Operating Regimes of Large Wireless Networks

Operating Regimes of Large Wireless Networks

Ayfer Özgür
Olivier Lévêque

Ecole Polytechnique Fédérale de Lausanne
Faculté Informatique et Communications
Lausanne 1015
Switzerland
{ayfer.ozgur,olivier.leveque}@epfl.ch

David Tse

University of California at Berkeley
Berkeley, CA 94720
USA
dtse@eecs.berkeley.edu

now
the essence of knowledge
Boston – Delft

Full text available at: http://dx.doi.org/10.1561/1300000016

Foundations and Trends® in Networking

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is A. Özgür, O. Lévêque and D. Tse, Operating Regimes of Large Wireless Networks, Foundation and Trends® in Networking, vol 5, no 1, pp 1–107, 2010

ISBN: 978-1-60198-502-6
© 2011 A. Özgür, O. Lévêque and D. Tse

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc. for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Foundations and Trends® in Networking

Volume 5 Issue 1, 2010

Editorial Board

Editorial Scope

Foundations and Trends® in Networking will publish survey and tutorial articles in the following topics:

- Ad Hoc Wireless Networks
- Sensor Networks
- Optical Networks
- Local Area Networks
- Satellite and Hybrid Networks
- Cellular Networks
- Internet and Web Services
- Protocols and Cross-Layer Design
- Network Coding
- Energy-Efficiency Incentives/Pricing/Utility-based
- Games (co-operative or not)
- Security
- Scalability
- Topology
- Control/Graph-theoretic models
- Dynamics and Asymptotic Behavior of Networks

Information for Librarians

Foundations and Trends® in Networking, 2010, Volume 5, 4 issues. ISSN paper version 1554-057X. ISSN online version 1554-0588. Also available as a combined paper and online subscription.

Foundations and Trends® in
Networking
Vol. 5, No. 1 (2010) 1–107

DOI: 10.1561/1300000016

Operating Regimes of Large Wireless Networks

Ayfer Özgür[1], Olivier Lévêque[2] and David Tse[1]

[1] *Ecole Polytechnique Fédérale de Lausanne, Faculté Informatique et Communications, Lausanne, 1015, Switzerland, {ayfer.ozgur,olivier.leveque}@epfl.ch*
[2] *University of California at Berkeley, Department of EECS, Berkeley, CA, 94720, USA, dtse@eecs.berkeley.edu*

Abstract

Multi-hop is the current communication architecture of wireless mesh and *ad hoc* networks. Information is relayed from each source to its destination in successive transmissions between intermediate nodes. A major problem regarding this architecture is its poor performance at large system size: as the number of users in a wireless network increases, the communication rate for each user rapidly decreases. Can we design new communication architectures that significantly increase the capacity of large wireless networks?

In this monograph, we present a scaling law characterization of the information-theoretic capacity of wireless networks, which sheds some light on this question. We show that the answer depends on the parameter range in which a particular network lies, namely the operating regime of the network. There are operating regimes where the information-theoretic capacity of the network is drastically higher than the capacity of conventional multi-hop. New architectures can provide

substantial capacity gains here. We determine what these regimes are and investigate the new architectures that are able to approach the information-theoretic capacity of the network. In some regimes, there is no way to outperform multi-hop. In other words, the conventional multi-hop architecture indeed achieves the information-theoretic capacity of the network. We discuss the fundamental factors limiting the capacity of the network in these regimes and provide an understanding of why conventional multi-hop indeed turns out to be the right architecture.

The monograph is structured as follows: In Section 2, we discuss the role of interference in wireless networks. We show that while current communication architectures are fundamentally limited by interference, new architectures based on distributed MIMO communication can overcome this interference limitation, yielding drastic performance improvements. Section 3 discusses the impact of power. We show that in power-limited regimes, distributed MIMO-based techniques are important not only because they remove interference but also because they provide received power gain. We identify the power-limited operating regimes of wireless networks and define the engineering quantities that determine the operating regime of a given wireless network. We show that unless the wireless network operates in a severely power-limited regime, distributed MIMO communication provides significant capacity gain over current techniques. Finally, in Section 4, we study how the area of the network, i.e., space, impacts the capacity of the network. This study enriches the earlier picture by adding new operating regimes where wireless networks can be moderately or severely space-limited. We see that unless the network is severely limited in space, distributed-MIMO-based communication continues to provide drastic improvements over conventional multi-hop.

Contents

1

Introduction

In wired networks, a source can send information to a destination by routing it along a path, where intermediate nodes forward the information towards the destination. The application of this strategy to wireless networks has been the subject of a large body of research in the past two decades. Similar to wired networks, packets are sent here from each source to its destination via multiple intermediate nodes acting as relays. Each relay decodes the packets sent from the previous relay and forwards them to the next.

Multi-hop is a natural fit for wired networks; however, it is not clear whether it provides a good premise for wireless. It is based on point-to-point communication between nodes. Wired networks are already composed of point-to-point links over which signals travel in isolation. However, the notion of a point-to-point link is vague in the case of wireless.

Wireless signals are not isolated and they interact in complex ways. The signal transmitted by a given user is heard not only by its intended receiver but also by all the receivers in the vicinity of the transmitter. When there are multiple simultaneous transmissions over the same frequency band, each receiver observes a mixture of all the transmitted

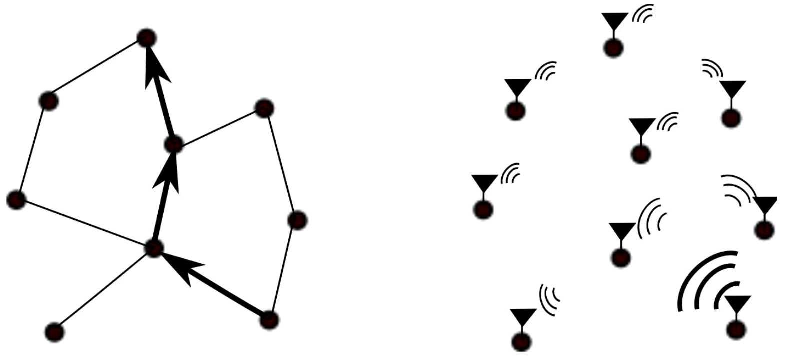

Fig. 1.1 Wired vs. wireless networks.

signals. Therefore, signals of interest to a receiver mix together with overheard signals from other transmissions. As wired and wireless networks are so different in their fundamental nature, it is not clear whether an architecture rooted in the practice of the former can provide a good premise for the latter. (Figure 1.1).

Today, there is an increasing need to connect a massive number of wireless devices and to support various resource-intensive applications. This leads us to especially discuss the performance of the conventional multi-hop architecture in large wireless networks: Can multi-hop efficiently support communication in large wireless networks or do we need new architectures for the rapidly growing wireless networks of the future? In particular, can new architectures tailored for wireless significantly outperform multi-hop in large networks? In this monograph, we study the information-theoretic capacity of large wireless networks, to shed some light on these questions.

1.1 Interference

In this section, we argue that the performance of the conventional multi-hop architecture is fundamentally limited by the interference between simultaneous transmissions in the shared wireless medium. However, this interference limitation is not fundamental and can be overcome with new architectures tailored for wireless. We discuss hierarchical cooperation, an architecture that constructively uses interference for communication. As a result, it offers significant performance gains in large networks. This section provides a summary of Section 2 of this monograph.

1.1.1 Multi-hop is Interference Limited

Multi-hop is based on relaying information from sources to destinations via successive point-to-point transmissions between intermediate nodes. To do the point-to-point transmissions, we need to designate nodes in the network as transmitter–receiver pairs. Each receiver is to decode the message from its designated transmitter. Overheard signals from other transmitters constitute harmful *interference* corrupting the desired signals and are treated as additional noise at the receivers. The choice of these transmitter–receiver pairs in the network is a major optimization problem determining the throughput performance of the multi-hop architecture.

In order to achieve high overall throughput, it is desired to choose the transmitter–receiver pairs such that many of them can communicate simultaneously without interfering too much with each other. This would provide a dense mesh for relaying information inside the network. On the other hand, it is also desirable to have a large separation between every transmitter–receiver pair so that messages advance by a large distance towards their destinations in every hop. The interference between simultaneous transmissions poses a fundamental trade-off between these two trends. Each transmission creates strong interference for other receivers around its transmitter. The radius of this strong interference zone is proportional to the transmitted power, which is in turn proportional to the range of the targeted transmission (Figure 1.2). Therefore, the larger the separation between the transmitter-receiver pairs in the network, the fewer of them can communicate at the same time.

In particular, if we allow for direct transmissions from the source nodes in the network to their destinations, only few of these source-destination pairs can communicate at a time, as source-destination pairs in a network are typically separated by large distances. Consider a network with a large number of users n, where users are randomly

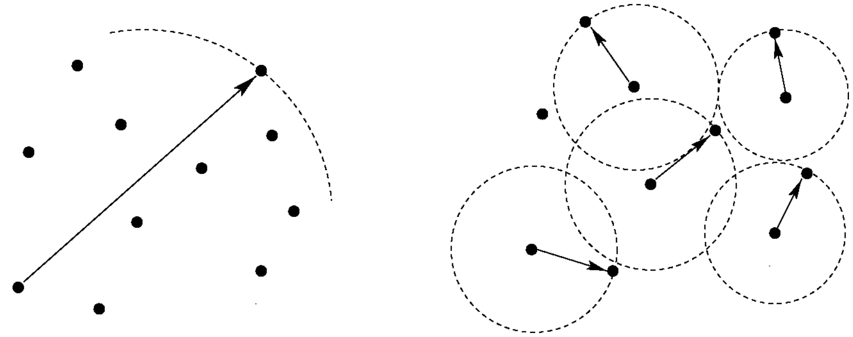

Fig. 1.2 Long vs. short-range communication in wireless networks. The nodes inside each circle are subject to interference from the corresponding transmission.

paired into $n/2$ source-destination pairs. Each source wants to communicate to its corresponding destination node. Such a random pairing will lead to $\Theta(n)$ pairs separated by a distance of the order of the diameter of the network. If source-destination pairs are to communicate directly with each other, these $\Theta(n)$ pairs should go one at a time. The per-pair communication rate with such a time-sharing strategy decreases as $\Theta(1/n)$ with increasing number of users n. Note that each pair gets to transmit once in $\Theta(n)$ time slots.[1]

The other extreme is to confine to nearest-neighbor communication inside the network. As wireless signals get attenuated with distance, many local communications can be simultaneously active without interfering too much with each other (spatial reuse). See Figure 1.2. In particular, confining to nearest-neighbor communication maximizes the number of simultaneous transmissions inside the network. However, to cover long distances in short hops, each packet now has to be retransmitted many times before getting to its final destination. This relaying burden limits the achievable throughput.

In their seminal work [9] in 2000, Gupta and Kumar showed that confining to nearest-neighbor transmissions maximizes the throughput of multi-hop and provides an aggregate throughput of order $\Theta(\sqrt{n})$ in a network of n users. This corresponds to a per-user rate that

---

[1] Here, transmissions are orthogonalized over time so that they do not interfere. Equivalently, transmissions can be orthogonalized in frequency or in code space.

scales as $\Theta(1/\sqrt{n})$. Note that this scaling is significantly better than the $\Theta(1/n)$ scaling with direct communication (single-hop) between source-destination pairs. Nevertheless, it still decreases quite rapidly to zero with an increasing number of users n. This limitation is precisely due to the fact that in a nearest-neighbor multi-hop architecture, most users have to relay information for $\Theta(\sqrt{n})$ source-destination pairs on average.

The sub-linear scaling of the system throughput is fundamentally due to the need to reduce interference between point-to-point transmissions. If transmissions were not interfering, we could have many simultaneous long distance transmissions in the network, ideally every source could directly and simultaneously communicate to its destination. It is because of the interference that we need to confine to short distance communication, in which case the resulting relaying burden limits the system throughput.

1.1.2 Constructive Use of Interference: Hierarchical Cooperation

A natural question is whether we can surpass the interference barrier by allowing more sophisticated cooperation between the nodes, in particular by removing the restriction to point-to-point communication. Can we design cooperation architectures whose performance *scales* with system size? In Section 2, we present a hierarchical cooperation architecture that achieves an aggregate throughput of $\Theta(n^{1-\epsilon})$ for any $\epsilon > 0$. An aggregate throughput scaling arbitrarily close to linear in the number of nodes means that there is essentially no interference limitation: The rate for each source-destination pair does not degrade significantly, even if the network serves a growing number of users. This result demonstrates that the fundamental capacity of wireless networks can be significantly higher than the capacity of multi-hop and that more sophisticated cooperation architectures can provide substantial performance gains in large networks.

The key to this result is distributed MIMO (multi-input multi-output) communication. MIMO is a physical-layer technique, which was originally developed in the classical point-to-point setting. In this setting, multiple antennas are installed on both the transmitter and the receiver. This allows to simultaneously send an independent stream of data from each transmit antenna. Each receive antenna observes a different combination of the transmitted signals. Jointly processing the vector of received observations at the antennas allows the receiver to remove the interference between the transmitted signals and recover the original data streams [5, 29]. A natural approach to apply this concept to the network setting is to have nodes cooperate in *clusters* to form distributed transmit and receive antenna arrays. In this manner, mutually interfering signals can be turned into useful ones that can be jointly decoded at the receive cluster and spatial multiplexing gain can be realized.

One way to incorporate distributed MIMO communication is to transfer the packets of each source node to its destination in three consecutive phases: The packets of a source node are first distributed among a cluster of nodes in its vicinity. In a second phase, the nodes in this source cluster simultaneously transmit these packets to a group of nodes around the destination node. These simultaneous transmissions can be regarded as distributed MIMO communication if the observations of the various nodes in the destination cluster can be jointly processed. Therefore, in a third phase, the distributed MIMO observations should be collected at the actual destination node, which can then jointly process these observations and recover the packets from its source node.

The above strategy potentially offers performance gain via the simultaneous long distance transmissions in the second phase. The interference between these transmissions is not anymore harmful, as they are jointly decoded at the end of the third phase. However, the overhead introduced by the first and the third phases to establish the necessary transmit and receive cooperation can drastically reduce the useful throughput. The key to efficient cooperation in the first and third phases is a digital and *hierarchical* architecture that makes use of distributed MIMO communication at increasing scales. Cooperation first takes place between nodes within small local clusters. These small clusters can operate simultaneously, as the decay of signals with distance allows spatial reuse. The cooperation facilitates MIMO

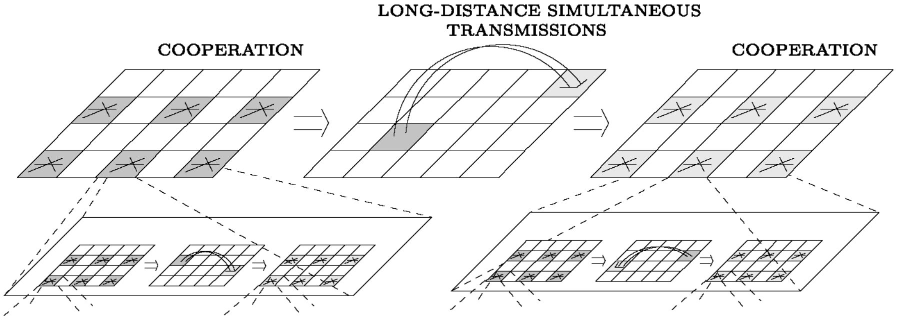

Fig. 1.3 The salient features of the hierarchical cooperation architecture.

communication over a larger spatial scale. This can then be used as a communication infrastructure for cooperation within larger clusters at the next level of the hierarchy. Continuing in this fashion, cooperation can be achieved at an almost global scale. At the highest level of the hierarchy, long-range MIMO communications can be performed between clusters almost as large as the whole network. By increasing the number of levels in such a hierarchical architecture, one can get arbitrarily close to linear aggregate throughput scaling. Figure 1.3 illustrates the hierarchical architecture with a focus on the top two levels.

The distributed MIMO-based approach summarized above is closely related to physical-layer network coding. Physical-layer network coding [11, 34] is another recent paradigm in wireless networking, based on the same motivation to embrace the wireless interference instead of avoiding it. Physical-layer network coding allows for two strategically picked transmissions to interfere at a relay node, which then forwards the mixture of the two signals. The fundamental difference between distributed MIMO-based hierarchical cooperation and physical-layer network coding is the scale over which wireless interference is embraced. Physical-layer network coding maintains the multi-hop architecture at the global scale and allows two local transmissions to interfere at each hop. Such an approach has the potential to double the throughput of the network, but no more (this was shown precisely in Ref. [17]). In the hierarchical cooperation architecture, communication is organized so that wireless interference can be embraced at the global scale.

It can be viewed as an aggressive form of physical-layer network coding, where $\Theta(n)$ transmissions are allowed to interfere instead of only two. Consequently, the gain is more substantial: instead of doubling the aggregate throughput, we can elevate its scaling from $\Theta(\sqrt{n})$ to linear in n.

1.2 Power

Interference is not the only factor that can potentially limit performance in wireless networks. Power can be another limiting factor. In some wireless networks, the reason to confine to short-range communication and relay packets via multiple transmissions may not be the interference that would be caused by long distance communication. The attenuation of wireless signals with distance may not allow sufficient received SNR (signal-to-noise power ratio) to directly reach far-away destinations. This can be the case due to a number of reasons:

(a) The power available at the nodes can be limited.
(b) The network can be distributed over a large geographical area.
(c) The attenuation in the environment can be high.
(d) The network can be operating on a large bandwidth (wideband system).

The objective in such wireless networks is not only to deal with interference but also to transfer power efficiently to the receivers. In particular, in an extremely power-limited network interference may be far below the noise level at the receivers. In such a regime, the strategies that provide the best throughput would be the ones that utilize power most efficiently. In this section, we discuss the question we raised in the previous section by also putting *power* into play: Is the traditional multi-hop architecture able to efficiently transfer power in large wireless networks? Can more sophisticated architectures, for example, hierarchical cooperation, provide significant capacity improvement in *power-limited* wireless networks?

The restriction to point-to-point communication in the traditional multi-hop setting can now be questioned from the *power* point of view.

In point-to-point communication, the signals received from a particular transmission are treated as noise at all but one receiver inside the network. In the previous section, we have seen that with distributed MIMO-based communication, we are able to turn mutually interfering signals into useful ones. By exploiting the broadcast nature of the wireless medium, such techniques can provide a received power gain, in addition to the spatial multiplexing gain emphasized in the previous section. This power gain can translate into a significant capacity gain in certain power-limited networks. The impact of power is discussed in detail in Section 3. We provide below a short summary of the conclusions of this section.

1.2.1 Impact of Power in the Point-to-point Wireless Channel

To understand the impact of power in wireless networks, let us first review how the amount of available transmission power impacts the capacity of the point-to-point additive white Gaussian noise channel. The capacity of this channel is given by Shannon's famous formula

$$C = W \log\left(1 + \frac{P}{N_0 W}\right) \tag{1.1}$$

in terms of the bandwidth of the channel W in Hz, received power P in Watts, and noise power spectral density $N_0/2$ in Watts/Hz.

The most important engineering parameter we associate with this channel is SNR defined as

$$\mathrm{SNR} = \frac{P}{N_0 W}.$$

This parameter determines the operating regime of the channel. When $\mathrm{SNR} \ll 0\,\mathrm{dB}$, the channel is in a power-limited regime: the capacity is approximately linear in the power, and the performance depends critically on the power available, but not so much on the bandwidth. In this regime, if we double the transmit power, we can approximately double the channel capacity; however, doubling the bandwidth only marginally improves capacity. In the bandwidth-limited (or high-SNR) regime, where $\mathrm{SNR} \gg 0\,\mathrm{dB}$, we have the opposite situation: the capacity is approximately linear in the bandwidth and the performance depends critically on the bandwidth, but not so much on the power. These two observations can be immediately verified from the capacity formula in Equation (1.1), noting that when $\mathrm{SNR} \ll 0\,\mathrm{dB}$, $\log(1+x) \approx x$ and when $\mathrm{SNR} \gg 0\,\mathrm{dB}$, the logarithm function gets saturated and increases very slowly in its argument.

These two fundamentally different operating regimes have two completely different implications in terms of communication system design. For a bandwidth-limited channel, the least we would expect from a good communication strategy for this channel is that its performance is approximately linear in the bandwidth, i.e., able to follow the trend of the capacity. On the contrary, for a power-limited channel, we should design a strategy whose performance increases linearly in the power. In the sequel, we will call a strategy scaling optimal or simply optimal for a certain regime, if its performance exhibits approximately the same dependence to system parameters as the information-theoretic capacity of the system.[2] Note that there is no guarantee that a strategy which is scaling optimal for a certain regime, meaning that its performance exhibits approximately the right behavior in terms of system parameters in this regime, would also be optimal for another regime.

1.2.2 Impact of Power in Wireless Networks

The interference discussion of the earlier section was implicitly based on a regime where the capacity of the wireless network is bandwidth-limited. The basis for the discussion was the scaling law approach of Gupta and Kumar [9], which looks at how the capacity of the network scales with the number of users. As the number of users in the wireless network increases, the other parameters of the network, such as area, bandwidth, per-user power, are kept fixed. This scaling results in a large network whose information-theoretic capacity is approximately given by nW. While the capacity of multi-hop in this regime behaves as $\sqrt{n}W$, the capacity of the new hierarchical cooperation strategy behaves as nW. This makes hierarchical cooperation scaling optimal in this regime.

---

[2] The approximation is within a poly-logarithmic factor.

The discussion on the operating regimes of the point-to-point wireless channel suggests that we could also have power-limited operating regimes in wireless networks. In this case, however, the capacity exhibits a completely different behavior. Indeed, power turns out to be a more sophisticated player in wireless networks than in the point-to-point case. There are a number of fundamentally different power-limited regimes in wireless networks. This is first due to the fact that the power limitation is jointly determined by a number of independent parameters (a)–(d) listed above. These parameters have different impact and their interplay creates a number of qualitatively different cases. For example, a network that suffers power limitation due to high attenuation in the environment is not equivalent to (cannot be translated to) a network that suffers from limited power available at the wireless nodes. Second, a wireless network can be power-limited in different degrees. For example, in a severely power-limited wireless network, channels between all pairs of nodes in the network are weak (of low SNR). In less severe cases, only the channels between far-away pairs are weak, whereas close-by nodes are connected via strong channels (of high SNR).

The backbone of the hierarchical cooperation architecture introduced in the previous section is distributed MIMO communication: at the highest level of the hierarchy, we perform simultaneous long distance transmissions from a source cluster of $\Theta(n)$ nodes to a destination cluster of $\Theta(n)$ nodes. The transmissions from each node in the source cluster are heard by all the nodes inside the destination cluster, though these $\Theta(n)$ simultaneous transmissions interfere with each other. When the interference between these transmissions is removed via joint decoding at the destination node, power-wise, it is as if we were able to observe each transmission interference-free at $\Theta(n)$ different receivers. In other words, for each transmission, the hierarchical cooperation architecture collects the power received by the $\Theta(n)$ nodes inside the destination cluster.

This leads to the following interesting fact: *A priori*, we may expect to observe some sort of power limitation in a wireless network if the received SNR between some pair of nodes in the network is not sufficient for direct communication, most notably between far-away pairs. However, the information-theoretic capacity of the network is bandwidth-limited and not power-limited, approximately given by nW, as long as n *times* the SNR between far-away pairs is larger than 0 dB. We define this quantity as the long distance SNR of the network, denoted as SNR_l: it is n *times* the received SNR of a point-of-point channel with the transmitter and receiver separated by a distance equal to the diameter of the network. Note that the diameter defines the largest geographical scale for communication inside the network. As long as $\mathrm{SNR}_l \gg 0\,\mathrm{dB}$, the wireless network is bandwidth-limited and hierarchical cooperation is scaling optimal. This is not only because hierarchical cooperation can handle interference efficiently as discussed in the earlier section but also because it is able to efficiently exploit the broadcasting nature of the wireless medium in this regime.

A wireless network starts to experience power limitation when the long distance SNR drops below 0 dB. When $\mathrm{SNR}_l \ll 0\,\mathrm{dB}$, the network is power-limited over the largest geographical scale but can still be bandwidth-limited over a shorter communication scale. The optimal cooperation architecture is determined by two parameters in this case. The first one is the power path loss exponent of the environment, α. It describes how fast signal power decays with distance: signals transmitted from one node to another at distance r apart are subject to a power loss of $r^{-\alpha}$, where typically $2 \leq \alpha \leq 6$. $\alpha = 2$ corresponds to free-space propagation and larger α to more lossy environment. The power path loss exponent defines a dichotomy: when $2 \leq \alpha < 3$, the hierarchical cooperation architecture transfers power optimally inside the network and achieves the information-theoretic capacity scaling of the network. Signal power decays slowly with distance in this case and hierarchical cooperation yields maximal received power by collecting the received signals of $\Theta(n)$ nodes around each destination node.

When $\alpha \geq 3$, signal power decays fast with distance and long distance communication in the network is not preferable, even with its additional $\Theta(n)$ power gain. The optimal architecture depends on the strength of the power limitation in the network, which is captured by a second SNR parameter, the short distance SNR, denoted by SNR_s. SNR_s is the received SNR in a point-to-point transmission over the typical nearest-neighbor distance inside the network. The nearest-neighbor distance is the shortest scale for communication inside the network. When $\mathrm{SNR}_s \ll 0\,\mathrm{dB}$, communication over even the shortest geographical scale is limited in power. In this case, the conventional nearest-neighbor multi-hop architecture is the fundamentally right strategy for transferring power; it indeed achieves the information-theoretic scaling of the network capacity. The broadcasting nature of the wireless media plays insignificant role in such severely power-limited networks and therefore, confining to point-to-point communication is not anymore suboptimal.

When $\alpha \geq 3$, but $\mathrm{SNR}_s \gg 0\,\mathrm{dB}$, the nearest-neighbor scale is bandwidth-limited. Note that $\mathrm{SNR}_l \ll 0\,\mathrm{dB}$; hence, the network is still power-limited over distances of the order of the network diameter. In this case, the broadcasting of wireless signals is significant up to an intermediate geographical scale determined by the precise value of SNR_s and α. There is therefore the potential to improve performance with long distance communication up to this particular geographical scale. Beyond this scale, the network is power-limited and power attenuates rapidly for $\alpha > 3$; hence, communication over longer distances is inefficient. The optimal solution is to form MIMO clusters of an intermediate size and then multi-hop across several clusters to get to the final destination cluster. Each hop between adjacent clusters is now performed using distributed MIMO transmissions of the corresponding intermediate scale. This hybrid architecture is illustrated in Figure 1.4.

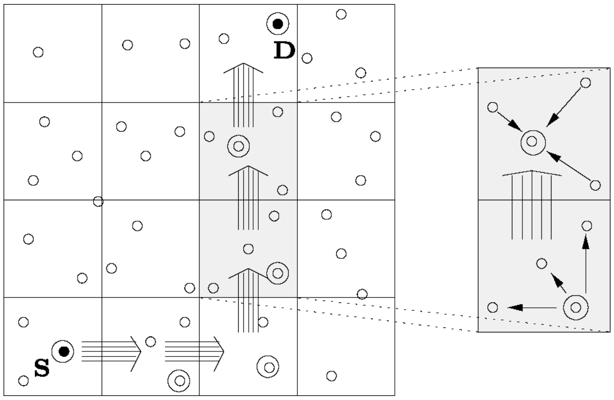

Fig. 1.4 Cooperate locally multi-hop globally: A generic optimal architecture for wireless networks.

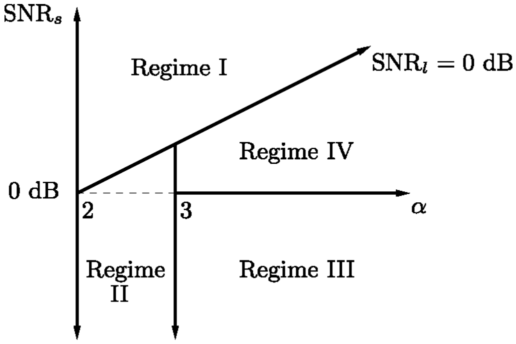

Fig. 1.5 The four operating regimes. The optimal schemes in these regimes are I–II: Hierarchical cooperation, III — Multi-hop, IV — Hybrid Multi-hop + Hierarchical Cooperation.

The two extremes of this architecture are precisely traditional multi-hop, where the cluster size is 1 and the number of hops is $\Theta(\sqrt{n})$, and hierarchical cooperation, where the cluster size is $\Theta(n)$ and the number of hops is 1. This hybrid architecture combining hierarchical cooperation with multi-hop provides a generic optimal solution for all wireless networks. For optimality, the cooperation scale should be adjusted according to the power available in the network and the power path loss exponent of the environment. The resultant four operating regimes and the corresponding optimal schemes for each regime are illustrated in Figure 1.5.

1.3 Space

The geographical area of the network not only plays a role in determining the received powers in the network, but also has an independent impact on capacity. It determines the number of independent spatial channels available for communication inside the wireless network. Information is communicated in the form of electromagnetic waves and the area of the network determines the diversity available in the physical channel. Consider the $\Theta(n)$ simultaneous long distance transmissions between the source and the destination clusters in the hierarchical cooperation architecture. Each node in the destination cluster observes a linear combination of the transmitted electromagnetic signals, each scaled and shifted according to the loss and delay in the corresponding path. The destination node can only remove the interference between these transmissions via joint decoding, if the linear combinations of the signals are independent. When the $\Theta(n)$ nodes in the source cluster and the $\Theta(n)$ nodes in the destination cluster are packed together in small geographical areas, the linear combinations of the transmitted signals cannot be anymore independent. In this section, we reconsider the question raised in the earlier sections by concentrating on the impact of space on the capacity of wireless networks.

As we discuss in detail in Section 4, there are $\Theta(\sqrt{A}/\lambda)$ spatial degrees of freedom in a wireless network of area A, operating on a carrier wavelength λ. This is the number of independent spatial channels available for communication inside the network. Limited by interference, the multi-hop architecture can only achieve $\Theta(\sqrt{n})$ degrees of freedom. If the number of spatial degrees of freedom in the network is already as small as $\Theta(\sqrt{n})$, then multi-hop is fundamentally optimal, as it is able to achieve the full degrees of freedom of the network. When the available degrees of freedom in the network are more than $\Theta(\sqrt{n})$, there is the potential to exploit these additional degrees of freedom by more sophisticated cooperation. We will discuss in Section 4 that when the number of spatial degrees of freedom in the network is larger than $\Theta(\sqrt{n})$, the hierarchical cooperation architecture is able to achieve the full degrees of freedom in the network given by

$$\min(n, \sqrt{A}/\lambda).$$

In particular in wireless networks where $\sqrt{A}/\lambda \gg n$, there is no space limitation, as there are sufficient spatial degrees of freedom for all users. Hierarchical cooperation achieves linear aggregate throughput scaling in this case.

1.4 Operating Regimes

In this monograph, we discuss three factors that can potentially limit performance in wireless networks. We have already seen that the first one, interference, usually thought of to be a major performance limitation in wireless networks, can be overcome with cooperation between nodes. The latter two, power and space, impose fundamental limitations on communication in wireless networks.

Can more sophisticated cooperation techniques provide significant capacity gains over the conventional multi-hop architecture in large wireless networks? We have seen that the answer to this question depends on the parameter range in which a particular wireless network lies. This naturally fits in a framework of operating regimes. Each operating regime corresponds to a subset of the parameter space where the optimal architecture for cooperation is different. The underlying reason is that the information-theoretic capacity of the network exhibits a qualitatively different behavior in each of these regimes. We have seen that there are many operating regimes where the information-theoretic capacity of the network is significantly higher than the capacity of conventional multi-hop and where architectures better tailored for wireless networks, hierarchical cooperation in particular, can provide substantial capacity gains. In certain regimes, most notably when the network is severely limited in either power or space, there is no way to outperform multi-hop. In other words, the conventional multi-hop architecture is able to achieve the information-theoretic scaling of the network capacity and is fundamentally optimal.

Which of these operating regimes are most relevant to practice? The above discussion also identifies the engineering quantities that determine the operating regime of a wireless network, such as short-range SNR, long-range SNR, area, power path loss exponent, etc. Note that these quantities can be easily computed or directly measured in the network. In Example 1.1 below, we plug in some typical values for the parameters of the network to get some insight on the most relevant operating regimes for various applications in practice.

Example 1.1. Suppose that, as a communication systems engineer, you need to suggest a communication architecture for a wireless network which will operate on a university campus. The campus has an area of $A = 1\,\mathrm{km}^2$ and will operate around $3\,\mathrm{GHz}$ ($\lambda = 0.1\,\mathrm{m}$). According to the discussion in Section 1.3, the number of spatial degrees of freedom in the network is given by $\sqrt{A}/\lambda = 10'000$. Therefore, if there are up to 10'000 students, we expect to have no space-limitation in the network: there are sufficient spatial degrees of freedom for all users to communicate. When there are more than 10'000 users, the network is space-limited. However, multi-hop can achieve all the degrees of freedom only when the number of users in this network of area $1\,\mathrm{km}^2$ are larger than 10^8, a humongous number. Up to this size, we need hierarchical cooperation to exploit the available degrees of freedom in the network. This suggests that although in practice, we might have wireless networks that are space-limited, severely space-limited networks where multi-hop is the right architecture are very unlikely.

In addition, we would most often expect such a network to be bandwidth-limited and not power-limited. Under free-space propagation, the transmitted power P and the received power P_r are related by the Friis formula:

$$P_r = \frac{G_{Tx} \cdot G_{Rx}}{(4\pi r/\lambda)^2} P,$$

where r is the distance between the transmitter and the receiver and G_{Tx} and G_{Rx} are the transmit and receive antenna gains. Assuming unit transmit and receive antenna gains, the attenuation factor $(G_{Tx} \cdot G_{Rx} \cdot \lambda^2)/16\pi^2$ in the formula is 10^{-6}. Assume transmitter power P of $100\,\mathrm{mW}$ per node, thermal noise N_0 at $-174\,\mathrm{dBm}$, a bandwidth W of $10\,\mathrm{MHz}$ and noise figure $\mathrm{NF} = 10\,\mathrm{dB}$. The SNR between a transmitter and receiver pair separated by the maximal distance of $1\,\mathrm{km}$ is $54\,\mathrm{dB}$. With 10'000 users in the network, the long distance SNR is $\mathrm{SNR}_l = 104\,\mathrm{dB}$, very much in the high SNR regime. Note that even if the transmit power per node is $1\,\mathrm{mW}$, a value more typical for sensor nodes, we still have $\mathrm{SNR}_l = 84\,\mathrm{dB}$, a bandwidth-limited network. In a lossy environment, SNR_l will be smaller, but can still be expected to be well above $0\,\mathrm{dB}$.

Therefore, with 10'000 students on the campus, we do not expect to observe any power or space-limitation in the network. In this case, while traditional multi-hop can achieve a total throughput of the order of 100 bits/s/Hz, hierarchical cooperation promises an aggregate throughput of the order of 10'000 bits/s/Hz.

1.5 Problem Formulation

The results presented in this monograph are based on a scaling law characterization of the information-theoretic capacity of wireless networks. This scaling law formulation, developed mainly in Section 3.1, is used as a mathematical tool to identify the operating regimes of large wireless networks, without having to exactly characterize their capacity. It is based on identifying the parameters of wireless networks that have large operational range in practical applications, such as the area of the network, the transmit power available at the users and the bandwidth. Note that these are independent parameters, each of which can be large or small in different applications. As there are no typical values for these parameters, a thorough understanding of the capacity requires to study the interplay between these parameters. We model the interplay through a coupling to the number of users. Characterizing the scaling exponent of the capacity with the number of users for all possible couplings accounts for all possible interplay between these system parameters. Such a scaling law study allows not only to identify the operating regimes of wireless networks but also to approximately characterize the dependence of the information-theoretic capacity of the network to major system parameters.

There are two aspects to such a scaling exponent characterization: upper and lower bounds. Upper bounds on the best possible scaling exponent are derived using tools from information theory. Lower bounds are obtained by constructing explicit cooperation architectures and computing the scaling exponents they achieve. An architecture is called scaling optimal for a certain regime if it is able to achieve the best possible scaling exponent in this regime. This means that the performance of the architecture exhibits the same dependence to system parameters as the capacity itself. Such an optimality definition has an engineering significance: it guarantees that the gap to the information-theoretic capacity of the network does not explode rapidly with any of the system parameters.

The current text is slightly biased in detail towards lower bounds, as we believe the architectures themselves are of higher engineering interest than the theoretical proofs of their optimality. However, without going into too much technical detail, we also tried to give the main intuition behind the information-theoretical upper bounds on capacity.

1.6 Historical Notes

The line of research that leads to the results summarized in this paper was initiated by the seminal work of Gupta and Kumar in 2000 [9]. The work of Gupta and Kumar was stimulating from several points of view. First, it initiated the study of the *scaling* of the capacity of wireless networks with the number of users. Such a scaling law formulation puts the emphasis on large system size and is useful to devise architectural guidelines for large wireless networks. The formulation turned out to be more amenable to analysis than the long-sought capacity region in information theory for a given number of users. Second, it introduced a simple random network model that captures the essential aspects of the problem: the spatial distribution of nodes over the network area and the traffic requirement between them, the attenuation of wireless signals with distance and the broadcasting and superposition nature of wireless media. Most importantly, using this model, Gupta and Kumar identified the interference-limited nature of the conventional multi-hop architecture, showing that in the best case, it achieves a $\Theta(\sqrt{n})$ scaling of the system throughput. A scheme achieving exactly $\Theta(\sqrt{n})$ throughput for generic random wireless networks was then proposed in Ref. [6].

The work of Gupta and Kumar inspired the research tackling the main question of interest in this monograph: Can we do better by more sophisticated physical-layer processing? This question was first addressed by Xie and Kumar [31]. They showed that whenever $\mathrm{SNR}_s \ll 0\,\mathrm{dB}$ and the power path loss exponent α of the environment is greater than 6, the nearest-neighbor multi-hop architecture is in fact order-optimal. The work [31] was followed by several others [1, 10, 16, 26, 32, 33]. Successively, they improved the threshold on the path loss exponent α for which multi-hop is order-optimal due to the severe power limitation $\mathrm{SNR}_s \ll 0\,\mathrm{dB}$ ($\alpha > 5$ in [10], $\alpha > 4.5$ in [1], $\alpha > 4$ in [32], and $\alpha > 3$ in [26]). The work of Franceschetti et al. [7] established the optimality of multi-hop under severe space-limitation (when $\sqrt{A}/\lambda \ll \sqrt{n}$). A similar conclusion was earlier obtained in [25] by modeling the space limitation through a degenerate physical channel model.

Aeron and Saligrama were the first to show that the interference limitation suffered by conventional multi-hop can be surpassed with more sophisticated cooperation: they exhibited a scheme that yields a throughput scaling of $\Theta(n^{2/3})$. The hierarchical cooperation architecture achieving aggregate throughput $\Theta(n^{1-\epsilon})$ for any $\epsilon > 0$ has been introduced by the authors [26]. Both the scheme proposed by Aeron and Saligrama and the hierarchical cooperation architecture are based on combining MIMO communication [5, 29] with cooperative relaying ideas from network information theory. In particular, the hierarchical cooperation scheme critically employs compress-and-forward relaying, a strategy introduced in [4] (see also Refs. [12, 13]). The hybrid architecture combining hierarchical cooperation with multi-hop was introduced in Ref. [23]. The same paper also shows that this hybrid architecture surpasses multi-hop when the network is not severely power-limited, either when $\alpha < 3$ or when $\mathrm{SNR}_s \gg 0$ dB. The same hybrid architecture was independently proposed in Ref. [20] to deal with arbitrary placement of nodes inside the network area. The optimality of hierarchical cooperation when the network is partially space-limited was established in Refs. [14, 15, 27].

The characterization of wireless networks presented in this paper is based on the operating regimes framework developed in Ref. [23]. This framework offers a unified perspective on various fragmented or even seemingly contradicting results in this field. More importantly, it allows the deduction of concise engineering principles from the theory.

There are many interesting ideas we have not included in this monograph. In Refs. [20, 21], Niesen et al. extend some of the ideas in this monograph to networks with arbitrary node placement and arbitrary traffic demand. The work [24] investigates the throughput-delay tradeoff of the hierarchical cooperation scheme and [8] provides a refined analysis of its performance.

In an independent line of research, Cadambe and Jafar [3] and Nazer et al. [18] showed that interference alignment techniques provide an alternative way of dealing with interference in wireless networks and achieving high throughput. The scaling performance of these techniques in wireless networks has been discussed in Refs. [19, 28]. The basic idea behind these techniques is fundamentally different from the schemes discussed in this monograph, in the sense that communication between order n source-destination pairs should be established in one shot, without intermediate relaying. Instead, by making use of sophisticated signaling at the transmitters, each destination receives a signal with two orthogonal components, one of which is the intended signal, whereas the other contains the interfering signals transmitted by all the other users. The intended signal can then recovered at each destination by a simple projection.

One of the major differences between interference alignment and cooperative schemes is therefore that interference alignment schemes do not rely on spatial reuse, which makes them superior in the regime when SNR is extremely large. On the other hand, they heavily rely on transmit channel state information, which is challenging to get in practice, while the techniques discussed in this monograph require channel state information only at the receiver side. Also, interference alignment techniques are less efficient in terms of power transfer than distributed MIMO transmissions: we have indeed seen above that distributed MIMO transmissions benefit from a significant power gain, of the order of the number of nodes participating to the transmission; this power gain is simply absent in interference alignment schemes.

1.7 Notation

To describe limiting behavior of functions, we often adopt the following notation: For two functions $f(n)$ and $g(n)$, the notation $f(n) = O(g(n))$ means that $|f(n)/g(n)|$ remains bounded as n increases. We express $g(n) = \Theta(f(n))$ to denote that $f(n) = O(g(n))$ and $g(n) = O(f(n))$. Finally, $f(n) = \Omega(g(n))$ if $|g(n)/f(n)|$ remains bounded as n increases.

References

[1] S. Ahmad, A. Jovicic, and P. Viswanath, "Outer bounds to the capacity region of wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2770–2776, June 2006.

[2] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*. Cambridge University Press, 2009.

[3] V. Cadambe and S. Jafar, "Interference alignment and the degrees of freedom for the K-user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, August 2008.

[4] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, September 1979.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *AT&T Bell Labs Tech. Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[6] M. Franceschetti, O. Dousse, D. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1009–1018, March 2007.

[7] M. Franceschetti, M. D. Migliore, and P. Minero, "The capacity of wireless networks: Information-theoretic and physical limits," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3413–3424, August 2009.

[8] J. Ghaderi, L.-L. Xie, and X. Shen, "Hierarchical cooperation in ad hoc networks: Optimal clustering and achievable throughput," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3425–3436, August 2009.

[9] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 388–404, March 2000.

[10] A. Jovicic, P. Viswanath, and S. R. Kulkarni, "Upper bounds to transport capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2555–2565, November 2004.

[11] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," in *Proceedings of the ACM SIGCOMM*, Kyoto, Japan, August, 2007.

[12] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, September 2005.

[13] G. Kramer, I. Maric, and R. Yates, "Cooperative Communications," *Foundations and Trends in Networking*, vol. 1, no. 3–4, pp. 271–425, 2006.

[14] S.-H. Lee and S.-Y. Chung, "Effect of channel correlation on the capacity scaling in wireless networks," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, July, 2008.

[15] S.-H. Lee and S.-Y. Chung, "On the capacity scaling of wireless ad hoc networks: Effect of finite wavelength," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Austin, TX, July, 2010.

[16] O. Lévêque and E. Telatar, "Information theoretic upper bounds on the capacity of large, extended ad-hoc wireless networks," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 858–865, March 2005.

[17] J. Liu, D. Goeckel, and D. Towsley, "Bounds on the throughput gain of network coding in unicast and multicast wireless networks," *IEEE Transactions on Selected Areas in Communications*, vol. 27, no. 5, pp. 582–592, May 2009.

[18] B. Nazer, M. Gastpar, S. Jafar, and S. Vishwanath, "Ergodic interference alignment," in *Proceedings of the IEEE International Symposium on Information Theory*, Seoul, South Korea, July, 2009.

[19] U. Niesen, "Interference alignment in dense wireless networks," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2889–2901, May 2011.

[20] U. Niesen, P. Gupta, and D. Shah, "On capacity scaling in arbitrary wireless networks," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 3959–3982, September 2009.

[21] U. Niesen, P. Gupta, and D. Shah, "The balanced unicast and multicast capacity regions of large wireless networks," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2249–2271, May 2010.

[22] A. Özgür, "Fundamental limits and optimal operation in large wireless networks," PhD Thesis Nr 4483, EPFL, 2009.

[23] A. Özgür, R. Johari, O. Lévêque, and D. Tse, "Information theoretic operating regimes of large wireless networks," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 427–437, January 2010.

[24] A. Özgür and O. Lévêque, "Throughput-delay tradeoff for hierarchical cooperation in ad hoc wireless networks," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1369–1377, March 2010.

[25] A. Özgür, O. Lévêque, and E. Preissmann, "Scaling laws for one and two-dimensional random wireless networks in the low attenuation regime," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3573–3585, October 2007.

[26] A. Özgür, O. Lévêque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad-hoc networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3549–3572, October 2007.

[27] A. Özgür, O. Lévêque, and D. Tse, "Linear capacity scaling in wireless networks: Beyond physical limits?," in *Proceedings of the IEEE Information Theory and Applications Workshop*, San Diego, CA, February, 2010.

[28] A. Özgür and D. Tse, "Achieving linear scaling with interference alignment," in *Proceedings of the IEEE International Symposium on Information Theory*, Seoul, South Korea, 2009.

[29] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, November 1999.

[30] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[31] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communications: Scaling laws and optimal operation," *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[32] L.-L. Xie and P. R. Kumar, "On the path-loss attenuation regime for positive cost and linear scaling of transport capacity in wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2313–2328, June 2006.

[33] F. Xue, L.-L. Xie, and P. R. Kumar, "The transport capacity of wireless networks over fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 834–847, March 2005.

[34] S. Zhang, S. Liew, and P. Lam, "Physical layer network coding," in *Proceedings of the ACM MobiCom*, Los Angeles, CA, September, 2006.