Analysis

In the previous section we have analytically derived the expression of the coverage area of a single CDMA cell where cooperation is applied by means of a fixed number Q of RSs. As was pointed out, this area is given by Q circles, whose radii are defined as in (3.55). Since d_j depends not only on n_j but also on n, in our scenario the allocation of the MTs must be well planned. As a matter of fact, when a new MT begins a transmission using the j–th RS as cooperator, not only is d_j modified, as in the non cooperative scheme, but also the radii related to the other RSs shrink, as depicted in Figure 3.17.

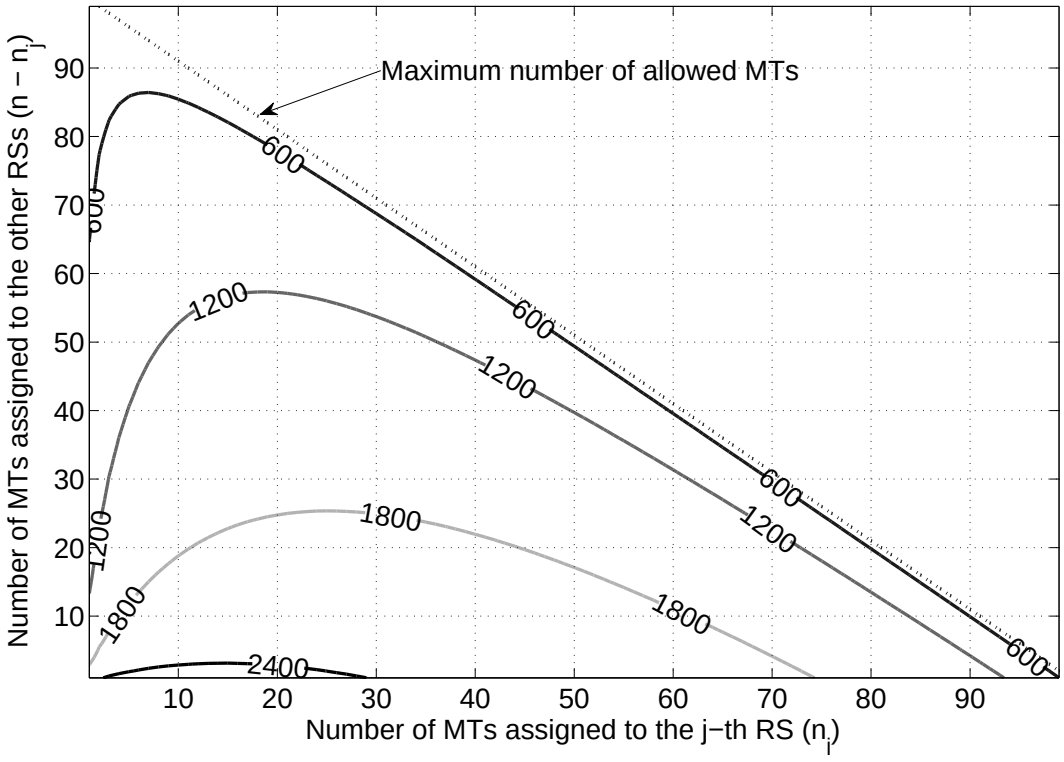


Figure 3.17. *Contour lines of the allowed distance of a MT from its assigned RS. It is plotted as a function of the number of MTs assigned to the same RS and of those assigned to the other ones. Here 4 RSs are deployed, with* $d_R = 1800m$.

The particular choice of the power allocation policy at the RSs determines how each radius varies: focusing on RS j, when the number of MTs assigned to other RSs is raised, then d_j becomes lower. On the contrary, d_j increases with n_j, due to the higher power used by the RS. However, when n_j grows beyond a certain value, then d_j begins to decrease, due to the excessive interference introduced in the system.

From Figure 3.17 it can also be inferred that there exists a maximum number N_c of MTs that can be served by the RSs. By putting $d_j = 0$ in (3.55) the exact value is found to be

$$N_c = \left\lfloor \left(\frac{C_g}{\Gamma} + 1 \right) - \frac{N_0}{M_g G_b \Psi d_R^{-\alpha}} \right\rfloor \tag{3.56}$$

Numerically, it is found that this value is only slightly lower than in the non cooperative case, due to the term $N_0/(M_g G_b \Psi d_R^{-\alpha})$, usually much smaller than the ratio C_g/Γ. This is an effect of the noise amplification caused by the A&F scheme. Anyway, with a proper network design, the enhanced coverage range granted by cooperation results in the possibility for allocating a larger overall number of MTs. In fact, in a traditional CDMA cell the coverage area shrinks as the number of MTs is raised. This means that the MTs farther from the BS are

likely to be dropped when several communications are taking place. The use of two or more frequency bands can mitigate this effect, but the same reasoning holds for each band. When using RSs, on the contrary, the coverage range still decreases for higher number of MTs, but now the covered area shrinks around the RSs, which are deployed far from the BS. Hence, also MTs not close to the BS can still be served, although with half the bandwidth.

In order to evaluate this effect, we calculate the probability $\mathcal{P}_c(d)$ that a MT placed at distance d from the BS is within coverage range, given the number of MTs already present in the cell, the number of RSs and their distance from the BS. Exploiting the symmetry of the problem, we consider that the n MTs are equally divided among the RSs, that is $n_j = n/Q \ \forall j$. Moreover, given the interference assumptions behind equation (3.54), we assume that each cooperating MT is always linked to the closest RS, even if its position is within the coverage area of two or more RSs. Under this assumption, $\mathcal{P}_c(d)$ can be calculated based on geometric considerations. Consider a cartesian plane with the BS placed at the origin; for each value of d we can define a circle centered at the origin with radius d. The fraction of this circle that intersects the coverage areas around the RSs gives $\mathcal{P}_c(d)$:

$$\mathcal{P}_c(d) = \begin{cases} 0 & d_j < |d - d_R| \\ 1 & d_j > d + d_R \\ \min\left(\frac{Q}{\pi d} \arccos\left(\frac{d^2 + d_R^2 - d_j^2}{2 d d_R}\right), 1\right) & otherwise \end{cases} \tag{3.57}$$

where the minimum is taken when the coverage area of two or more RSs overlap.

Another interesting evaluation is based on the dropping probability $\mathcal{P}_d$. When a new transmission from a MT starts, it may happen that other far MTs fall outside the reduced coverage area, and consequently can no longer transmit to the BS. We define $\mathcal{P}_d$ as the probability that at least one communication must be dropped when a MT begins a new transmission. This probability depends on the number N of already active MTs, as well as on d_R and Q. In the non cooperative scenario, we assume that the N MTs are randomly deployed within the allowed coverage range $d_T(N/2)$ (in order to make a fair comparison, we assume that two frequency bands are available, and that $N/2$ MTs are assigned to each one), and that the new MT is also randomly placed within a reference distance ξ from the BS. This MT can choose each either of the two frequency bands; nonetheless, we do not allow the other MTs to change band via frequency handover.

The dropping probability is the sum of two terms. The former takes into account the probability that the new MT appears at a distance $\rho > d_T(N/2 + 1)$, meaning that it cannot

be served [5]; the latter on the contrary is for the case $\rho < d_T(N/2+1)$. In this case, the new MT can be served, but no communication is dropped only if for at least one of the two frequency bands all the $N/2$ MTs allocated are placed at a distance lower than $d_T(N/2+1)$ from the BS. This event has probability

$$p_a = 1 - 2\left(\frac{r_{N/2+1}}{r_{N/2}}\right)^N + \left(\frac{r_{N/2+1}}{r_{N/2}}\right)^{2N} \tag{3.58}$$

where we define $r_i = d_T(i)$. The dropping probability is hence

$$\mathcal{P}_d = 1 + (p_a - 1)\left(\frac{r_{N/2+1}}{\xi}\right)^2 \tag{3.59}$$

Consider now the cooperative scenario, with two frequency bands still available, one for the non cooperating MTs and one for the MTs assigned to one of the Q available RSs. Assume again that N MTs are present, equally divided among the two bands, and that the $N/2$ cooperating nodes are equally distributed among the RSs (that is, $n_j = N/(2Q)\ \forall j$). These nodes are uniformly distributed within the corresponding coverage areas. When a MT, whose position is uniformly chosen within the reference distance ξ from the BS, starts a new transmission, four disjoint events may occur:

1. The MT may appear in the area $\mathcal{A}_{nc}$ where only direct transmission to the BS is possible. In this case, no transmission is dropped only if all the $N/2$ non cooperating MTs are within a circle of radius $r_{N/2+1}$ from the BS.

2. The MT may appear within the fraction of the coverage area $\mathcal{A}_j$ of the j–th RS where only a cooperative transmission is possible. Hence, no transmission is dropped if all the cooperating MTs assigned to all the RSs fall inside the new coverage areas determined by the addition of a new MT (which generally causes the radii of the coverage areas around the other RSs to shrink).

3. The MT may appear in the fraction of $\mathcal{A}_j$ where both cooperative and non cooperative transmission are possible. In this case, each choice leads to a dropping probability described as in the previous two cases, and the best one is chosen.

4. The MT may appear in the area $\mathcal{A}_{out}$ where neither cooperative nor non cooperative transmission is possible, and in this case the dropping probability is 1.

[5] In this case we consider that the transmission from the new MT is dropped, although actually it does not start at all.

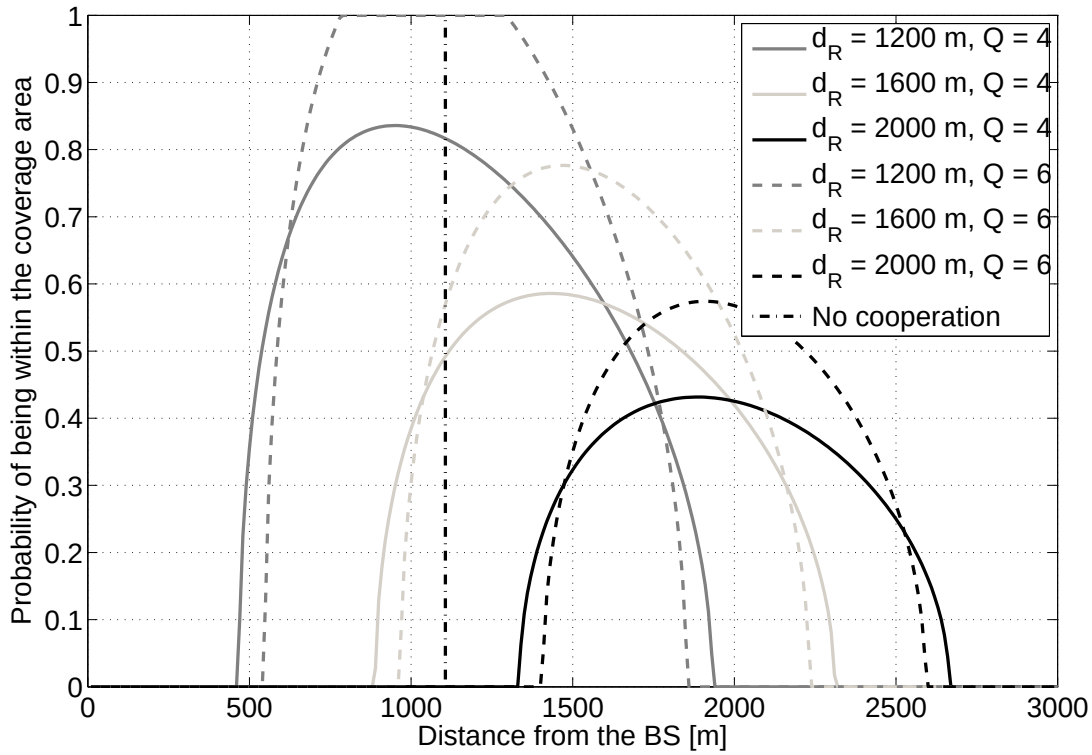


Figure 3.18. *Probability $\mathcal{P}_c(d)$ that a MT located at distance d from the BS is within the coverage area, for different values of d_R and Q. Here 96 MTs are active in the cell.*

The exact expression of $\mathcal{P}_d$ in the cooperative case is not reported here, but can be found in closed form, since all the probabilities can be derived as ratios between areas, which in turn can be found with simple geometrical considerations.

3.4.4 Numerical Results

In this Section we report some results on the performance of the cooperative scheme proposed, compared with that of the non cooperative one.

In Figure 3.18 we show the probability $\mathcal{P}_c(d)$ that a MT located at distance d from the BS is within the coverage area either of the BS or of one of the RSs. The number n of total active MTs in the cell is equal to 96, close to the maximum value allowed with the used setup parameters. It can be noticed that while the non cooperative scheme guarantees the transmission up to about one kilometer, the use of RSs grants a much higher coverage range. However, since only a limited number of RSs are deployed, the region far from the BS is not fully covered, resulting in a reduced value of $\mathcal{P}_c(d)$. In addition, increasing d_R can give advantages in terms of maximum transmission range, but far MTs can be served with lower

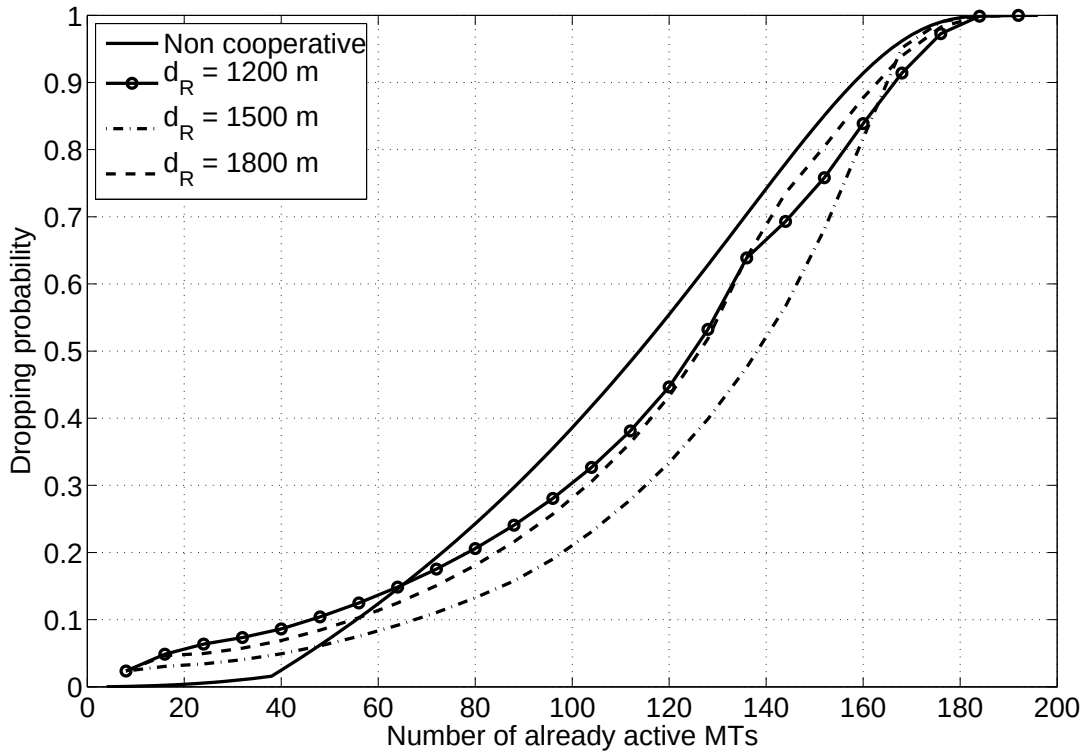


Figure 3.19. *Dropping probability as a function of the already active MTs in the cell, for various values of d_R. Here the reference radius ξ of the area where the new MT can appear is set equal to $d_T(20)$.*

probability, since the areas covered by the RSs are only shifted, and not enlarged. A higher number of RSs can increase $\mathcal{P}_c(d)$ also for far MTs, but the covered areas of the RSs are slightly shrunk, due to the increased interference.

The dropping probability is depicted in Figure 3.19. The non cooperative scheme performs better when few MTs are present. In this case, the coverage area of the BS is large, and includes most of the covered areas of the RSs, which therefore are not useful. On the contrary, when the cell is high loaded, the intersections between the coverage area of the BS and those of the RSs becomes smaller. Hence, MTs closer to the BS can be served without cooperation, whereas the farther ones can be assigned to one of the RSs, thus lowering the dropping probability. This effect depends on the position of the RSs: if they are deployed too close to the BS, most of their benefit is lost, since their coverage area is again included within that of the BS; otherwise, if they are located too far, their coverage areas becomes disjoint from that of the BS, resulting in a fragmented cell.

3.4.5 Discussion

In this part we have proposed a practical implementation of a cooperative scheme based on A&F in CDMA cellular networks. In such scheme, each MT can be helped by a fixed Relay Station by means of a TDMA transmission. The benefits granted by cooperation have been evaluated in terms of coverage range and dropping probability. Results show that the use of cooperation can not only increase the area covered by the BS, but also reduce the probability that an ongoing transmission must be dropped when a MT starts a new communication. These promising results are useful, since they show that even with A&F, whose performance has been proved to be worse than that achievable with D&F, a significant benefit can be obtained. However, some assumptions are still making the simulation setup different from a real cellular system. The main points that have not been covered yet are the following:

- Inter-cell interference has not been taken into account. In a cooperative scenario, where Relay Stations are used to increase the coverage area, this issue may have a strong impact. Actually, the Relay Stations are placed closer to the cell edge, as well as the relayed Mobile Terminals. This in turn means that the interference created by these mobiles to the Relay Stations of the neighboring cells may be higher (depending on the carrier configuration).

- Directional antennas may further improve the performance of the Relay Stations. For instance, if two carriers are not available, this solution may dramatically reduce the interference at the Relay Stations from mobiles that are not relayed. In addition, if directional antennas are used for both the Relay access link and the backhaul link, the isolation could be high enough to remove the constraint of half-duplex transmission.

- Soft Handover plays an important role in cellular networks. If it is applicable also among Base Stations and Relay Stations, it could somehow reproduce the benefits coming from MRC. If on the contrary it cannot be used for relayed mobiles, this could limit the performance improvement of the cooperative scheme.

- Different distributions of the mobiles could be taken into account. Since the cooperators can significantly reduce the path loss, they may be used also to improve the transmission reliability when coverage holes are present. For example, Mobile Terminals inside a building, whose transmissions are hampered by an additional penetration

loss, could take huge advantage from a Relay Station placed near a window. In this case, the locations of the Relay Stations and those of the mobiles are no longer uniform nor independent. Moreover, such a realistic scenario is likely to be quite attractive for operators.

3.5 Full system simulation

This part of the work presents the results obtained through extensive simulations of an entire cellular network with relayed UL transmissions. These results have been developed during a six-months internship in the Corporate Research Labs of Qualcomm Incorporated.

In the previous sections, the application of simple relaying techniques, like A&F and D&F, in CDMA cellular networks showed a large potential performance improvement in terms of coverage range (or, equivalently, required transmission power). However, even when considering a whole cell in the analyzed scenario, many features cannot be captured by our theoretical model. For instance, power control effectiveness should be checked, as well as the impact of relaying on soft handover mechanisms. In addition, channel allocation should be studied, and more realistic traffic models taken into account. Focusing only on the capabilities of the relays, directional antennas are likely to severely affect the performance, since on the one side they allow high self-interference attenuation, but on the flip side they are more susceptible to inter-cell interference. If D&F is used, as in the present Section, queueing delays at the relays may also have a relevant impact on the overall throughput. The user allocation is another important factor which requires a proper relay deployment in the cell. Finally, from a practical perspective, a more realistic channel model should be used, in order to capture the effect of correlated shadowing, penetration loss and multipath fading.

3.5.1 System description

We describe here the scenario used in the simulations to measure the actual benefit that can be granted by cooperation through the use of relays employing D&F. We stress that the simulation is comprehensive of all the layers, from physical to higher layers.

We focus on a scenario with 21 cells, located in 7 sites, as in Figure 3.20(a), since the hexagon around each site is covered by three directional antennas. The number of users per cell is n, which may be deployed according to the two following distributions:

- *Uniform distribution*, meaning that for each cell n users are uniformly distributed in the cell area;

- *Clustered distribution*, meaning that a given percentage of users (50% or 80%) is clustered, and distributed in small circular areas around a given number k of locations, whereas the remaining percentage of users is uniformly distributed in the whole area of the cell.

In both cases, however, the mobiles are required to satisfy some conditions on the minimum allowed distances among mobiles and Base Stations, as reported in Table 3.5. When mobiles are dropped, if one of them does not match these conditions, it is moved to another randomly selected location. Mobiles positions are considered fixed during the simulations.

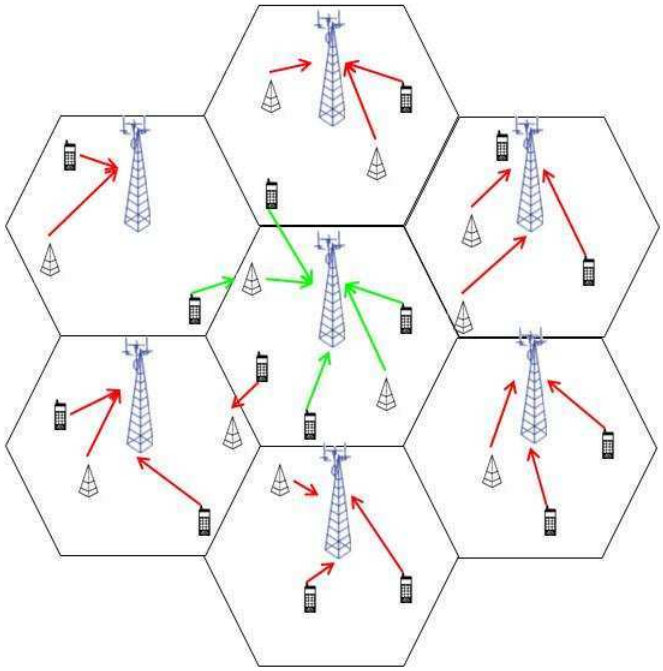
It is clear that the clustered distribution may appear more realistic, since it may capture the effect of areas with a high concentration of mobile terminals, for instance houses, offices, meeting rooms. When clustered distribution is adopted, *coverage holes* may be simulated around the clusters. In practice, this means that an additional penetration loss is applied to the path loss from the mobile terminal to the corresponding base station. Also this augmented penetration loss is added to better simulate the case of a number of mobile terminals distributed in a small indoor space, which is likely to be covered and shielded by walls or roofs.

The traffic is modeled as bursty. This means that packets with a fixed size L are generated at each mobile terminal as a Poisson process with given intensity, or, equivalently, with a given average interarrival time t_p. The packets are generated at the higher layers, and then passed to the lower ones, fragmented, and sent over the UMTS channels towards the Base Station.

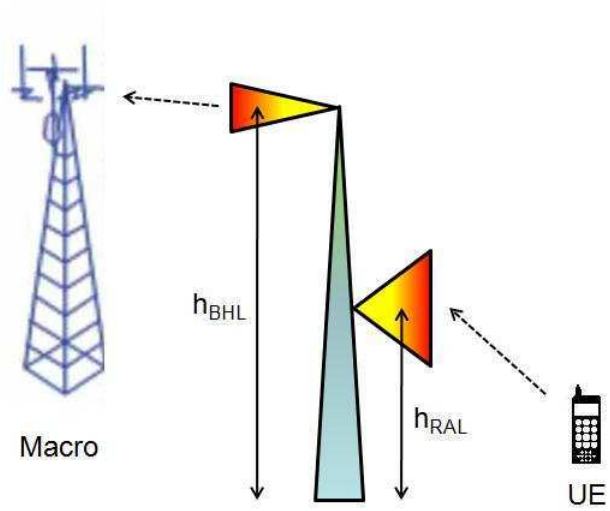
The mobility metric, which determines the Base Station each mobile is allocated to, is a function of the received downlink SINR. This means that the mobile is allocated to the Base Station whose downlink signal has the highest SINR.

As regards the channel model, path loss decays according to the distance between transmitter and receiver, with given offset A and exponent α. Fast fading is modeled as a PA3 model. An additional penetration loss δ is added; shadowing is modeled as a lognormal random variable, with fixed standard deviation σ_s, and inter-site shadowing is correlated with correlation factor ρ_s. All the used parameters are collected in Tables 3.4 and 3.6.

When cooperation is used, a fixed number k of relays are uniformly deployed in each



(a) Cell layout



(b) Relay model

Figure 3.20. *In a), the cell layout used in simulations, where at each site three Base Stations are colocated (sectorization). Since allocation is based on downlink SINR, it is possible that some users are not allocated to the closest Base Station. In b), the relay model: in our simulations,* $h_{BHL} = 10m$ *and* $h_{RAL} = 5m$*; directional antennas are used for both RAL and BHL.*

Scenario Configuration	Used Values
Relays Layout	Uniform
Mobiles Layout	Uniform, Clustered (50%, 80%)
Macro Cell Layout	21 cells (2 rings)
Soft Handover	Only Macro
Traffic Type	Bursty Traffic
Page Size (L)	250 kb
Interarrival time (t_i)	5 s
Aggregation over BHL	All
Coverage Holes	In clustered only
Relays per Macro (k)	8
Mobiles per Macro (n)	8, 16
Mobility Metric	E_c/I_o
Monte Carlo statistics	5

Table 3.4. *List of parameters used in the full system simulations.*

	Macro	Relay	Mobile
Macro	1000	75	35
Relay	75	75	10
Mobile	35	10	0

Table 3.5. *Minimum distances [m] among relays, mobiles and base stations. The distance between two base stations is fixed to 1000 m.*

Parameter	Value
I_{AC}	45 dB
Relay Isolation	100 dB
Shadowing Standard Deviation (σ_s)	8 dB
Shadowing Correlation Factor ρ_s	0.5
Penetration Loss (δ)	10 dB

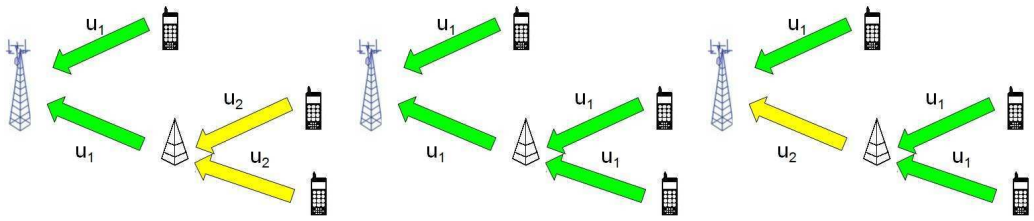
Table 3.6. *Channel parameters.*

cell, under the constraints on the minimum distances among terminals reported in Table 3.5. If clustered user distribution is used, each relay is placed at the center of each cluster area. This is reasonable, since relays are meant to be located in areas where traffic is congested and a high number of users is active. In order to make a per-user comparison, mobiles in simulations are kept in the same positions when relays are added. We call Relay Access Link (RAL) the channel towards the relay, and Backhaul Link (BHL) the one from the relay to the base station.

Relays are considered to be indoor-outdoor, meaning that the RAL channel is indoor, whereas the BHL channel is outdoor. In practice, if relays are used to help communications in congested indoor areas, they are placed such that their receiving antenna is indoor, while their transmitting antenna is outdoor. This implies not only that no penetration loss is present in either of the two links, but also that an additional isolation between the two antennas can be granted at the relay. This is important, since we consider full duplex relays. A contribution to the required isolation is also given by the distance between receiving and transmitting antennas, fixed to 5 meters, by the use of directional antennas for both RAL and BHL, and, in some scenarios, by the use of different frequencies (an additional isolation I_{AC} between adjacent channels is considered in this case). A scheme of the relay model is depicted in Figure 3.20(b).

We distinguish three different cooperative scenarios, depending on carrier assignment:

- **Dedicated RAL**: one carrier is used by the users transmitting to a relay, and the second one by the relays and by the users transmitting directly to the Base Station;
- **One Carrier**: one carrier is used for all transmissions;
- **Dedicated BHL**: one carrier is used by all the mobile terminals, whereas the second one is used by the relays.



(a) Scenario 1, Dedicated RAL (b) Scenario 2, One Carrier (c) Scenario 3, Dedicated BHL

Figure 3.21. *The three scenarios analyzed, which differ in the carrier assignment. While scenario 1 and scenario 3 require two carriers, only one is necessary in Scenario 2.*

The three cases are depicted in Figure 3.21. It is clear that the second scenario requires only half the bandwidth, but also implies more restrictive constraints in relay design in order to avoid overwhelming self-interference. Finally, the maximum transmission power of each relay is 6 dB higher than that of a mobile terminal, since it is likely to happen that multiple mobiles are allocated to the same relay.

3.5.2 Aggregation and Mobility Algorithm

In this section, we describe two additional features that were developed to enhance the performance of relayed uplink transmission. The concept of aggregation derives from the fact that each relay may be required to help the transmissions of several mobile terminals. However, all these transmissions share the same backhaul link towards the Base Station, meaning that a lower amount of information should be sent on the BHL. For example, information on channel conditions or pilot sequences can be used for all the relayed packets rather than repeated for each of them. Aggregation can be implemented in a number of ways. In this work, in consideration of the traffic model, we use the model depicted in Figure 3.22: after being received at the relay, an additional overhead is added to each packet, for instance an IP address. The packets are then queued, and sent from the relay according to a FIFO policy. Note that in this manner the relay is acting as a common mobile terminal, using only one pilot and one control channel. Similarly, the Node B receives each packet as if it comes from a simple mobile terminal, decodes it, and sends it to a Gateway, where the original packet is recovered, its header is read, and it is consequently sent to the original destination. The main advantage of this implementation is that instantaneous interference is reduced, since one packet is transmitted at each time from a single relay. The main drawback is that if the number of mobiles allocated to a single relay is too high, or the traffic load

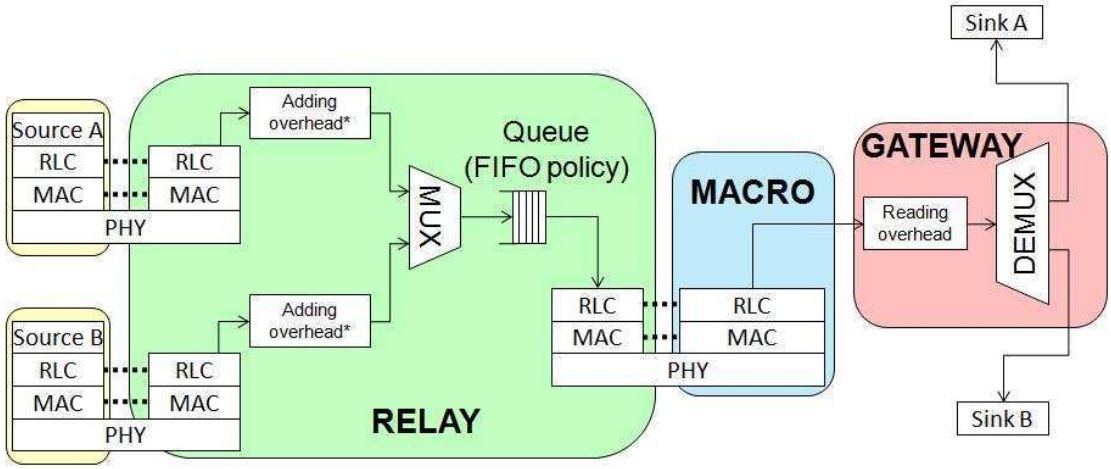


Figure 3.22. *Schematic view of the used aggregation model on the BHL.*

increases, long queueing delays at the relays may considerably degrade the performance in terms of throughput.

A second element that should be taken into account is the selected mobility metric. Actually, the mobile terminal is only able to estimate the quality of the RAL channel. It may happen that the RAL channel is much better than the direct link towards the base station, and the mobile is hence allocated to the relay. However, if the BHL channel is bad, thus becoming the bottleneck of the relayed transmission, then the allocation to the relay is not optimal, and may on the contrary badly affect the quality of the transmission. It is then necessary to develop a way to let the mobile know also the condition of the BHL channel, hence allowing the best choice between direct and relayed transmission. In this work, we used the Relay Mobility Algorithm (RMA), which was developed for downlink transmission. The details are not reported here due to lack of space, but the rationale behind it has been clarified. Although this algorithm is not optimized for uplink, it was proven to grant high benefits in downlink relayed transmissions, which may compensate minor losses in uplink performance when cooperation is used for both uplink and downlink.

3.5.3 Main goal and performance metrics

The main aim of this work is to verify whether relays can grant significant performance improvements when used to help uplink transmissions. In addition, results should also show which kinds of benefits are achieved through cooperation, and how they depend on system settings, like mobiles and relays deployment, traffic load and carrier assignment.

In fact, relays are expected to considerably improve the path loss of mobiles close to

the cell edge, as well as to reduce the transmission power used by the relayed mobiles. In addition, aggregation should reduce the instantaneous interference at the base station, hence also helping the non relayed mobile terminals.

The improvement is likely to depend on the mobile positions: in clustered scenarios the mobiles are usually much closer to the relay, resulting in lower transmission power, and if coverage holes are applied, the use of relays can reduce the number of mobiles in outage (that is, mobiles that cannot satisfy the condition on the received SINR at the base station even when transmitting with maximum power).

As was observed above, also the traffic load has an impact, since it affects the queue length at the relay, and therefore the packet delivery time. Finally, the use of two carriers aims at reducing both intra-cell and inter-cell interference at the relays.

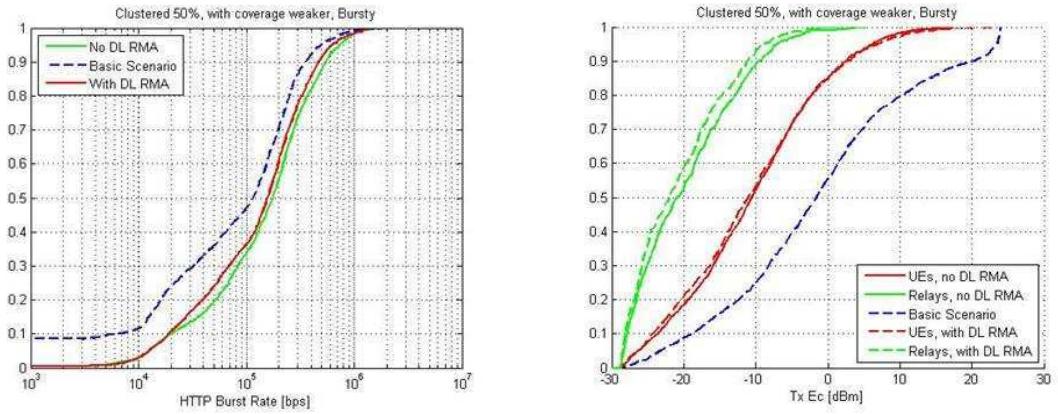
We focus in the results Section on two performance metrics, namely burst rate, defined as the ratio between page size and time required for its complete reception at the base station, and mobile terminal transmission power.

3.5.4 Results

We present in this Section the results obtained through use of extensive C++ simulations. All the graphs are presented in the form of CDFs, for both throughput and mobile terminal transmission power. Only a subset of the obtained graphs are reported due to lack of space.

Scenario 1: Dedicated RAL

In Figure 3.23 the burst rate and the mobile terminal transmission power for a scenario with $n = 16$ users per cell and 50% clustering are reported; coverage holes are also present. It can be observed that the use of relays grants a significant burst rate improvement, up to approximately 30%, especially for mobiles in worse conditions. Moreover, mobiles which appear to be in outage without cooperation (since the CDF does not approach zero as the burst rate decreases) can transmit successfully by means of relay help. In fact, this is the main purpose of the relay deployment in this scenario, where the additional penetration loss due to the coverage hole is avoided in relayed transmission. The DL RMA shows lower improvement. When the Downlink Relay Mobility Algorithm is used, a lower number of mobiles is in general allocated to the relays, due to the fact that relays with bad BHL channel are used less often. However, this also means that the number of users per relay is decreased, limiting the beneficial effect of aggregation. The same reasoning explains the results about



(a) Burst Rate in Scenario 1

(b) Transmission power in Scenario 1

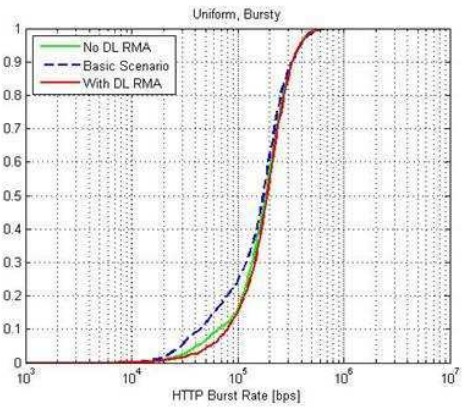
Figure 3.23. *Performance improvement through use of relaying in Scenario 1 (Dedicated RAL), with 16 users per cell. Here users are clustered (50% clustering) and coverage holes are present.*

transmission power. Here again a huge improvement is shown for almost all the mobiles, with the exception of the ones with best channels. The improvement is slightly lower for mobiles with very bad channels, near the cell edge. This is due to the fact that soft handover is not allowed for relayed users, and its benefits are hence lost. Even higher gains are achievable in scenarios with more clustered mobiles.

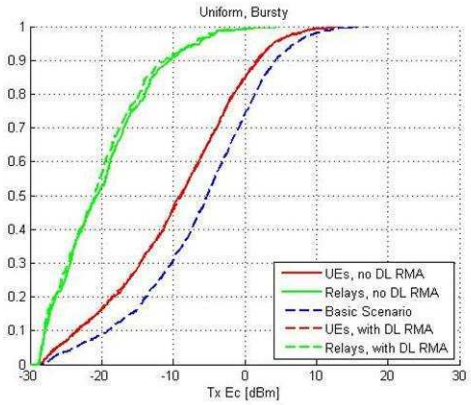
Another observation regards the impact of inter-cell interference. When two carriers are used, inter-cell interference is likely to be different for the two carriers. More precisely, relays suffer from high inter-cell interference, since not only are relayed users of other cells usually closer to the cell edge, but also the relays are far from the center of the cell, and their directional antennas (for the access link) are directed away from the Base Station. This is even more evident in scenarios with uniformly distributed users, because the relayed users are farther from the relays, thus using higher transmission power. A possible solution to limit this effect is to increase the distance among the Base Stations when cooperation is used.

Scenario 2: One Carrier

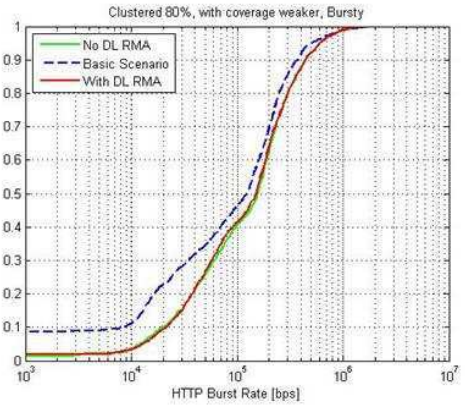
Results for Scenario 2 are depicted in Figure 3.24. Performance improvement here is lower than in Scenario 1, but also the required bandwidth is halved. As in the previous case, the main benefits come from the reduced path loss for mobiles in outage, which are able to transmit when relays are used, and from aggregation, which instead reduces interference at the macro, which affects also the non relayed mobiles.



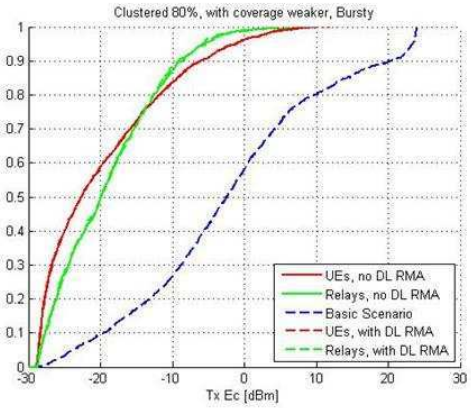
(a) Burst Rate, uniform user distribution



(b) Transmission power, uniform user distribution



(c) Burst Rate, clustered user distribution



(d) Transmission power, clustered user distribution

Figure 3.24. *Performance improvement through use of relaying in Scenario 2 (One Carrier), with 16 users per cell. In a) and b) users are uniformly distributed, whereas in c) and d) they are clustered (80% clustering) and coverage holes are present.*

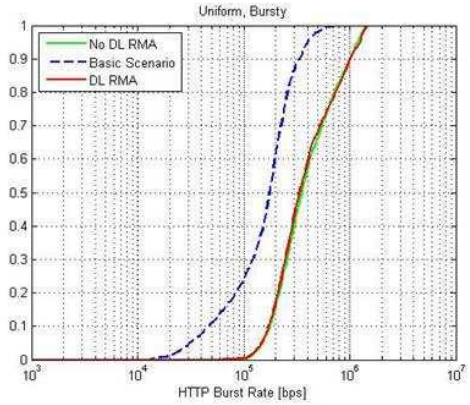
Given the selected mobility metric, however, it can be shown that the relayed mobiles tend to be much closer to the relay, meaning that the choice of a single carrier is more suitable for scenarios with clusters of mobiles concentrated in smaller areas. This also explains the higher gain in transmission power, compared to the ones shown for the previous scenario. The main problem is interference. On the one side, as in Scenario 1, relays suffer from inter-cell interference, especially in scenarios with uniformly deployed mobiles; on the flip side, they also suffer from high intra-cell interference from non relayed mobiles, which transmit on the same frequency. Finally, isolation at the relays is reduced due to the use of the same band for the RAL and the BHL.

Scenario 3: Dedicated BHL

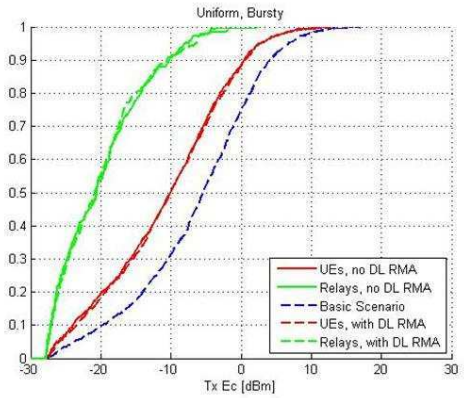
This third scenario puts together the benefits of the previous ones. The use of two carriers makes the effect of aggregation more evident, since the number of simultaneous transmissions on the BHL is highly reduced, especially in clustered scenarios, when the number of users per relay is higher. However, as in the second scenario, the fact that the RAL and the direct link are on the same carrier reduces the average distance between relay and relayed mobiles, leading to strong reduction of the transmission power as well. Moreover, since the BHL is usually very good, the DL RMA here increases the number of relayed mobiles, enhancing the effect of aggregation and hence improving the performance of uplink transmission (and also of the downlink ones, for which it was designed).

The burst rate and transmission power improvement are reported in Figure 3.25. It can be observed that the burst rate increase is much higher than in the previous two cases, especially for cell edge mobiles. Again, the highest performance is for clustered distribution with coverage holes. Moreover, the use of DL RMA further boosts the burst rate, due to the higher number of users per relay and, consequently, to the augmented effectiveness of aggregation. Table 3.7 shows the comparison among the effects of cooperation in the three scenarios described above.

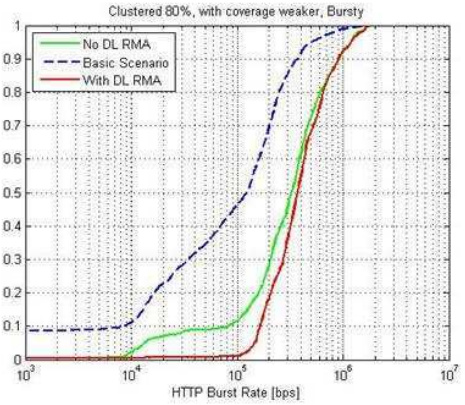
It is interesting to notice that the various carrier assignments have a different impact on the relayed and non relayed users, in terms of burst rate. If the RAL has a dedicated carrier the higher improvement is offered to the non relayed terminals. This is due to the fact that their transmissions experience a lower interference at the macro: this reduction derives from the use of aggregation, although the relays transmit on the same carrier. For relayed users, despite the good quality of the RAL (few users are transmitting to each relay), the BHL still



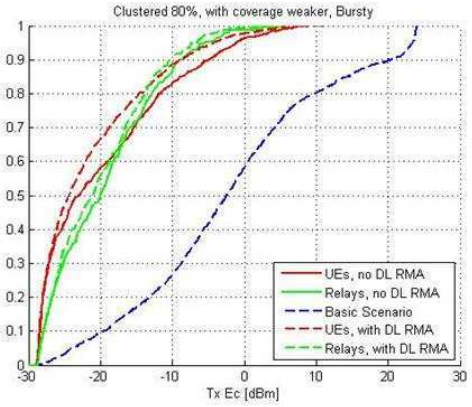
(a) Burst Rate, uniform user distribution



(b) Transmission power, uniform user distribution



(c) Burst Rate, clustered user distribution

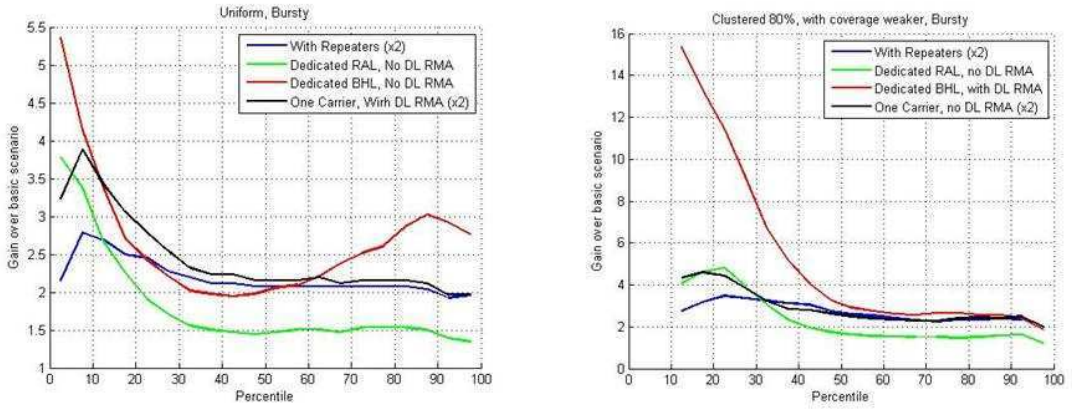


(d) Transmission power, clustered user distribution

Figure 3.25. *Performance improvement through use of relaying in Scenario 3 (Dedicated BHL), with 16 users per cell. In a) and b), users are uniformly distributed, whereas in c) and d) they are clustered (80% clustering) and coverage holes are present.*

	Dedicated RAL	One Carrier	Dedicated BHL
Carriers required	2	1	2
Better in clustered scenarios	High radius	Small radius	Small radius
Improvement Non Relayed MTs	Higher	Similar	Lower
Improvement Relayed MTs	Lower	Similar	Higher
DL RMA Performance Gain	Limits	Almost no effect	Enhances

Table 3.7. *Comparison among effects of cooperation in the three considered scenarios.*



(a) Uniform user distribution

(b) Clustered user distribution (80%)

Figure 3.26. *Performance improvement comparison by percentile with repeaters and with relays. The three different scenarios are reported for D&F scenario. The gains are normalized to the required bandwidth.*

suffers from interference coming from non relayed terminals, and queueing delays are also possible.

If, on the contrary, BHL has a dedicated carrier, things are much different. With this configuration, in fact, non relayed terminals see a slightly higher improvement than in case 1, since the few relay transmissions are now on a different carrier, but the relayed users see a much more pronounced burst rate enhancement. The reason is that now the BHL is very good, since only the relays are transmitting on it, and they use aggregation. As regards the RAL, the use of directional antennas and the small cluster radius make the interference from non relayed users negligible, granting an overall burst rate increase.

3.5.5 Repeater-Relay comparison

We finally present the comparison between the performance achievable with relays and those granted by the use of simple repeaters (that is, using A&F). Repeaters were deployed in the same locations of the relays, to make a per user comparison possible. They require only one carrier, since isolation is dependent on the use of directional antennas (same characteristics as those used for relays). The amplification factor is $A_p = 70$ dB, up to a maximum trasmission power $P_{rep} = 30$ dBm. The saturation effect was also considered in simulations. The burst rate improvement over the basic scenario is reported in Figure 3.26, where the gains are normalized to the required bandwidth (gains of repeater case and scenario 2 of relay case are doubled).

The fairest comparison is clearly the one between the scenario with repeaters and the one

with relays employing only one carrier. It can be observed that repeaters usually perform worse than relays, especially for cell edge users. This is mainly due to the fact that, despite the much lower delay introduced at the repeater, interference and noise are also amplified and sent over the BHL, thus degrading the throughput of users far from the base station. However, these are only preliminary results: other traffic models and users deployments are still to be considered, and may lead to different results.

3.5.6 Discussion

In this part of the work, we have tested through detailed simulations the effectiveness of relaying techniques in improving the uplink performance in CDMA cellular networks. The simplifying assumptions used in the theoretical analysis were removed, and a full multi-cell multi-layer scenario was entirely simulated. Results were obtained for different user distributions and carrier assignments, showing that the use of two carriers can lead to a significant increase of burst rate and to a considerable power saving. Relays appear to be extremely useful in some realistic scenarios, where a fraction of the users is concentrated in small indoor areas, and can hence be seen as a promising direction in the design of 4G cellular networks. A preliminary comparison with the performance offered by repeaters was also provided. As predicted by theoretical analysis, the use of A&F, though requiring much less complexity and allowing lower delays, usually achieves lower performance.

3.6 Circle Intersection Algorithm

In deriving the coverage range of a cooperative CDMA cellular network, where Relay Stations are deployed around the Base Stations, it has been found in Section 3.4 that the covered area depends on the positions of the RSs. Given the constraints on the mobile transmission power and on the QoS of the transmission, a relayed mobile must be located within a given distance from the RS in order to achieve the required transmission reliability. In addition, it was proved that the coverage range of each RS depends on how the mobiles are distributed among the relays.

If the capacity of the system, in terms of allocated users, is to be investigated, it may be necessary to calculate the probability of specific configurations (stations-mobiles associations), which in turn depend on the values of the areas covered by one or more Relay/Base Stations. Each configuration may in fact determine a different partition of the plane, and

under the approximation that the coverage areas have a circular shape, this leads to the problem of computing the intersection areas of a potentially huge number of circles.

Motivation: applications to Wireless Networks

Many frameworks for the evaluation, analysis and simulation of wireless communications are based on a signal propagation model where the attenuation incurred by a transmitted signal is a monotonically decreasing function of the distance from its source. Thus the performance of a receiver is a function of its distance from the wanted source, and this leads to a characterization of wireless networks based on the concept of *coverage range*.

The coverage range of a transmission can be defined in many different ways. For instance, once a target bit error rate (BER) is fixed, the coverage range is the maximum distance from the source which guarantees that a receiver is able to decode the wanted signal achieving the target BER on average.

In a scenario with multiple sources, a region covered by a set of circles corresponds to the positions in the plane where a receiver achieves the target performance when decoding the signals sent by the subset of sources associated to those circles.

The computation of the areas of the regions of the plane generated by the coverage ranges of various sources finds several applications in the area of wireless networking, ranging from routing to user allocation. An example of work studying the allocation of mobile users in wireless cellular networks where coverage areas are modeled as intersecting circles is [63].

However, the computation of the intersection areas is widely used in many other infrastructured network scenarios. For instance, in a heterogeneous network, the areas covered by different network infrastructures (GSM, UMTS, local area networks, and so on) may intersect. Thus, in order to allocate users and compute the average performance we need to know which are the areas covered by multiple technologies.

The distribution of the number of circles that cover a point in an area can be useful in many other applications. Localization in sensor networks relies on the reception of beacons sent by nodes whose positions are known. The accuracy achieved by the localization algorithm depends on the distribution of the number of beacon sources that the node can hear. This requires the computation of the probability that a node falls within an area covered by a certain number of circles. Furthermore, intersections of multiple circles are also found

when addressing the problem of preserving complete sensing coverage of a certain area and connectivity [64].

Another important example in which intersection areas are a fundamental aspect of analysis is routing [65]. The intersection of the circles may represent the area in which a user can provide connectivity to some nodes of the network (corresponding to the centers of the various circles). When considering geographic packet forwarding [66], intersection areas may be helpful to derive the distribution of the advancement and the success probability of the communication.

The ability to compute the area of regions of the plane generated by circles may also be useful to model the performance of a receiver in multiple concurrent access frameworks, where the distance from the wanted/interfering active sources can be quantized in order to define regions where a receiver achieves a certain performance.

Despite the wide range of applications of this geometric problem, it has been addressed in the literature only for three circles and only for some specific configurations [67], and a systematic approach is still lacking. In the following section, we state in detail the addressed geometric problem and the contribution of the work.

3.6.1 Introduction

The computation of the intersection area of many circles is a challenging problem. While the intersection of two circles is straightforward, even three circles admit several configurations, each resulting in a different expression for the intersection area. Given the centers and the radii of the circles, the automatic discrimination among the various cases requires involved condition testing. If we consider cases with several circles the problem may appear unsolvable, as the close-form expressions for the intersection areas become more and more involved and depend on the specific configuration among the huge number of possibilities.

In this part of the work, we present an algorithm that efficiently computes the intersections of an arbitrary number of circles. The algorithm works in an iterative fashion and is based on a trellis structure. In each iteration, the existence of any intersection is checked based on the areas computed in the previous steps, thus highly reducing the computational load. Moreover, we show that only the first three steps involve geometric considerations, whereas, when the number of circles is higher than three, all the areas can be found via simple algebraic calculations. Our proposal allows to efficiently solve configurations with many tens of circles, without any assumption on the centers and radii of the considered circles.

Problem Statement and Contribution

Consider a set $\mathcal{C} = \{\gamma_1, \gamma_2, \ldots, \gamma_{N_c}\}$ of N_c circles, whose centers and radii are known. The circles in $\mathcal{C}$ may partially overlap. We denote with

$$\boldsymbol{\mathcal{I}}^{(n)}=\{\mathcal{I}^{(n)}_{\{i_1,\ldots,i_n\}}, i_1,\ldots,i_n\in\{1,\ldots,N_c\}, i_j\neq i_u, \text{for} j\neq u\} \tag{3.60}$$

the set of all the possible intersection regions generated by n circles, where we denote $\mathcal{I}^{(n)}_{\{i_1,\ldots,i_n\}}=\bigcap_{i\in\{i_1,\ldots,i_n\}}\gamma_i$ as the set of the points that belong to all circles in $\{\gamma_{i_1},\gamma_{i_2},\ldots,\gamma_{i_n}\}\subseteq\mathcal{C}$. The set $\boldsymbol{\mathcal{I}}^{(1)}$ contains the circles in $\mathcal{C}$. We also define the notation $\mathcal{I}^{(n)}(i_1,\ldots,i_{N_c-n})$ to denote the intersection of n circles out of the N_c in $\mathcal{C}$ where circles $i_1,\ldots,i_{N_c-n}$ are not considered, i.e., $\mathcal{I}^{(n)}(i_1,\ldots,i_{N_c-n})=\bigcap_{i\in\{1,\ldots,N_c\}\setminus\{i_1,\ldots,i_{N_c-n}\}}\gamma_i$.

However, these intersections are not disjoint regions of the plane. See for instance Fig. 3.27, where a configuration with three circles is depicted. In the figure, $\mathcal{I}^{(3)}_{\{1,2,3\}}=\mathcal{A}_1$, $\mathcal{I}^{(2)}_{\{1,2\}}=\mathcal{A}_1\cup\mathcal{A}_4$, $\mathcal{I}^{(2)}_{\{1,3\}}=\mathcal{A}_1\cup\mathcal{A}_2$ and $\mathcal{I}^{(2)}_{\{2,3\}}=\mathcal{A}_1\cup\mathcal{A}_3$.

We call the regions $\mathcal{A}_i$ in the figure *exclusive intersection regions*, as they correspond to the intersection of a certain subset of circles, excluding the regions covered by the other circles in $\mathcal{C}$. We denote these regions as $\mathcal{E}^{(n)}(i_1,\ldots,i_{N_c-n})$, where

$$\mathcal{E}^{(n)}(i_1,\ldots,i_{N_c-n})=\mathcal{I}^{(n)}(i_1,\ldots,i_{N_c-n})\setminus\bigcup_{e\in\{i_1,\ldots,i_{N_c-n}\}}\gamma_e. \tag{3.61}$$

For instance, $\mathcal{E}^{(1)}(2,3)$ is the region of the plane covered by γ_1 and that does not overlap with any other circle of $\mathcal{C}$ ($\mathcal{A}_5$ in Fig. 3.27), and $\mathcal{E}^{(2)}(3)=\mathcal{A}_4$ is the intersection of γ_1 and γ_2, excluding the area covered by γ_3. We define the set $\boldsymbol{\mathcal{E}}^{(n)}$ as the set of all the exclusive intersection regions generated by n circles. Let us define the set $\boldsymbol{\mathcal{E}}=\bigcup_{n=1,\ldots,N_c}\boldsymbol{\mathcal{E}}^{(n)}$. Then, the elements of $\boldsymbol{\mathcal{E}}$ are disjoint regions that tessellate the overall region covered by $\mathcal{C}$. We also define the measure ϕ, where $\phi(\mathcal{A})$ corresponds to the area of the region $\mathcal{A}$.

In our work, we address the problem of computing the measure of the regions $\boldsymbol{\mathcal{E}}^{(n)}$, $n=1,\ldots,N_c$, given the centers and the radii of the circles in $\mathcal{C}$. This appears to be an extremely complex geometrical problem. In fact, while the area covered by the intersection of two circles has a simple measure, even the intersection area of three circles has a rather involved form, that depends on the mutual positions of the circles [67]. When more than three circles are considered, the number of configurations grows larger, and the complexity of the geometric conditions and the associated expression of the intersection area become difficult to handle.

The contribution of this work is as follows:

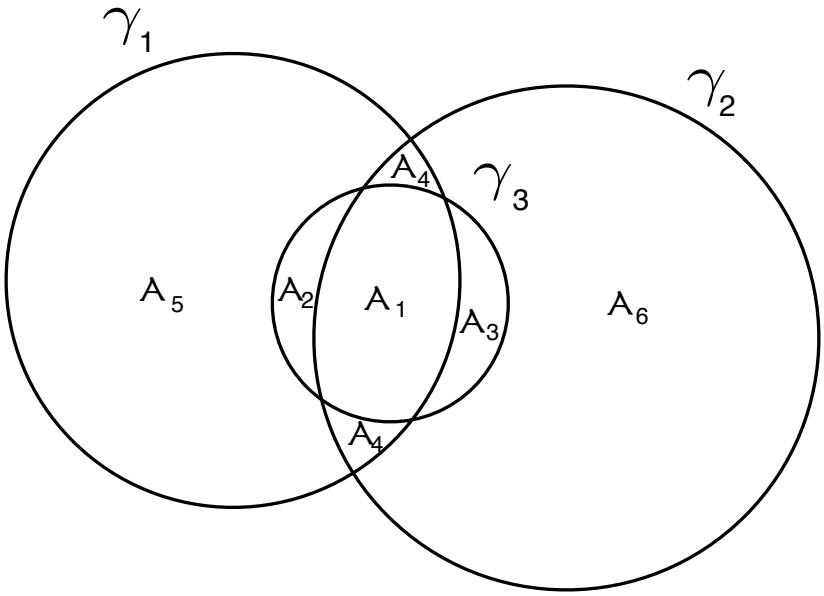


Figure 3.27. *Example of configuration with three circles and identification of the regions of interest.*

- we derive two theorems, that provide an easy way to check the existence and calculate the area of the intersection region of more than three circles, once the existence and the area of the intersections involving a smaller number of circles are known;

- we present a trellis-based iterative algorithm that allows an easy computation of the wanted areas even for configurations with a large number of circles

- as an example of application, we address a user allocation problem in code division multiple access (CDMA) cellular networks. In particular, we derive the probability that a certain number of mobile users can be allocated in a multi-cell environment.

3.6.2 Geometric results

In this section, we present the geometric results that represent the core of the proposed algorithm. A key observation is the following: a necessary (but in general not sufficient) condition for the existence of the intersection of n circles is the existence of the intersection of any subset of $n-1$ circles.[6] This consideration may be very useful, since the calculation of several areas among circles which are not all intersecting can be avoided. Nonetheless, a stronger result can be stated if the number of considered circles is greater than 3.

[6]We say that an intersection *exists* if it is non-empty.

Theorem 1. *Consider a subset of $n \leq N_c$ circles $\mathcal{S}=\{\gamma_{i_1}, \dots, \gamma_{i_n}\} \subseteq \mathcal{C}$ and the associated intersection region $\mathcal{I}^{(n)}_{\{i_1,\dots,i_n\}}$. With a slight abuse of notation, we refer to the considered subset of circles when denoting the intersection regions, and we write $\mathcal{I}^{(n-\ell)}_{\{i_1,\dots,i_n\}\setminus\{e_1,\dots,e_\ell\}} = \mathcal{I}^{(n-\ell)}(e_1, \dots, e_\ell)$.*

Then, $\forall \{i_1, \dots, i_n\}, i_j \neq i_u$, for $j \neq u$, if $n \geq 4$ the following holds: if

$$\phi(\mathcal{I}^{(n-1)}(k)) > 0 \;\; \forall k \in \{i_1, \dots, i_n\}, \tag{3.62}$$

then also $\phi(\mathcal{I}^{(n)}_{\{i_1,\dots,i_n\}}) > 0$.

In other words, a sufficient condition for the intersection among the n circles in $\mathcal{S}$ to exist is the existence of the n intersections among any subset of $n-1$ circles in $\mathcal{S}$

Proof. Consider a set of n circles $\{\gamma_{i_1}, \dots, \gamma_{i_n}\}$. By hypothesis $\phi(\mathcal{I}^{(n-1)}(k)) > 0, \forall k \in \{i_1, \dots, i_n\}$. Fix the index $k = \overline{k}$ and define the set $\mathcal{S}_{\overline{k}} = \{i_1, \dots, i_n\} \setminus \{\overline{k}\}$. As $\phi(\mathcal{I}^{(n-1)}(\overline{k})) > 0$, then $\mathcal{I}^{(n-1)}(\overline{k}) \neq \emptyset$. Call Δ the polygon[7] delimiting $\mathcal{I}^{(n-1)}(\overline{k})$ and whose sides are m arcs of circumference α_j, $j = 1, \dots, m$, with $1 \leq m \leq 2(n-2)$. Note that more than one arc α_j may belong to the same circle γ_i.[8] We denote with $\boldsymbol{\alpha}(i)$ the set of arcs belonging to the circle γ_i.

We have to consider three cases:

1. $\exists \gamma_i \in \mathcal{S}_{\overline{k}} : \; \boldsymbol{\alpha}(i) = \emptyset$;
2. $\forall \gamma_i \in \mathcal{S}_{\overline{k}} \; |\boldsymbol{\alpha}(i)| = 1$;
3. $\forall \gamma_i \in \mathcal{S}_{\overline{k}} \; \boldsymbol{\alpha}(i) \neq \emptyset$ and $\exists j : \; |\boldsymbol{\alpha}(j)| > 1$

In the first case γ_i fully contains the whole intersection but that its circumference does not hit it. Thus $\mathcal{I}^{(n-1)}(\overline{k}) = \mathcal{I}^{(n-2)}(\overline{k}, i) \cap \gamma_i = \mathcal{I}^{(n-2)}(\overline{k}, i)$. Since by assumption $\mathcal{I}^{(n-1)}(k) \neq \emptyset$, $\forall k \in \{i_1, \dots, i_n\}$, then also $\mathcal{I}^{(n-1)}(i) = \mathcal{I}^{(n-2)}(\overline{k}, i) \cap \gamma_{\overline{k}} \neq \emptyset$. Therefore, $\mathcal{I}^{(n-1)}(\overline{k}) \cap \gamma_{\overline{k}} \neq \emptyset$, and thus $\mathcal{I}^{(n)}_{\{i_1,\dots,i_n\}} \neq \emptyset$ and $\phi(\mathcal{I}^{(n)}_{\{i_1,\dots,i_n\}}) > 0$.

In the second case, the polygon Δ is delimited by exactly $n-1$ arcs of circumference, each belonging to one of the $n-1$ circles in $\mathcal{S}_{\overline{k}}$. This situation is depicted in Figure 3.28. Consider two non-consecutive arcs of Δ, namely α_t, α_r, and assume, without any loss of generality, that they belong to the circles γ_{i_1} and γ_{i_2}, with $\overline{k} \neq i_1, i_2$. By assumption, as the intersection of any combination of $n-1$ circles exists, then $\gamma_{\overline{k}}$ must contain at least one point $P \in \mathcal{I}^{(n-2)}(i_1, \overline{k})$ and one point $Q \in \mathcal{I}^{(n-2)}(i_2, \overline{k})$. Since a circle is a convex figure, then the whole segment joining P and Q must be contained in $\gamma_{\overline{k}}$. Moreover, P and Q both belong

[7] Δ is a particular polygon whose sides are arcs of circumference.

[8] We say that an arc belongs to a circle when it corresponds to a portion of its circumference

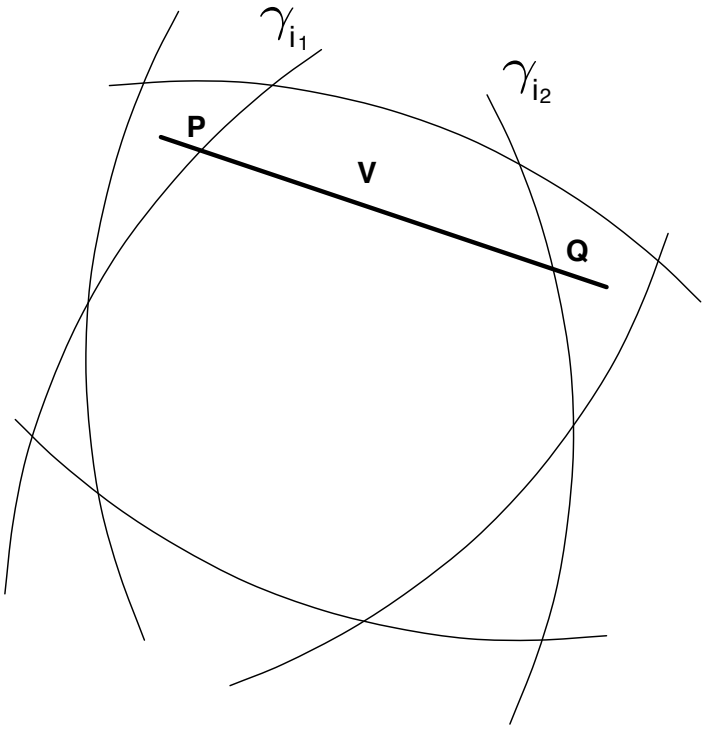


Figure 3.28. *Intersection of n circles. Here each circumference contains one arc belonging to the circular polygon Δ, which delimits the intersection.*

to $\mathcal{I}^{(n-3)}(i_1, i_2, \overline{k})$, which is also a convex set (since it is the intersection of convex sets), and therefore the segment $V{=}\overline{PQ}{\subset}\mathcal{I}^{(n-3)}(i_1, i_2, \overline{k})$. Since $P{\in}\gamma_{i_2}$ and $Q{\notin}\gamma_{i_2}$, then$V{\cap}\gamma_{i_2}{\neq}\emptyset$. It follows that $V{\subset}\mathcal{I}^{(n-2)}(i_1, \overline{k})$. Furthermore, since α_t and α_r are not consecutive sides of Δ, V intersects α_r. As a side of Δ, $\alpha_r{\subset}\mathcal{I}^{(n-1)}(\overline{k})$, then also $V{\subset}\mathcal{I}^{(n-1)}(\overline{k})$. Since $V{\subset}\gamma_{\overline{k}}$, we can conclude that $V{\subset}\mathcal{I}^{(n)}_{\{i_1,\ldots,i_n\}}$, which therefore is a non-empty set. Finally, we observe that the found set has non-zero measure by construction, and therefore $\phi(\mathcal{I}^{(n)}_{\{i_1,\ldots,i_n\}}){>}0$. Note that this proof does not hold for $n{=}4$, where all the arcs of Δ are consecutive. However, in this case the thesis can be proved in an analogous way.

In the third case, there is a circle γ_j whose associated set of arcs $\boldsymbol{\alpha}(j)$ contains at least two arcs of Δ. It can be shown that two consecutive arcs cannot belong to the same circle. Once observed this, the proof is analogous to that of the previous case by choosing α_t and α_r as non-consecutive arcs belonging to $\boldsymbol{\alpha}(j)$. □

The previous theorem represents a powerful tool for testing the existence of the intersection regions generated by n circles, once those generated by $n{-}1$ are known. In fact, the intersection of a set $\mathcal{S}$ of $n{\geq}4$ circles is non-empty if and only if all the intersections among

any subset of $n-1$ of $\mathcal{S}$ is non-empty. This simple principle substitutes involved and time-demanding geometric considerations.

The following theorem, provides an effective way to compute any intersection of n circles as a function of some specific intersections generated by $n-1$ and $n-2$ circles.

Theorem 2. *Consider a subset of n circles $\mathcal{S}=\{\gamma_{i_1},\dots,\gamma_{i_n}\}\subseteq\mathcal{C}$, with $4\leq n\leq N_c$, and the associated intersection region $\mathcal{I}^{(n)}_{\{i_1,\dots,i_n\}}$, with $\phi(\mathcal{I}^{(n)}_{\{i_1,\dots,i_n\}})>0$. Denote with m the number of arcs of circumference that delimit $\mathcal{I}^{(n)}$. Then, if $m\geq 4$, there exist two circles $\gamma_t,\gamma_r\in\mathcal{S}$ such that:*

$$\bigcap_{i\neq t,r}\gamma_i\subset\gamma_t\cup\gamma_r \tag{3.63}$$

Proof. Analogously to the proof of the previous theorem, we distinguish three cases:

1. $\exists\{\gamma_{u_1},\dots,\gamma_{u_w}\}\subseteq\mathcal{S}$:$\boldsymbol{\alpha}(u_1),\dots,\boldsymbol{\alpha}(u_w)=\emptyset$;
2. $\forall\gamma_i\in\mathcal{S}\ |\boldsymbol{\alpha}(i)|=1$;
3. $\forall\gamma_i\in\mathcal{S}\ \boldsymbol{\alpha}(i)\neq\emptyset$ and $\exists j:\ |\boldsymbol{\alpha}(j)|>1$.

In the first case, there exist w circles containing $\mathcal{I}^{(n)}_{\{i_1,\dots,i_n\}}$, with $1\leq w\leq n-4$. Define the subset $\mathcal{S}^{(n-w)}=\mathcal{S}\setminus\{\gamma_{u_1},\dots,\gamma_{u_{n-w}}\}\subset\mathcal{S}$ that contains the circles γ_j such that $\boldsymbol{\alpha}(j)\neq\emptyset$, and their intersection $\mathcal{I}^{(n-w)}(u_1,\dots,u_{n-w})$. Thus, there are $n-w$ circles with a non-empty set of arcs $\boldsymbol{\alpha}(j)$ and w circles that fully contain $\mathcal{I}^{(n)}_{\{i_1,\dots,i_n\}}$. This case can be referred to the second or third cases, where we consider only the circles with a non-empty set of arcs of Δ. We then refer the proof for this case to that of the following cases. We observe that if the theorem holds, in this case we have that $\exists\gamma_t,\gamma_r\in\mathcal{S}^{n-w}:\ \mathcal{I}^{(n-w-2)}(t,r)\subset\gamma_t\cup\gamma_r$. Therefore,

$$\mathcal{I}^{(n-2)}(r,t)\subset\mathcal{I}^{(n-w-2)}(r,t)\subset\gamma_r\cup\gamma_t, \tag{3.64}$$

In fact, $\mathcal{I}^{(n-2)}(r,t)$ is equal to the intersection between $\mathcal{I}^{(n-w-2)}(r,t)$ and $\bigcap_{i\in\mathcal{S}\setminus\mathcal{S}^{(n-w)}}\gamma_i$, and is thus a subset of $\mathcal{I}^{(n-w-2)}(r,t)$. That proves the theorem in this case. Note that $w\leq n-3$, in fact in order to have $m\geq 4$ arcs, we need Δ to be bounded by at least three circles.

In the second case, each circle γ_i has one arc belonging to Δ. Since $m\geq 4$, there exist two circles γ_t, $\gamma_r\in\mathcal{S}$ such that $\alpha_t\in\boldsymbol{\alpha}(t)$ and $\alpha_r\in\boldsymbol{\alpha}(r)$ are non-consecutive sides of Δ. Let us denote as P and Q the points of intersection of γ_t and γ_r. Note that the existence of P and Q is guaranteed by hypothesis, as $\mathcal{I}^{(n)}\neq\emptyset$. We also define α_t^r as the arc of Δ belonging to γ_t and fully contained in γ_r, and α_r^t as the arc of Δ belonging to γ_r and fully contained in γ_t.

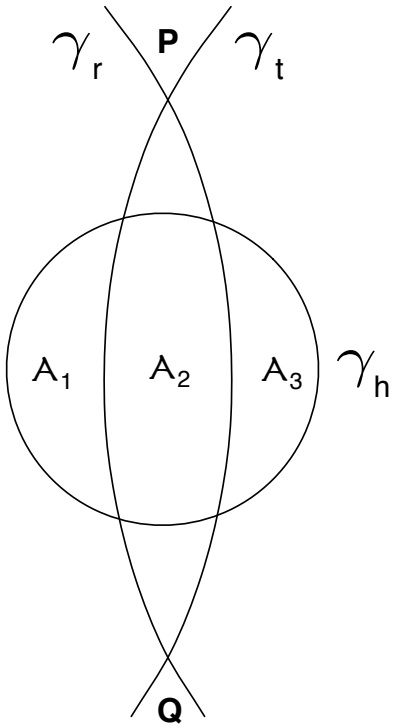


Figure 3.29. *First configuration of the two possible when $m{=}n$, and each of the n intersecting circles has exactly one arc of circumference belonging to the polygon Δ. In the figure $\mathcal{A}_2$ contains Δ.*

Moreover, there exists a circle $\gamma_h{\in}\mathcal{S}$ such that $P{\notin}\gamma_h$. In fact, if this circle did not exist, then $P{\in}\mathcal{I}^{(n)}$ and the arcs α_r and α_t would be consecutive arcs of Δ.

We need to distinguish two configurations for point Q:

- In the first configuration, Q does not belong to γ_h (see Fig. 3.29). Therefore, the intersection of the circumferences of γ_r and γ_h belongs to the arc α_r^t and the intersection of the circumferences of γ_t and γ_h belongs to α_t^r. As a consequence, γ_h is divided into three disjoint areas $\mathcal{A}_1{\subset}\gamma_r$, $\mathcal{A}_2{\subset}\gamma_r{\cap}\gamma_t$ and $\mathcal{A}_3{\subset}\gamma_r$. We thus conclude that $\gamma_h{\subset}\gamma_r{\cup}\gamma_t$. Since, $\mathcal{I}^{(n-2)}(r,t)$ is contained in γ_h, then it is also included in $\gamma_r{\cup}\gamma_t$.

- In the second configuration, Q belongs to γ_h (see Fig. 3.30). In this case there exists a γ_x such that $Q{\notin}\gamma_x$. If $P{\notin}\gamma_x$, the theorem can be proved as in the previous configuration. If $P{\in}\gamma_x$, then α_r^t and α_t^r belong to $\gamma_h{\cap}\gamma_x$, as they are sides of Δ and hence are part of $\mathcal{I}^{(n)}$ Thus, the points of intersection between the circumferences of γ_h and γ_x cannot belong to $\gamma_r{\cap}\gamma_t$. It follows that $\gamma_r{\cap}\gamma_t$ is divided into three regions: $\mathcal{A}_1{\subset}\gamma_h$, $\mathcal{A}_2{\subset}\gamma_h{\cap}\gamma_x$ and $\mathcal{A}_3{\subset}\gamma_x$. Therefore, $\gamma_r{\cap}\gamma_t{\subset}\gamma_h{\cup}\gamma_x$, and since $\mathcal{I}^{(n-2)}(h,x){\subset}\gamma_r{\cap}\gamma_t$ the thesis is proved.

In the third case, consider the circle γ_j with $|\boldsymbol{\alpha}(j)|{>}1$. In the proof of Theorem 1, we

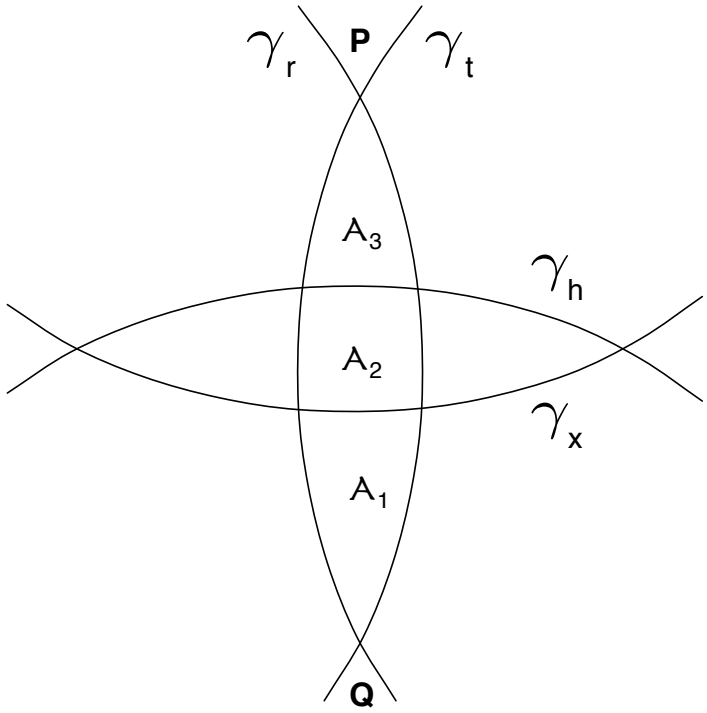


Figure 3.30. *Second configuration of the two possible when $m{=}n$, and each of the n intersecting circles has exactly one arc of circumference belonging to the polygon Δ.*

showed that the two arcs Δ adjacent to $\alpha_t{\in}\boldsymbol{\alpha}(j)$ must belong to the circumference of two different circles, say γ_h and γ_x. By construction, the arc α_t is contained in $\gamma_h{\cap}\gamma_x$. Moreover, also $\alpha_r{\in}\boldsymbol{\alpha}(j)$, with $\alpha_r{\neq}\alpha_t$, is contained in $\gamma_h{\cap}\gamma_x$, as α_r is an arc of Δ. Therefore, γ_j can be divided into three disjoint regions analogous to those of the previous case, namely $\mathcal{A}_1{\subset}\gamma_x$, $\mathcal{A}_2{\subset}\gamma_x{\cap}\gamma_h$ and $\mathcal{A}_3{\subset}\gamma_h$, which implies $\gamma_j{\subset}\gamma_x{\cup}\gamma_h$. Since $\mathcal{I}^{(n-2)}(x,h){\subset}\gamma_j$, the thesis in the third case is proved. □

Theorem 2 requires Δ to have a number of arcs greater than 3. If this does not hold, $\mathcal{I}^{(n)}$ collapses into the intersection of two or three circles, which can be obtained geometrically. If four or more circles are involved, Theorem 2 ensures that the area of $\mathcal{I}^{(n)}$ can be found by algebraic manipulation of the intersections among $n{-}1$ and $n{-}2$ circles.

In fact, if Theorem 2 holds, the intersection area $\mathcal{I}^{(n-2)}(t,r)$ among all circles but γ_t and γ_r is fully included in the union between γ_t and γ_r. The area of this intersection can be calculated by considering the partition of $\gamma_t{\cup}\gamma_r$ into three regions, one belonging only to γ_t, one to both γ_t and γ_r, and the third only to γ_r. All these three regions exist, the second one by hypothesis, the first and the third one because otherwise the required intersection $\mathcal{I}^{(n)}$ would be fully included in one circle, and the problem could be reduced to an analogous

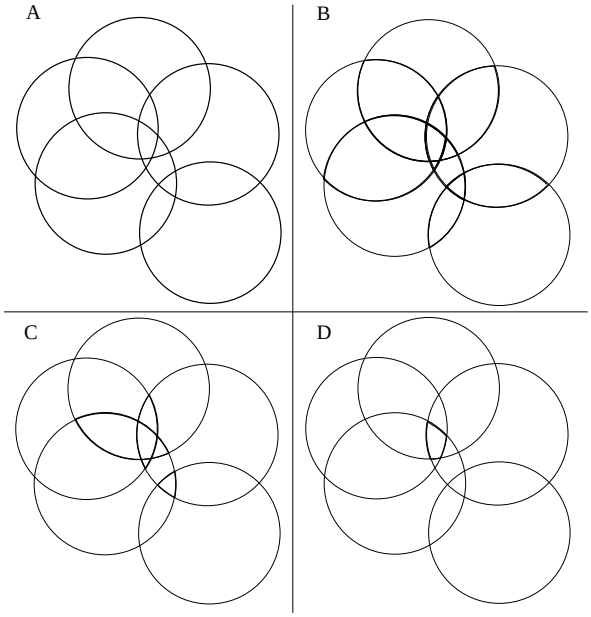


Figure 3.31. *Example of circles configuration and areas considered in the first four steps of the algorithm.*

one with a lower number of circles. Call these areas $\mathcal{B}_1$, $\mathcal{B}_2$ and $\mathcal{B}_3$. $\mathcal{I}^{(n-2)}(t,r)$ intersects $\mathcal{B}_2$ by hypothesis, and it may also intersect one area between $\mathcal{B}_1$ and $\mathcal{B}_3$ or even both. Assume that it intersects all the three areas, and call the intersections respectively $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$. It is clear that $\mathcal{A}_2 = \mathcal{I}^{(n)}$. These three areas are all unknown. However, it derives from their definition that $\mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{I}^{(n-1)}(r)$, $\mathcal{A}_2 \cup \mathcal{A}_3 = \mathcal{I}^{(n-1)}(t)$, and $\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 = \mathcal{I}^{(n-2)}(r,t)$. Then

$$\phi(\mathcal{A}_2) \;=\; (\phi(\mathcal{A}_1)+\phi(\mathcal{A}_2))+(\phi(\mathcal{A}_2)+\phi(\mathcal{A}_3))-(\phi(\mathcal{A}_1)+\phi(\mathcal{A}_2)+\phi(\mathcal{A}_3)) \tag{3.65}$$

and, therefore:

$$\phi\left(\mathcal{I}^{(n)}\right) \;=\; \phi\left(\mathcal{I}^{(n-1)}(t)\right)+\phi\left(\mathcal{I}^{(n-1)}(r)\right)-\phi\left(\mathcal{I}^{(n-2)}(t,r)\right), \tag{3.66}$$

It can be easily shown that even if $\mathcal{I}^{(n-2)}(t,r)$ does not interect $\mathcal{B}_1$, $\mathcal{B}_3$ or both, the results still holds.

3.6.3 The Algorithm

In this Section, we describe in detail the proposed algorithm. We remark that the number of areas to be calculated increases exponentially with the number of considered circles. The algorithm, exploiting the previously derived results, iteratively checks the existence and computes the intersection areas with a greatly reduced complexity. In fact, as soon as

the number of circles becomes greater than three, the existence-checks and the computation of the intersection areas in the following steps are based on simple algebraic calculations instead of extremely complex geometric considerations. Moreover, also the number of required operations is reduced thanks to a particular structure that allows the identification of which intersections exist and thus which areas actually need to be computed.

As stated before, the algorithm is based on a trellis structure, which is built iteration by iteration with a simple procedure. The output of the algorithm is a set of auxiliary variables, that are then used to compute the desired areas.

Trellis Structure

The algorithm is based on a trellis structure. In each iteration, the algorithm takes as input the structure and the variables built at the previous steps, in order to update the trellis.

Let us clarify how the algorithm works via a graphical example. Consider the circles in Fig. 3.31, where the first four steps of the algorithm are depicted in the subfigures A, B, C and D, respectively. In the first step, the algorithm simply calculates the areas of all the circles. In the second step, the algorithm checks the existence and computes the intersection areas of any pair of circles via simple geometric calculations. In the third step, the intersection areas of any existing intersection between any triplet of circles are computed. This is the last step involving geometric computations, as in the subsequent steps the existence and the measure of the intersection regions of any set of $4 \leq n \leq N_c$ circles are carried out via simple algebraic manipulation of the areas computed at the previous steps, exploiting the two theorems presented in Section 3.6.2.

In order to effectively keep track of the existence and value of the regions computed in each step we build a proper graph, in which the various intersection areas correspond to the vertices of the graph. In this way, the relationships can be represented as the edges of the graph. As a byproduct, the graph is also useful to list and order all the areas to be computed.

Consider a set $\mathcal{C}$ of N_c circles, labeled as γ_i, with $1 \leq i \leq N_c$. Our trellis structure is given by a set of vertices $\mathcal{V}$ and a set of edges $\mathcal{R}$. Each vertex corresponds to an intersection among some of the circles belonging to the set $\mathcal{C}$ considered. Hence, there is a one-to-one relationship among the vertices and all possible subsets of $\mathcal{C}$, except the empty set. We define as $\mathcal{S}(v) \subseteq \mathcal{C}$ the set of circles whose intersection corresponds to vertex v in the graph.

The vertices are divided into the subsets $\mathcal{V}(n)$, with $n = 1, 2, \ldots, N_c$. The subset $\mathcal{V}(n)$ is the

set of all the intersections among n circles out of the N_c considered (that is, the set of all the subsets of $\mathcal{C}$ with exactly n elements).

The obtained subsets can be ordered from $\mathcal{V}(1)$ to $\mathcal{V}(N_c)$. Each edge can connect only two vertices belonging to two consecutive subsets $\mathcal{V}(n)$ and $\mathcal{V}(n+1)$. An edge connecting $i \in \mathcal{V}(n)$ and $j \in \mathcal{V}(n+1)$ exists if and only if $\mathcal{S}(i) \subset \mathcal{S}(j)$. In other words, j is reachable from i if it corresponds to the intersection of all the circles associated with i plus another one.

Finally, also the elements of each $\mathcal{V}(n)$ can be ordered. Vertex i of the graph can be uniquely identified by a binary sequence of N_c elements $\mathbf{b}_i = [b_1^i, b_2^i, \ldots, b_{N_c}^i]$ such that $b_t^i = 1$ if $\gamma_t \in \mathcal{S}_i$. For instance, if $N_c = 5$, the vertex corresponding to the intersection of γ_2, γ_3 and γ_5 can be labeled as 01101. An equivalent labeling is obtained using the decimal representations of the binary sequences. In the example above, the same vertex is hence labeled as 13. The vertices of $\mathcal{V}(n)$ are then ordered with decreasing labels. An example of the full graph for $N_c = 5$ is also reported in Fig. 3.32.

The idea behind the trellis is straightforward: a vertex belonging to $\mathcal{V}(3)$, for instance, corresponds to the intersection of three circles. It follows that this area is related to the three intersection areas among any pair of the three considered circles (represented by three vertices belonging to $\mathcal{V}_2$), since it can be derived from each one of them by adding the missing circle. The exact way through which this can be done is explained below, and relies on Theorem 2.

With the ordering described above, the graph can be fully described by $N_c - 1$ transition matrices. We define the transition matrix $\mathbf{M}_{n,n+1}^{(N_c)}$, for $1 \leq n < N_c$, as a binary matrix containing the information about which edges exist between the vertices in $\mathcal{V}(n)$ and the ones in $\mathcal{V}(n+1)$. $\mathbf{M}_{n,n+1}^{(N_c)}(i,j) = 1$ if there exists an edge connecting the i–th vertex of $\mathcal{V}(n+1)$ and the j–th vertex of $\mathcal{V}(n)$, and 0 otherwise.

The computation of the transition matrices can be performed recursively, if we consider an additional subset $\mathcal{V}(0)$, which contains a single element corresponding to the empty set $\emptyset$, that is, to the region not covered by any circle. This "virtual" vertex is not included in the trellis in Fig. 3.32, though it can be inserted on the left side and connected to all the vertices belonging to $\mathcal{V}(1)$.

With this modification, the following properties about the transition matrices hold:

- $\mathbf{M}_{n,n+1}^{(N_c)}$ is a $p \times q$ matrix, where

$$p = \binom{N_c}{n+1}, \; q = \binom{N_c}{n} \tag{3.67}$$

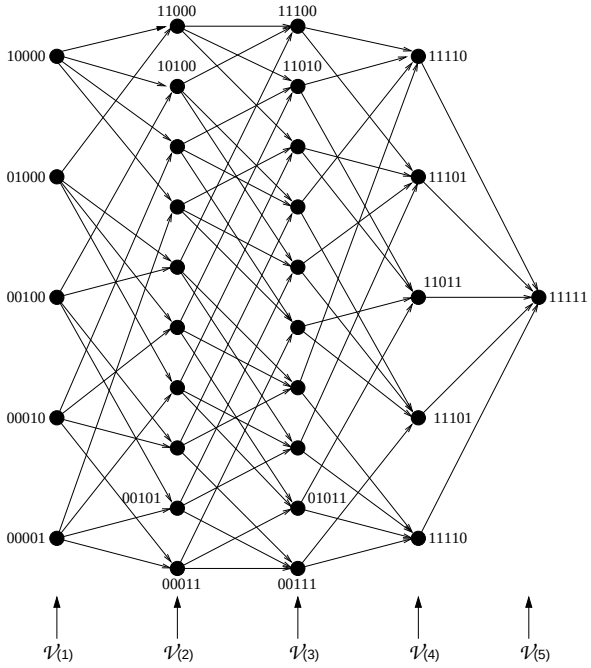


Figure 3.32. *The trellis structure for $n = 5$. Some of the binary labels are reported. The i–th column of vertices corresponds to the subset $\mathcal{V}_i$.*

- $\mathbf{M}_{0,1}^{(N_c)}$ is a $N_c \times 1$ column vector whose elements are all equal to 1.

- If $n \geq \lceil N_c/2 \rceil$, then $\mathbf{M}_{n,n+1}^{(N_c)} = \left(\mathbf{M}_{N_c-n-1,N_c-n}^{(N_c)}\right)^c$, where ^{c} indicates the transposition along the secondary diagonal, i.e., the element of the matrix of the i–th row and j–th column is moved to the $q-j+1$–th row and $p-i+1$–th column.

- Each transition matrix can be derived recursively as:

$$\mathbf{M}_{n,n+1}^{(N_c)} = \begin{bmatrix} \mathbf{M}_{n-1,n}^{(N_c-1)} & \mathbf{I} \\ \mathbf{0} & \mathbf{M}_{n,n+1}^{(N_c-1)} \end{bmatrix} \tag{3.68}$$

  where $\mathbf{I}$ is the identity matrix. We point out that, in any case, it must be $n < N_c$. However, for $n \geq \lceil N_c/2 \rceil$, due to the third property of the transition matrix, we can compute instead $\mathbf{M}_{N_c-n-1,N_c-n}^{(N_c)}$. The recursion appears in the first and in the fourth term of the matrix on the right side in Eq. (3.68). In the first term both N_c and n are reduced at each iteration, until we get to a matrix with $n = 0$, which can be derived using the second property of the transition matrices. As regards the fourth term, at each recursion only N_c is reduced. In this case, the recursion ends when $N_c - 1 = n + 1$. In fact, with the exception of the trivial case $n = 0$, the matrix $\mathbf{M}_{n,n+1}^{(N_c-1)}$ can be now

derived from $\mathbf{M}_{0,1}^{(N_c-1)}$, due to the third property, which is in turn computed according to the second property.

Different transition matrices can be multiplied together, thus giving information about which vertices of a set $\mathcal{V}(n)$ are connected to vertices belonging to $\mathcal{V}(n+k)$, with $k>1$. The elements of the transition matrices are binary variables, and we define the transition matrix between $\mathcal{V}(n)$ and $\mathcal{V}(n+k)$ as

$$\mathbf{M}_{n,n+k}^{(N_c)} = \frac{1}{k!} \prod_{i=0}^{k-1} \mathbf{M}_{n+k-i-1,n+k-i}^{(N_c)} \tag{3.69}$$

Auxiliary Variables

As stated before, the algorithm at each step n finds the intersection areas represented by the vertices belonging to $\mathcal{V}(n)$. In order to obtain these areas, it exploits the areas belonging to $\mathcal{V}(n-1)$, found in the previous step, and applies Theorem 2 based on the relationships expressed by the transition matrix $\mathbf{M}_{n-1,n}^{(N_c)}$.[9] During the subsequent steps, however, some information must be collected and stored. Before describing the steps in detail, we list the auxiliary variables which are computed at each step. At the end of the algorithm, they are used to retrieve the required intersection areas.

- We define the n–th label vector $\mathbf{L}_n$ as the vector containing the decimal labels of all the elements of the set $\mathcal{V}(n)$. Therefore, $\mathbf{L}_n$ has length $\binom{N_c}{n}$, and its elements are sorted in decreasing order. These vectors are computed using the transition matrices, since $\mathbf{L}_1 = [2^{N_c-1}, 2^{N_c-2}, \ldots, 2, 1]^T$, and

$$\mathbf{L}_n = \frac{1}{n-1} \mathbf{M}_{n-1,n} \mathbf{L}_{n-1} \tag{3.70}$$

- The n–th existence vector $\mathbf{E}_n$ is a binary vector containing information on the intersection areas of n circles; therefore, its size is equal to the cardinality of $\mathcal{V}(n)$. The i–th element of $\mathbf{E}_n$ is equal to 1 if the area of the intersection among the circles associated with the i–th vertex in $\mathcal{V}(n)$ is not empty. Assuming that all the N_c circles belonging to $\mathcal{C}$ have positive radius, then $\mathbf{E}_1$ is an $N_c \times 1$ column vector whose elements are all equal to 1. The other vectors can be recursively computed with the following rule:

$$\mathbf{E}_{n+1} = \max\left(\mathbf{M}_{n,n+1}\mathbf{E}_n - n\,\mathbf{1},\, \mathbf{0}\right) \tag{3.71}$$

[9] In the following we drop the superscript of the transition matrices for notation clarity.

where **1** and **0** are here column vectors of length $\binom{N_c}{n+1}$, with all elements equal to 1 and 0 respectively, and the maximum is taken element-wise. Equation (3.71) can be explained as follows. The intersection area $\mathcal{A}_w$ of $n+1$ circles, represented by vertex $w \in \mathcal{V}(n+1)$, can exist only if all the intersection areas among any n of those circles also exist. These areas are in turn represented by $n+1$ vertices belonging to $\mathcal{V}(n)$, which are all connected to w in the graph. Assume that w is the i–th vertex of $\mathcal{V}(n+1)$: the existence of $\mathcal{A}_w$ is indicated by $\mathbf{E}_{n+1}(i)$. Note now that the multiplication of the i–th row of $\mathbf{M}_{n,n+1}$ and $\mathbf{E}_n$ is equal to the number of existing intersections among n circles out of the $n+1$ associated with w. Therefore, in order for $\mathcal{A}_w$ to exist, this product must be equal to $n+1$. This holds for each element of $\mathbf{E}_{n+1}$, which is then reduced to a binary vector by subtracting n and nulling the negative elements. When we are considering set of up to two circles we need to do an additional check. In particular, This means that even when $\mathbf{E}_{n+1}(i)=1$, the corresponding area may not exist, and the existence vector must therefore be modified accordingly.

- We define the n–th area vector $\mathbf{A}_n$ as the vector containing the values of all the intersection areas among n circles. As $\mathbf{L}_n$ and $\mathbf{E}_n$, it is a column vector of $\binom{N_c}{n}$ elements. The values corresponding to non existing areas are equal to 0. We remark that the outcome of these calculations are all the regions $\boldsymbol{\mathcal{I}}^{(n)}$, $n=1,\ldots,N_c$, while our goal is to compute the exclusive intersection regions, i.e., $\boldsymbol{\mathcal{E}}^{(n)}$. We will show how to compute those areas later.

Besides those vectors, the algorithm also keeps track of the vectors $\mathbf{r}$, $\mathbf{x}_c$ and $\mathbf{y}_c$, containing the radii and the coordinates of the centers of the n circles. Finally, the symmetric matrix $\mathbf{D}$ contains the distances among the centers: $\mathbf{D}(i,j)$ is the distance between the centers of the circles γ_i and γ_j.

3.6.4 Computation of the Intersection Areas

Based on the trellis structure and the auxiliary variables introduced above, we now illustrate how the algorithm works.

In the inizialization phase, the auxiliary variables $\mathbf{L}_1$, $\mathbf{E}_1$ and $\mathbf{A}_1$ are calculated as stated in Section 3.6.3. Also the transition matrices are recursively derived. We assume than all the N_c circles involved have finite and positive radius, and thus the measure of the region covered by each circle is strictly positive. At each step n, the aim of the algorithm is to calculate

and store the corresponding area vector $\mathbf{A}_n$. To this purpose, it performs the following operations:

- generation of $\mathbf{L}_n$ from $\mathbf{L}_{n-1}$. This variable is useful, since the binary labels, which can be easily obtained from the decimal ones, provide an effective way to recognize which vertices of $\mathcal{V}(n-1)$ are connected to each vertex of $\mathcal{V}(n)$.

- generation of $\mathbf{E}_n$ from $\mathbf{E}_{n-1}$ using the transition matrix $\mathbf{M}_{n-1,n}$. The number and positions of the nonzero elements of $\mathbf{E}_n$ provide information on which intersection areas must be calculated. For $n \leq 3$, however, a geometric check is necessary for each of these elements, since a nonzero value does not necessarily mean that the intersection represented by the corresponding vertex exists. The check is done based on the centers and the radii of the involved circles (identified through the label vector $\mathbf{L}_n$), and it is then possible to obtain $\mathbf{E}_n$. No checks are needed for $n > 3$, thanks to Theorem 1;

- calculation of $\mathbf{A}_n$. This calculation is more involved: if $n \leq 3$, the existing areas, according to the information given by $\mathbf{E}_n$, can be calculated geometrically, applying known formulas with the centers and the radii of the intersecting circles, again retrieved through $\mathbf{L}_n$. If on the contrary $n > 3$, then we rely on Theorem 2. The calculation is done element by element. We focus at first on the case $n = N_c$, meaning that $\mathbf{A}_n$ is a scalar. By assumption, all the $\mathbf{E}_n$, for $n \leq N_c$, and all the $\mathbf{A}_n$, for $n < N_c$, are known.

We remark that the elements of the vectors $\mathbf{A}_n$ correspond to the intersection regions $\boldsymbol{\mathcal{I}}^{(n)}$. We define the vectors $\hat{\mathbf{A}}_n$, for $1 \leq n \leq N_c - 1$ whose elements correspond to the exclusive intersection regions $\boldsymbol{\mathcal{E}}^{(n)}$ as

$$\hat{\mathbf{A}}_n = \begin{cases} \mathbf{A}_n - \sum_{j=n+1}^{N_c-1} (-1)^{j-n+1} \mathbf{M}_{n,j}^T \mathbf{A}_j & n < N_c - 1 \\ \mathbf{A}_n & n = N_c - 1 \end{cases} \tag{3.72}$$

This vector would contain the exact values of the intersection areas if the sums were up to N_c. This is not possible, since $\mathbf{A}_{N_c}$ is not known. As explained above, if the hypothesis $m > 3$ of Theorem 2 holds, then it is possible to compute this area starting from the ones already derived. Unfortunately, the Theorem cannot be applied directly, since it is not known a priori which two circles γ_r and γ_t satisfy the statement of the Theorem. Moreover, since in equation (3.72), that is used to compute the areas of the exclusive intersections, the summation is up to $N_c - 1$, we do not have the exact values of those areas yet. However, $\mathbf{A}_{N_c}$ can be still

retrieved. More precisely, it can be shown that if $N_c > 4$, then $\mathbf{A}_{N_c}$ is given by the maximum value of $-\hat{\mathbf{A}}_{N_c-2}$. This holds also for $N_c = 4$, unless the minimum values of $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{A}}_3$ are equal, and greater than the maximum value of $-\hat{\mathbf{A}}_2$. In this special case, the value of $\mathbf{A}_{N_c}$ may be instead equal to the minimum value of $\hat{\mathbf{A}}_{N_c-1}$, depending on the mutual positions of the circles. In this special case, an additional geometric check is necessary.

A formal proof of this result is reported below.

The case $n<N_c$ can be handled in a similar manner. If $n<N_c$, each element of $\mathbf{A}_n$, corresponding to the intersection of exactly n circles out of the N_c considered, can be retrieved by simply neglecting the circles that are not involved (which can be done by using either the label vectors or the transition matrices to null all the elements of the existence vectors and of the area vectors that are not of interest). Actually, this is the same as running a reduced version of the algorithm with only n circles, where $\mathbf{A}_n(p)$ plays exactly the same role as $\mathbf{A}_{N_c}$ in the case $n=N_c$.

A more detailed description of how the algorithm works and guidelines for its implementation can be found in [68].

3.6.5 Proof of algorithm correctness

We prove here that, if the vectors $\mathbf{A}_i$ and $\hat{\mathbf{A}}_i$ are known, for $1 \leq i < N_c$, then $\mathbf{A}_{N_c}$, which is a scalar, can be found as the maximum value of $-\hat{\mathbf{A}}_{N_c-2}$. This holds for $N_c > 4$, and very often also for $N_c = 4$ where, however, in some special cases the required value is instead equal to the minimum value of $\hat{\mathbf{A}}_{N_c-1}$. Finally, for $N_c \leq 3$, all the areas can be found geometrically.

We recall that the vectors $\mathbf{A}_i$ contain the non exclusive intersection areas, whereas the vectors $\hat{\mathbf{A}}_i$ can be obtained using (3.72). The proof is structured as follows:

- we first determine an expression for the elements of the vectors $\hat{\mathbf{A}}_i$. We focus on the first element of $\hat{\mathbf{A}}_1$, since all the other ones can be retrieved in an analogous manner. We express them as sums of exclusive intersection areas.
- using the expression of $\hat{\mathbf{A}}_i$ and applying Theorem 2, we prove the statement for $N > 4$.
- we point out in which cases the statement is not true for $N_c = 4$, and determine how the correct value of $\hat{\mathbf{A}}_{N_c}$ can be found in these special cases.

In the previous Section, it has been stated that the two geometric Theorems 1 and 2 cannot be applied directly in the proposed algorithm. This is because when the algorithm

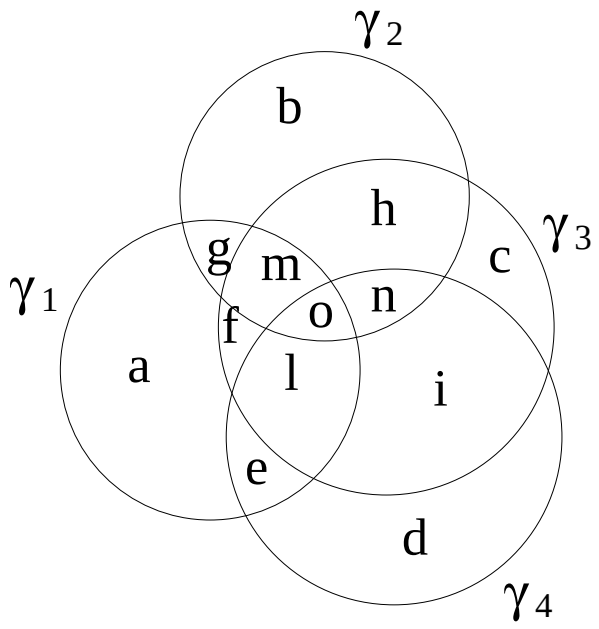


Figure 3.33. *Example of disposition of 4 circles.*

is executed, it is not known a priori how the circles intersect with each other (which would require an exponentially complex conditions check). The only available data at the i-th step are the numerical values of the areas computed in the previous $i-1$ steps. More precisely, at step N_c, the vectors $\mathbf{A}_i$ for $1 \leq i < N_c$ are available. However, the elements of these vectors are not the exclusive intersection areas, as explained before. Vectors $\hat{\mathbf{A}}_i$, for $1 \leq i < N_c$, can also be computed, according to (3.72), but also these vectors do not contain the values of the exclusive intersection areas, since $\mathbf{A}_{N_c}$ is not known, and consequently the sums in (3.72) are up to $N_c - 1$.

Therefore, it is not straightforward how the value of $\mathbf{A}_{N_c}$ can be obtained starting from the vectors $\mathbf{A}_i$ and $\hat{\mathbf{A}}_i$, for $1 \leq i < N_c$. We focus on $\mathbf{A}_{N_c}$ since, in order to compute each element of $\mathbf{A}_i$, for $3 < i < N_c$, it is sufficient to run the algorithm while considering only i circles. In this reduced version of the algorithm, $\mathbf{A}_i$ has only one element, and its role is exactly the same as $\mathbf{A}_{N_c}$ in the non reduced version of the algorithm. Finally, the elements of $\mathbf{A}_i$ for $i \leq 3$ can be derived via geometric computation.

The unknown value $\mathbf{A}_{N_c}$ is the intersection area of all the N_c circles. In order to explain how this value can be retrieved, we first clarify which are the areas corresponding to the elements of the known vectors. Each element of $\mathbf{A}_i$ is the non exclusive intersection area of i circles. In other words, it is the value of the area whose points belong to all the i considered circles, but which may be included also in other circles. In order to distinguish them, we

call $\mu_{i_1,i_2,...,i_k}$ the element of $\mathbf{A}_k$ corresponding to the non exclusive intersection of circles γ_{i_1}, γ_{i_2}..., γ_{i_k}. In Figure 3.33, we depict an example of deployment of four circles. In this scenario, for example, $\mu_{1,2}$ is given by $g + m + o$, whereas $\mu_{1,3,4}$ is equal to $l + o$. We simply call μ the intersection of all the N_c circles, which is the only element of $\mathbf{A}_{N_c}$, and, thus, the value we want to find. Analogously, we call $\mu^*_{i_1,i_2,...,i_k}$ the exclusive intersection of circles γ_{i_1}, γ_{i_2}..., γ_{i_k}. With reference to Figure 3.33, we have $\mu^*_{1,2} = g$, and $\mu^*_{1,3,4} = l$. Note that these values are not necessarily contained in any of the vectors $\mathbf{A}_i$ or $\hat{\mathbf{A}}_i$.

We now want to express the elements of the vectors $\hat{\mathbf{A}}_i$, for $1 \leq i < N_c$ as a function of the exclusive intersection areas, and we start with $\hat{\mathbf{A}}_1$, which has exactly N_c elements. The rationale behind equation (3.72) is the following. Since $\mathbf{A}_1$ contains the whole areas of all the circles, in order to find the areas covered by exactly one circle we need to subtract the fractions of these areas which are shared with other circles. We first subtract the areas shared by two circles, which are contained in $\mathbf{A}_2$. In doing so, since $\mathbf{A}_2$ does not contain the exclusive intersection areas, we are subtracting the areas shared by three circles more than once. Hence, we need to sum them again. They are contained in $\mathbf{A}_3$. For the same reason, we then need to subtract again the areas shared by four circles, contained in $\mathbf{A}_4$, and so on and so forth. If we had the intersection area of all the N_c circles, at the end of this sum we would get the exclusive intersection areas of one circle, that is, with a slight abuse of expression, the areas covered by exactly one circle.

For the sake of clarity, we recall again the example of Figure 3.33. The first element of $\mathbf{A}_1$ contains the area of γ_1, which we call μ_1. According to equation (3.72), the first element of $\hat{\mathbf{A}}_1$ is computed as:

$$\begin{aligned}
\hat{\mathbf{A}}_1(1) &= \mu_1 - \mu_{1,2} - \mu_{1,3} - \mu_{1,4} + \mu_{1,2,3} + \mu_{1,2,4} + \mu_{1,3,4} \\
&= (a + e + f + g + l + m + o) - (g + m + o) - (f + l + m + o) - (e + l + o) + \\
&\quad +(m + o) + (o) + (l + o) \\
&= a + o
\end{aligned} \tag{3.73}$$

As can be observed, the result is not the area covered only by circle γ_1, which would be instead if $\mathbf{A}_4 = o$ was subtracted by the result. Since $\mathbf{A}_4$ is the area we are looking for, it is necessary to find a way to determine its value.

We first derive an expression for the elements of the vectors $\hat{\mathbf{A}}_k$. It is clear from (3.72) that the number of elements of $\hat{\mathbf{A}}_k$ is equal to the number of elements of $\mathbf{A}_k$ for each k. Since the derivation is analogous for every value of k, we focus on $\hat{\mathbf{A}}_1$.

The first element of this vector is obtained as:

$$\begin{aligned}\hat{\mathbf{A}}_1(1) &= \mu_1 + \sum_{n=2}^{N_c-1} (-1)^{n-1} \sum_{1<i_1<1_2<\ldots<i_{n-1}\leq N_c} \mu_{1,i_1,i_2,\ldots,i_{n-1}} \\ &= \sum_{n=1}^{N_c-1} (-1)^{n-1} \lambda_n \end{aligned} \tag{3.74}$$

where we define

$$\begin{aligned}\lambda_1 &= \mu_1; \\ \lambda_n &= \sum_{1<i_1<i_2<\ldots<i_{n-1}\leq N_c} \mu_{1,i_1,i_2,\ldots,i_{n-1}} \end{aligned} \tag{3.75}$$

Since we are computing the first element of $\mathbf{A}_1$, we are focusing on circle γ_1. The first term of the sum is the area of the circle. The second one, λ_2, is the sum of all the non exclusive intersection areas between γ_1 and another circle, λ_3 is the sum of all the non exclusive intersections between γ_1 and two other circles, and so on. These areas are not disjoint, since they are the non exclusive intersections. We want to rewrite them in terms of disjoint areas, in order to simplify the expression in (3.74). Define the following sums of disjoint areas:

$$\begin{aligned}\lambda_1^* &= \mu_1^* \\ \lambda_n^* &= \sum_{1<i_1<i_2<\ldots<i_{n-1}\leq N_c} \mu^*_{1,i_1,i_2,\ldots,i_{n-1}} \end{aligned} \tag{3.76}$$

According to the above definitions, λ_1^* is the area covered only by γ_1, λ_2^* is the sum of all the areas covered only by γ_1 and another circle, that is, the sum of all the exclusive intersection areas between γ_1 and another circle. With reference to Figure 3.33, we can write for instance:

$$\begin{aligned}\lambda_1 &= a+e+f+g+l+m+o, \\ \lambda_1^* &= a, \\ \lambda_2 &= e+f+g+2l+2m+3o, \\ \lambda_2^* &= e+f+g. \end{aligned}$$

We can now express the non exclusive intersection areas, which appear in (3.74), as functions of the exclusive ones, by writing relationships between the λ_i's and the λ_i^*s. In order to do so, we note the following facts:

- λ_1 is the area of γ_1, and is hence given by the sum of all the exclusive intersection areas involving γ_1, that is:

$$\lambda_1 = \sum_{i=1}^{N_c} \lambda_i^* \tag{3.77}$$

- λ_2 is the sum of all the non exclusive intersection areas between γ_1 and another circles. In doing this sum, the areas covered by more than two circles (one of which is, by definition, γ_1), are counted more than once. More precisely, the areas covered by n circles are counted $n-1$ times, and hence:

$$\lambda_2 = \sum_{i=2}^{N_c} (i-1)\, \lambda_i^* \tag{3.78}$$

- in general, when computing λ_k, the areas covered by $n \geq k$ circles are counted as many times as the number of possible extractions of $k-1$ circles out of $n-1$. On the contrary, the areas covered by $n < k$ circles are never counted. Therefore:

$$\lambda_k = \sum_{n=k}^{N_c} \binom{n-1}{k-1} \lambda_n^* \tag{3.79}$$

We can order the expressions of the λ_i in an $N_c - 1 \times N_c$ matrix $\mathbf{L}$, whose i–th row contains the terms of the sum defining λ_i. The sign of the even rows is then changed, to get:

$$\mathbf{L} = \begin{bmatrix} \binom{0}{0}\lambda_1^* & \binom{1}{0}\lambda_2^* & \binom{2}{0}\lambda_3^* & \binom{3}{0}\lambda_4^* & \dots & \binom{N_c-1}{0}\lambda_{N_c}^* \\ 0 & -\binom{1}{1}\lambda_2^* & -\binom{2}{1}\lambda_3^* & -\binom{3}{1}\lambda_4^* & \dots & -\binom{N_c-1}{1}\lambda_{N_c}^* \\ 0 & 0 & \binom{2}{2}\lambda_3^* & \binom{3}{2}\lambda_4^* & \dots & \binom{N_c-1}{2}\lambda_{N_c}^* \\ 0 & 0 & 0 & -\binom{3}{3}\lambda_4^* & \dots & -\binom{N_c-1}{3}\lambda_{N_c}^* \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & (-1)^{N_c}\binom{N_c-1}{N_c-2}\lambda_{N_c}^* \end{bmatrix} \tag{3.80}$$

With this representation, according to (3.74), the first element of $\hat{\mathbf{A}}_1$ is simply given by the sum of all the elements of $\mathbf{L}$. However, while in (3.74) the sum is computed row by row, we now observe that it is simpler to consider the columns. The sum of the elements of the first column is clearly equal to λ_1^*. For $2 \leq n \leq N_c - 1$, the sum s_n of the elements of column n is expressed as:

$$\begin{aligned} s_n &= \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \lambda_n^* \\ &= \lambda_n^* (1-1)^{n-1} = 0 \end{aligned} \tag{3.81}$$

The sum of all the columns is hence 0, with the exception of the last one, due to the fact that the sum in (3.72) is up to $N_c - 1$. The calculation of s_{N_c} is straightforward:

$$\begin{aligned} s_{N_c} &= \sum_{k=0}^{N_c-2} (-1)^k \binom{N_c-1}{k} \lambda^*_{N_c} \\ &= \sum_{k=0}^{N_c-1} (-1)^k \binom{N_c-1}{k} \lambda^*_{N_c} - (-1)^{N_c-1} \lambda^*_{N_c} \\ &= (-1)^{N_c} \lambda^*_{N_c} \end{aligned} \tag{3.82}$$

Putting all together, we have that

$$\begin{aligned} \hat{\mathbf{A}}_1(1) &= \sum_{n=1}^{N_c} s_n \\ &= s_1 + s_{N_c} \\ &= \lambda^*_1 + (-1)^{N_c} \lambda^*_{N_c} \\ &= \mu^*_1 + (-1)^{N_c} \mu^*_{1,2,3,\ldots,N_c} \\ &= \mu^*_1 + (-1)^{N_c} \mu \end{aligned} \tag{3.83}$$

where μ, as defined above, is the intersection area of all the N_c circles. It is clear that the same derivation can be done for all the elements of $\hat{\mathbf{A}}_1$, such that $\hat{\mathbf{A}}_1(k) = \mu^*_k + (-1)^{N_c}\mu$. It can be shown also that an analogous procedure can be followed to determine the elements of the other vectors $\hat{\mathbf{A}}_n$, with $2 \leq n < N_c$. In this case, for each element $\hat{\mathbf{A}}_n(k)$ the sums λ_i and λ^*_i can be defined in the same manner, for $n \leq i \leq N_c$, and an $N_c - n - 1 \times N_c - n$ matrix can be constructed, finally resulting in

$$\hat{\mathbf{A}}_n(k) = \mu^*_{i_1,i_2,\ldots,i_n} + (-1)^{N_c-n+1}\mu \tag{3.84}$$

In the expression above, the indices $i_1, i_2, \ldots, i_n$ depend on the element which is being calculated. Reporting again the example in Figure 3.33, we have:

$$\hat{\mathbf{A}}_1 = \begin{bmatrix} \mu^*_1 + \mu \\ \mu^*_2 + \mu \\ \mu^*_3 + \mu \\ \mu^*_4 + \mu \end{bmatrix}, \; \hat{\mathbf{A}}_2 = \begin{bmatrix} \mu^*_{1,2} - \mu \\ \mu^*_{1,3} - \mu \\ \mu^*_{1,4} - \mu \\ \mu^*_{2,3} - \mu \\ \mu^*_{2,4} - \mu \\ \mu^*_{3,4} - \mu \end{bmatrix}, \; \hat{\mathbf{A}}_3 = \begin{bmatrix} \mu^*_{1,2,3} + \mu \\ \mu^*_{1,2,4} + \mu \\ \mu^*_{1,3,4} + \mu \\ \mu^*_{2,3,4} + \mu \end{bmatrix} \tag{3.85}$$

The vectors above can be computed by the algorithm at each step. We stress again that the area μ, as well as all the exclusive intersection areas, are unknown, and cannot in general

be retrieved from the vectors $\hat{\mathbf{A}}_i$. This is true for $N_c > 3$, as stated before, since otherwise all the intersection areas can be found geometrically. We skip for now the special case $N_c = 4$. For $N_c \geq 5$, we can now exploit Theorem 2. The theorem holds only if the number m of circular arcs that delimit the intersection area of all the circles (in this case equal to μ) is greater or equal to 4.

If this is not true, then there is a circle γ_k that fully contains the intersection area of all the other $N_c - 1$ circles. This area also belongs to γ_k, which implies that there is at least one exclusive intersection area among $N_c - 1$ circles which is empty. Looking at the example for $N_c = 4$, whose vectors are reported in (3.85), this means that one of the values $\mu^*_{i,j,k}$ in $\mathbf{A}_3$ is zero. As a consequence, in order to retrieve μ it is sufficient in general to take the minimum element of $\hat{\mathbf{A}}_{N_c-1}$.

If on the contrary the hypothesis of Theorem 2 holds, then it means that there are two circles γ_j and γ_k that fully contain the intersection area of all the other $N_c - 2$ circles. Following the same reasoning as above, it can be concluded that there exists at least one exclusive intersection area among $N_c - 2$ circles which is empty, and the value of μ is the maximum of $-\hat{\mathbf{A}}_{N_c-2}$.

Unfortunately, this is not enough, since it is not known a priori if the hypothesis of Theorem 2 holds or not. If it does not, however, we can use the following argument. If $m = p$, with $1 \leq p \leq 3$, there are p circles whose intersection is fully included in any other circle (for $p = 1$, there is one circle which is contained in any other circle). We call $\mathcal{P}$ the set of these circles, whereas $\mathcal{Q}$ is the set of the remaining $N_c - p$ circles. Consider $N_c - p - 2$ other circles belonging to $\mathcal{Q}$. The intersection of these circles and the ones belonging to $\mathcal{P}$ is fully contained in the remaining two circles of $\mathcal{Q}$. As a consequence, at least one of the exclusive intersection areas among $N_c - 2$ circles is empty, and again μ can be obtained as the maximum of $-\hat{\mathbf{A}}_{N_c-2}$, as in the case where the hypothesis of Theorem 2 holds. The abovementioned consideration clearly requires that N_c is at least equal to 5, otherwise, if $p = 3$, it is not possible to take $N_c - p - 2$ circles from $\mathcal{Q}$ [10].

The only case that must be studied separately is $N_c = 4$. The available vectors in this case are $\hat{\mathbf{A}}_1$, $\hat{\mathbf{A}}_2$ and $\hat{\mathbf{A}}_3$, whose expressions have been reported in (3.85). It is useful to consider $\hat{\mathbf{A}}^*_2 = -\hat{\mathbf{A}}_2$. In this manner, all the elements of $\hat{\mathbf{A}}^*_2$ are smaller or equal to μ, whereas all the

[10] if $N_c = 5$ and $p = 3$, the intersection of the circles belonging to $\mathcal{P}$ is fully included in the two circles belonging to $\mathcal{Q}$, and the same reasoning still holds.

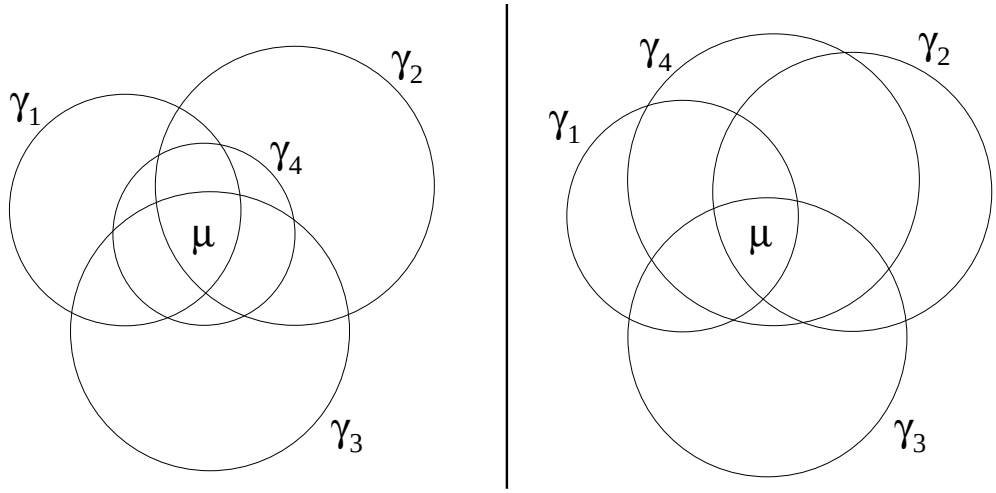


Figure 3.34. *Possible deployments of 4 circles when m is equal to 3. On the left, γ_4 is included in the union of the other three circles, whereas on the right this is not true.*

elements of $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{A}}_3$ are greater than or equal to μ. Therefore, we define

$$\begin{aligned} a_\gamma &= \min\left(\hat{\mathbf{A}}_1\right) \\ b_\gamma &= \max\left(\hat{\mathbf{A}}_2^*\right) \\ c_\gamma &= \min\left(\hat{\mathbf{A}}_3\right) \end{aligned} \tag{3.86}$$

According to the considerations reported above, if $m \geq 4$ at least one of the exclusive intersection areas between 2 circles is 0, and μ is simply equal to b_γ. The same holds also if $m = 1$ and $m = 2$, so the only case that must be studied is when m is equal to 3. Assume, without loss of generality, that the circle γ_4 contains the intersection of γ_1, γ_2 and γ_3. Then, there are two possible cases, as reported also in Figure 3.34:

- γ_4 is contained in the union of the other three circles. In this case, $\mu_4^* = 0$, and hence $\mu = a_\gamma$;
- γ_4 is not contained in the union of the other three circles. In this case, it can be shown that it fully contains the exclusive intersection of two circles, meaning that $\mu = b_\gamma$.

Moreover, since the intersection area of γ_1, γ_2 and γ_3 is included in γ_4, also $\mu_{1,2,3}^* = 0$, and in both cases also $\mu = c_\gamma$. In this manner, we have proved that in any case the unknown value of μ is equal to one of the three values a_γ, b_γ or c_γ, as summarized also in Table 3.8. However, it is still to be determined how the algorithm can recognize which one of the three terms is the actual value. In most cases, this can be inferred by the relationships among their values. We can distinguish the following exhaustive possibilities:

Table 3.8. *Relationships among a_γ, b_γ and c_γ, depending on the positions of the circles.*

Case	Ordering	Value of μ
$m \neq 3$	$a_\gamma, c_\gamma \geq b_\gamma$	b_γ
$m = 3$ and $\gamma_4 \subset \bigcup_{i=1}^{3} \gamma_i$	$a_\gamma = c_\gamma \geq b_\gamma$	$a_\gamma (= c_\gamma)$
$m = 3$ and $\gamma_4 \not\subset \bigcup_{i=1}^{3} \gamma_i$	$a_\gamma \geq b_\gamma = c_\gamma$	$b_\gamma (= c_\gamma)$

- $a_\gamma \neq b_\gamma \neq c_\gamma$: in this case, looking at Table 3.8, it follows that $m \neq 3$, and therefore $\mu = b_\gamma$;
- $a_\gamma = b_\gamma \neq c_\gamma$ or $a_\gamma \neq b_\gamma = c_\gamma$: in both these cases, recalling that $a_\gamma, c_\gamma \geq \mu$ whereas $b_\gamma \leq \mu$, it follows that necessarily $\mu = b_\gamma$;
- $a_\gamma = b_\gamma = c_\gamma$: as in the previous case;
- $a_\gamma = c_\gamma \neq b_\gamma$: this is the only case where it is not possible to determine whether $\mu = b_\gamma$ or $\mu = a_\gamma$. In fact, looking at Table 3.8, this may happen both if $m \neq 3$ and if $m = 3$ and $\gamma_4 \subset \bigcup_{i=1}^{3} \gamma_i$: in the former case, $\mu = b_\gamma$, in the latter instead $\mu = a_\gamma$.

It is clear that, in the last case, it is very unlikely that $m \neq 3$, since this would mean that the smallest exclusive intersection among 1 circle has the same (nonzero) area as the smallest exclusive intersection of three circles (since $a_\gamma = c_\gamma$). Anyway, since this may happen, it is necessary to determine the value of m in a different way.

We first of all calculate which values of m may effectively result in $a_\gamma = c_\gamma \neq b_\gamma$. Since it can be shown that $1 \leq m \leq 2(N_c - 1)$, for $N_c = 4$ we have $1 \leq m \leq 6$. We can exclude the following possibilities:

- $m = 1$: in this case $\exists i$: $\gamma_i \subset \gamma_k$, $\forall k \in \{1, 2, 3, 4\}$, $k \neq i$. This implies that $\mu_i^* = 0$. Moreover, $\forall j \neq i$, we have $\gamma_i \cap \gamma_j \subset \gamma_k$, $\forall k \in \{1, 2, 3, 4\}$, $k \neq i, j$. Hence, also $\mu_{i,j}^* = 0$, which implies that $a_\gamma = b_\gamma$.
- $m = 2$: in this case $\exists i, j$, $i \neq j$: $\gamma_i \cap \gamma_j \subset \gamma_k$, $\forall k \in \{1, 2, 3, 4\}$, $k \neq i, j$. This means that $\mu_{i,j}^* = 0$. In addition, $\forall k \neq i, j$, it is also true that $\gamma_i \cap \gamma_j \cap \gamma_k \subset \gamma_p$, $\forall p \in \{1, 2, 3, 4\}$, $p \neq i, j, k$, which in turn implies that also $\mu_{i,j,k}^* = 0$. As a consequence, $b_\gamma = c_\gamma$.
- $m = 4$, with 2 sides belonging to the same circle: in this case Theorem 1 holds, meaning that there exists one exclusive intersection among two circles with zero area. Moreover, it is clear that there exists a circle γ_i that fully contains the intersection of the other

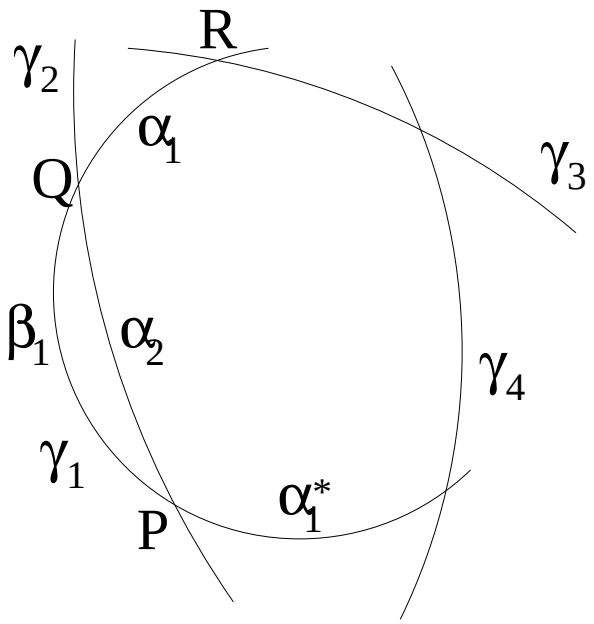


Figure 3.35. *Intersection of 4 circles, with* $m = 5$.

three circles (the ones whose arcs delimit the intersection area of all the four circles), meaning that $\mu^*_{j,k,p} = 0$, with $j, k, p \in \{1, 2, 3, 4\}$, and $i \neq j \neq k \neq p$. Therefore $b_\gamma = c_\gamma$.

- $m = 5$: here again Theorem 1 holds, meaning that $\exists i, j \in \{1, 2, 3, 4\}$, $i \neq j$: $\mu^*_{i,j} = 0$. In addition, as reported in Figure 3.35, it is clear that two non consecutive arcs α_1 and α^*_1 delimiting the intersection area of the four circles belong to the same circle, say γ_1. Assume that the arc between them, namely α_2, belongs to γ_2. This arc divides the circle γ_1 in two parts. The one containing the intersection of the four circles is fully contained in γ_2, since the two circumferences $\hat{\gamma}_1$ and $\hat{\gamma}_2$ cannot intersect in more than two points, and they already intersect in P and Q. The other part of γ_1 is contained in γ_3 (as well as in γ_4). In fact, if this was not true, since P and Q both belong to γ_3, the arc β_1 should intersect the circumference $\hat{\gamma}_3$ in two points. Since $\hat{\gamma}_1$ and $\hat{\gamma}_3$ already intersect in R, this would cause them to intersect in more than two points,which is not possible. Therefore, $\gamma_1 \subset \gamma_2 \cup \gamma_3$, and $\mu^*_1 = 0$, meaning that $a_\gamma = b_\gamma$.

- $m = 6$: the same reasoning as for $m = 5$ can be done in this case, again resulting in $a_\gamma = b_\gamma$.

Having excluded all the above listed cases, there are only two possible deployments that may result in $a_\gamma = c_\gamma \neq b_\gamma$: either $m = 3$ or $m = 4$ with all the four arcs belonging to different circles. They are reported in Figure 3.36.

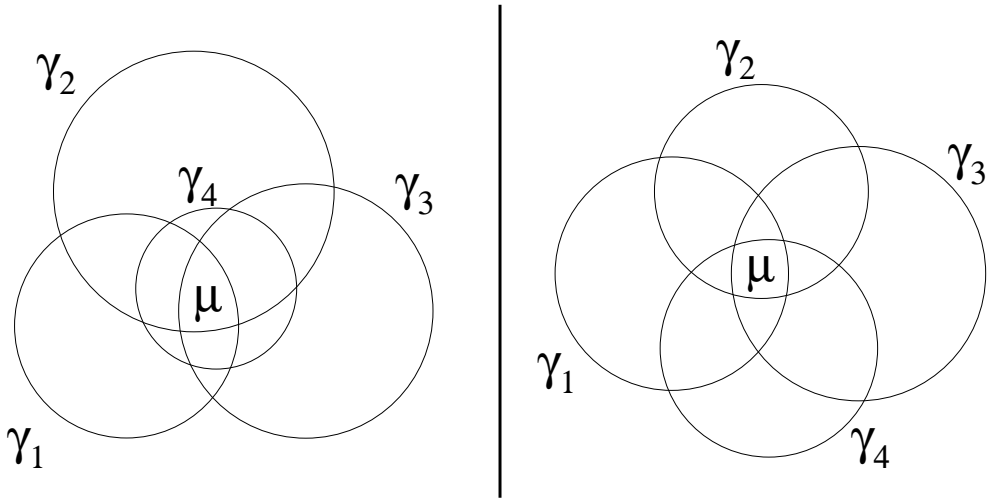


Figure 3.36. *The only two possible deployments of 4 circles such that $a_\gamma = c_\gamma \neq b_\gamma$.*

Recall that if $m = 3$, the only possible case is the one reported in Figure 3.34 on the left. The deployment depicted in Figure 3.34 on the right is instead not possible, since it implies $b_\gamma = c_\gamma$, as reported also in Table 3.8.

From the previous investigation, we can conclude that when $a_\gamma = c_\gamma \neq b_\gamma$, the value of μ is a_γ if the circles are deployed as in Figure 3.36 on the left, and is equal to b_γ if the circles are deployed as in Figure 3.36 on the right. No other deployments are compatible with the given inequality among a_γ, b_γ and c_γ. The straightforward way to distinguish between the two cases is to calculate all the twelve intersection points between the four circles. In both situations each circle contains exactly three points of intersection between the other three circles. More precisely: $\forall i \in \{1, 2, 3, 4\}$, $\exists P, Q, R$: $P \in \hat{\gamma_j} \cap \hat{\gamma_k}$, $Q \in \hat{\gamma_j} \cap \hat{\gamma_p}$ and $R \in \hat{\gamma_k} \cap \hat{\gamma_p}$, with $j, k, p \in \{1, 2, 3, 4\}$, and $i \neq j \neq k \neq p$, such that $P, Q, R \in \gamma_i$, where again $\hat{\gamma}_i$ indicated the circumference of circle γ_i.

However, if $m = 3$ there is one circle that contains the intersection area of the other three circles, meaning that there is one (and only one) circle γ_i such that, using the same notation as above, $P \in \gamma_p$, $Q \in \gamma_k$ and $R \in \gamma_j$. This is not true when $m = 4$, since in that case, one of the three points contained in each circle γ_i, given by the intersection of two circumferences $\hat{\gamma}_j$ and $\hat{\gamma}_k$, does not belong to the third circle γ_p. Simple geometric comparisons among the distances between centers and intersection points and the radii of the circles are then enough to distinguish the two cases and, finally, determine the correct value of μ. Note that this check is necessary only for $N_c = 4$, and when $a_\gamma = c_\gamma \neq b_\gamma$.

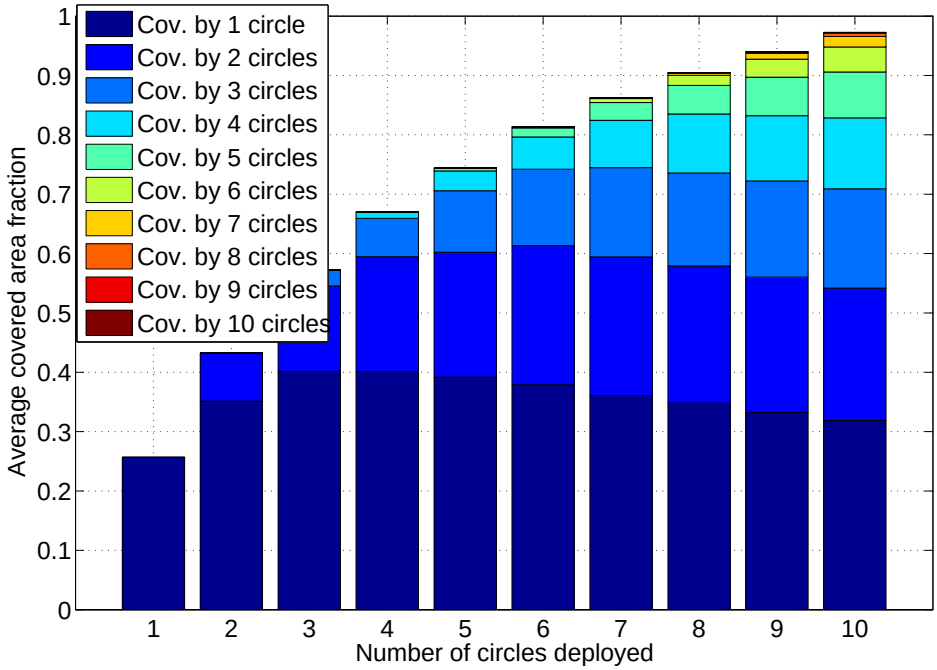


Figure 3.37. *Percentage of the area covered with an increasing number of circles. Here,* L=200, R=80 *and* L=$L-2R$.

3.6.6 Examples of Application

As discussed in the Introduction, the computation of the exclusive intersection areas finds many applications in wireless networking analysis, modeling and performance evaluation.

As a first example of application, we show a straightforward result of our algorithm, that would require an overwhelming computational complexity if addressed via traditional geometric approaches. Assuming that we want to cover a certain area with circles of fixed radius, it is clear that the percentage of covered area grows with the number of circles. However, the fraction of area covered by a certain number of circles would require the resolution of an extremely difficult geometric problem. Note that this distribution corresponds to the probability that a randomly placed user falls within the area covered by a certain number of other nodes. For instance, this result can be applied to localization in sensor networks, where the accuracy of localization depends on the number of sensors within range of the target node. From this result one can derive how many sensors are required to cover a certain area in order to obtain a wanted average localization accuracy. Our algorithm allows a simple computation of this rather involved distribution.

In Figure 3.37, we plot the fraction of area covered by a given number of randomly placed

circles of fixed radius equal to R in a square area with side S, as a function of the overall number of deployed circles N_c. In order to avoid border effects, the centers are uniformly distributed within a square area with a smaller side $L{=}S{-}2R$. The percentage is obtained by averaging over 1000 random deployments for each value of N_c.

The example presented before is a direct application of the proposed algorithm. However, there are many non-trivial interesting applications. As an example, in the following we show how the capability to compute the intersection areas enabled by the presented algorithm can be applied to a rather involved multi-user allocation problem.

Consider a certain number of base stations,[11] whose coverage areas may overlap. In order to evaluate their optimal placement, we can define as performance metric the probability that a certain number of mobile users can successfully communicate with one of the base stations. Note that this represents an upper bound for the resource allocation, as it consider any possible allocation for the users.

We focus on the uplink transmission in a code division multiple access (CDMA) multiple cell scenario. We assume that in order to communicate with a base station (BS), the signal from a mobile terminal (MT) is required to be received by the former with a target average signal to interference plus noise ratio (SINR) S_t. Let us assume that each BS is assigned a single frequency channel and that those channels do not interfere with each other. Then, a MT communicating with a certain BS interferes only with the other MTs allocated to that BS. We consider power controlled transmissions by the MTs with a maximum transmit power equal to $P_{\max}$. The coverage range is defined as the maximum distance from the BS at which a signal from a MT can be received with an average SINR equal S_t without violating the constraint on the maximum power. It is clear that the coverage range is a function of the number of MTs communicating with the BS, due to the mutual interference.

Assuming that the attenuation incurred by a signal at distance d from its transmitter is proportional to $d^{-\alpha}$, and that all the signals arrive at the BS with the same power, i.e., ideal power control, the coverage range of a BS with u MTs and target SINR $S_t{=}s$ is

$$\rho(u,s) = \left(\frac{C_g G_b P_{\max}/s - u}{\sigma^2} \right)^{\frac{1}{\alpha}}, \tag{3.87}$$

where σ^2 is the noise power, $C_g{=}3840/T_r$ is the code gain (T_r is the transmission rate in

[11] We refer here to base stations. In a heterogeneous network we may have other different technologies. Another example where we have intersection of coverage areas is the deployment of cooperative networks, where we have many relay stations that help user communications.

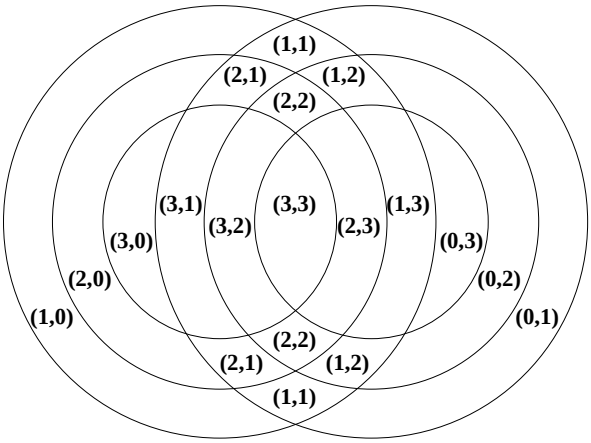


Figure 3.38. *Example of Labeling with two base stations and* 3 *users.*

kbps), and $G_b{=}MB_g/A$ accounts for the shadowing margin, the antenna gain of the BS and the fixed attenuation factor.

Once the wanted SINR S_t is fixed, the coverage range of a BS shrinks for every further allocated MT and we can draw a set of circles centered in each BS and whose radii are $\rho(u, S_t)$ computed for the different numbers of MTs. Thus, u MTs inside the circle with radii $\rho(u, S_t)$ and centered on a BS can be allocated to it. The coverage areas of the BS may intersect, generating a set of regions covered by multiple BSs.

Consider Fig. 3.38, where the coverage regions of two BS for $u{=}1, 2, 3$ MTs are depicted. We label each region defined by the circumferences of the circles with (u_1, u_2), meaning that in that region at most u_1 and u_2 MTs can be allocated to the left-hand and right-hand BS, namely BS 1 and 2, respectively. Thus, assuming we have 4 MTs in the region $(2, 3)$, we may allocate either 1 to the first and 3 to the second or 2 MTs to the first and 2 MTs to the second BS. In the former case we can still allocate another MT to the first BS, while in the latter we can allocate another user to the second BS. Note that we do not need to distinguish among regions with the same label, and that the overall set of regions corresponds to the set of all the exclusive intersection regions investigated in this work.

The probability that a fixed number of MTs, randomly deployed in a certain area, can be allocated to the set of existing BSs can be found by properly setting a rather involved combinatorial problem, whose goal is to assess the existence of a valid allocation. In the following, we present a novel and efficient way to address this kind of combinatorial problem. For the sake of simplicity, we assume that each MT is allocated to the closest BS. This reduces the number of possible configurations, but the proposed structure can be easily applied to the

Table 3.9. *System parameters.*

Maximum power $P_{\max}$	-9 dBW
Noise power	-103dBm
Transmission rate T_r	12.2 kbit/s
Path-loss exponent α	3.5
Shadowing margin M	8 dB
Antenna gain B_g	14 dB
Fixed attenuation parameter A	1000

case where this constraint is removed. In order to derive the wanted probability, we need to evaluate the probability that a certain number of MTs fall within each of the various regions. Therefore, the capability to evaluate the areas corresponding to the exclusive intersection regions granted by our algorithm provide a solution to the previously discussed allocation problem.

Let N_b be the number of BSs deployed in a circular area with radius ρ_a centered in the origin of the axis of the plane. Thus, each BS is the center of a set of circles, corresponding to the coverage ranges for the various numbers of allocated MTs. We denote as $\gamma_i(u)$ the circle associated with the i–th BS when u (radius equal to $\rho(u, S_t)$) MTs are allocated to it.

Each of the regions defined by the exclusive intersection of a subset of circles is thus labeled by the vector $(\ell_1, \ldots, \ell_{N_b})$. The element ℓ_i corresponds to the maximum number of MTs that can be allocated to the i–th BS. Thus, $\ell_i{=}u{>}0$ if the region is contained in $\gamma_i(u)$, but not in $\gamma_i(u{+}1)$. We also set $\ell_i{=}0$ if the considered region falls out of all the circles calculated for $u{>}0$ MTs.

Assume that N_m MTs are deployed in the overall area. Thus the allocation of the MTs can be summarized by a vector whose N_m elements correspond to the region where the MT is actually dropped. We collect these vectors in the $N_m{\times}N_b$ matrix $\mathbf{Z}$. The number of possible deployments is $(N_m^{\max}{+}1)^{N_b}$, where $N_m^{\max}$ is the maximum number of MTs that can be allocated to a BS.

Consider the matrix corresponding to a specific deployment where each row corresponds to a mobile and each column to a BS. The mobile corresponding to the m–th row should be allocated to the BS associated to the column j such that $\mathbf{Z}_{m,b}{=}\max_k \mathbf{Z}_{m,k}$. The same reasoning holds for each row. If we assume that the maximum value appears only once in each row, then the required allocation is unique. Now, in order to check if this allocation is al-

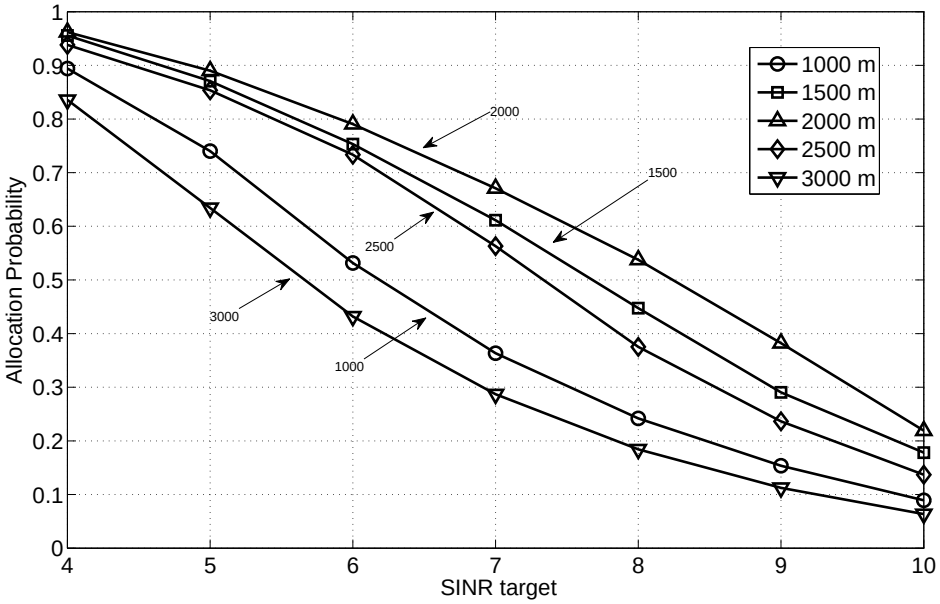


Figure 3.39. *Probability that the considered MTs are successfully allocated.*

lowed, consider the columns of $\mathbf{Z}$. For each column b consider the rows whose maximum value is located on that column (which correspond to the mobiles that should be assigned to the b–th BS), call the set of these values $\mathcal{T}_b$, and call its cardinality $\tau_b{=}|\mathcal{T}_b|$. It can be shown that the minimum values among those belonging to $\mathcal{T}_b$ must be equal or greater than τ_b for each b in order for the deployment to be admissible. Consider for example the deployment represented by the matrix:

$$\mathbf{Z} = \begin{bmatrix} 2 & 1 & 0 \\ 3 & 1 & 2 \\ 0 & 1 & 0 \end{bmatrix} \tag{3.88}$$

In this example mobiles 1 and 2 should be allocated to BS 1, and mobile 3 to BS 2. The allocation is possible since $\mathcal{T}_1{=}\{2,3\}$, and its minimum value is not lower than $\tau_1{=}2$; analogously, $\mathcal{T}_2{=}\{1\}$, whose minimum value is again not lower than $\tau_2{=}1$. If in one row the maximum value appears more than once, then different checks must be done, each considering one of the highest values as the maximum. If at least one choice leads to an admissible solution, then the deployment is acceptable.

The verification of the deployment can be quickly done in a recursive way. Nonetheless, the number of possible deployments grows dramatically with the number of mobiles and Base Stations. Possible approximations can be applied, such as considering only some of the coverage areas around each Base Station. In the example below, we consider a scenario

with three BSs, and we randomly deploy three clusters of ten users. We assume that all the mobiles of the same clusters fall in the same subregion, and consider for the three BSs only three coverage ranges, $\rho(10, S_t)$, $\rho(20, S_t)$ and $\rho(30, S_t)$. These ranges depend on the SINR target, which in turn may be related to the rate used for transmission. The mobiles are randomly deployed in a circular area with radius r_d. We calculate the probability that the clusters can be assigned to the closest BSs as a function of the target SINR S_t.

Please note that the same labeling strategy can be applied to find the probability when the assumption that the MTs are allocated to the closest BS is removed. In this case, however, the verification that a deployment is admissible becomes much more computationally expensive, since much more possible choices can be made for each mobile.

In Fig. 3.39 we plot the probability that all the three clusters are allocated to the closest BS, as a function of the target SINR S_t. The three BSs are deployed at a fixed distance d_B from the center of the region in which the MTs are randomly deployed. The different curves in Fig. 3.39 are for different values of this distance. The parameters used are summarized in Table 3.9. Results show that the best scenario is for d_B approximately equal to 2 km. If the distance increases, the probability is lower, since part of the area covered by each BS falls outside the region where the MTs are deployed. The probability however decreases also for lower values of d_B. This is due to the fact that the number of regions covered by more BSs is greater. Many of these regions cannot be used if another cluster is already allocated to the BS closest to them, thus resulting in a lower number of admissible deployments.

Chapter 4

Conclusions

This thesis investigates the practical application of cooperative techniques in wireless networks, in order to point out which kinds of benefits are achievable via cooperation and under which conditions and constraints.

More specifically, the work is divided into two main parts. In the first part, wireless Ad Hoc networks are considered, and a full Coded Cooperation based protocol is designed and tested. The second part is instead focused on cellular networks, and well known schemes like Amplify and Forward (A&F) and Decode and Forward (D&F) are applied and compared. We underscore the differences between the two environments, which led us to different choices in the application of cooperation, as well as the main contribution added by the present work in a very rich and flourishing research area.

Wireless Ad Hoc networks are intrinsically infrastructureless. This means that transmissions are usually asynchronous, there are multiple sources and multiple receivers, and resources must be assigned and shared in a distributed fashion. In large networks, it is also possible that a single message is relayed through a number of wireless links before reaching its intended destination, which in turns requires a routing table or, alternatively, some sort of updated geographical knowledge at each node. In these kinds of networks, cross layer protocols have shown excellent performance; cooperation can be also implemented at different layers, and with different purposes. The main constraint is the fact that a finite amount of resources must be shared not only by normal transmissions, but also by cooperative ones, which in general also require additional signalling and coordination.

The main contribution of the thesis in this direction is twofold. On the one side, we analyzed how cooperation can be implemented and how cooperators can be used, not only

to improve the quality of bad links through the use of idle neighboring nodes, but also to dynamically select, in multihop networks, alternative paths and avoid congested areas. On the other side, we analyzed the impact of cooperation from a network perspective, trying to assess if, despite the additional interference introduced by the cooperative transmissions, cooperation can improve the overall throughput of the network.

It is well known that when resource multiplexing is not possible, either in time or in frequency, interference is a main issue. In ad hoc networks, where transmissions are not synchronized or supervised by a central node, interference may significantly vary during a single transmission, leading to a sudden change in the channel condition of a given link. In order to counteract its detrimental effect, coding or adaptive modulation techniques are usually adopted. The help of cooperating nodes, if realized in a flexible way, can also grant a remarkable improvement by exploiting spatial diversity.

In our study, we designed a protocol for MIMO ad hoc networks, which relies on a distributed version of HARQ. If a transmission fails, nodes which decoded the packet from the source can send additional redundancy to the destination, following a Coded Cooperation scheme. We adopted Linear Erasure Codes to generate redundancy. The protocol allows the concurrence of multiple transmissions, which can be decoded through the use of spreading codes and BLAST receivers. Signalling packets are kept only to help channel estimation and cooperator selection, when necessary.

We also added an opportunistic routing feature: since cooperators intervene in the transmission when interference is high, they may also decide to forward a packet to its final destination on a different path, if this results in a shorter delivery time. In fact, cooperators are usually required in congested areas, and their capability of re-routing packets may significantly help in reducing congestion, with a local and global benefit.

For testing our protocol, we chose a full network simulator. The reason for this choice is the evaluation of cooperation from a global point of view. A higher number of nodes are transmitting, when cooperation is used, and interference may be higher and undergo higher variations, thus adversely affecting other communications in the same area. The presented results clearly show that cooperation not only improves performance locally, but also increases the overall throughput. Both opportunistic routing and coded cooperation increase the aggregate throughput, the former by strengthening bad links, the latter by dynamically changing the routing paths through less congested regions. In networks with low density the first approach is preferable, since fewer alternative paths are available, whereas for high

density coded cooperation may generate excessive interference, thus making oppostunistic routing a better choice. Finally, we combined the two schemes together, showing that their benefits can be added, thus achieving even better performance.

The second research direction of this thesis focuses on cellular networks. In this scenario, transmissions are controlled by a central node, and resources are assigned in a centralized fashion. As a consequence, each transmission is assigned a given resource, which is in general guaranteed for the duration of the transmission. Moreover, when considering a single cell, there is only one source (downlink) or one destination (uplink), transmitter and receiver may have different capabilities, and usually several other mechanisms (power control, handover...) are also involved in each transmission. A number of standards exist in these kinds of networks, meaning that the impact of any modification should be reduced as much as possible. Therefore, when cooperation is applied to cellular networks, several constraints have to be considered. This is the reason which determined our choice of simpler cooperative techniques for this environment.

We started our analysis of Uplink performance by comparing A&F and D&F in a simple three-node scenario. The aim of this first part was to determine the impact of various factors, like the cooperator position and its capabilities, on the effectiveness of both the cooperative schemes taken into account. This preliminary study was essential to give some design directions, to decide which scheme is preferable, whether mobile terminals can act as cooperators, how cooperators should be assigned to each transmission. We derived the outage probability of cooperative schemes and compared them. From these first results some conclusions were inferred: D&F is more effective, although less practical, since a higher delay is required for decoding; relay stations are much more preferable than mobile terminals as cooperators, although they are fixed and in lower number. This in turn requires a proper allocation of the helped users, in order to maximize the overall performance.

Since power control is a key issue in cellular networks, we also designed and proposed some simple cooperative power control algorithms. The main task is to control the power of two sources in order to combat fast fading on three different channels, in such a way that the global SNR at the destination is kept above a target value. Four algorithms were described, each relying on different information about the channels or the nodes positions. Preliminary results, presented in this thesis, show that their performance is good, although their practical implementation has not been considered in detail.

We then proceeded in our analysis by considering a whole cell scenario. Also in cellular

networks, in fact, interference plays a fundamental role in determining the effectiveness of cooperation. In addition, since we chose to use relay stations, each cooperator is required to help more than one node. The proposed scheme requires two frequencies, to avoid overwhelming interference at the relays, and a TDMA scheme for relayed transmissions, since full duplex relays are considered not available. We derived expressions for the coverage area of a single cell when A&F is used at the relay stations, as a function of the number of relays and of their position. Our study shows that the coverage area can be substantially enlarged by properly placing the relay stations and allocating users to each of them, and the dropping probability can be lowered as well.

Finally, we tested the effectiveness of relaying schemes in a multi-cell simulator. The simplifying assumptions adopted in the theoretical analysis were removed, in order to verify whether the predicted benefits are achievable in real applications. In this part of the study, practical issues, regarding the channel model, soft handover schemes, aggregation over backhaul, allocation policy, were taken into account. Simulation results showed that a significant improvement is actually offered by cooperation, in terms of both throughput and power saving. D&F performs better than A&F, although improvements strongly depend on the users distribution as well as on the carrier assignment.

In an effort to find a theoretical expression for the system capacity in terms of number of users, we developed, as a side result, an iterative algorithm to compute the intersection areas of any number of circles. Based on two geometrical results, this algorithm may be used in a variety of applications, ranging from routing to localization to user allocation.

Further investigation directions regarding cooperation in cellular networks include the analysis of D&F relay stations, interference cancellation receivers, directional antennas and different users distributions.

Appendix A

Distributed estimation in Wireless Sensor Networks

Sensor networks are wireless networks whose nodes perform some kind of measurement of a given quantity. Due to the presence of several nodes, a distributed measurement is possible, and, through exchange of information among nodes, a distributed estimation of the desired quantity is available. In this sense, it is clear that in these networks all nodes cooperate in a peculiar way to achieve a common goal. Nevertheless, it is clear that a different paradigm should be applied to sensor networks, where usually different constraints, as well as different transmission techniques and communication protocols, are likely to be considered. Hence, we present in this appendix a side work, which focuses on practical distributed estimation in sensor networks through use of gossip algorithms.

A.1 Related work and main contribution

The increased availability of inexpensive nodes carrying a wide range of sensing capabilities continues to create interest in developing large-scale sensing platforms, which could be used to measure a variety of physical phenomena [69–71]. However, the energy-constrained nature of the devices making up these sensor networks continues to be a challenge to application and protocol designers. Additionally, these networks are often bandwidth constrained. This combination of energy and bandwidth constraint has motivated the design of various data fusion and estimation algorithms [72,73], where the goal is to perform some amount of in-network processing to minimize the amount of data that needs to be transmitted to the sink while still maximizing the accuracy of the measurements.

To design effective estimation algorithms for wireless sensor networks, an underlying data dissemination protocol must also be used. A great deal of work has been performed on both distributed estimation and dissemination algorithms; however, these two aspects are typically analyzed separately. Good implementations of distributed estimation filters often do not take into account problems linked to the transmission over a wireless network, for instance collisions, energy consumption, impact of packet losses on the overall performance; furthermore, the broadcast nature of the medium is rarely exploited, whereas point-to-point communications are usually assumed [74–79]. On the contrary, detailed dissemination protocols are usually not designed to support the deployment of a distributed estimation algorithm [80–82].

One effective technique for filtering and estimation is the Kalman filter [83], which is the optimal filter that minimizes a quadratic cost function that penalizes the covariance of the estimation error. Kalman filters can be effectively implemented as a recursive algorithm: at each step the estimate (output) is obtained by updating the previous estimate using the information carried by the new sample of the input. Kalman filters have been shown to be very effective at producing accurate estimates even in distributed environments [84–87].

The main contribution of this work is the design of a simple communication protocol, leveraging the broadcast nature of the wireless medium. In fact, many authors have deeply investigated issues related to distributed estimation, mostly through a properly designed Kalman Filter, and interesting results have been found and proved. Nevertheless, the practical implementation of these schemes usually relies on point to point communications, thus not exploiting the broadcast potential in terms of information spreading. On the other side, broadcast communications cause a large amount of interference and consequently increase the packet loss, thus worsening estimation accuracy. Henceforth, our aim is to develop a realistic network protocol, based on gossip communications, through which a simple distributed estimation scheme can be applied to a Wireless Sensor Network. A comparison between the expected results given by well-known theoretical approaches and those obtained using our protocol is carried out.

A.2 A Broadcast Solution for Distributed Estimation of a Constant Signal

In this section, as a first step in the analysis, we propose a distributed algorithm to solve the estimation problem focusing on the simple case of estimating a constant in the presence of noise. We assume that the communication medium is broadcast in nature and want each node to use information from its neighbors to compute its estimate, which it then broadcasts. Neighboring nodes do not answer broadcasts, but merely use the information to update their own estimates.

A.2.1 Problem Formulation

Consider the problem of estimating a constant process $x(t) \equiv X \; \forall \; t$ from N noisy measurements, collected by N agents:

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_N(t) \end{bmatrix} = \begin{bmatrix} X \\ X \\ \vdots \\ X \end{bmatrix} + \begin{bmatrix} w_1(t) \\ w_2(t) \\ \vdots \\ w_N(t) \end{bmatrix} = X\mathbb{1} + w(t), \tag{A.1}$$

where $w(t)$ represents N independent noise sources, which corrupt the agents' measurements.

In the following, we assume that all noise is i.i.d. with zero mean, therefore:

$$E[ww^T] = \sigma_w^2 \, I_N.$$

Since we are interested in a distributed solution, each agent is going to compute an estimate. In the following, we refer to the estimate of the i-th agent with the notation $\hat{x}_i$ and we collect the estimates in a vector:

$$\hat{x} = \begin{bmatrix} \hat{x}_1(t) \\ \hat{x}_2(t) \\ \vdots \\ \hat{x}_N(t) \end{bmatrix}.$$

We assume that agents can communicate according to a communication graph $\mathcal{G} = (V, E)$, *i.e.*, nodes i and j can communicate only if $(i, j) \in E$. We call $\mathcal{N}(i) \subseteq V$ the set of neighboring nodes of i.

A.2.2 High-Level Algorithm Description

Our algorithm is divided into two phases as follows:

Phase 1: Measure and estimate update.

As in traditional estimators, the old estimate, $\hat{x}(t)$, is updated by means of a weighted average between the old estimate and the new data. The estimate we obtain is called **local**, $\hat{x}_{loc}$.

In the simple case we are dealing with, the update algorithm can be written as the convex combination of the old estimate and the new measurement:

$$\hat{x}_{loc}(t+1) = \lambda \hat{x}(t) + (1-\lambda) y(t).$$

Please note that, although this approach is not the most effective for this simple scenario, it is well suited for the more interesting case developed in the next Section. We investigate such a solution here in order to introduce the approach used in the rest of the work. From a practical point of view, one would use a value of λ as close as possible to 1, in order to increase the low pass effect of the resulting filter. Such a choice, however, would increase the transient duration. Since the choice of λ is not the focus of this work, we simply use $\lambda = 0.9$.

Phase 2: Information exchange among neighboring nodes.

Once every sample period, a node, i, randomly turns on and broadcasts its estimate to its neighbors. All neighboring nodes update their local estimate with a weighted average to obtain an improved estimate, called **regional**, $\hat{x}_{reg}$, or simply $\hat{x}$

$$\hat{x}_{reg_k} = \alpha_{i,k} \hat{x}_{loc_k} + (1-\alpha_{i,k}) \hat{x}_{loc_i} \qquad k \neq i,\ k \in \mathcal{N}(i), \tag{A.2}$$

where $1-\alpha_{i,k}$ is a parameter that weighs how much node k should rely on node i's measurement.

The N update equations in (A.2) can be rewritten in vector form as:

$$\hat{x}_{reg}(t) = F(t) \hat{x}_{loc}(t), \tag{A.3}$$

where $F(t) = F_i$ and i is the node that is active at time t. Essentially, a matrix is extracted from the alphabet set $\mathcal{A} = \{F_1, \ldots, F_N\}$.

Obviously F_i depends on the set of i's neighbors: If, for instance, $\mathcal{N}(1) \equiv V$, then:

$$F(t) = F_1 = \begin{bmatrix} 1 & 0 & 0 & & 0 \\ 1-\alpha_{1,2} & \alpha_{1,2} & 0 & & 0 \\ 1-\alpha_{1,3} & 0 & \alpha_{1,3} & & 0 \\ \vdots & & & \ddots & \\ 1-\alpha_{1,N} & & & & \alpha_{1,N} \end{bmatrix}.$$

While, if $\mathcal{N}(1) = \{1, 2, 5\}$, then:

$$F(t) = F_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & & 0 \\ 1-\alpha_{1,2} & \alpha_{1,2} & 0 & 0 & 0 & & 0 \\ 0 & 0 & 1 & 0 & 0 & & 0 \\ 0 & 0 & 0 & 1 & 0 & & 0 \\ 1-\alpha_{1,5} & 0 & 0 & 0 & \alpha_{1,5} & & 0 \\ \vdots & & & & & \ddots & \\ 0 & & & & & & 1 \end{bmatrix}.$$

It is important to notice that $F(t)$ is always a stochastic matrix.

Summarizing, the algorithm we propose is described by the equations:

$$\hat{x}_{loc}(t+1) = \lambda\hat{x}(t) + (1-\lambda)y(t) \tag{A.4}$$

$$\hat{x}(t+1) = F(t)\hat{x}_{loc}(t+1), \tag{A.5}$$

which can be written in short as:

$$\hat{x}(t+1) = F(t)\Big(\lambda\hat{x}(t) + (1-\lambda)y(t)\Big). \tag{A.6}$$

A.2.3 Generalization

In the algorithm description presented above, we assumed that in each sample period, all nodes collect one output measurement and that one node broadcasts its estimate to its neighbors. In a general setup we might choose to measure more often than communicate or vice versa. Note that equations (A.4) and (A.5) can still be used to describe the algorithm:

- When measurements are made more often than transmissions (*e.g.*, one transmission every A measurements), it holds that:

$$F(t) = I \quad \forall\, t \neq nA, \qquad n = \{1, 2, 3, \dots\},$$

- When transmissions are made more often than measurements (*e.g.*, M transmissions between two consecutive measurements), it holds that:

$$F(t) = F_{i_1} F_{i_2} \dots F_{i_M}, \qquad F_{i_k} \in \mathcal{A},$$

i.e., M matrices are extracted from $\mathcal{A}$ and multiplied together. Note that $F(t)$ is a stochastic matrix in this case as well.

A.2.4 Algorithm Analysis

Algorithm Correctness

In this Section, we present some results that are useful for the analysis of the proposed algorithm. Due to space limitations, we omit the proofs in this appendix, although they may be not trivial since we deal with a time varying system.

Consider N noisy measurements of a constant process as in (A.1) and the estimator (A.6). If $F(t)$ is stochastic $\forall\ k$ and $|\lambda| < 1$, then the estimate $\hat{x}(t+1)$ has the property:

$$\lim_{t \to +\infty} E[\hat{x}(t)] = X\mathbb{1}. \tag{A.7}$$

Moreover it holds that the estimation error variance remains bounded as $t \to +\infty$ which means that the estimation error does not diverge, and therefore gives a meaningful estimate.

Estimate Variance

It is interesting to assess the performance of our estimation algorithm and compare its performance with other algorithms. The aim of our algorithm is to obtain an estimate of the observed process that is as close as possible to the true value X. If we consider the class of the estimation algorithms with the property (A.7), a natural performance assessment parameter is the estimation error variance.

We present now some results on the estimation error variance of our algorithm and introduce two other non-distributed solutions that we will use for comparison.

The estimate error $\tilde{x}(t) = x(t) - \hat{x}(t)$ is:

$$\tilde{x}(t+1) = \lambda F(t)\tilde{x}(t) + (1-\lambda)w(t).$$

Moreover it holds that

$$E[\tilde{x}(t)\ w(t)] = 0,$$

Therefore:

$$\text{Var } \tilde{x}(t+1) = \lambda^2 F(t)\text{Var } \tilde{x}(t)F^T(t) + (1-\lambda)^2\sigma^2 F(t)IF^T(t).$$

Let us consider three possible situations:

- *Local solution*:

  Each agent can use only its own measurements.

  This is the case in which each node has no neighbor, *i.e.*, $\mathcal{N}(i) = i$. Therefore, there is no information exchange between nodes and we run N independent estimators, which is the algorithm described by equation (A.6) with $F(t) = I\ \forall t$. We call the estimate error variance in this case v_1.

- *Centralized solution*:

  Each agent knows the measurements collected by all other agents. In this setup, the most efficient way to aggregate the information coming from the sensors is to compute the average between all the estimates in the nodes, which is the algorithm described by equation (A.6) with $F(t) = \frac{1}{N}\mathbb{1}\mathbb{1}^T\ \forall t$. We call the estimate error variance in this case v_2.

- *Distributed solution*:

  This case describes our algorithm. We call the estimate error variance in this case v_3.

Local and Centralized solutions can be seen as two boundary cases to our algorithm with respectively minimum and maximum information from other nodes. In these two cases, we can easily compute the estimation error variance. In the first case, i.e., $F(t) = I,\ \forall t$, we have that:

$$\text{Var } \tilde{x}(t+1) = \frac{1-\lambda}{1+\lambda}\sigma^2 I.$$

while in the second one, i.e. $F(t) = \frac{1}{N}\mathbb{1}\mathbb{1}^T,\ \forall t$,

$$\text{Var } \tilde{x}(t+1) = \frac{1}{N}\frac{1-\lambda}{1+\lambda}\sigma^2\mathbb{1}\mathbb{1}^T.$$

Therefore we can reduce the estimation error variance by N times by means of N agents collecting N independent measurements and sharing them.

The analysis of the distributed solution becomes more complex as finding a general expression or some bounds that hold for a general topology and for every possible choice of the broadcaster is difficult. Since $F(t)$ is a matrix that depends on the random choice of the

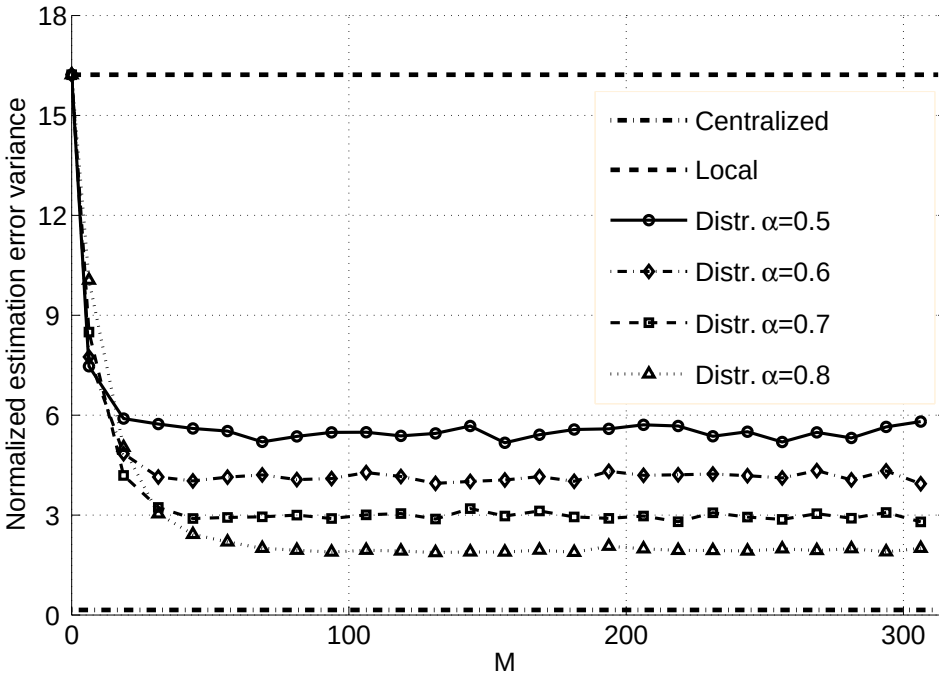


Figure A.1. *Estimation error variance for different values of α, with $\lambda = 0.9$.*

broadcaster node, we consider the expected value of the estimation error variance over all possible node selections, under the natural assumptions that such selections are independent events and that all nodes have the same probability of being selected.

In Fig. A.1, we report the estimation error variance obtained by simulation. It is shown that with our simple strategy it is possible to improve the performance with respect to the local approach, even with very few communications. However, it is not possible to achieve the performance of the centralized solution, not even as $M \to +\infty$. This is due to the fact that we are using a very simple estimation scheme. Better performance could be achieved using more refined solutions, as in other works [76, 88]. Nonetheless, such schemes often require full duplex communication, and are complex to be analyzed and usually less capable of handling unknown topology changes. In Fig. A.1 the effect of the weight α is also shown: lower values of α lead to a slower convergence but smaller mean error variance.

A.3 Estimation of a Random Process by Means of a Distributed Kalman Filter

The assumption in the previous section that the signal we want to measure is a constant is obviously very restrictive and unrealistic, and was made to simplify the discussion and

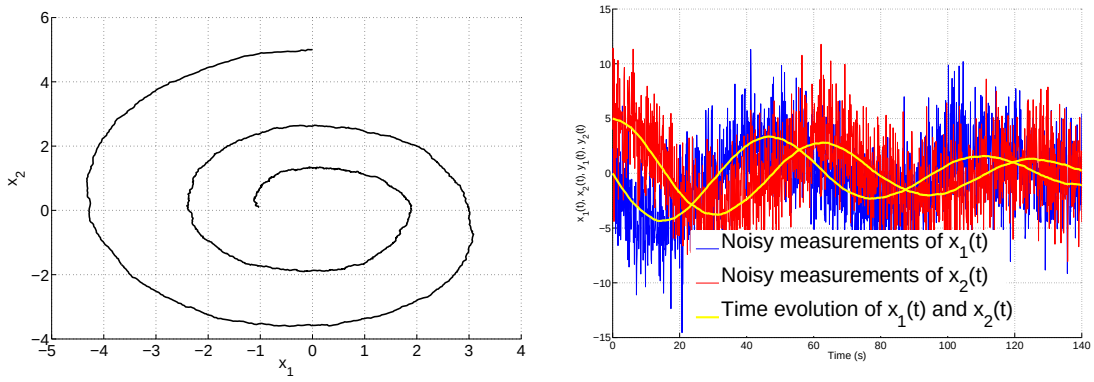
to introduce the notation. In this section, we present an analysis, analogous to the one previously introduced, of a distributed estimation algorithm that handles time varying signals. We first present a modification of the problem formulation.

A large class of stochastic processes can be described as follows:

$$\begin{cases} x(t+1) = Ax(t) + Bv(t) \\ y(t) = Cx(t) + w(t) \end{cases}, \tag{A.8}$$

where $x(t) \in \mathbb{R}^n$ is the signal to be estimated and $y(t) \in \mathbb{R}^p$ is the variable that represents the observable signal used to estimate the process x. $y(t)$ is corrupted by a white noise $w(t)$. $x(t)$ itself is not a deterministic signal, because it depends on the white noise $v(t)$.

As an example, we consider here a two variable random process and we assume A to be a stable matrix. For our choice of A, and for $v \equiv 0$, the trajectory of x would be a perfect spiral but the presence of the noise causes $x(t)$ to fluctuate around this spiral (see Figure A.2(a)). Figure A.2(b) shows that we are considering significantly noisy measurements of the process. To solve the problem of estimation and filtering for processes described by



(a) Trajectory (b) Time evolution

Figure A.2. *Evolution of the random process considered*

(A.8), we use a Kalman filter, which is an optimal linear filter that minimizes the estimation error variance [72,73].

As we did previously, we refer to the situation in which each node uses just its own measurement as the **Local Kalman Filter**, and the situation in which every node can use every other nodes' measurements as the **Centralized Kalman Filter**.

The algorithm we present here to implement the Kalman filter in a distributed fashion works in a similar way to the one we previously described. It is divided in two phases:

Phase 1: Measurement and estimate update.

As in traditional Kalman filters, the old estimate, $\hat{x}(t)$, is updated by a weighted average of the old estimate and the prediction error. The estimate we obtain is called **local**, $\hat{x}_{loc}$.

Phase 2: Information exchange among neighboring nodes.

Once every sample period, a random node, say node i, turns on and broadcasts its estimate to its neighbors. Neighboring nodes update their local estimate with a weighted average to obtain an improved estimate, which we will call **regional**, $\hat{x}_{reg}$, or simply $\hat{x}$

$$\hat{x}_{reg_k} = \alpha_{i,k}\hat{x}_{loc_k} + (1 - \alpha_{i,k})\hat{x}_{loc_i} \qquad k \neq i,\ k \in \mathcal{N}(i), \tag{A.9}$$

where $\alpha_{i,k}$ is a parameter that weighs how much node k should rely on node i's measurement.

A theoretical analysis becomes rather complex, therefore in this work we study the problem via simulation. A deeper theoretical investigation is beyond the purpose of the work at this stage, which is to study how Kalman filtering can be applied in a realistic wireless networking scenario, and it would be an interesting refinement of the analysis left for future work. The comparison between $x(t)$ and $\hat{x}(t)$ is reported in Figure A.3, and snapshots of the process state and the estimates in the local case are depicted in Figure A.4. Here, the process trajectory has a spiral shape: the process state is shown for four different instants, and the corresponding estimates computed by the single nodes, as can be observed, are always quite close to the real value.

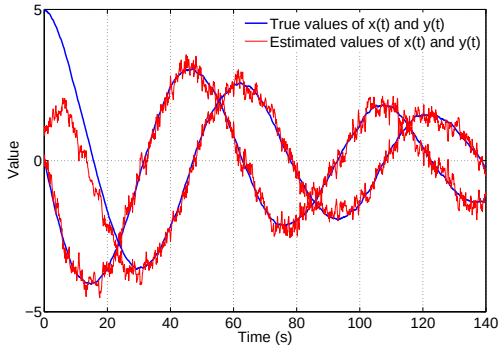


Figure A.3. *Comparison of true and estimated values.*

We verified that the algorithm gives an unbiased estimate of the process. Additionally, as in the constant case previously presented, the distributed Kalman filter has an error variance smaller than the local filter but larger than the centralized one as expected. Due to space constraints, we will present performance results for this case in Section A.4 where we will take into account also lower level issues, such as interference and packet loss.

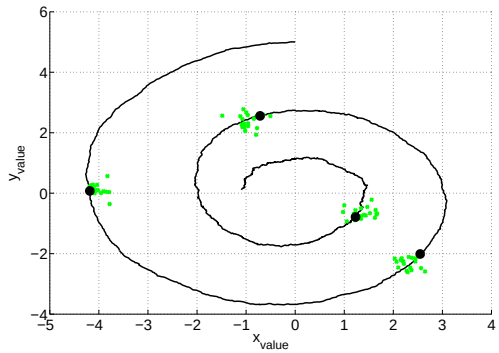


Figure A.4. *Snapshot of process state and estimates*

We conclude this section by remarking that the computational load of the Kalman filter in this application is extremely light, as it only involves vector multiplication with matrix gains and vector summation, and the proposed technique is therefore applicable to a Wireless Sensor Network scenario [1].

A.4 Kalsip Communication Protocol

While the theoretical results demonstrated so far show that our proposed algorithm effectively provides accurate estimates in a distributed manner, we now present our protocol, called Kalsip, which is designed to operate in a wireless sensor network. The evaluations in this section consider communications effects such as MAC layer collisions, packet loss, and interference, as well as the estimator properties considered in the previous sections. Additionally, nodes may not share a common schedule, as they use some mechanisms such as backoff and the choice of a random time between two successive measurements, which results in a completely asynchronous behavior.

The main assumption in the design of our protocol is that the measurement and transmission functions on a sensor node can occur simultaneously, which we believe is not unreasonable given the state of current sensor node hardware.

In this section we first describe how our protocol functions, highlighting the effects of its different parameters, and then present simulation results demonstrating the performance of the full protocol suite.

[1] Note that, accepting a suboptimal behavior during the initial transient, one can use the asymptotic Kalman gain in place of the (optimal) time-varying gain. The former can be computed off–line, while the latter has to be recomputed at each iteration. In this way on–line matrix inversion can be avoided [72, 73].

A.4.1 Protocol Design

Kalsip is not synchronous, hence nodes can measure and transmit at any time, with no need for a common clock. It is possible in real networks, however, that measurements are taken almost regularly, and this is why we introduced a backoff time before each transmission.

Nodes can only transmit in broadcast fashion, but they can detect if other transmissions are ongoing within their sensing range. Therefore, we define two different areas around each source node: the first is the area within which other nodes must be in order to correctly receive messages from the source with high probability. In the graph representing the network, all nodes in this area are linked to the source node. The second area is larger, and contains all nodes whose transmissions generate interference at the source node. This area is theoretically infinite, but can be constrained by considering only interference above a certain minimum threshold. These two areas are both idealized as circles centered at the source node, whose radii are called respectively detection range and interference range [2].

Each node begins its operations in the *idle state*, which lasts until a new measurement is performed or a new transmission is detected. We model the time intervals between successive measurements made by a single node as exponential random variables.

If a measurement is completed before any transmission is detected, the node modifies its estimate according to (A.4) and broadcasts the new value with probability p_r, otherwise it remains in the *idle state* until a new event occurs. If it decides to transmit, the source node first enters the *backoff state*, *i.e.*, it waits a random amount of time before starting transmission. The backoff time is uniformly chosen in the interval between one TX-time and ten TX-times, where a TX-time is the time needed for a single packet transmission. If a transmission is detected while the node is in *backoff state*, the node enters the *receiving state* and freezes the backoff countdown. When backoff time expires, the node begins its transmission. Upon completing the transmission, the node returns to the *idle state*.

If a node detects a transmission while in the *idle state*, it enters the *receiving state*. After completely receiving the packet, it updates its estimate with the new information according to (A.2) and transmits the new estimate with probability p_t, otherwise, the node returns to the *idle state*.

[2]Note that in the performance evaluation we consider the cumulative effect of interference, rather than the commonly used (but inaccurate) "protocol interference model"

If a node performs a new measurement while in the *backoff state*, it updates its value and will transmit this new value at the end of the countdown.

Several parameters may have an impact on the performance of our Kalsip protocol. Topology is certainly one of the key factors, since the degree of the graph (*i.e.*, how many nodes are located in the detection area of a source node) determines the number of nodes that receive a single broadcast transmission. The higher the degree, the higher the number of transmissions, because every time a node receives a packet it has the possibility to transmit its updated value. The number of nodes deployed, the network area and the transmission range of the nodes all influence the degree of the related graph, and consequently also the number of transmissions. However, the number of nodes may also affect the precision of the estimate, as we have already discussed.

The values p_r and p_t must also be tuned in order to get a good trade-off between the number of transmissions and the quality of the estimate. Simulations show that even small values (around 0.1) are enough to provide very accurate estimates. Nonetheless, the aim of our investigation is also to analyze the joint effect of these probabilities and of the other setting parameters.

Finally, the backoff time must be taken into account. Its effect should be similar to those of the two gossip probabilities because we exploit the possibility of updating values during the countdown. Therefore, the information transmitted is always quite recent, and either low p_r and p_t or longer average backoff time should result in fewer transmissions.

A.4.2 Kalsip in Realistic Scenarios

In this section, we present the simulation of Kalsip under realistic network conditions. To perform our tests, we varied the values of node degree, p_r, p_t and the sensing threshold, to study the effects of the network parameters. Additionally, the most important parameter in a gossip-like protocol, such as Kalsip, is the probability that a node will forward data. In Kalsip, each time a node completes a measurement, it decides whether or not to transmit with probability p_r. Analogously, when a node receives a packet from one of its neighbors, it updates its own estimate and broadcasts it with probability p_t. For the sake of simplicity, we consider $p_t = p_r = p_{gossip}$.

It can be seen that very small values of p_{gossip} should be chosen in order to limit the number of transmissions in the network. In fact, while in other gossip-like protocols the nodes which receive the message can decide only once whether to retransmit it or not, in

Kalsip the message is slightly modified by each node which receives it. Hence, the same node can receive multiple transmissions generated by a single measurement by one of its neighboring nodes. Our simulations on networks of 100 nodes show that good values of p_{gossip} are in the range between 0.01 and 0.1.

A second parameter which strongly affects the performance of Kalsip is the degree of the communication graph, since it determines the average number of nodes that receive a broadcast communication from a single node. In our network simulations we can change the degree in two different ways: we can reduce the detection range by simply limiting the transmission power of the nodes, or we can increase the area in which the nodes are placed. Since the two ways are equivalent, we fix the transmission power and place the nodes in square areas of different sizes in order to study the impact of this parameter.

For fixed values of both transmission range and network area, the number of nodes also affects the average degree of the communication graph. However, a larger number of nodes also implies a higher number of different measurements, which also influences the overall performance. For this reason, we decided to keep the number of nodes constant.

Finally, a parameter of lesser importance is the sensing range. Its variations can affect the number of transmissions in the network and reduce collisions and packet losses, but simulations show that this influence is always small. Even if a packet is lost, most of its information can be retrieved by the ones which are sent almost immediately by neighboring nodes. Hence, we fixed the sensing range to a constant value.

As can be seen in Table A.1, the performance of the algorithm described in Section A.2 is somewhat reduced by its application to a realistic scenario. More specifically, as the number M of transmissions between two subsequent measurements increases, the reduction in error variance granted by communication among the nodes is lower, about 60%, with respect to the 85% predicted by the analysis. The reason is that as the number of transmissions becomes higher, the interference is greater, leading to higher packet loss, which was not taken into account in our theoretical analysis. As a consequence, a fraction of the communication is lost, and the overall performance improves more slowly.

A.4.3 Impact of Synchronization

Apart from the physical parameters analyzed in the previous section, a deeper study is required on the synchronization of measurements. Kalsip, as described in section A.4.1, is totally asynchronous. Each node chooses a random amount of time before taking a new

	Estimation error variance reduction (v_1/v_3)	
M	no collision or packet loss	collision or packet loss may occur
1	1,0767	1,0733
3	1,2247	1,1601
5	1,3774	1,2414
7	1,5284	1,328
11	1,7979	1,4995
18	2,3333	1,5865
30	2,9926	1,7769
56	4,1736	2,0905
97	5,6072	2,4961
155	7,0009	3,1612

Table A.1. *Impact of collision and packet loss on the efficiency of the estimation algorithm.*

measurement. This random value has a Poisson distribution, and the consequence is that in large networks, measurements can be considered uniformly distributed, as proved in [74]. However, the backoff mechanism allows another choice.

A solution that requires synchronization among nodes can be proposed as follows: time is divided into slots within which all nodes take only one measurement following a fixed sequence. We will refer to this variant of the protocol as *Constant Kalsip*, since the intervals between two successive measurements are always constant. We will show that, as expected, there is little difference in the performance of this variant and the standard asynchronous Kalsip.

It is also interesting to analyze the case in which all nodes take their measurements together. In this variant, which we call *Synchronous Kalsip*, time is still divided into slots of fixed length, but all measurements are now taken at the beginning of each slot. Although this may appear a worse choice, since all transmissions tend to be synchronized, leading to a stronger interference, it allows the introduction of a more sophisticated way of updating the estimates at the nodes.

The estimate transmitted by a node that has already received several estimates by its neighbors should have a higher weight when compared to one from a node that has only taken its own measurement. In order to take this element into account, in *Synchronous Kalsip*,

nodes can send not only their estimate, but also their weight, *i.e.*, the number of packets already received since the beginning of the current time slot. In this manner, the receiver can compare the number of measurements collected by the source with the number of its own, thus making a weighted average between its current estimate and the one just received. When a new time slot begins, at each node a new measurement is taken, the estimate is updated according to (A.4), and the weight is set to 1.

In the following section, we analyze the performance of all the variants of Kalsip in terms of error variance, ratio between number of transmissions and number of measurements, and probability of correct packet reception.

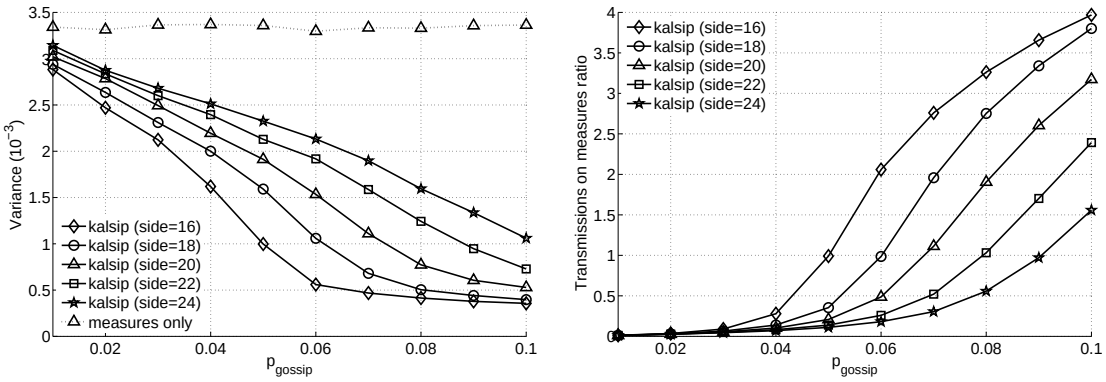
A.4.4 Data Analysis and Results

For all the scenarios described above, simulations were performed in order to compare and understand the behavior and performance of the variants of Kalsip. Each set of simulations consists of 20 random topologies with similar degree. In every simulation, each node acquires an average of 200 measurements. We studied three important metrics: the average variance of the estimation error, the ratio between transmissions and measurements, and the probability of correct reception at the nodes. The first metric gives us the precision of our algorithm. The second one represents how many transmissions affect the precision of the estimate and therefore the relationship between energy consumption and precision. The last metric explains why the performance improvement is not linear, but rather tends to saturate.

In the following figures, we compare the behavior of single nodes that estimate a value only with their own measurements with Kalsip. With very few transmissions (*i.e.*, very low p_{gossip}), the variance of the estimation error is less than that of a single node that does not transmit anything (see Figure A.5(a)). Comparing the different scenarios of Kalsip, we can see that the degree of the network has a large effect on the variance. With a high degree, a single transmission is received by many nodes, which in turn update their estimate. So, with the same number of transmissions, there are more receptions, which means in turn more data exchange.

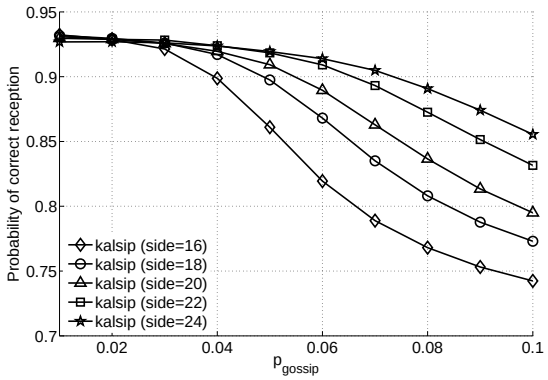
However, as p_{gossip} grows to the saturation point of the channel, the effect of the number of neighbors per node decreases, since nodes send more data with high probability, making up for the fewer receptions per transmission.

In Figure A.5(b), it is clear that for high values of p_{gossip} the topologies with larger degree



(a) Average variance of estimation error for different values of P gossip

(b) Number of transmissions over measurements for different values of P gossip



(c) Probability of correct reception

Figure A.5. *Comparison between measure only and Kalsip protocol*

have more transmissions. This can be explained by the spatial reuse of the channel: if the network is larger, with low p_{gossip} it is more difficult to maintain the gossip due to the lower degree and this agrees with the first part of the graph. But a larger degree means more simultaneous communications, so the total capacity of the network, reached when p_{gossip} grows, is higher.

More neighbors and higher p_{gossip} also imply greater interference. Figure A.5(c) shows how the network performance worsens as long as more packets cannot be decoded by the receiver. In a scenario where energy is a very important metric, this can be a problem and a trade-off between energy and precision must be found.

From the previous graphs (Figure A.5), the large effect of the number of transmissions on the quality of the estimator is clear. Thus we want to know how the performance of the protocol improves by changing the measurement schedule.

Figure A.6 compares *Kalsip* with its variant called *Constant Kalsip*. It can be seen that the

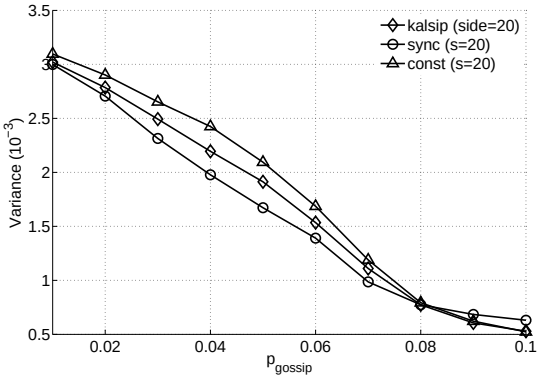


Figure A.6. *Comparison of average variance of estimation error for different values of P gossip between Kalsip and a constant sampling Kalsip protocol*

difference is not substantial. This is due to the large number of nodes. In such a scenario, in a finite time interval, arrivals of several Poisson random processes are distributed in a uniform manner, so they are quite similar to the constant sampling case.

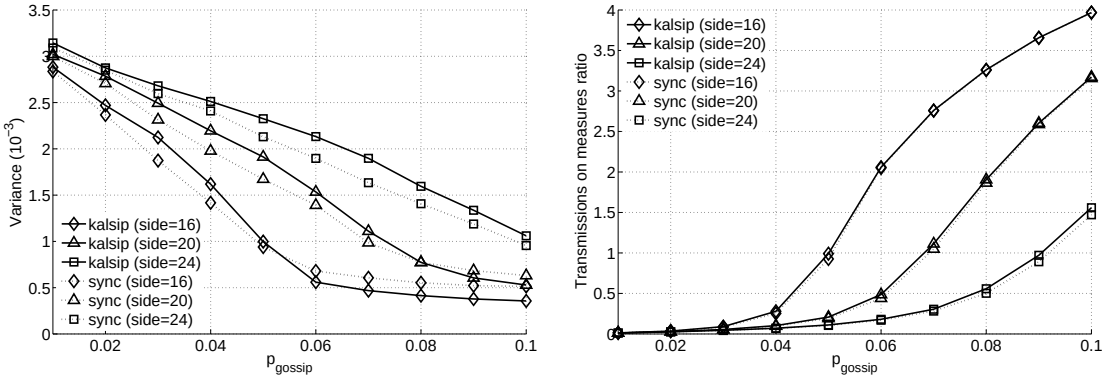
Figure A.7 compares *Synchronous Kalsip* (sync in the figure) with Kalsip. Recall that *Synchronous Kalsip* is an algorithm with synchronous measurements and weighted exchanges of the estimates. Figure A.7(a) shows that this variant yields some improvement in estimation performance but it is worse in network performance when p_{gossip} is quite small. If nodes transmit few times and they measure asynchronously with the same sample period, all transmissions are approximately equally distributed in the *timeslot*, *i.e.*, the time between two consecutive measurements, so it is less likely that transmissions overlap. If all nodes measure in a synchronous way, all the transmissions are concentrated in the first part of each slot and this causes more interference in the network and hence more errors (see Figure A.7(c)).

The weighted solution is better than the simple Kalsip protocol, but improvements are not so large, whereas the computational complexity is increased, so the simple Kalsip appears as the best solution.

A.4.5 Implementation of Kalman Filters over Kalsip

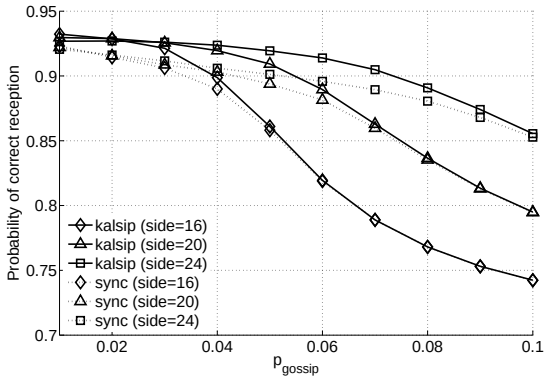
The final component of our solution is the distributed Kalman filter over our Kalsip communication protocol. As in the previous sections, we are interested in analyzing the performance while changing the degree of the network and the value of p_{gossip}.

Figure A.8 and Table A.2 show that our complete protocol performs very well. With $p_{gossip} = 0.02$, so with $M = 3$, we are able to halve the error variance of the estimation.



(a) Average variance of estimation error for different values of P_{gossip}

(b) Number of transmissions over measures for different values of P gossip



(c) Probability of correct reception

Figure A.7. *Comparison between Kalsip and Synchronous Kalsip with weights*

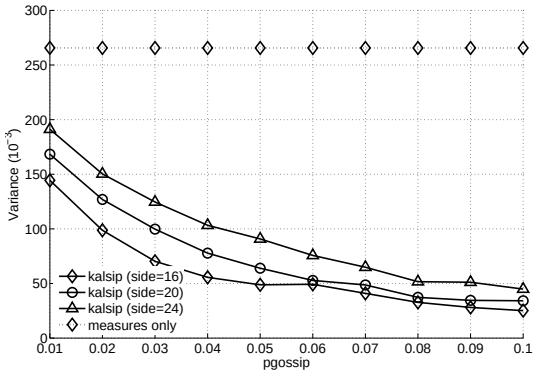


Figure A.8. *Average variance of estimation error for different values of P_{gossip} with Kalman Filter*

p_{gossip}	0,01	0,02	0,04	0,05	0,06	0,1
$E[M]$	1,256	3,4835	25,076	150,13	700,22	1576,1
v_1/v_3	1,7929	2,7178	4,9525	6,0688	7,5812	9,7185

Table A.2. *Ratio between the estimation error variance without and with node communication*

Increasing M, the performance increases until the theoretical limit is reached when $M = 1500$, which is obviously an unrealistic setup, but confirms that for large M the behavior is similar to the centralized algorithm.

A.5 Discussion

The wide variety of applications for wireless sensor networks combined with the energy constrained nature of sensor nodes has motivated research on algorithms for in-network estimations of measured physical quantities to reduce the amount of data that needs to be transmitted to the sink.

In this work, we presented a distributed estimation solution that leverages the broadcast nature of the wireless channel in sensor networks. We present a data dissemination protocol we call Kalsip, which is designed to support the implementation of our distributed estimation algorithm, based on Kalman filters. We show through analysis and simulation that our protocol provides accurate estimations of physical properties while minimizing the number

of transmissions needed and requiring nodes to only overhear estimates broadcast by their neighbors.

Future work includes implementing our protocol on a sensor network testbed to further validate its viability for use in actual sensing applications, such as environmental monitoring.

Appendix B

List of Publications

Part of the work presented here appeared also in the following articles:

- M. Levorato, F. Librino, M. Zorzi, **"Distributed Cooperative Routing and Hybrid ARQ in MIMO-BLAST Ad Hoc Networks"**, presented at *GLOBECOM 2007*, Washington (Washington DC, USA) November 26-30, 2007.
- S. Del Favero, F. Librino, A. Harris, F. Zorzi, M. Zorzi, **"A Distributed Solution to Estimation Problems in Wireless Sensor Networks Leveraging Broadcast Communications"**, presented at *Wons 2009*, Salt Lake City (Utah, USA), February 2-4, 2009;
- M. Levorato, F. Librino, M. Zorzi, **"Cooperation in UMTS Cellular Networks: a Practical Perspective"**, presented at *Pimrc 2008*, Cannes (France), September 15-18, 2008.
- M. Levorato, F. Librino, M. Zorzi, **"Performance Analysis and Resource Allocation in CDMA Cellular Networks with Relay Stations"**, presented at *ICC 2009*, Dresden (Germany), June 14-18, 2009.
- M. Levorato, F. Librino, M. Zorzi, **"An Algorithmic Solution for Computing Circles Intersection Areas and its Applications to Wireless Communications"**, presented at *WiOpt 2009*, Seoul (South Korea), June 23-27, 2009.
- L. Badia, M. Levorato, F. Librino, M. Zorzi, **"Cooperation Techniques for Wireless Systems from a Network-wide Perspective"**, accepted for publication on *Wireless Communication Magazine*, to appear in 2010;
- M. Levorato, F. Librino, M. Zorzi, **"Integrated Cooperative Opportunistic Packet Forwarding and Distributed Error Control in MIMO Ad Hoc Networks"**, under review for *Transactions on Communications*, revised version submitted on December 2009;

Bibliography

[1] IEEE Standards Department, *ANSI / IEEE Standard 802.11*. IEEE Press, 1999.

[2] F. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part II–The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution," *IEEE Trans. Commun.*, vol. 23, no. 12, pp. 1417–1433, Dec. 1975.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[4] A. Khandani, J. Abounadi, E. Modiano, and L. Zheng, "Cooperative Routing in Static Wireless Networks," *IEEE Trans. Commun.*, vol. 55, no. 11, pp. 2185 – 2192, Nov. 2007.

[5] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[6] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity–part I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[7] ——, "User cooperation diversity–part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[8] F. Librino, M. Levorato, and M. Zorzi, "Distributed cooperative routing and hybrid ARQ in MIMO-BLAST ad hoc networks," in *Proc. IEEE GLOBECOM 2007*, Washington DC, USA, Nov. 26 – 30 2007, pp. 657–662.

[9] A. Hedayat, T. E. Hunter, and A. Nosratinia, "Cooperative communications in wireless communications," *IEEE Commun. Mag.*, vol. 42, pp. 74–80, Oct. 2004.

[10] B. Zhao and M. C. Valenti, "Practical relay networks: a generalization of hybrid–ARQ," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, pp. 7–18, Jan. 2005.

[11] M. Levorato, S. Tomasin, and M. Zorzi, "Cooperative spatial Multiplexing for Ad Hoc Networks with Hybrid ARQ: Design and Performance Analysis," *IEEE Trans. Commun.*, vol. 56, no. 9, pp. 1545–1555, Sept. 2008.

[12] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative Strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[13] M. Felegyhazi, J. P. Hubaux, and L. Buttyan, "Nash equilibria of packet forwarding strategies in wireless ad hoc networks," *IEEE Commun. Mag.*, vol. 5, no. 5, pp. 463–476, May 2006.

[14] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V–BLAST: an architecture for realizing very high data rates over the rich–scattering wireless channel," in *Proc. of IEEE ISSSE*, Pisa, Italy, Sept. 1998, pp. 295–300.

[15] R. R. Choudhury, X. Yang, R. Ramanathan, and N. H. Vaidya, "On designing MAC protocols for wireless networks using directional antennas," *IEEE Trans. Mobile Comput.*, vol. 5, no. 5, pp. 477–491, May 2006.

[16] R. R. Choudhury and N. H. Vaidya, "Deafness: a MAC problem in ad hoc networks when using directional antennas," in *Proc. of IEEE ICNP*, Oct. 2004.

[17] R. Ramanathan, J. Redi, C. Santivanez, D. Viggins, and S. Polit, "Ad hoc networking with directional antennas: a complete system solution," *IEEE J. Select. Areas Commun.*, vol. 23, no. 3, pp. 496–506, Mar. 2005.

[18] A. Nasipuri, S. Ye, J. You, and R. E. Hiromoto, "A MAC protocol for mobile ad hoc networks using directional antennas," in *Proc. of IEEE WCNC*, vol. 2, Chicago, IL, Sept. 2000, pp. 1214–1219.

[19] T. Korakis, G. Jakllari, and L. Tassiulas, "A MAC protocol for full exploitation of directional antennas in ad hoc wireless networks," in *Proc. of ACM MobiHoc*, Annapolis, MD, June 2003.

[20] F. Rossetto and M. Zorzi, "A low-delay MAC solution for MIMO ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 130–135, Jan. 2009.

[21] H. Jafarkhani, *Space–Time Coding: Theory and Practice*. Cambridge University Press, Sept. 2005.

[22] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space–Time Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003.

[23] M. Park, S.-H. Choi, and S. M. Nettles, "Cross–layer MAC design for wireless networks using MIMO," in *Proc. of IEEE GlobeCom*, vol. 2, St. Louis, MO, Nov. 2005, pp. 938–942.

[24] P. Casari, M. Levorato, and M. Zorzi, "MAC/PHY Cross-Layer Design of MIMO Ad Hoc Networks with Layered Multiuser Detection," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4596–4607, Nov. 2008.

[25] 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*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[26] T. E. Hunter, S. Sanayei, and A. Nosratinia, "Outage analysis of coded cooperation," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 375–391, Feb. 2006.

[27] A. Host-Madsen, "Capacity bounds for cooperative diversity," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1522–1544, Apr. 2006.

[28] L. Lai, K. Liu, and H. E. Gamal, "The three–node wireless network: achievable rates and cooperation strategies," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 805–828, Mar. 2006.

[29] X. Qin and R. A. Berry, "Distributed approaches for exploiting multiuser diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 392–413, Feb. 2006.

[30] P. Liu, Z. Tao, and S. Panwar, "A Cooperative MAC Protocol for Wireless Local Area Networks," in *Proc. IEEE ICC*, Seoul, Korea, May 16 – 20 2005, pp. 2962–2968.

[31] P. Liu, Z. Tao, Z. Lin, E. Erkip, and S. Panwar, "Cooperative Wireless Communications: A Cross-Layer Approach," *IEEE Trans. Wireless Commun.*, vol. 52, no. 8, pp. 84–92, Aug. 2006.

[32] T. E. Hunter and A. Nosratinia, "Cooperative diversity through coding," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Lausanne, Switzerland, June 2002, p. 220.

[33] A. Stefanov and E. Erkip, "Cooperative coding for wireless networks," *IEEE Trans. on Commun.*, vol. 52, no. 9, pp. 1470–1476, Sept. 2004.

[34] Z. Lin, E. Erkip, and A. Stefanov, "Cooperative regions and partner choice in coded cooperative systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1323–1334, July 2006.

[35] R. Liu, P. Spasojevic, and E. Soljanin, "User cooperation with punctured turbo codes," in *Proc. 41st Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Oct. 2003.

[36] A. Stefanov and E. Erkip, "Cooperative space-time coding for wireless networks," *IEEE Trans. on Commun.*, vol. 53, no. 11, pp. 1804–1809, Nov. 2005.

[37] P. Coronel, R. Doss, and W. Schott, "Location–based cooperative relaying in wireless sensor networks," in *IEEE/ACM SensorWare*, Bangalore, India, Jan. 2007.

[38] M. H. Cheung and T. M. Lok, "Cooperative routing in UWB wireless networks," in *Proc. IEEE Int. Symp. on Wireless Communications and Networking Conference (WCNC) 2007*, Hong Kong, China, Mar. 11 – 15 2007, pp. 1740–1744.

[39] C. Chen, B. Zheng, X. Zhao, and Z. Yan, "A Novel Weighted Cooperative Routing Algorithm Based on Distributed Relay Selection," in *Proc. IEEE 2nd International Symposium on Wireless Pervasive Computing*, San Juan, Puerto Rico, Feb. 5 – 7 2007, pp. 224 – 229.

[40] S. Zhu and K. Leung, "Distributed Cooperative Routing for UWB Ad-Hoc Networks," in *Proc. IEEE ICC*, Glasgow, Scotland, UK, June 24 – 28 2007, pp. 3339 – 3344.

[41] X. Fang, T. Hui, Z. Ping, and Y. Ning, "Cooperative routing strategies in ad hoc networks," in *Proc. IEEE 61st Vehicular Technology Conference, VTC Spring*, vol. 4, Stockholm, Sweden, May 30 – Jun. 1 2005, pp. 2509 – 2512.

[42] Z. Yang, J. Liu, and A. Host-Madsen, "Cooperative routing and power allocation in ad-hoc networks," in *Proc. IEEE GLOBECOM 2005*, vol. 5, S. Louis, Missouri, USA, Nov. 28 – Dec. 2 2005, pp. 2730 – 2734.

[43] L. Liu and H. Ge, "Space-time coding for wireless sensor networks with cooperative routing diversity," in *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, vol. 1, Pacific Grove, California, USA, Nov. 7 – 10 2004, pp. 1271 – 1275.

[44] F. Li, K. Wu, and A. Lippman, "Energy-efficient cooperative routing in multi-hop wireless ad hoc networks," in *IEEE Proc. 25th International Performance, Computing, and Communications Conference IPCCC 2006*, Phoenix, Arizona, USA, Apr. 10 – 12 2006, pp. 215 – 222.

[45] A. S. Ibrahim, Z. Han, and K. J. R. Liu, "Distributed energy-efficient cooperative routing in wireless networks," in *Proc. IEEE GLOBECOM 2007*, Washington DC, USA, Nov. 26 – 30 2007, pp. 4413–4418.

[46] S. Sfar, R. D. Murch, and K. B. Letaief, "Layered space–time multiuser detection over wireless uplink systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 653–668, July 2003.

[47] G. Ginis and J. M. Cioffi, "On the relation between V-BLAST and the GDFE," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 364–366, Sept. 2001.

[48] M. Levorato, S. Tomasin, P. Casari, and M. Zorzi, "Physical layer approximation for cross–layer performance analysis in MIMO–BLAST ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 12, pp. 4390–4400, Dec. 2007.

[49] R. E. Blahut, *Theory and practice of error control codes*. MA: Addison Wesley, 1984.

[50] L. Rizzo, "Effective erasure codes for reliable computer communications protocols," *ACM Computer Commun. Review*, vol. 27, no. 2, pp. 24–36, Apr. 1997.

[51] H. H. st Madsen, "Capacity bounds and power allocation for wireless relay channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2020–2040, June 2005.

[52] S. A. Zummo, "Performance analysis of coded cooperation diversity in wireless networks," in *Proc. IEEE ICC*, Istanbul, Turkey, June 11 – 15 2006, pp. 4560–4565.

[53] E. Yeh and R. Berry, "Throughput optimal control of cooperative relay networks," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Adelaide, Australia, Sept. 4 – 9 2005, pp. 1206–1210.

[54] T. E. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 283–289, Feb. 2006.

[55] ——, "Cooperation in wideband channels," in *Proc. IEEE 6th Workshop on Signal Processing Advances in Wireless Communications*, New York, June 5 – 8 2005, pp. 955–959.

[56] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: Space–time transmission and iterative decoding," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 362–371, Feb. 2004.

[57] P. Liu, T. Zao, Z. Lin, E. Erkip, and S. Panwar, "Cooperative wireless communications: A cross–layer approach," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 84–92, Aug. 2006.

[58] A. K. Sadek, Z. Han, and K. J. R. Liu, "An efficient cooperation protocol to extend coverage range area in cellular networks," in *Proc. IEEE WCNC 2006*, Las Vegas, Nevada, Apr. 3 – 6 2006, pp. 1687–1692.

[59] J. Zhang and Q. Zhang, "Cooperative routing in multi-source, multi-destination, multi-hop wireless networks," in *Proc. IEEE INFOCOM 2008*, Phoenix, Arizona, USA, Apr. 13 – 18 2008, pp. 2369–2377.

[60] W. Chen, L. Dai, K. B. Letaief, and Z. Cao, "A unified cross-layer framework for resource allocation in cooperative networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3000–3012, Aug. 2008.

[61] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Vehicular Technology*, vol. 29, no. 3, pp. 317–325, Aug. 1980.

[62] H. Holma and A. Toskala, *WCDMA for UMTS*. England: John Wiley & Sons, Ltd, 2004.

[63] R. Mathar and T. Niessen, "Optimum positioning of base stations for cellular radio networks," *Wireless Networks*, vol. 6, no. 6, pp. 421–428, Dec. 2000.

[64] H. Zhang and J. C. Hou, "Maintaining sensing coverage and connectivity in large sensor networks," *Ad Hoc & Sensor Wireless Networks*, vol. 1, pp. 89–124, Mar. 2005.

[65] M. Heissenbüttel, T. Braun, T. Bernoulli, and M. Wälchli, "BLR: beacon-less routing algorithm for mobile ad hoc networks," *Computer Communications*, vol. 27, no. 11, pp. 1076–1086, July 2004.

[66] P. Casari, P. Nati, C. Petrioli, and M. Zorzi, "Efficient non-planar routing around dead ends in sparse topologies using random forwarding," in *Proc. of IEEE ICC*, Glasgow, Scotland, June 2007.

[67] M. P. Fewell, "Area of common overlap of three circles," Maritime Operations Division, Australian Government, Department of Defence, Tech. Rep. DSTO-TN-0722, Oct. 2006. [Online]. Available: http://hdl.handle.net/1947/4551

[68] F. Librino and M. Levorato and M. Zorzi, "An Algorithmic Solution for Computing Circle Intersection Areas and its Applications to Wireless Communications," Technical report, 2009. [Online]. Available: \emph{www.dei.unipd.it/{$\sim$}\!levorato/techrep/techrep-intersections.pdf}

[69] "Special issue on sensor networks and applications," *Proc. IEEE*, Aug. 2003.

[70] "Special issue on sensor networks," *Int. Journal High Performance Comp. Application*, vol. 16, no. 3, fall 2002.

[71] "Special issue on sensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 3, Mar. 2002.

[72] Brian D.O. Anderson and John B. Moore, *Optimal Filtering*. Dover Publications, Jan. 2005.

[73] Robert F. Stengel, *Optimal Control and Estimation*. Dover Publications, reissue, Sept. 1994.

[74] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[75] R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Proc. IEEE 44th CDC-ECC*, Seville, Spain, Dec. 12 – 15 2005, pp. 6698–6703.

[76] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *Proc. IEEE 44th CDC-ECC*, South Bend, IN, Dec. 2005, pp. 8179–8184.

[77] P. Alriksson and A. Rantzer, "Distributed Kalman filtering using weighted average," in *Proc. 17th International Symposium on Mathematical Theory of Networks and Systems*, Kyoto, Japan, July 24 – 28 2006.

[78] A. Speranzon, C. Fischione, and K. H. Johansson, "Distributed and collaborative estimation over wireless sensor networks," in *Proc. IEEE 45th Conference on Decision and Control*, San Diego, California, Dec. 13 – 15 2006, pp. 1025–1030.

[79] C. Moallemi and B. V. Roy, "Consensus propagation," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 4753–4766, Nov. 2006.

[80] X. Li, K. Moaveninejad, and O. Frieder, "Regional gossip routing for wireless ad hoc networks," in *Proc. IEEE 28th Annual International Conference on Local Computer Networks*, Königswinter, Germany, Oct. 20 – 24 2003, pp. 274–275.

[81] R. Karp, S. S. C. Schindelhauer, and B. Vöcking, "Randomized rumor spreading," in *Proc. IEEE 41st Annual Symposium on Foundations of Computer Science*, Redondo Beach, California, Nov. 12 – 14 2000, pp. 565–574.

[82] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. IEEE 44th Annual Symposium on Foundations of Computer Science*, Cambridge, Massachussets, Oct. 11 – 14 2003, pp. 482–491.

[83] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filter based on consensus strategies," *IEEE J. Select. Areas Commun.*, vol. 26, no. 4, pp. 622–633, May 2008.

[84] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks – part I: gaussian case," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[85] ——, "Bandwidth-constrained distributed estimation for wireless sensor networks – part II: unknown probability density function," *IEEE Trans. Signal Processing*, vol. 54, no. 7, pp. 2784–2796, July 2006.

[86] ——, "Distributed Kalman filtering based on severely quantized WSN data," in *Proc. IEEE 13th workshop on Statistical Signal Processing*, Bordeaux, France, July 17 – 20 2005, pp. 1250–1255.

[87] ——, "SOI–KF: distributed Kalman filtering with low–cost communications using the sign of innovation," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 4782–4795, Dec. 2006.

[88] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Approximate distributed Kalman filtering in sensor networks with quantifiable performance," in *Proc. IEEE 4th International Symposium on Information Processing in Sensor Networks*, Los Angeles, CA, Apr. 25 – 27 2005, pp. 133–139.