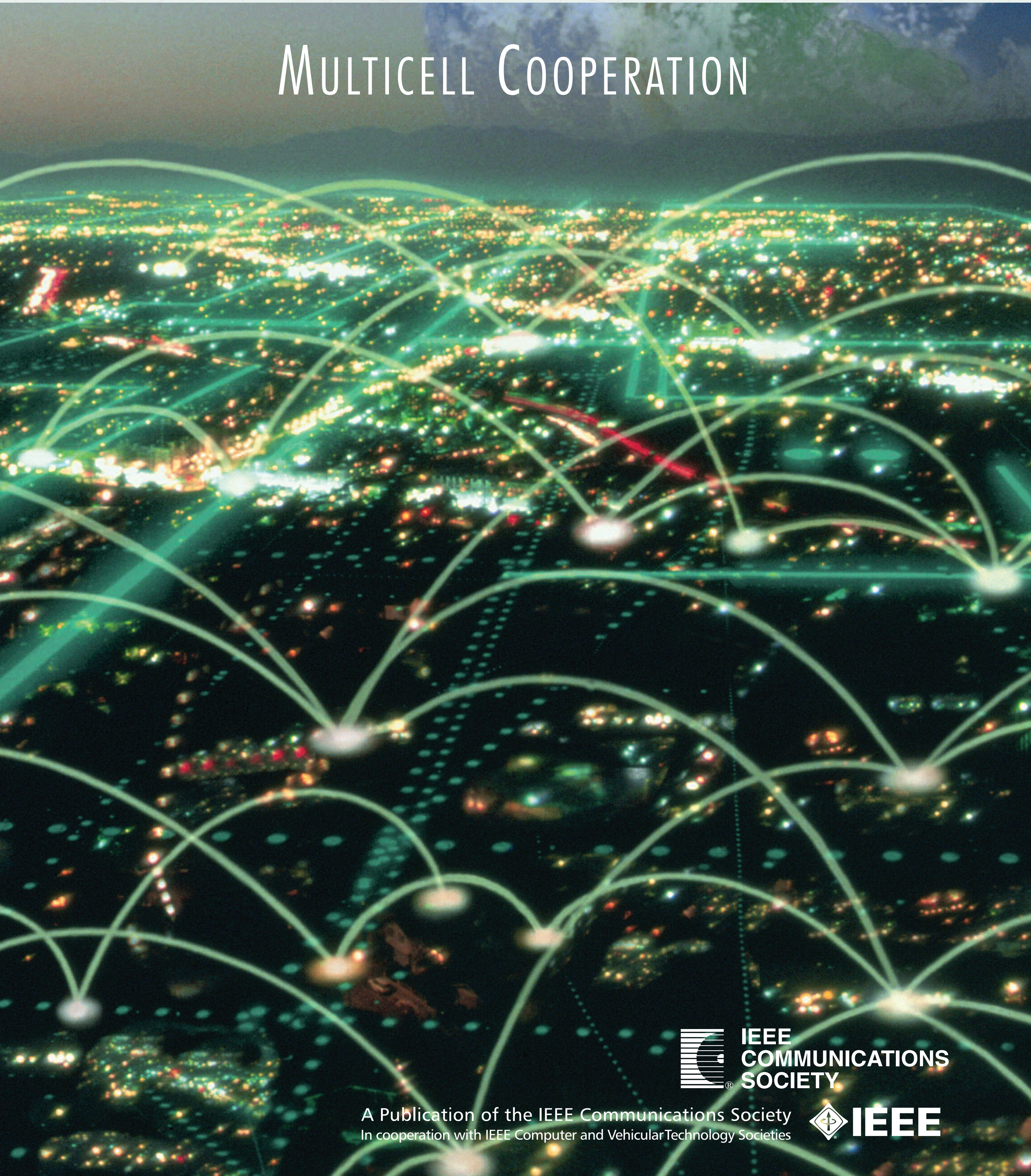
IEEE
WIRELESS COMMUNICATIONS

February 2013, Vol. 20, No. 1

- Integration of Local Area and Wide Area Systems
- Coordination and Cooperation in Large-Scale Networks
- Multicell Cooperation for LTE-A Heterogeneous Network
- Practical Challenges of Interference Alignment
- CSI Sharing Strategies for Transmitter Cooperation
- Multicell Cooperation with Multiple Receive Antennas
- Interference Management in 3GPP LTE-A Networks

Multicell Cooperation



IEEE COMMUNICATIONS SOCIETY

A Publication of the IEEE Communications Society
In cooperation with IEEE Computer and Vehicular Technology Societies

IEEE

FEBRUARY 2013/VOL. 20, NO. 1

IEEE WIRELESS COMMUNICATIONS

SPECIAL ISSUE
MULTICELL COOPERATION

Cover image: Getty Images

MULTICELL COOPERATION FOR THE 4G MOBILE CELLULAR SYSTEMS

This issue features a feature topic on "Multicell Cooperation," which contains 11 excellent papers collected by our Guest Editors, Ekram Hossain and Zander (Zhongding) Lei. They worked extremely hard to edit this feature topic due to a fairly large number of submissions to the Call for Papers. Multicell cooperation is a newly emerging communication technology that can help to significantly improve the overall system capacity of a cellular system through mitigating or coordinating intercell interference. A multicell cooperation system works based largely on cooperative communications at both the base station (BS) and mobile station (MS) levels, and it has become a viable technology to address various critical issues in throughput and coverage enhancement in fourth generation mobile cellular systems. As a matter of fact, multicell cooperation has been identified as one of the key technologies adopted in the LTE-A standard, which offers a peak data transmission rate up to 1 Gb/s. Therefore, the importance and timeliness of this feature topic is evident. Due to the excellent work done by the Guest Editors, this feature topic offers you a collection of 11 articles, contributed by experts from both academia and industry. We do hope that you will enjoy reading them.

HSIAO-HWA CHEN

BIOGRAPHY

HSIAO-HWA CHEN [S'89, M'91, SM'00, F'10] (hshwchen@ieee.org) is currently a Distinguished Professor in the Department of Engineering Science, National Cheng Kung University, Taiwan. He obtained his B.Sc. and M.Sc. degrees from Zhejiang University, China, and a Ph.D. degree from the University of Oulu, Finland, in 1982, 1985, and 1991, respectively. From 2001 to 2003, he served as the founding director of the Institute of Communications Engineering at National Sun Yat-Sen University, Taiwan, which was the first communication research institute established in southern Taiwan. Due to his instrumental role and strong leadership, this institute has educated thousands of telecommunication postgraduate degree holders for Taiwan, contributing to the fast-growing local telecommunication industry. He has authored or co-authored over 300 technical papers in major international journals and conferences, six books, and more than 10 book chapters in the areas of telecommunications, including the books *Next Generation Wireless Systems and Networks* (Wiley, 2005) and *The Next Generation CDMA Technologies* (Wiley, 2007). He has been an active volunteer with IEEE in various technical activities for over 22 years. He has served as General Chair, TPC Chair, and Symposium Chair for many international conferences. Recently, he served as lead TPC Chair for IEEE SmartGridComm 2012,Taiwan, November 2012, and is currently serving as Vice TPC Chair for IEEE WCNC 2013, Shanghai, China.

IEEE Wireless Communications (ISSN 1536-1284) is published bimonthly by The Institute of Electrical and Electronics Engineers, Inc. Headquarters address: IEEE, 3 Park Avenue, 17th Floor, New York, NY 10016-5997; Tel: (212) 705-8900; Fax: (212) 705-8999; E-mail: j.porcello@comsoc.org. Responsibility for the contents rests upon authors of signed articles and not the IEEE or its members. Unless otherwise specified, the IEEE neither endorses nor sanctions any positions or actions espoused in *IEEE Wireless Communications.*

Annual subscription: Member subscription: $40 per year; Non-member subscription: $250 per year. Single copy: $50.

Editorial correspondence: Manuscripts for consideration may be submitted to the Editor-in-Chief: Hsiao-Hwa Chen, Department of Engineering Science, National Cheng Kung University, Tainan 70101, Taiwan. Electronic submissions may be sent to: hshwchen@mail.ncku.edu.tw.

Postmaster: Send address changes to *IEEE Wireless Communications*, IEEE, 445 Hoes Lane, Piscataway, NJ 08855-1331; or E-mail to address.change@ieee.org. Printed in USA. Periodicals postage paid at New York, NY and at additional mailing offices. Canadian GST #40030962. Return undeliverable Canadian addresses to: Frontier, PO Box 1051, 1031 Helena Street, Fort Eire, ON L2A 6C7.

Subscriptions: Send orders, address changes to: IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08855-1331; Tel: (908) 981-0060.

THE PURSUIT OF TENS OF GIGABITS PER SECOND WIRELESS SYSTEMS

CARLOS CORDEIRO, INTEL CORP.

INTRODUCTION

More than any time ever before, today technology has a significant impact on people's lives. The proliferation of slim, mobile, and portable devices such as notebooks, ultrabooks, tablets, and smartphones is a clear testament to the importance of computing and communication in modern society. In the particular case of wireless systems, this explosive growth has led to an insatiable need to constantly boost capacity so that innovative uses can be offered to consumers. This vicious cycle has brought to market many innovative wireless solutions including the IEEE 802.11 family of standards, third generation (3G), and fourth generation (4G), to name a few.

When it comes to wireless systems with data rates greater than 1 Gb/s at a few meters, the most notable examples include the IEEE 802.11ac and IEEE 802.11ad amendments to the base IEEE 802.11 standard [2], with key parameters summarized in Table 1. Several companies have announced products implementing these technologies, with a few of those products already available, or soon to be available, to consumers.

The data rates offered by IEEE 802.11ac and IEEE 802.11ad can meet the needs of many applications, with replacement of wired digital interface (WDI) cables arguably the most prominent new use of these technologies. This includes replacing WDI cables used between a notebook/tablet/smartphone and certain display (e.g., HDMI, DisplayPort) and IO (e.g., USB) devices, thereby enabling a fully wireless user experience [4].

THE EVOLVING LANDSCAPE IN WDIS

The vicious cycle of technology permeates both wired and wireless interfaces. Since replacement of WDI cables is the primary use of this new wave of multi-gigabit per second wireless systems, it is important to understand how prominent WDIs are expected to evolve over time. This is illustrated in Table 2, which focuses primarily on the approximate maximum data rates currently supported by each of the corresponding WDI. Future versions of these WDIs are under development and promise to offer even higher data rates.

Supporting the data rates outlined in Table 2 over a wireless system is nontrivial. While IEEE 802.11ac and IEEE 802.11ad are able to support degrees of these WDIs, when it comes to supporting, say, the high-end data rates for certain WDIs shown in Table 2, it becomes apparent that there is still a performance gap. This gap is only expected to grow larger, given the rapid evolution in WDI speeds – clearly, this is one of the major drivers of the vicious cycle behind multi-Gb/s wireless technologies. As such, we can conclude that next generation wireless systems needs to support data rates in the order to tens of Gb/s to be able to keep up with the evolution of WDI speeds.

THE WAY TOWARD TENS OF GIGABIT PER SECOND WIRELESS

So far today, there has been little discussion in the industry on ways to achieve tens of Gb/s wireless speeds with ranges of a few meters. Despite that, there is growing consensus being formed that the only practical way of getting there is through the use of higher frequencies, with 60 GHz being the starting point.

The amount of unlicensed spectrum in frequencies below 6 GHz is very limited. Even though there are ongoing efforts to allocate additional unlicensed spectrum in the 5 GHz band, the added capacity will be far short of allowing tens of gigabits per second rates to be offered at ranges of 5–10 m. As such, the 60 GHz frequency band, with its vast bandwidth of around 7 GHz, is currently seen as the only viable solution to achieve those goals. Referring back to Table 1, this implies that reusing and extending IEEE 802.11ad is the way toward tens of gigabits per second wireless, since this technology operates in the 60 GHz frequency band. Arguably, unlicensed bands higher than 60 GHz (e.g., terahertz) could also be used to achieve tens of gigabits per

Parameter		IEEE 802.11ac	IEEE 802.11ad
Approx. max. theoretical data rate (Gb/s)	Per device	3.5 (4× 4, 160 MHz channel)	6.8
	Aggregate	6.95	
Range at max data rate (m)		Few (< 10)	
Channel bandwidth (MHz)		20/40/80/160	2160
Number of channels in the US		Five 80 MHz	Three
Frequency band (GHz)		5	60
Expected publication		Feb/14	Dec/12

Table 1. *Brief comparison between two > 1 Gb/s systems.*

WDI	Version	Approx. max. data rate (Gb/s)
DisplayPort	1.0–1.1	8.6
	1.2	17.3
HDMI	1.0–1.2a	5
	1.3–1.4a	10.2
USB	2.0	0.48
	3.0	5
Thunderbolt	1.0	10

Table 2. *Evolution of prominent WDIs.*

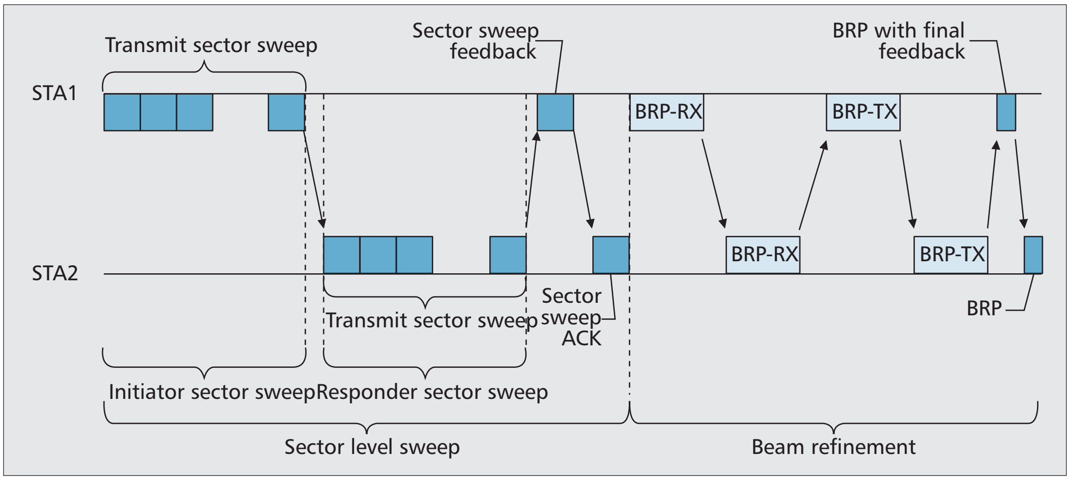


Figure 1. *Beamforming protocol in IEEE 802.11ad.*

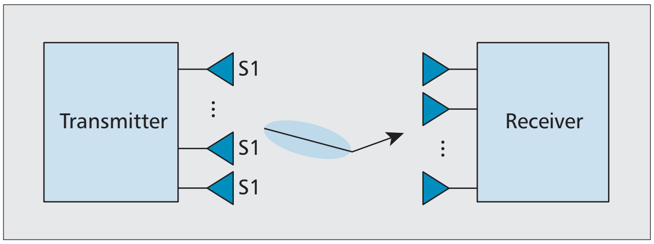


Figure 2. *IEEE 802.11ad beamforming: single RF path.*

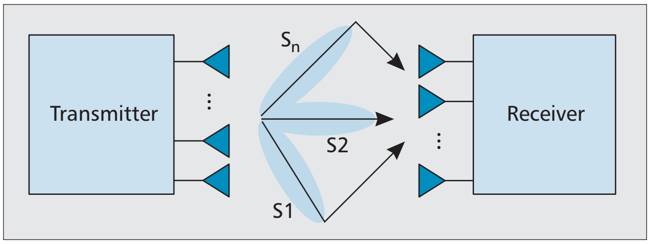


Figure 3. *Next generation 60 GHz beamforming: > 1 RF path.*

second wireless, but technologies operating at these frequencies are nowhere near commercialization.

REVIEW OF BEAMFORMING IN IEEE 802.11AD

An overview of the IEEE 802.11ad standard is given in [3]. The unique feature of this standard is the fact that communication in 60 GHz requires the use of high gain directional antennas in order to compensate for the large path loss [1]. While directional communication addresses the link budget issue, it introduces the need for finding the direction of communication with a neighbor. This is addressed by a novel beamforming (BF) protocol.

BF is the process used by a pair of stations to achieve the necessary link budget for subsequent communication. In IEEE 802.11ad, BF consists of two phases, as depicted in Fig. 1: sector level sweep (SLS) and beam refinement phase (BRP). SLS enables a pair of devices to establish the initial (coarse) direction of communication. At the end of SLS, a pair of devices determines the best transmit sector for communication with each other. BRP is used for receive antenna training and to fine tune antenna settings to improve the quality of directional communication, and hence obtain a multi-gigabit-per-second link.

One of the most important characteristics of the BF protocol in IEEE 802.11ad is that it is capable of discovering a single radio frequency (RF) path for communication between a pair of devices. Thus, this means that even if a device possesses more than one antenna array, only one of those antenna arrays is used for communication between any transmitter-receiver device pair. This is depicted in Fig. 2.

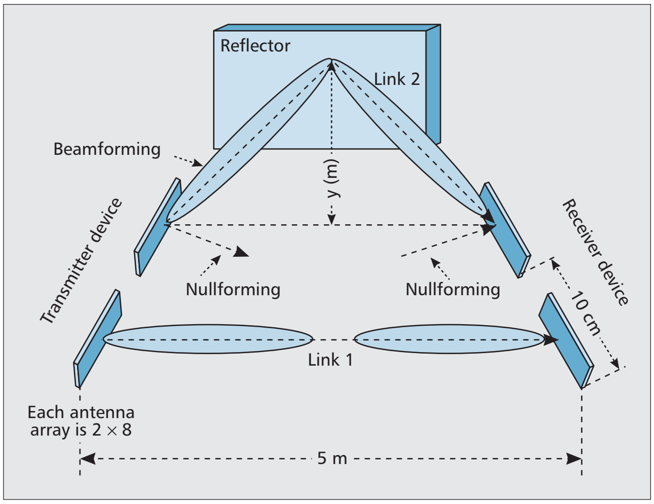


Figure 4. *Simulation scenario.*

ACHIEVING TENS OF GIGABITS PER SECOND WIRELESS

Similar to the evolution of IEEE 802.11a/b/g toward IEEE 802.11n, we believe that the next step in the evolution of 60 GHz standards beyond IEEE 802.11ad will include the introduction of multiple-input multiple-output (MIMO) and channel bonding capabilities. With MIMO, different signals are transmitted from different 60 GHz antenna arrays, as shown in Fig. 3. This results in an improvement in diversity and performance.

To illustrate the gains possible with MIMO in 60 GHz, we have simulated the scenario shown in Fig. 4, wherein

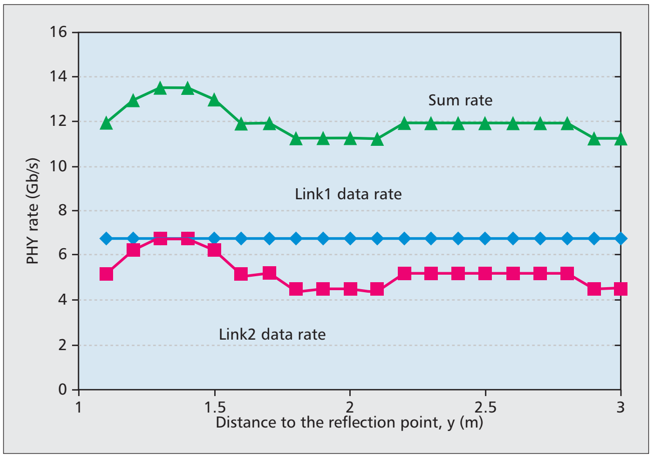


Figure 5. *Simulation results of MIMO with 60 GHz.*

two devices are separated by 5 m, and each device is equipped with two antenna arrays each with 2× 8 elements. The center of the two antenna arrays in a device are separated by 10 cm. A reflector with a reflection coefficient of 2 dB is placed next to the two devices to form a reflected path between the two devices. The distance to the reflection point (y) is varied from 1 to 3 m. The total transmit power of each antenna array is set to 10 dBm, and the noise figure and implementation loss are set to 15 dB. Both beamforming and nullforming techniques are used to enable two simultaneous data transmissions between the two devices over the direct path (i.e., link 1) and the reflected path (i.e., link 2). The MCS is chosen from MCS 13 (693 Mb/s) to MCS 24 (6.76 Gb/s) of IEEE 802.11ad based on the received signal-to-interference-plus-noise ratio (SINR). Figure 5 shows the results obtained, where an aggregate data rate of at least 11 Gb/s can be achieved with two simultaneous transmissions.

When coupled with channel bonding, the aggregate data rate is essentially multiplied. Therefore, MIMO with channel bonding can lead to data rates greater than 20 Gb/s for two bonded channels and 30 Gb/s for three bonded channels. The use of higher-order MIMO schemes and larger numbers of elements per antenna array than those shown in Fig. 4 can lead to even higher aggregate data rates.

CONCLUSIONS AND FUTURE DIRECTIONS

The next few years will experience a significant growth in the availability of commercial wireless technologies capable of multi-gigabit-per-second data rates. While this will provide a major technological advancement compared to the capacity of today's short-range wireless systems based on IEEE 802.11 [2], these speeds fall short of meeting the rates of future WDIs. The use of MIMO and channel bonding at 60 GHz frequencies is expected to provide the next major bump in wireless speeds that can lead to data rates on the order of tens of gigabits per second, thus serving as the path toward supporting the envisioned evolution in WDI speeds. That said, research is still in its infancy when it comes to the use of MIMO in 60 GHz, and further work is needed on developing this capability and proving its commercial feasibility.

ACKNOWLEDGMENTS

The author would like to thank Minyoung Park from Intel for providing helpful contributions to this article.

REFERENCES

[1] C. Cordeiro, "Evaluation of Medium Access Technologies for Next Generation Millimeter-Wave WLAN and WPAN," *IEEE ICC*, 2009.
[2] IEEE 802.11-2012 Standard, http://www.ieee802.org/11/
[3] C. Cordeiro *et al.*, "IEEE 802.11ad: Introduction and Performance Evaluation of the First Multi-Gb/s WiFi Technology," *ACM mmCom*, 2010.
[4] C. Hansen, "WiGig: Multi-Gb/s Wireless Communications in the 60 GHz Band," *IEEE Wireless Comms*, Dec. 2011.

Book Reviews

Edited by Satyajayant Misra

Fundamentals of Wireless Sensor Networks: Theory and Practice, Waltenegus Dargie and Christian Poellabauer, Wiley Series on Wireless Communications and Mobile Computing; Series Editors: Xuemin (Sherman) Shen and Yi Pan; John Wiley and Sons, Ltd.; Edition 1, 2010; ISBN: 978-0-470-997-659; Hardcover, 332 pages.

Reviewer: Satyajayant Misra

Wireless sensor networks (WSNs) have been an area of sustained research for more than a decade. Although their initial application has been restricted to niche domains, such as in the military and for academic projects, recently, they are finding use in many fields, such as civil and mechanical engineering, biology, ecology, medicine, and individual civilian uses. Several recent efforts have also been launched to make wireless sensor nodes easy to assemble, using commercial off-the-shelf components, and easy to program using freely available open source software. Several commercial ventures have been started to help users set up their own WSNs, gather data and store it in the cloud, and also perform sophisticated analysis in the cloud. Looking into the future, these spirited efforts will result in WSNs becoming ubiquitous in our day-to-day life. This exciting phase of WSN development will enable users, practitioners, enthusiasts, and researchers attempting to apply WSNs in their domains. This calls for availability of literature and texts that provide an understanding of the fundamentals of WSNs from both the theoretical and the practical perspectives. These texts should not only be written with students and researchers in the area in mind, but also aim to benefit a wide spectrum of readers. This book, titled *Fundamentals of Wireless Sensor Networks: Theory and Practice*, is a good match for the needs of this wide spectrum of possible users, including students at the graduate and undergraduate levels. The book provides an introduction to the fundamental concepts and principles of WSNs, and a survey of protocols, algorithms, and technologies in subdomains, such as the network protocol stack, middleware, and applications.

Given its breadth of coverage of the topic of wireless sensor networks and its close following of the state of the art, the book would serve not only as a good textbook, but also as a very good reference for researchers and practitioners.

This book is divided into three broad sections: Introduction, Basic Architectural Framework, and Node and Network Management, containing a total of 12 chapters. The Introduction section contains four chapters. Chapter 1 provides the motivation for the use of WSNs by juxtaposing WSNs against traditional networks (e.g., wired networks). The chapter also provides some insights on the nature of typical sensors that are attached to sensor nodes, and the classification for these sensors. It also details the challenges and constraints with WSNs, in terms of the environment, wireless channel, and energy scarcity, among others. The second chapter provides details on several application scenarios for WSNs: structural health monitoring, traffic monitoring, healthcare, volcano monitoring, underground mining, pipeline monitoring, and precision agriculture. There are of course several other applications that may be envisioned for WSN application, such as habitat monitoring and ecological surveys of animals and their migration patterns. Chapter 3 presents the architecture of a wireless sensor node, which includes the sensing subsystem consisting of the analog-to-digital converter as the integral part; the processor/microcontroller subsystem, which uses different technologies, such as application-specific integrated circuits and field programmable gate arrays; and communication interfaces, such as the serial peripheral interface and the inter-integrated circuits. The chapter also presents some of the prototypical architectures, such as the Intel iMote node architecture, the XYZ architecture, and the Hogthrob architecture. Chapter 4 presents different features and functional aspects of operating systems for a wireless sensor node, and also presents and compares four OSs: TinyOS, SOS, Contiki, and LiteOS.

The second section of the book pertains to basic architectural frameworks, and provides a detailed discussion of protocols and algorithms used at different network protocol layers. Chapter 5 presents a discussion about the physical layer architectures and concepts. It starts by presenting the basic components of a wireless sensor system, and then presents source encoding, channel encoding, different modulation techniques for communication in the wireless channel, and signal propagation. Chapter 6 starts with an overview of medium access control (MAC) protocols: the paradigms of contention-based and contention-free MAC; then presents MAC protocols used in the ad hoc wireless domain, and the requirements of a MAC protocol in WSNs. The chapter also presents several contention-free, contention-based, and hybrid MAC protocols for WSNs. Chapter 7 presents an overview of the network layer, the routing metrics, and also the different routing paradigms, such as flooding and gossiping, data-centric routing, proactive routing, on-demand routing, hierarchical routing, location-based routing, QoS-based routing, and the corresponding protocols.

The third section deals with node and network management, presenting several additional techniques and solutions for dealing with system and network dynamics and WSN-specific requirements. Chapter 8 deals with power management in WSNs: local power management aspects and dynamic power management. Chapter 9 deals with time synchronization primitives for the nodes in the WSN. It presents the basics for time synchronization in WSNs and the corresponding challenges. The chapter further presents several time synchronization protocols, such as lightweight tree-based synchronization, the timing-sync protocol for sensor networks, the flooding time synchronization protocol, and protocols for multihop time synchronization. Chapter 10 deals with localization — ranging techniques, different techniques for range-based localization, range-free localization, and event-based localization. Chapter 11 deals with WSN security, that is, the security fundamentals necessary for proper operation of a WSN, the challenges to security, and potential security attacks. The chapter also presents protocols and mechanisms to improve security and provide reusable primitives. Chapter 12 deals with sensor network programming: the challenges to programming wireless sensor nodes; the node-centric programming constructs in the event-driven programming paradigm; the idea of macroprogramming, where a node is no longer the focus, but a group of nodes or a whole network; the idea of dynamic reprogramming of sensor nodes post-deployment; and the available sensor network simulators and tools.

The book provides a holistic view of WSNs from the perspectives of both theory and practice. Given its breadth of coverage of the topic of wireless sensor networks and its close following of the state of the art, the book would serve not only as a good textbook, but also as a very good reference for researchers and practitioners. The book also has several interesting exercise questions at the end of each chapter. The references provided at the end of each chapter are fairly representative of the state of the art.

Smart Grid Communications and Networking, Edited by Ekram Hossain, Zhu Han, and H. Vincent Poor, Cambridge University Press, Edition 1, May 2012, ISBN-13: 9781107014138, Hard cover, 481 pages.

Reviewer: Satyajayant Misra

Due to its huge economic, environmental, and societal benefits, the smart grid has become a critical agenda worldwide. The smart grid is an enhanced power grid system, in which grid entities can cooperate to improve, the response of the power grid, reduce peak demand, and make power generation, distribution, and consumption economically. A major enabler of this cooperation is efficient communication and information exchange between the grid entities, the producers, distributors, policy makers, and the consumers. The communication and information exchange will be facilitated by several wireless and wire-line technologies, such as IEEE 802.11, ZigBee, 6LoWPAN, Ethernet, Power-Line Communication, etc. The smart grid will provide information and intelligence to the power grid to enable grid automation, active operation, and efficient demand response. This makes the need for research and development in communication protocols for the smart grid an area needing immediate attention and thus, research in the development of smart grid communications and networking systems is gaining momentum. Smart Grid Communications and Networking, edited by Ekram Hossain from the University of Manitoba, Canada, Zhu Han from the University of Houston, USA, and H. Vincent Poor from the Princeton University, USA—a recently published book by Cambridge University Press on smart grid communications has been released for enthusiasts, practitioners, and engineers at the right time.

Even though this is an edited book, the book has a very good flow and the materials are coherently presented. All of the major aspects related to smart grid communications and networking have been covered.

The book contains twenty chapters, which are organized into six parts. Part one consists of four chapters covering the basics of communication architectures, models, and technologies, network control systems, demand-side management models, as well as the vehicle-to-grid systems in the future smart grid. Part two onwards; the book deals in greater detail with communication, namely existing paradigms, technologies, and security and privacy issues on the smart grid. Part two, consisting of four chapters, deals with the physical data communications, access, detection, and estimation techniques for the smart grid systems. Chapter 5 presents communication and access technologies from legacy systems, to technologies that will find application in smart grids of the future. The emerging paradigm of machine-to-machine (M2M) communications is reviewed in Chapter 6. Chapter 7 deals with the important area of bad-data detection in the smart grid especially in a distributed manner. Chapter 8 deals with distributed state estimation with emphasis on a learning-based framework.

Part three of the book consists of two chapters and focuses on wireless networks and performance evaluation of network architectures and protocols for the smart grid systems. Chapter 9 presents networking technologies for wide-area power-systems measurement, monitoring, protection, and control. Chapter 10 presents factors to take into account when using wireless technologies (especially wireless sensors and actuators) for smart grid applications. Wireless sensor and actuator networks will play a very important role for the generation systems, transmission and distribution systems, as well as in the consumers' premises in the smart grid environments. Part four of the book, consisting of four chapters, focuses on the applications of the technologies for these sensor networks, major design challenges, as well as implementation and performance evaluation of sensor networking protocols for the smart grid systems. Chapter 11 deals with research challenges and potential application of wireless sensor networks for smart grid. Chapter 12 presents sensor techniques and network protocols customized to the needs of the smart grid. Chapter 13 presents potential methods, such as pervasive service-oriented network and content-aware intelligent control, for development/deployment of sensor and actuator networks in smart grid. Chapter 14 implementation and performance evaluation of wireless sensor networks for smart grid. Security is a significant issue in the smart grid research and it has attracted substantial research interest from both academia and industry. Part five of this book is dedicated to security issues in the smart grid. Five chapters in this part deal with issues such as cyber-attack impact analysis, jamming-based attack and bad-data injection attack, cross-layer design for security provisioning, and intrusion detection, and application-driven design approach for secured smart grid applications. Lastly, part six overviews several smart grid field trials.

The articles in the book are well-organized and well-articulated, with access to relevant references at the end of each chapter. It will be a valuable source of information for people new to the area of smart grid communications and networking and the smart grid systems in general. Also, scientists and engineers already involved in smart grid research will find this book a very good reference source. This is one of the few books on the smart grid available today which is useful to the researchers in the area of communications, networking, and signal processing who intend to work in the smart grid area.

Even though this is an edited book, the book has a very good flow and the materials are coherently presented. All of the major aspects related to smart grid communications and networking have been covered. This book may be also useful to readers from the power engineering community seeking to optimize the communication systems for smart grid.

Scanning the Literature

Edited by Pan Li

Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing

Dejun Yang, Guoliang Xue, Xi Fang, and Jian Tang, ACM MobiCom, Istanbul, Turkey, August 2012

Mobile phone sensing is a new paradigm which takes advantage of the pervasive smartphones to collect and analyze data beyond the scale of what was previously possible. In a mobile phone sensing system, the platform recruits smartphone users to provide sensing service. Existing mobile phone sensing applications and systems lack good incentive mechanisms that can attract more user participation. To address this issue, the authors design incentive mechanisms for mobile phone sensing. They consider two system models: the platform-centric model where the platform provides a reward shared by participating users, and the user-centric model where users have more control over the payment they will receive. For the platform-centric model, the authors design an incentive mechanism using a Stackelberg game, where the platform is the leader while the users are the followers. They show how to compute the unique Stackelberg Equilibrium, at which the utility of the platform is maximized, and none of the users can improve its utility by unilaterally deviating from its current strategy. For the user-centric model, the authors design an auction-based incentive mechanism, which is computationally efficient, individually rational, profitable, and truthful. Through extensive simulations, they evaluate the performance and validate the theoretical properties of our incentive mechanisms.

A Picasso: Flexible RF and Spectrum Slicing

Steven Hong, Jeff Mehlman, and Sachin Katti, ACM SIGCOMM, Helsinki, Finland, August 2012

This paper presents the design, implementation and evaluation of Picasso, a novel radio design that allows simultaneous transmission and reception on separate and arbitrary spectrum fragments using a single RF front end and antenna. Picasso leverages this capability to flexibly partition fragmented spectrum into multiple slices that share the RF front end and antenna, yet operate concurrent and independent PHY/MAC protocols. The authors show how this capability provides a general and clean abstraction to exploit fragmented spectrum inWiFi networks and handle coexistence in dense deployments. They prototype Picasso, and demonstrate experimentally that a Picasso radio partitioned into four slices, each concurrently operating four standard WiFi OFDM PHY and CSMA MAC stacks, can achieve the same sum throughput as four physically separate radios individually configured to operate on the spectrum fragments. The authors also demonstrate experimentally how Picasso's slicing abstraction provides a clean mechanism to enable multiple diverse networks to coexist and achieve higher throughput, better video quality and latency than the best known state of the art approaches.

Combined Relay Selection and Cooperative Beamforming for Physical Layer Security

Junsu Kim, Aissa Ikhlef, and Robert Schober, Journal of Communications and Networks, 14(4), 364-373, August 2012

This paper proposes combined relay selection and cooperative beamforming schemes for physical layer security. Generally, high operational complexity is required for cooperative beamforming with multiple relays because of the required information exchange and synchronization among the relays. On the other hand, while it is desirable to reduce the number of relays participating in cooperative beamforming because of the associated complexity problem, doing so may degrade the coding gain of cooperative beamforming. Hence, the authors propose combined relay selection and cooperative beamforming schemes, where only two of the available relays are selected for beamforming and data transmission. The proposed schemes introduce a selection gain which partially compensates for the decrease in coding gain due to limiting the number of participating relays to two. Both the cases where full and only partial channel state information is available for relay selection and cooperative beamforming are considered. Analytical and simulation results for the proposed schemes show improved secrecy capacities compared to existing physical layer security schemes employing cooperative relays.

Outage Probability and Outage-Based Robust Beamforming for MIMO Interference Channels with Imperfect Channel State Information

Juho Park, Youngchul Sung, Donggun Kim and H. Vincent Poor, IEEE Transactions on Wireless Communications, 11(10), 3561 – 3573, Octorber 2012

In this paper, the outage probability and outage-based beam design for multiple-input multiple-output (MIMO) interference channels are considered. First, closed-form expressions for the outage probability in MIMO interference channels are derived under the assumption of Gaussian-distributed channel state information (CSI) error, and the asymptotic behavior of the outage probability as a function of several system parameters is examined by using the Chernoff bound. It is shown that the outage probability decreases exponentially with respect to the quality of CSI measured by the inverse of the mean square error of CSI. Second, based on the derived outage probability expressions, an iterative beam design algorithm for maximizing the sum outage rate is proposed. Numerical results show that the proposed beam design algorithm yields significantly better sum outage rate performance than conventional algorithms such as interference alignment developed under the assumption of perfect CSI.

DOF: A Local Wireless Information Plane

Steven Hong and Sachin Katti, ACM SIGCOMM, Toronto, ON, Canada, August 2011

The ability to detect what unlicensed radios are operating in a neighborhood, their spectrum occupancies and the spatial directions their signals are traversing is a fundamental primitive needed by many applications, ranging from smart radios to coexistence to network management to security. This paper presents DOF, a detector that in a single framework accurately estimates all three parameters. DOF builds on the insight that in most wireless protocols, there are hidden repeating patterns in the signals that can be used to construct unique signatures, and accurately estimate signal types and their spectral and spatial parameters. The paper shows via experimental evaluation in an indoor testbed that DOF is robust and accurate, and achieves greater than 85 percent accuracy even when the Signal-to-Noise Ratios (SNRs) of the detected signals are as low as 0 dB, and even when there are multiple interfering signals present. To demonstrate the benefits of DOF, the authors design and implement a preliminary prototype of a smart radio that operates on top of DOF, and show experimentally that it provides a 80 percent increase in throughput over Jello, the best known prior implementation, while causing less than 10 percent performance drop for co-existing WiFi and Zigbee radios.

Interference Channel with Constrained Partial Group Decoding

Chen Gong, Ali Tajer, and Xiaodong Wang, IEEE Transactions on Communications, vol. 59, no. 11, November 2011, pp. 3059–71

This paper proposes novel coding and decoding methods for a fully connected K-user Gaussian interference channel. Each transmitter encodes its information into multiple layers and transmits the superposition of those layers. Each receiver employs a constrained partial group decoder (CPGD) that decodes its designated message along with a part of the interference. In particular, each receiver performs a twofold task by first identifying which interferers it should decode and then determining which layers of them should be decoded. Determining the layers to be decoded and decoding them are carried out in a successive manner, where in each step a group of layers with a constraint on its group size is identified and jointly decoded while the remaining layers are treated as Gaussian noise. The decoded layers are then subtracted from the received signal and the same procedure is repeated for the remaining layers. The authors provide a distributed algorithm, tailored to the nature of the interference channels, that determines the transmission rate at each transmitter based on some optimality measure and also finds the order of the layers to be successively decoded at each receiver. They also consider practical design of a system that employs the quadrature amplitude modulations (QAM) and rateless codes. Numerical results are provided on the achievable sum-rate under the ideal case of Gaussian signaling with random codes as well as on the system throughput under practical modulations and channel codes. The results show that the proposed multilayer coding scheme with CPGD offers significant performance gain over the traditional un-layered transmission with single-user decoding.

No Time to Countdown: Migrating Backoff to the Frequency Domain

Souvik Sen, Romit Roy Choudhury, and Srihari Nelakuditi, ACM MobiCom, Las Vegas, NV, USA, September 2011

Conventional WiFi networks perform channel contention in time domain. This is known to be wasteful because the channel is forced to remain idle while all contending nodes are backing off for multiple time slots. This paper proposes to break away from convention and recreate the backing off operation in the frequency domain. The basic idea leverages the observation that OFDM subcarriers can be treated as integer numbers. Thus, instead of picking a random backoff duration in time, a contending node can signal on a randomly chosen subcarrier. By employing a second antenna to listen to all the subcarriers, each node can determine whether its chosen integer (or subcarrier) is the smallest among all others. In fact, each node can even determine the rank of its chosen subcarrier, enabling the feasibility of scheduled transmissions after every round of contention. The authors develop these ideas into a Back2F protocol that migrates WiFi backoff to the frequency domain. Experiments on a prototype of 10 USRPs confirm feasibility, along with consistent throughput gains over 802.11. Trace based simulations affirm scalability to larger, real-world network topologies.

PACP: An Efficient Pseudonymous Authentication-Based Conditional Privacy Protocol for VANETs

Dijiang Huang, Satyajayant Misra, Mayank Verma, and Guoliang Xue, IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 3, September 2011, pp. 736–46

This paper proposes a new privacy preservation scheme, named pseudonymous authentication-based conditional privacy (PACP), which allows vehicles in a vehicular ad hoc network (VANET) to use pseudonyms instead of their true identity to obtain provably good privacy. In the proposed scheme, vehicles interact with roadside units to help them generate pseudonyms for anonymous communication. In the scheme setup, the pseudonyms are only known to the vehicles but have no other entities in the network. In addition, the proposed scheme provides an efficient revocation mechanism that allows vehicles to be identified and revoked from the network if needed. Thus, the authors can provide conditional privacy to the vehicles in the system, that is, the vehicles will be anonymous in the network until they are revoked, at which point, they cease to be anonymous.

Achieving MAC-Layer Fairness in CSMA/CA Networks

Ying Jian, Ming Zhang, and Shigang Chen, IEEE/ACM Transactions on Networking, vol. 19, no. 5, October 2011, pp. 1472–84

CSMA/CA networks, including IEEE 802.11 networks, exhibit severe fairness problem in many scenarios, where some hosts obtain most of the channel's bandwidth while others starve. Most existing solutions require nodes to overhear transmissions made by contending nodes and, based on the overheard information, adjust local rates to achieve fairness among all contending links. Their underlying assumption is that transmissions made by contending nodes can be overheard. However, this assumption holds only when the transmission range is equal to the interference range, which is not true in reality. As this paper reveals, the overhearing-based solutions, as well as several nonoverhearing AIMD solutions, cannot achieve MAC-layer fairness in various settings. The authors propose a new rate control protocol, called Proportional Increase Synchronized multiplicative Decrease (PISD). Without relying on overhearing, it provides fairness in CSMA/CA networks, particularly IEEE 802.11 networks, by using only local information and performing localized operations. It combines several novel rate control mechanisms, including synchronized multiplicative decrease, proportional increase, and background transmission. The authors prove that PISD converges and achieves (weighted) fairness. They further introduce Queue Spreading (QS) to achieve MAC-layer fairness when there are multiple contention groups, in which case PISD will fail.

MULTICELL COOPERATION

EKRAM HOSSAIN ZANDER (ZHONGDING) LEI

It is well recognized that the conventional cellular networks can no longer meet the user performance requirements in terms of throughput and coverage, which is due to the significant increase in the number of mobile users along with their traffic demands and the limited spectrum resource. This has stimulated the search for new approaches to alleviate this problem. Recently, cooperative communications at both the base station (BS) and mobile station (MS) levels has been proposed as a key technology to address this problem. Multicell cooperation is a new and emerging communication paradigm promising significant improvement in system capacity by mitigating intercell interference. It has been shown that, with the aid of multicell cooperation, a significant improvement of cellular system performance can be achieved, and the increase in throughput can be as large as an order of magnitude. Although the advantages of multicell cooperation have been demonstrated, many problems related to the research and development of practical and effective schemes for multicell cooperation exist. They have attracted considerable research attention in both academia and industry. This special issue of *IEEE Wireless Communications* puts together some recent results in the area of multicell cooperation for the next generation cellular wireless networks. It includes 11 articles, a brief account for each of which is provided below.

In the first article, "Future Steps of LTE-A: Evolution toward Integration of Local Area and Wide Area" by Kishiyama *et al.*, the authors focus on the importance of integrating local area and wide area in the future of LTE-Advanced (LTE-A) networks. The concept of a macro-assisted small cell is introduced for this integration in a frequency-separated deployment scenario. The authors also identify some potential technologies such as 3D MIMO/beamforming, receiver interference cancelation, and dynamic TDD, to enhance the spectrum efficiency of higher frequency bands.

In the next article, "Multicell Cooperation: Evolution of Cooperation and Cooperation in Large-Scale Cellular Networks" by Burchardt and Hass, the authors emphasize the importance of multicell cooperation in large-scale cellular networks. They present different techniques to perform such cooperation including user-based cooperation, system-wide optimization, and large-scale multiple-antenna systems. A recently proposed Pareto-based optimal OFDMA transmission scheduling method is introduced as an example in which both the concept of user-based cooperation and multiple-antenna systems are combined.

In the third article, "Multicell Cooperation for LTE-Advanced Heterogeneous Networks Scenarios," Soret *et al.* present two scenarios in which simple multicell cooperation is used for LTE-A heterogeneous networks. In the co-channel deployment scenario, the cross-tier interference is mitigated by using enhanced intercell interference coordination, which also ensures that the number of offloaded users is sufficient. In the other scenario, different carriers are assigned to the deployed macro and small cells while using collaborative intersite carrier aggregation to fully utilize the fragmented spectrum. In addition, the authors point out the importance of explicit terminal support as well as distributed coordination among the base stations for both the scenarios to perform optimally.

Interference alignment is emerging as a key transmission strategy to facilitate interference cancellation for communication in interference channels. In the article "The Practical Challenges of Interference Alignment," El Ayach, Peters, and Heath, Jr. give an overview on interference alignment and its challenges. These challenges include the system performance in realistic propagation environments, how to obtain the channel state information at the transmitter, and the practicality of interference alignment in large-scale wireless networks.

Cooperation among multiple transmit antennas has been shown to be an effective interference management tool for communications over interference channels. In the fifth article, "CSI Sharing Strategies in Wireless Networks," de Karet and Gesbert emphasize the cost of exchanging the channel state information (CSI) among the collaborating transmitters. They present two scenarios, the interference alignment and network MIMO scenarios, in which the need for full CSI sharing could be compromised to get more practical and scalable cooperation while maintaining the same performance. To reduce the amount of CSI sharing in the interference alignment scenario, the authors use specific antenna configurations, whereas in the network MIMO scenario they exploit the decay property of signal power. Then they present the challenges of precoding design in the network MIMO channels.

The sixth article, "Multicell Cooperative Systems with Multiple Receive Antennas" by Hwang *et al.*, emphasizes the importance of using multiple receive antennas in a multicell cooperative system. The authors review recent developments in multicell advanced receiver-processing algorithms to mitigate the interference. Next, they outline the main limitations to achieve cooperation in multicell cooperative systems and provide an overview of recent research efforts to overcome such limitations using multiple receive antennas.

Coordinated multipoint transmission/reception (CoMP), in which base stations cooperate to mitigate or eliminate intercell interference, has been specified as one of the enabling techniques for LTE-Advanced networks (in Release 11). In the article "Interference Management through CoMP in 3GPP LTE-Advanced Networks," Sun *et al.* discuss the performance benefits of CoMP in both downlink and uplink scenarios defined in the 3GPP LTE-Advanced standard. In downlink CoMP, they focus on four different techniques: transmission scheme (which includes dynamic point selection, dynamic point blanking, joint transmission, and coordinated scheduling/beamforming), CSI report, interference management, and reference signal design. Next, the authors explain briefly why it is easier to support CoMP in the uplink than in the downlink. Then the efficiency of CoMP is evaluated for four different scenarios defined in the standard, and several unsolved problems as well as their possible solutions are also outlined.

In the next article, "How Do We Design CoMP to Achieve Its Promised Potential?," Yang *et al.* first discuss the features of CoMP systems that are different from those of single-cell MIMO systems. Then the challenges related to CSI acquisition are discussed for both time-division duplexing (TDD) and frequency-division duplexing (FDD) systems. Also, the challenges that arise due to the limited backhaul capacity for the cooperating base stations in a CoMP system are discussed, and possible approaches to tackle these challenges are outlined.

In the ninth article, "On the Impact of Network Geometric Models on Multicell Cooperative Communications Systems," Tukmanov *et al.* emphasize the trade-off between complexity and precision in modeling multicell cooperative networks. The authors describe the Wyner model as a well-known deterministic approach to model multicell cooperative networks, and then they present the stochastic geometry approach to model the random multicell cooperative wireless networks. The authors give examples of some common tools that can be used to model multicell cooperative networks considering fixed base stations and random user locations as well as random base stations and user locations.

An interesting overview on how to improve the energy efficiency in multicell cooperative cellular networks is given by Han and Ansari in their article "On Greening Cellular Networks via Multicell Cooperation." The authors discuss how the number of active base stations can be reduced and thus energy saved by using traffic-demand-aware multicell cooperation. Also, the authors introduce the concept of energy-aware multicell cooperation in which the power consumption of on-grid base stations is reduced by offloading traffic to off-grid base stations powered by renewable energy. To this end, the authors discuss how CoMP can be used to improve the energy efficiency of cellular networks as well as the challenges of energy-efficient multicell cooperation in next generation cellular networks.

In the last article, "Optimal Trade-off between Power Saving and QoS Provisioning for Multicell Cooperation networks," Zhang *et al.* develop a new scheme to achieve an optimal trade-off between the power saving and QoS provisioning in multicell cooperation. To obtain the optimal power scheduling policy that minimizes the power consumption at the base stations while guaranteeing QoS to the users, the authors formulate a stochastic optimization scheme based on a semi-Markov decision process.

The articles included in this Special Issue consider different aspects of multicell cooperative cellular wireless networks which will be very significant in the context of the evolving LTE-Advanced (and beyond) networks. We hope that you enjoy reading this issue and find the articles useful. We would like to thank all the authors for submitting their work and the reviewers for their time in reviewing the articles. Our special thanks to Dr. Hongyang Chen who first conceived the idea to organize this Special Issue and has helped us throughout the process. Also, we acknowledge the support from Professor Hsiao Hwa Chen who has guided us in this endeavor.

BIOGRAPHIES

EKRAM HOSSAIN [S'98, M'01, SM'06] (ekram@ee.umanitoba.ca) is a professor in the Department of Electrical and Computer Engineering at the University of Manitoba, Winnipeg, Canada. He received his Ph.D. in electrical engineering from the University of Victoria, Canada, in 2001. His current research interests include design, analysis, and optimization of wireless/mobile communications networks, cooperative and cognitive wireless networks, and green radio technology. He has authored/edited several books in these areas (http://www.ee.umanitoba.ca/~ekram). He serves as the Editor-in-Chief for *IEEE Communications Surveys and Tutorials* (for the term 2012–2013), and an Editor for *IEEE Journal on Selected Areas in Communications - Cognitive Radio Series* and *IEEE Wireless Communications*. Also, he serves on the IEEE Press Editorial Board (for the term 2013–2015). Previously, he served as Area Editor for *IEEE Transactions on Wireless Communications* in the area of Resource Management and Multiple Access from 2009 to 2011 and an Editor for *IEEE Transactions on Mobile Computing* from 2007 to 2012. He has won several research awards including the University of Manitoba Merit Award in 2010 (for Research and Scholarly Activities), the 2011 IEEE Communications Society Fred Ellersick Prize Paper Award, and the IEEE Wireless Communications and Networking Conference 2012 Best Paper Award. He is a Distinguished Lecturer of the IEEE Communications Society for the term 2012–2013. He is a registered Professional Engineer in the province of Manitoba, Canada.

ZANDER (ZHONGDING) LEI [M'01, SM'04] is a senior scientist and laboratory head, Institute for Infocomm Research, Singapore. He received his B.S. (Honor, 1st upper) and Ph.D. degrees from Beijing Jiaotong University in 1991 and 1997, and his M.Eng. degree from Tong Ji University (Shanghai Railway Institute) in 1994. From 1997 to 1999, he was a post-doctoral fellow at the National University of Singapore. Since 1999 he has been with the Institute for Infocomm Research Singapore, leading research teams working on next generation cellular and local area networks. He was a technical leader in the IEEE 802.22 standard working group that published the first IEEE standard for cognitive radio systems. His research interests include next generation/emerging wireless communication systems/technologies, cooperative communications, and cognitive radio, among others. He has 18 patents granted and over 90 technical papers published. He is a co-recipient of the Institution of Engineering of Singapore Prestigious Engineering Achievement Awards in 2005 and 2007. He has been the PHY Technical Editor for the IEEE 802.22 standard since 2006. He is an Editor for *IEEE Communications Letters* and *IEEE Communications Magazine*, a Lead Guest Editor for a *EURSIP WCN* Special Issue on Multicell Cooperation, and a Guest Editor for a *EURSIP WCN* Special Issue on Cognitive Radio.

FUTURE STEPS OF LTE-A: EVOLUTION TOWARD INTEGRATION OF LOCAL AREA AND WIDE AREA SYSTEMS

YOSHIHISA KISHIYAMA, ANASS BENJEBBOUR, AND TAKEHIRO NAKAMURA, NTT DOCOMO, INC.
HIROYUKI ISHII, DOCOMO INNOVATIONS, INC.

ABSTRACT

In the future steps of 3GPP LTE-Advanced, we will need to ensure the sustainability of 3GPP radio access technologies in order to respond to the anticipated challenging requirements of the future. To this end, this article presents our views on the evolution concept and candidate technologies for future steps of LTE-A taking into account the ever increasing importance of local area (small cells) and the need for further spectrum extension to higher frequency bands. In our evolution concept, we emphasize the integration of the local area with the wide area as a new form of cooperation between conventional macrocells in lower frequency bands and small cells deployed in higher frequency bands. Furthermore, we identify the potential key technologies for further spectrum efficiency enhancements for both the local area and the wide area, and for the efficient integration of the local and wide areas while assuming frequency separation between macrocells and small cells. In particular, multicell cooperation based on macrocell assistance of small cells (referred to as Phantom cell) is introduced, and the related benefits are discussed. Finally, in order to demonstrate the potential capacity gains of our evolution concept, system-level simulation results are provided for using outdoor small cells in higher/wider frequency bands.

INTRODUCTION

The Third Generation Partnership Project (3GPP) standardized the radio interface specifications for the next generation mobile communications system, called Long-Term Evolution (LTE) [1, 2]. The initial specifications for LTE (i.e., Release 8) were finalized in 2008, and commercial LTE service in Japan was launched in December 2010. Standards development for LTE is continued toward establishing an enhanced LTE radio interface called LTE-Advanced (LTE-A; LTE Release 10 and beyond), and now specifications for LTE Release 11 are in their finalization phase.

In order to continue to ensure the sustainability of 3GPP radio access technologies over the coming decade, 3GPP standardization will need to identify and provide new solutions that can respond to future challenges. Toward this end, the 3GPP initiated discussions on further steps in the evolution of LTE toward the future (i.e., Release 12 and onward) [3]. The general consensus in 3GPP so far is that the further standardization of LTE will continue to follow step-by-step evolutionary phases in the form of LTE Releases 12, 13, and so on. We refer to these new steps in the LTE evolution as LTE-B, LTE-C, and so on to indicate the next steps of LTE-A. Here, LTE-B refers to the next set of releases, LTE Releases 12 and 13 [3].

This article provides our views on future requirements, the evolution concept, and candidate solutions for the next steps after LTE-A. The goal is to establish a long-term evolution concept and identify the candidate technologies required to satisfy the future requirements of the next decade. The envisioned evolution concept consists of two key aspects:

- The efficient integration of wide and local area enhancements
- The efficient utilization of higher and wider frequency bands for small cells along with existing lower frequency bands for macrocells

To concretize this evolution concept, potential spectrum-efficiency-enhancing candidate technologies are identified for both the wide area and the local area, on both the transmitter and receiver sides. Furthermore, we introduce a macro-assisted small cell, called the *Phantom cell*, as a key solution to efficiently support the integration of the local area with the wide area. A macro-assisted small cell, the Phantom cell indeed establishes a new form of multicell cooperation between macrocells and small cells required for further network densification and spectrum extension into higher frequency bands. In LTE-A, technologies for multicell cooperation between macrocells and small cells are standardized such as coordinated multipoint (CoMP)

 1536-1284/13/$25.00 © 2013 IEEE

interference coordination (eICIC) [4]. Such LTE-A technologies, however, were developed mainly assuming single-frequency operation for macrocells and small cells with co-channel intercell interference between them. In our evolution concept, the frequency-separated deployment between the wide area and the local area is identified as an important key scenario toward the efficient utilization of higher frequency bands. For such a frequency-separated scenario, there are no issues related to macro-to-small intercell interference, but other new issues related to coverage, mobility, and energy/battery savings arise. We first explain these new issues, and then discuss the benefits associated with the Phantom cell solution. Finally, we provide the system-level simulation results for using outdoor small cells in higher/wider frequency bands in order to demonstrate the potential gains associated with our evolution concept.

transmission/reception and enhanced intercell

The rest of the article is organized as follows. We first present high-level requirements envisioning the next decade. We present the proposed evolution concept for the next steps after LTE-A. We identify and discuss the candidate technologies related to the proposed evolution concept. We present results for system-level evaluations. Finally, we present our conclusions.

Future Requirements

The most important requirement in the future is the provision of higher system capacity [5]. Over the past few years, there has been significant growth in the volume of mobile data traffic following the proliferation of smartphones and new mobile devices that support a wide range of applications and services. Last year alone, the volume of mobile data traffic grew 2.3 times, with a nearly threefold increase in the average smartphone usage rate [6]. There is now a general consensus in the industry that such growth will continue in the future as well. Many recent forecasts project mobile data traffic to grow beyond 24 times between 2010 and 2015, and thus beyond 500 times in 10 years (2010–2020) assuming the same pace of growth is maintained. Thus, the capacity of future systems needs to be increased significantly so that they can support such an order of magnitude of growth in traffic volume. Another important requirement is efficient support for a variety of traffic types including reduction in the impact of the ever increasing volume of signaling traffic on both the network and handheld device sides, the support of lower latency for the currently proliferating cloud applications, and the handling of simultaneous connections from the large number of small data packets for machine-to-machine (M2M) traffic.

The next requirement is to provide better user experience by improving both the achievable data rates and fairness in user throughput. The target we set here is to achieve a tenfold improvement in the data rate over the next decade. We also aim to achieve gigabit-per-second-order user experience throughput in the wide area.

One more requirement is to design flexible and cost-efficient network deployments. This becomes of particular importance as future spectrum utilization becomes more diverse, ranging from lower and narrower frequency bands to higher and wider frequency bands, as the network density is further increased by deploying more small cells, and as the simultaneous support of diverse environments and deployment scenarios becomes more important.

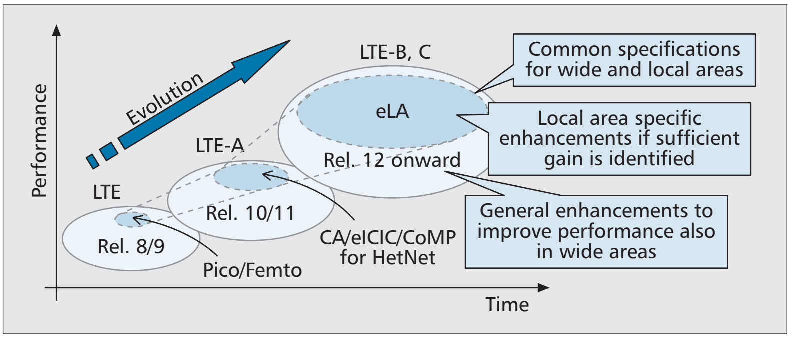


Figure 1. *Future steps for LTE-A (i.e., LTE-B and C) based on the evolution concept of eLA as an integrated part of general LTE enhancements.*

There are also other requirements regarding energy/battery savings and system robustness against emergencies (e.g., tsunamis and earthquakes). These requirements, however, are an implementation issue for which non-standardized solutions would generally be more effective. Thus, we do not set them as a primary target, but rather as an add-on that complements other requirements as part of the whole system design.

Conceptual Views on Solutions

Prior to explaining in detail the candidate solutions to satisfy future requirements, we describe the proposed concept for future evolution. In the future, we believe that wide area enhancement will continue to be important, but local area enhancement will become more important than in the past. Here we use the term *local area* to mainly refer to outdoor dense deployments and hotspots where a high volume of traffic needs to be supported, and better user experience throughput needs to be provided. The key challenge we identify here resides in how to enhance and integrate the local area with the wide area in such a way that future requirements can be satisfied.

The proposed concept for future evolution consists of two aspects:
- Efficient integration of wide and local area enhancements
- Efficient utilization of both lower and higher frequency bands through frequency-separated deployments between wide and local areas

Integration of Wide Area and Local Area Enhancements

In Releases 8 to 11, some local area technologies were introduced to LTE and LTE-A. As we move toward Release 12 onward, the performance of LTE-B, C, and so on needs to be improved further, and local area technologies will play a more important role in the future steps of LTE-A. Thus, the performance of the

The proposed way forward is to use lower frequency bands such as existing cellular bands in macrocells to provide basic coverage and mobility and to use separate frequency bands such as higher frequency bands in local area for small cells to provide high-speed data transmission.

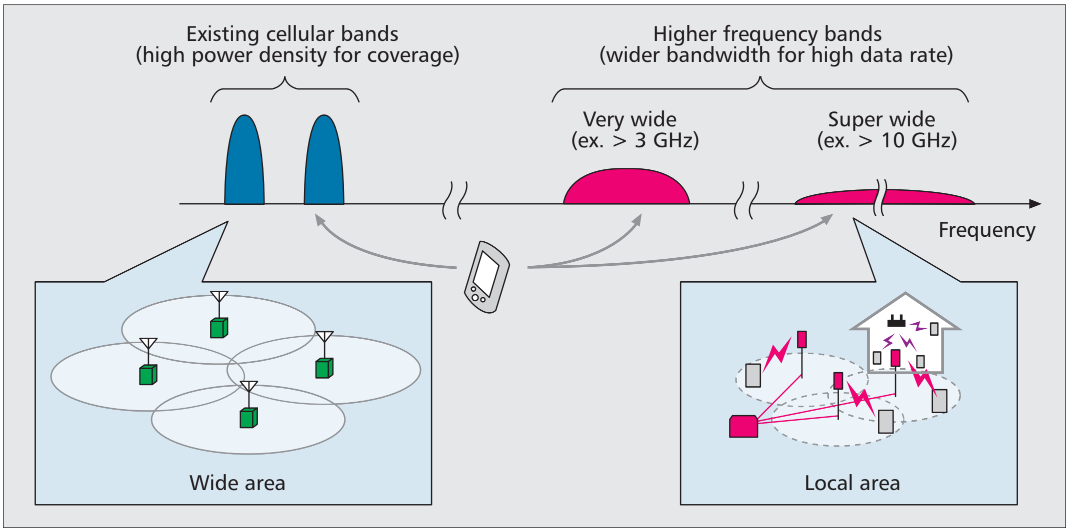


Figure 2. *Frequency-separated deployments between a wide area and a local area.*

local area should be further enhanced, and the required technologies should be specified. We refer to this as an enhanced local area (eLA), as shown in Fig. 1. However, eLA should be an integrated part of LTE Release 12 onward along with general LTE enhancements. We should retain commonality between wide and local areas from the specification point of view, and enhance it as an integrated part of the general enhancements for the wide area. Specific local area enhancements can be specified only if a sufficient performance gain can be identified. Thus, the proposed way forward is to retain common specifications between wide and local areas in terms of a fundamental LTE radio interface, and add eLA specific specifications (e.g., a new carrier type) only when necessary.

Efficient Utilization of Both Lower and Higher Frequency Bands through Frequency-Separated Deployments between Wide and Local Areas

From the spectrum utilization point of view, the spectrum in the lower frequency bands is becoming scarce. Thus, it is crucial to explore and utilize higher frequency bands in the future. However, higher frequency bands are difficult to accommodate in wide areas in macrocells because of either space limitations on the eNode B (eNB) side, for example, in terms of radio frequency (RF) equipment and antenna size, coverage limitations (e.g., higher path loss), or cost issues due to the need to alter the already established network infrastructure. The proposed way forward is to use lower frequency bands such as existing cellular bands in macrocells to provide basic coverage and mobility, and to use separate frequency bands such as higher frequency bands in the local area for small cells to provide high-speed data transmission, as shown in Fig. 2. The figure shows that in local the area, we can consider using a wider spectrum bandwidth in the higher frequency bands and local-area-specific technologies to support smaller and denser cell deployments.

Candidate Solutions

A set of radio access technologies is required in order to expand the current cellular capacity and performance, and thus be able to meet future requirements and challenges. Hereafter, we present our identified candidate solutions to address the future requirements along with three evolution dimensions: spectrum efficiency, spectrum extension, and network densification (referred to as "The Cube" in [3, 5]).

Candidate Technologies for Enhanced Spectrum Efficiency

General Enhancements — LTE radio access based on orthogonal frequency-division multiple access (OFDMA) achieves radio link performance that is close to the Shannon capacity limit for point-to-point communications. However, further overhead reduction or further enhancements for higher signal-to-interference-plus-noise ratio (SINR) regions could still be considered. In LTE-A, the focus of studies in recent releases has shifted to the study of interference solutions for multipoint/antenna/multi-user radio links in order to improve spectrum efficiency. The technical challenges for intracell access (i.e., multi-user multiple-input multiple-output [MU-MIMO) and intercell access (i.e., CoMP transmission technologies) will continue to be studied in LTE-B and later. To enhance the spectrum efficiency further, hopefully without significant increase in the amount of channel state information (CSI) feedback [7], we identify the following two key approaches.

Active Beamforming with More Antenna Elements — Three-dimensional (3D) MIMO/beamforming with active antenna system (AAS) is one of the major technologies that would be studied in LTE-B [8]. AAS, where not only horizontal antennas but also vertical antenna elements have independent RF units, allows flexible beam control in both horizontal and vertical directions, as shown in Fig. 3a1. From a specification perspective, 3D MIMO/beamforming will require the support of more than eight antenna elements. The current physical layer specifications (Release 10) only support up to eight antenna ports in the downlink (DL), for example, for the DL reference signals (RSs) and CSI feedback scheme. Furthermore, extending 3D MIMO/beamforming to MU-MIMO and CoMP operations would need to be investigated.

Massive MIMO, which to some extent is another variant of 3D MIMO/beamforming with AAS but with very large numbers of antennas (e.g., more than 100 antenna elements), is a key technology for small cells in future higher frequency bands (e.g., beyond 10 GHz) [9]. For higher frequency bands, antenna elements can be miniaturized, and more elements can be placed in the same space; thus, very narrow beamforming can be created, as shown in Fig. 3a2. We expect such very narrow beamforming to become essential in higher frequency bands in order to support practical coverage areas for small cells by compensating for the increased path loss. Assuming ideal beamforming gain, a two-dimensional mapping of antenna elements can compensate for the path loss with a frequency factor of 20 dB/decade. However, there are several technical issues toward massive antenna technologies that need to be resolved, such as how to achieve accurate beamforming or how to support control signaling for mobility and connectivity over highly directive links. One possibility to address the control signaling issue is to apply macro-assisted small cells (i.e., the Phantom cell), explained later.

Advanced Receiver Cancellation — Receiver cancellation technologies do not require much CSI feedback, and are generally beneficial in both wide and local areas. In LTE Release 11, an advanced receiver that employs interference rejection combining (IRC) was investigated for user equipment (UE) with two receiver (Rx) antennas [10]. Thus, further study on the IRC receiver with more Rx antennas (e.g., four or eight) will be a natural extension in LTE-B in order to utilize the increased degrees of freedom at the UE receiver for suppressing multiple interference resources, as shown in Fig. 3b1. CoMP technologies considering the IRC receiver (e.g., interference alignment [11]) may be a potential topic for future steps beyond LTE-B, although this might to some extent be achieved through eNB implementations.

In addition to the IRC receiver, we consider nonlinear interference cancellation, such as successive interference cancellation (SIC), as one key future advanced receiver technology for the future radio access. However, one issue regarding the SIC receiver for the UE terminal side is that the SIC receiver requires a condition in which interference signals to be cancelled are decodable at the UE receiver. Thus, it may be challenging to apply a SIC receiver to cancel intercell interference since those signals are typically difficult to decode. Compared to that, the application of SIC to intracell multiple access schemes can be more feasible. Initial investigations reported that intracell non-orthogonal multiple access using a SIC receiver improves spectrum efficiency and user fairness [12].

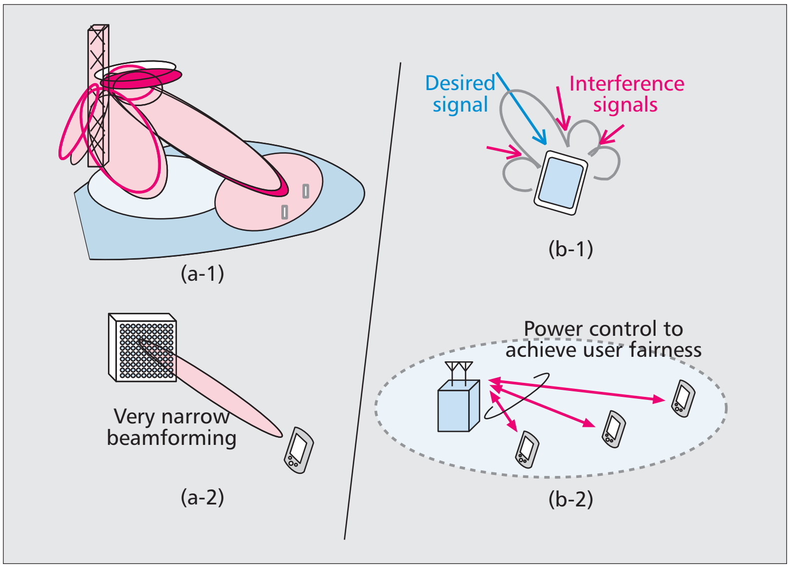


Figure 3. *General enhancements for spectrum efficiency: advanced transmitter and receiver: a1) 3D MIMO/beamforming; a2) massive MIMO in higher frequency bands; b1) advanced receiver with more Rx antennas; b2) non-orthogonal multiple access using a SIC receiver.*

eLA Focused Enhancements — The aforementioned approaches are generally applicable to both wide and local areas. The following enhancements can be additionally considered in order to satisfy the requirements specific to local area.

Dynamic Time-Division Duplex and Interference Coordination/Management for Dense Small Cells — In small cell deployments, it is expected that the number of UE devices per small cell will not be large, and the traffic pattern in each small cell will change widely over time depending on user applications. As a result, spectrum sharing between the uplink (UL) and the downlink (DL), so-called dynamic time-division duplex (TDD) [13], would provide gains in terms of spectrum efficiency compared to the semi-static partition of UL and DL. Study on dynamic TDD is already underway in Release 11. However, when small cells are densely deployed, interference issues such as UL-to-DL or DL-to-UL interference, as shown in Fig. 4a, become more important. Therefore, interference coordination and management schemes for dense small cells especially for dynamic TDD need to be established in LTE-B.

Enhancements for the Higher SINR Region: UL OFDMA and 256-QAM — In small cell deployments, especially in the aforementioned frequency-separated deployments between wide and local areas, it is expected that a relatively high SINR can often be observed at the eNB or UE receiver due to lower intercell interference compared to macrocellular deployments. Therefore, link-level performance enhancements for the higher SINR region (e.g., higher than 20 dB) is beneficial in improving the user-experience throughput in local area. Higher-order modulation such as 256-quadrature amplitude modulation (QAM) can be effective in improving the throughput perfor-

At a certain point in time in the future, it would become necessary to redesign the radio interface assuming the requirements for future newly identified spectrum in the World radiocommunication conferences (WRC-15) and beyond.

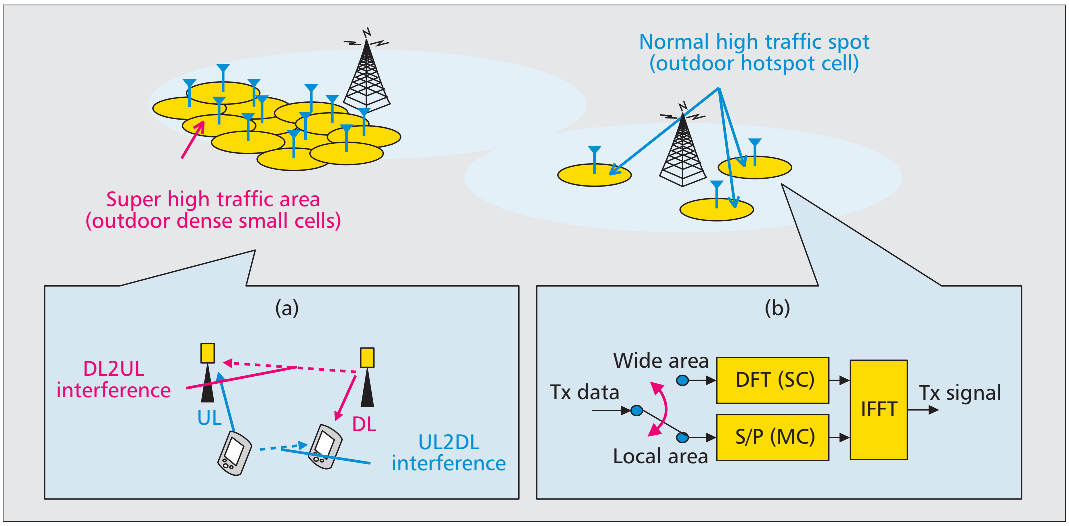


Figure 4. *eLA focused enhancements and deployment scenarios for higher frequency bands.*

mance in a high SINR region. For UL enhancements in a higher SINR region, the benefits of OFDMA compared to current single-carrier frequency-division multiple access (SC-FDMA) need to be revisited. Also, hybrid access between OFDMA and current SC-FDMA [14], as shown in Fig. 4b, can be considered for UL radio access.

Spectrum Extension Using Higher Frequency Bands

Currently, the spectrum below 2.5 GHz is mostly utilized by existing systems. Thus, spectrum extension to 3.5 GHz and higher frequency bands will be required in order to further expand the system capacity in the future steps of LTE-A. As described earlier, in our proposed concept frequency-separated deployments between wide and local areas are emphasized toward the efficient use of higher frequency bands. Thus, spectrum extension and network densification are highly correlated. However, when we look into higher frequency bands, (e.g., beyond 10 GHz), the current LTE radio interface, which was optimized for the existing cellular bands (i.e., around 2 GHz) would not be optimum. Therefore, at a certain point in the future, it would become necessary to redesign the radio interface assuming the requirements for future newly identified spectrum in the World Radiocommunication conferences (WRC-15) and beyond.

From another point of view, the market size for utilizing higher frequency bands needs to be sufficiently large. It is therefore desirable from the operator perspective that higher frequency bands can be used by many UE devices and for many service areas as much as possible. In this sense, higher frequency bands need to be utilized in various deployments, not only for indoor but also for outdoor deployments. Figure 4 shows two typical scenarios for outdoor small cell deployments, that is, normal high traffic spots such as railway stations in suburban areas in which small cells are sparsely deployed, and a super high traffic area such as the downtown area in a large city in which small cells are densely deployed to provide continuous or partial coverage. Furthermore, macrocell deployments using higher frequency bands should be allowed.

Network Densification with Phantom Cell Solution

Macro-Assisted Small Cell: Phantom Cell — In the current deployments, there are a number of capacity solutions for indoor environments such as WiFi, femtocells, and in-building cells using distributed antenna systems (DASs). However, there is a lack of capacity solutions for the outdoor environment that can also support good mobility and connectivity. Thus, we propose a macro-assisted small cell, called the Phantom cell, as a capacity solution that offers good mobility support while capitalizing on the existing LTE network. In the Phantom cell solution, the control (C)-plane/user data (U)-plane are split as shown in Fig. 5. The C-plane of UE in small cells is provided by a macrocell in a lower frequency band, while for UE in macrocells both the C-plane and U-plane are provided by the serving macrocell, the same as in a conventional system. On the other hand, the U-plane of UE in small cells is provided by a small cell using a higher frequency band. Hence, these macro-assisted small cells are called Phantom cells, as they are intended to transmit UE-specific signals only, and the radio resource control (RRC) connection procedures between the UE and a Phantom cell, such as channel establishment and release, are managed by the macrocell. The Phantom cells are not conventional cells in the sense that they are not configured with cell-specific signals and channels, that is, primary/secondary synchronization signals (PSSs/SSSs), cell-specific reference signals (CRSs), master information block and system information blocks (MIB/SIBs), and so on. Their visibility to the UE relies on macrocell signaling.

The Phantom cell solution comes with a range of benefits. One important benefit of macro assistance of small cells is that control signaling due to frequent handover between small cells and macrocell and among small cells can be significantly reduced, and connectivity can be maintained even when using small cells and higher frequency bands. In addition, by applying the new carrier type (NCT) that contains no or reduced legacy cell-specific signals, the Phantom cell is able to provide further benefits such as efficient energy savings, lower interference transmissions (and thus higher spectrum efficiency), and reduction in cell planning efforts for dense small cell deployments since Phantom cells do not rely on legacy cell-specific signals.

Technical Issues Related to the Phantom Cell

Network Architecture — To establish a network architecture that supports the C/U-plane split, interworking between the macrocell and the Phantom cell is required. A straightforward solution to achieve this is to support Phantom cells by using remote radio heads (RRHs) belonging to a single macro eNB. This approach can be referred to as intra-eNB carrier aggregation (CA) using RRHs [15]. However, such a tight CA-based architecture has some drawbacks as it requires single-node operation with low latency connections (e.g., optical fibers) between the macro and Phantom cells. In the future, more flexible network architectures should be investigated to allow for relaxed backhaul requirements between macro and Phantom cells and to support a distributed node deployment with separated network nodes for each (i.e., inter-eNB CA).

Macro-Assisted Discovery — In frequency-separated deployments, efficient small cell discovery in higher frequency bands is an important technical issue. To achieve efficient discovery (e.g., for UE battery savings), we propose to use a macro-assisted property in the Phantom cell concept and to introduce newly defined discovery signals, which are transmitted by small cells in a time-synchronized manner with macro downlink signals. Assisted by the C-plane provided by the macrocell, small cells simultaneously transmit the discovery signals with a relatively long transmission interval (e.g., longer than 100 ms), and the UE attempts the detection of discovery signals from small cells only in a short time interval. Thus, the interfrequency measurement effort and battery consumption of UE to detect the surrounding small cells in higher frequency bands can be reduced. Furthermore, network energy savings when there is no traffic are also expected due to the long transmission interval. The discovery signal should be designed to satisfy the requirements for small cell deployment such as robustness against intercell interference for the scenarios including user-deployed or closed subscriber group (CSG) cells, support for a large number of sequences to reduce the cell planning effort, and so on.

Flexible Duplex Support — In frequency-separated deployments, different duplex schemes, such as frequency-division duplex (FDD) and TDD, may be used for lower and higher frequency bands, respectively. In the future, it is desirable to support the Phantom cell (or inter-eNB CA) solution irrespective of the combination of duplex schemes in lower and higher frequency bands.

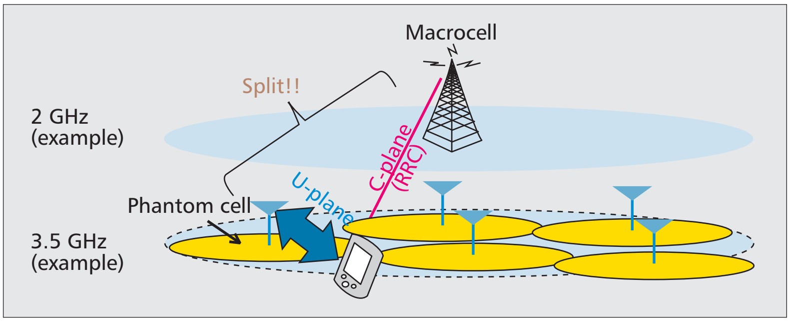


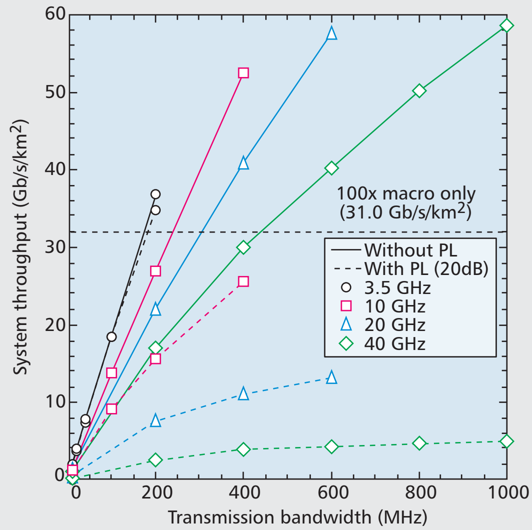
Figure 5. *Phantom cell: macro-assisted small cell.*

Simulation Evaluations

In this section, we present system-level simulation results to demonstrate the potential capacity gains of our evolution concept where outdoor small cell deployments are assumed using higher/wider frequency bands. The simulator we used is built in C language and consists of system-level modeling utilizing exponential SINR link-to-system-level mapping [16]. Figure 6 shows the system throughput performance per square kilometer as a function of the transmission bandwidth assuming variable carrier frequencies ranging from 3.5 to 40 GHz. The main simulation parameters are also depicted in the same figure. In this evaluation, we assume 12 small cells are randomly located within each sector, where each macrocell site consists of three sectors and intersite distance is set to 500 m. Thus, the size of each sector is about 0.072 km^2, and the density of small cells is 166.7 cells/km^2. We also assume that all U-plane related transmissions are conducted by small cells. In order to evaluate the impact of propagation loss increase in higher frequency bands, the distance-dependent path loss model in [17] is assumed. Considering that all small cells are deployed outdoors to further enhance system capacity, we evaluate two cases that differ in terms of the penetration loss. The first one is the case without penetration loss assuming all users are outdoors, and the other one is the case with penetration loss (= 20 dB) assuming all users are indoors. We note that the simulation results for the macrocell only case (without small cells) is 310 Mb/s/km^2 assuming a transmission bandwidth of 10 MHz in a 2 GHz carrier frequency.

In the case without penetration loss, the simulation results (solid lines) show that spectrum extension using higher frequency bands in small cells significantly increases the system throughput per square kilometer. For example, a system throughput of 31.0 Gb/s/km^2, which corresponds to a 100 times gain compared to that for the macrocell only case, is achieved by transmission bandwidth of 170, 230, 300, and 400 MHz for carrier frequency of 3.5, 10, 20, and 40 GHz, respectively. Meanwhile, for the case with a 20 dB penetration loss included, the simulation results (dotted lines) show that spectrum extension for the small cells using the higher frequencies beyond 10 GHz is not as effective due to the increase in the compound loss due to path loss and penetration loss. However, spectrum extension in 3.5 GHz carrier frequency remains very effective. Based on the results, we would like to express the following observations.

The establishment of the evolution concept is of particular importance in setting the evolution framework so that 3GPP standardization can focus the specification work on usage cases and deployment scenarios that really matter not only in the short term but also on the long term.



Macrocell layout	19 cells/3 sectors per cell intersite distance = 500 m
Number of small cells	12 per macro sector
Number of users	30 per macro sector
Moving speed of UE	3 km/h
Antenna configuration	2 × 2
Tx power of small cell	33 dBm
Distance dependent path loss	36.7log10(d) + 26log10(f) + 22.7 (d in m, f in GHz)
Shadowing standard deviation	8 dB
Channel mode	Pedestrian-A
Traffic model	Full buffer
Scheduling algorithm	Proportional fairness

Figure 6. *Throughput gains using higher frequency bands (12 small cells per macro sector).*

- In 3.5 GHz carrier frequency, outdoor small cell deployment combined with the spectrum extension is very effective to boost the system capacity.
- In higher frequency bands beyond 10 GHz, technology to overcome the propagation loss, such as beamforming (massive MIMO) is highly required to ensure outdoor small cells are useful not only for outdoor users but for indoor users as well.

Conclusion

In this article we present our views on the requirements, evolution concept, and candidate technologies for the future steps of LTE-A. The establishment of the evolution concept is of particular importance in setting the evolution framework so that 3GPP standardization can focus the specification work on usage cases and deployment scenarios that really matter not only in the short term but also in the long term. In the proposed evolution concept we emphasize on the importance of integrating local area into wide area enhancements. Also, we introduce the set of radio access technologies we have identified as the most promising for improving spectrum efficiency in both wide and local areas. In particular, we have identified frequency-separated deployment between wide and local areas as an important scenario toward efficient utilization of higher/wider frequency bands in local area while still maintaining basic mobility and connectivity in wide area over existing lower frequency bands. For this scenario, the macro-assisted small cell, called the Phantom cell, with a split of C-plane and U-plane between the macrocell and small (Phantom) cells, is introduced as a new form of multicell cooperation. In order to confirm the potential benefits of our evolution concept, simulation evaluation results assuming outdoor small cell deployments using higher/wider frequency bands are presented. Outdoor small cell deployment combined with spectrum extension is shown as an effective approach to boost the system capacity and achieve the future requirements.

References

[1] M. Tanno *et al.*, "Evolved UTRA — Physical Layer Overview," *Proc. IEEE SPAWC 2007*, June 2007.
[2] E. Dahlman, S. Parkvall, and J. Sköld, *4G –LTE/LTE-Advanced for Mobile Broadband*, Academic Press, 2011.
[3] NTT DOCOMO, "Requirements, Candidate Solutions & Technology Roadmap for LTE Rel-12 onward," 3GPP RWS-120010, June 2012.
[4] T. Abe *et al.*, "Radio Interface Technologies for Cooperative Transmission in 3GPP LTE-Advanced," *IEICE Trans. Commun.*, vol. E94-B, no. 12, Dec. 2011, pp. 3202–10.
[5] Y. Kishiyama, "LTE Enhancements and Future Radio Access," *Seminar on Future Wireless Technology*, Nov. 2010.
[6] Cisco Whitepaper, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011-2016," 15 Feb 2012.
[7] S. A. Ramprashad *et al.*, "Cooperative Cellular Networks Using Multi-User MIMO: Tradeoffs, Overheads, and Interference Control Across Architectures," *IEEE Commun. Mag.*, vol. 49, no. 5, May 2011, pp. 70–77.
[8] O. Yilmaz, S. Hamalainen, and J. Hamalainen, "System Level Analysis of Vertical Sectorization for 3GPP LTE," *Proc. ISWCS 2009*, Sept. 2009.
[9] T. L. Marzetta, "Non-Cooperative Cellular Wireless With Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, Nov. 2010.
[10] Y. Ohwatari *et al.*, "Performance of Interference Rejection Combining Receiver to Suppress Inter-Cell Interference in LTE-Advanced Downlink," *IEICE Trans. Commun.*, vol. E94-B, no. 12, Dec. 2011, pp. 3362–69.
[11] R. Cadambe and S. A. Jafar, "Interference Alignment and Degrees of Freedom of the *K*-User Interference Channel," *IEEE Trans. Info. Theory*, vol. 54, no. 8, Aug. 2008, pp. 3425–41.
[12] K. Higuchi and Y. Kishiyama, "Non-Orthogonal Access with Successive Interference Cancellation for Future Radio Access," *APWCS2012*, Aug. 2012.
[13] J. Li *et al.*, "Dynamic TDD and Fixed Cellular Networks," *IEEE Commun. Letters*, vol. 4, no. 7, July 2000, pp. 218–20.
[14] R. Dinis *et al.*, "A Multiple Access Scheme for the Uplink of Broadband Wireless Access," *Proc. IEEE GLOBECOM*, vol. 6, Dec. 2004, pp. 3808–12.
[15] M. Iwamura *et al.*, "Carrier Aggregation Framework in 3GPP LTE-Advanced," *IEEE Commun. Mag.*, vol. 48, no. 8, Aug. 2010, pp. 60–67.
[16] K. Brueninghaus *et al.*, "Link Performance Models for System Level Simulations of Broadband Radio Access Systems," *Proc. IEEE PIMRC '05*, vol. 4, Sept. 2005.
[17] ITU-R Rep. M.2135, "Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced," Nov. 2008.

MULTICELL COOPERATION: EVOLUTION OF COORDINATION AND COOPERATION IN LARGE-SCALE NETWORKS

HARALD BURCHARDT AND HARALD HAAS, UNIVERSITY OF EDINBURGH

ABSTRACT

Due to the large growth of mobile communications over the past two decades, the supporting cellular systems have continuously needed to expand and evolve in order to meet the ever increasing demand of wireless connections. While simple frequency reuse and power control were enough to manage the demand of these networks at first, multicellular cooperation and coordination have become paramount to the operation of larger and more highly utilized communication systems. This article discusses the development of different techniques for cooperation in large cellular networks, and offers insights into the need for such an evolution. We investigate various branches of multicell cooperation, including user-based cooperation, system-wide optimization, and the opportunity of multiple-base-station transmission. Furthermore, we offer an example of a recently proposed technique designed for multicell cooperation in full frequency reuse orthogonal frequency-division multiple access networks.

INTRODUCTION

Since the introduction of mobile technologies over two decades ago, wireless communication has evolved into a utility similar to water and electricity, needed by almost all people of today's modern society. To support the ever growing demand for mobile communications, cellular networks have had to evolve from simple local service providers to massively complex cooperative systems. While on one hand the underlying transmission methods have been systematically advanced with each generation of wireless technologies, that is:

- Time-/frequency-division multiple access (T/FDMA) in second generation (2G) networks
- Code-division multiple access (CDMA) in third generation (3G) networks
- Orthogonal frequency-division multiple access (OFDMA) in fourth generation (4G) networks

to improve the capacity and availability of these networks; on the other hand, the inherently enhanced complexity of these transmission schemes necessitates more intelligent methods of cooperation and system coordination. Ultimately, the dramatically increasing demand for mobile communications [1] requires more and more intelligent utilization of the radio frequency (RF) spectrum, and enhanced cooperation between cells.

Before we begin, the ultimate goals of multicell cooperation must be introduced. By coordinating resource and power allocation over multiple cells (or indeed a whole network), the system is able to provide simultaneous service to thousands of users. Thus, the main goals of cooperation can be compiled as:

- The optimal utilization of radio resources within a cellular network
- The minimization of interference to neighboring cells
- Hence, the maximization of the number of simultaneously transmitting users

In the rest of this introduction, we describe the most basic forms of network management without cooperation, and highlight the need for more advanced coordination techniques. Following this, we investigate three areas of multicell cooperation and how they intend to solve the problem of increased throughput demands for the limited bandwidth available for wireless systems. Finally, we conclude with a performance analysis of an example of such techniques, and the envisioned further evolution to future mobile networks.

FREQUENCY REUSE

The most basic form of interference mitigation utilized in wireless networks is that of frequency reuse. Consider two mobile stations (MSs) in neighboring cells transmitting on the same band of frequencies; these will surely interfere with each other and, consequently, degrade each other's signal. Hence, the logical conclusion would be to allocate orthogonal bandwidths to avoid interference. However, due to the quantity

 1536-1284/13/$25.00

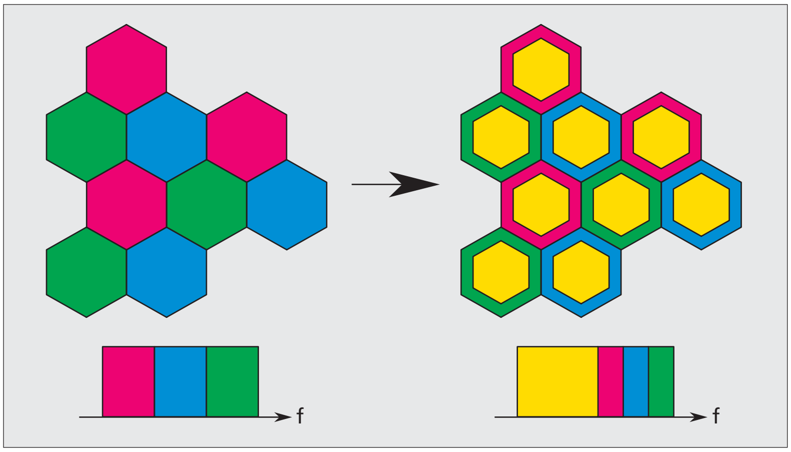


Figure 1. *On the left is the most basic from of frequency reuse, where the available bandwidth is divided into three equal bands, and distributed such that no neighboring cells utilize the same band. On the right, FFR improves the spatial reuse of resources by reusing the same band in the center of the cell, and protecting the edges through standard frequency reuse.*

of MSs requesting resources and the limited bandwidth available, it is clear that not all users can be allocated different frequencies (i.e., the frequencies must be reused). Utilizing the same frequency bands in neighboring cells can be detrimental to the achievable throughputs in each cell (as stated in the example); however, as the wireless signal attenuates directly as a function of the distance from the transmitter, a frequency band may be reused in cells that are sufficiently far away from each other. Thus, a frequency reuse scheme results (shown on the left of Fig. 1) in two neighboring cells never sharing the same set of transmission frequencies, and MSs can transmit virtually interference-free.

Fractional Frequency Reuse —While MSs in neighboring cells do not interfere with each other, it is clear that only a portion (1/3 in the example in Fig. 1) of the already limited system bandwidth is available in each cell; therefore, the number of users servable by the system is dramatically reduced. Therefore, fractional frequency reuse (FFR) is adopted in more modern wireless systems [2], where the center of all cells utilize the same frequency band, and only the cell edges (which are more vulnerable to interference from other cells as they are closer) employ frequency reuse. An example of this is shown on the right of Fig. 1, and it is clear that the bandwidth available in each cell has doubled in comparison to traditional frequency reuse. However, in future systems where full frequency reuse (i.e., the full bandwidth is used in all cells) is planned, cells will most certainly need to cooperate in order to ensure interference mitigation for their users.

Power Control

In general, a wireless base station (BS) will transmit at maximum power in the downlink in order to serve the MSs connected to it. However, in the uplink this can cause substantial difficulties such as large interference at a neighboring BS or massive received signal differences (from different users) at the same BS, referred to as the *near-far problem*. Therefore, transmit power control (TPC) is utilized in the uplink of most mobile communications systems to balance the signal-to-noise ratios (SNRs) of the MSs in a cell at the BS, and hence ease simultaneous reception and create a fairer system. In TPC, each MS transmits at a power level that is just large enough to overcome the path loss, or signal attenuation, from itself to its serving BS. Through this, nearby users will transmit with much lower power than users at the cell edge, as a high-power user near the BS would essentially drown out farther users. Furthermore, due to the significantly reduced transmit powers, the interference to neighboring cells is limited, additionally easing simultaneous transmission of users throughout the network. This will be essential for full frequency reuse networks.

While these basic forms of coordination were able to satisfy early networks' service demands, the vast increase in cellular activity cannot be compensated by these single-cell-management techniques. Hence, multicell cooperation has become more than necessary.

User-Based Cooperation

Due to the ever increasing complexity and utilization of modern mobile communications networks, it has become more and more difficult to serve the large number of MSs requesting service. Furthermore, resource allocation on a single-cell basis is no longer able to sustain the demands of these networks; hence, users often fall into outage. In this section, a class of techniques that utilize individual users' transmission requirements to perform multi-cell resource allocation is discussed. Here, cells adjust their resource selection and scheduling based on the MSs in neighboring cells already transmitting. A few examples of such techniques are discussed below.

Dynamic Frequency Reuse

The main drawback of static FFR is the inability to adapt to an ever changing environment and the consequent waste of wireless resources. Therefore, in dynamic frequency reuse (DFR) [3, 4] the allocation of frequency subbands to different cells is determined based on the contemporary resource allocation and consequent immediate interference environment in the cellular system. Ideally, we would prefer all frequencies to be available in all cells, so at minimum an FFR scheme with a common frequency band in the cell centers is applied. Subsequently, the frequencies allocated to the cell edge may be adapted based on the use of specific frequency bands in the neighboring cells, and the interference incident from these to the cell of interest. Clearly, this necessitates some signaling between BSs such that the possibility of close frequency reuse can be determined in each cell. Therefore, the frequency and consequent resource allocation in each cell is adapted dynamically to the behavior of the cellular system.

DFR may be implemented by a central or distributed approach. In centralized approaches, resources are assigned to BSs by means of a central controller, which generally achieves more efficient resource utilization, at the expense of additional signaling and higher complexity of the network infrastructure. In a distributed approach, each BS autonomously carries out the resource allocation, attempting to individually react to the interference environment. In distributed DFR methods, however, BSs generally may access only a predefined number of subbands, so these approaches provide little flexibility in subband reassignment when interference conditions change.

Successive Interference Cancellation

While frequency reuse schemes help protect cells from interference, they are unable to actively eliminate co-channel interference (CCI) at the MS. This challenge, however, can be met through *interference cancellation*. To accurately define successive interference cancellation (SIC), we use a definition given in [5], stating that "interference cancellation should be interpreted to mean the class of techniques that demodulate and/or decode desired information, and then use this information along with the channel estimates to cancel received interference from the received signal." In other words, signal processing is utilized after reception to improve the quality of the desired information, while removing the interference from the received signal. An example of a simple implementation of SIC is shown in Fig. 2.

In SIC, the strongest signal (it has, inevitably, the highest signal-to-noise-plus-interference ratio [SINR]) is detected and decoded first, then the next strongest, and so on. After the received signals have been reconstructed (through the respective channel estimations), these can be subtracted/cancelled from the composite received signal to improve the performance of the remaining MSs (i.e., their signal detection). Thus, with each removed signal, the interference in the composite received signal is reduced, and the remaining signal SINRs consequently enhanced.

It has been shown that SIC can nearly achieve the Shannon capacity for multi-user additive white Gaussian noise channels [5]. Many well-known examples of interference cancellation are utilized today, such as the decision feedback equalizer (DFE) and the original Bell Labs layered space-time (BLAST) system. In the DFE, previously decoded symbols are utilized to cancel intersymbol interference (ISI) of future received symbols, given that the channel is known. In BLAST decoding, spatial interference in multiple-input multiple-output (MIMO) systems can be removed in order to separate spatially multiplexed streams of data. Overall, there are many similarities [5] between the elimination of:
- ISI (equalizers)
- Spatial interference (MIMO receivers)
- Consequently, multi-user interference cancellation

Hence, while SIC for multicell cooperation may be rather novel, the idea of interference cancellation has been well proven over the years.

Of course, the great benefits that come from SIC in multi-user environments come at a price, which is not always satisfiable. In general, the concept of interference cancellation relies on the premise that the received signal can be reliably estimated. This means that not only can the desired signal be decoded, but also the received interfering signals. For this, a receiver must know both what was transmitted by an interferer (i.e., what data was sent with which modulation and coding scheme [MCS]) and the channel over which it was sent (i.e., how it has been modified). On one hand, the system may signal among the BSs the MCSs of all users simultaneously transmitting on the same frequencies such that this interference can be cancelled; this would, however, drastically increase the signaling burden on the network (e.g., in comparison to frequency reuse techniques). On the other hand, even if this information is known, the receiver must perfectly estimate the interfering channel to be able to decode the signal, which is highly improbable. Hence, while SIC provides great promise for multicell coordination, its requirements are very difficult to fulfill.

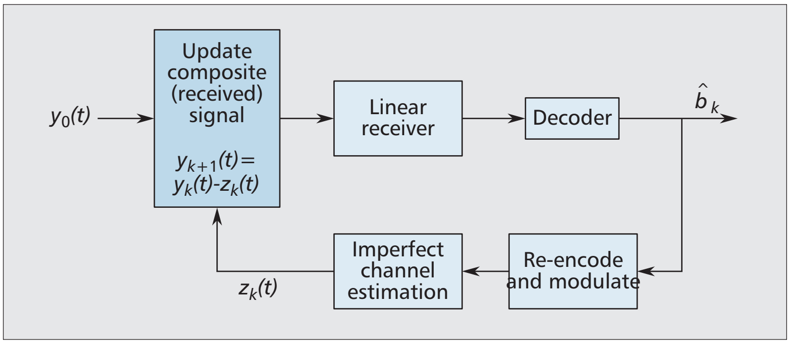


Figure 2. *An example of SIC, where the quality of the desired signal is improved by successively removing the detected interfering signals from the received signal.*

Uplink Interference Protection

It is clear that the high complexity of SIC and its heavy reliance on channel estimation make it difficult to implement in cellular communication networks. In addition, multi-cell cooperation and coordination are more difficult to achieve in the uplink; however, here is where it is usually most needed. Because each MS has a much more limited power budget than any BS, reducing interference in the uplink is paramount to modern mobile communications. Therefore, much work (e.g., TPC, briefly discussed earlier) is dedicated to alleviating the interference, especially that incident on cell edge users in the reverse link. One such technique recently proposed is uplink interference protection (ULIP), summarized here.

Whereas TPC determines MS transmit power based on local information, ULIP is a technique that actively combats CCI in the uplink of a cellular system [6]. The level of uplink interference originating from neighboring cells (affecting co-channel MSs in the cell of interest) can be effectively controlled by reducing the transmit power of the interfering MSs (i.e., rather than attempting to cancel the signal entirely). This is done based on the target SINR and tolerable interference of the vulnerable link. Bands are prioritized in order to differentiate those (vulnerable/ victim) MSs that are to be protected from interfer-

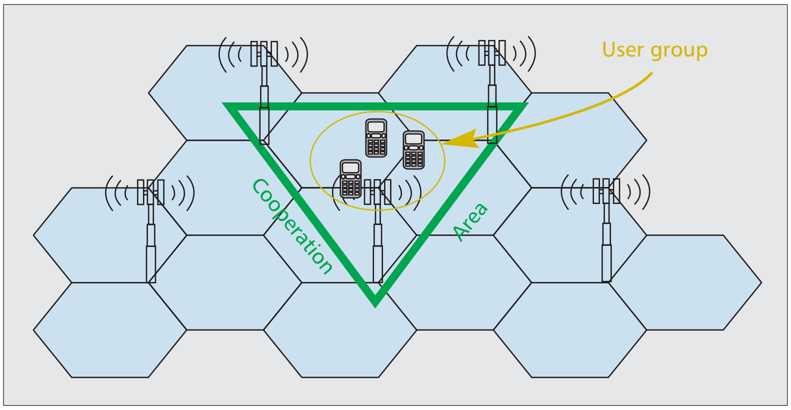


Figure 3. *A visualization of CoMP, where the users within neighboring cells that experience identical sets of strongest BSs can be served simultaneously and cooperatively by this set of BSs, eliminating interference from what would be the strongest interfering cells.*

ence and those (aggressor/interfering MSs) that are required to sacrifice transmission power to facilitate the protection. Furthermore, MSs are scheduled such that those users with poorer transmission conditions receive the highest interference protection, thus balancing the areal SINR distribution and creating a fairer allocation of the available resources. In addition to interference protection, the individual power reductions also serve to decrease the total system uplink power, resulting in a greener system.

However, while ULIP significantly protects cell edge users from uplink interference, it is clear that a substantial amount of signaling is necessary to facilitate this protection. Not only must the tolerable interference of each user be calculated and sent to its interferers; the interfering channel to that user must also be estimated. Fortunately, this can be done at the interfering MS using the reference signal received power (RSRP) of the cell with which it is interfering, and thus does not add any signaling complexity to the network.

System-Wide Optimization

In the user-target-based cooperation techniques discussed above, it is clear that a significant amount of information must be signaled between BSs and MSs in order to facilitate successful simultaneous transmission in multicellular systems. Furthermore, because these techniques rely on individual users' requirements, it is clear that they very rarely result in a system-optimal solution. Therefore, another class of techniques is considered where, assuming a central controller with all necessary information, the multicell cooperation and resource allocation problem can be expressed as an optimization problem. Through this, such techniques attempt to reach globally optimum solutions for a system, at the cost of near unlimited knowledge at the controller. Some examples of these techniques are discussed here.

Capacity Maximization

In wireless networks, the resource and power allocation problem of the system as a whole is often posed as an optimization problem where the objective is the maximization of the achievable throughput (i.e., sum rate) of the system [7]. Of course, individual user requirements must also be met, constraining the set of solutions to the problem. Therefore, a system optimization problem is generally constructed as follows:

$$\text{Objective}: \max \sum_i \text{UserRate}_i \quad i = 1, 2, \ldots N \text{ users}$$

$$\text{subject to:}$$

$$\text{Constraints}: \begin{cases} \text{UserRate}_i \geq \text{RateReq}_i & \forall i \\ \text{UserDelay}_i \leq \text{MaxDelay}_i & \forall i \\ \sum \text{UserPower}_i \leq \text{TotalPower} & \\ \vdots & \end{cases}$$

where the most common constraints posed are the user rate requirements and transmit power limitations. However, these may be extended to more detailed QoS requirements such as delay, buffer length, and priority. By solving the above problem, a system-wide optimal resource allocation for the network can be found.

Although such a method seems preferable to any individual-user-driven cooperation techniques, there are two significant drawbacks of system optimization that greatly reduce its applicability to wireless cellular networks. First of all, the biggest difficulty of such an optimization problem is that in almost all cases this problem becomes $\mathcal{NP}$-hard, and is therefore practically unsolvable using standard optimization techniques. In most cases, the scheduling problem is then broken down into multiple optimization problems, separating the resource allocation and transmit power determination, resulting in ultimately suboptimal solutions. Furthermore, because of the interdependency of, for example, subcarrier allocation and transmit power, solving the resource allocation does not guarantee a solution to the transmit power problem. Therefore, the optimality of these techniques is hardly ever achieved.

Second, not only is there a massive computational complexity involved, but also the large amount of system information that is required by the central controller to solve these system-wide problems must be communicated from each BS, implying a tremendous signaling burden on the network. Not only must the individual user requirements be signaled, but also the desired and interfering path gains of each MS in the system must be known at the central controller in order for an optimal resource and power allocation to be formed. Therefore, while in theory these problems present optimal network solutions, they are in practice almost impossible to implement.

Power Minimization

Another form of the above optimization problem is, rather than maximize the sum rate, to minimize the total transmit power utilized in the wireless network [8]. Here, clearly, the power constraint falls away as it implies the maximization of network performance with maximum power.

On the other hand, the individual user

requirements are the main constraints, where especially the rate of each user must be maintained; otherwise, a system without any transmission would result. This form of optimization is beneficial in two main ways:

- The minimization of interference to other cells eases power and resource allocation over the network.
- The minimization of transmit power results in much greener networks, a topic of fervent discussion for future wireless communications.

However, the same difficulties apply to this problem as to the rate maximization variation, and hence the practical implementation of such a system is highly complex.

LARGE-SCALE MULTIPLE-ANTENNA SYSTEMS

In the previous sections, we have discussed two contrasting forms of multicell cooperation and coordination: lower-complexity user-based cooperation and substantially complex system optimization. Furthermore, in a network with thousands of BSs and antennas, the signaling burden required by both of these coordination dogmas becomes immense. While this cannot necessarily be avoided, a new class of multicellular coordination that exploits this abundance of transmission elements is investigated, utilizing multiple antennas cooperatively to achieve not only spectral but also energy efficiency gains. This is discussed here.

The recently proposed concept of large-scale multiple-antenna systems (LSMASs) is a radical step toward a new way of conceiving MIMO systems for energy efficiency optimization [9]. LSMASs are extended MIMO systems with tens or hundreds of radiating elements, each one consuming a small amount of energy, which may be collocated at a BS, spread out on the face of a building, or geographically distributed over an entire network. LSMASs provide a plethora of advantages over conventional MIMO systems. In particular, these large-scale systems offer higher data rates, increased link reliability, and potential energy savings through the exploitation of the many degrees of freedom offered by the many antenna elements. Meanwhile, random impairments such as thermal noise and other-cell interference can be averaged out. While in the past multiple-antenna systems have been used to provide capacity gains through spatial diversity, energy-efficient networks can be achieved by using massive multi-antenna systems where the many antennas are used to form and focus directed beams to a multiplicity of terminals on the forward link, and selectively collect signals from these terminals on the reverse link. Furthermore, the proposed approach is scalable with the number of antennas, and it provides energy savings that increase with the number of radiating elements. On the other hand, there are several theoretical (capacity limits, power allocation strategies, signal transmission, channel estimation, amount of feedback, etc.) and practical (antenna design, compensation of mutual coupling and spatial correlation, etc.) issues that need to be addressed. Overall, however, LSMASs offer a radical new direction in multicellular cooperation to improve not only the spectral efficiency of wireless communications networks, but also generate large power savings on the existing infrastructure.

COORDINATED MULTIPOINT TRANSMISSION

One of the most fervently discussed topics for cooperative multicellular communication in future network is that of coordinated multipoint transmission (CoMP) [10], where neighboring BSs transmit together to multiple users in their cells. In CoMP, MSs in neighboring cells are grouped based on the received pilot signal strength from the surrounding BSs (i.e., users are grouped if the strongest BSs they see are identical). Once these groups are formed, each of the MSs in each group is served simultaneously by the BSs which provide the strongest signal strengths for that group (Fig. 3). Consequently, these BSs jointly precode their transmissions to the user group such that the combined received signal from the set of BSs at each MS provides the user with its desired signal (i.e., which would have been transmitted from its serving BS), without receiving any interference from the other cells in its set of strongest BSs.

In [11], the combination mechanisms of such preprocessing techniques are described and compared:

- **Joint TDMA (J-TDMA):** Only one MS in the user group is served in each time slot. Thus, transmission is automatically CCI-free, at the cost of reduced spectral efficiency.
- **Joint zero-forcing (J-ZF):** The joint preprocessing matrix $\mathbf{F} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1}$ is the pseudo-inverse of the aggregate channel matrix $\mathbf{H}$ such that each user only receives a noisy version of its intended data. However, just as ZF receivers amplify noise, J-ZF precoding generally increases the average transmit power.
- **Joint minimum mean square error (J-MMSE):** This preprocessing achieves a compromise between CCI cancellation and transmitter power efficiency. Based on the MMSE criterion, the precoding matrix is given by $\mathbf{F} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H + (N_0/\rho)\mathbf{I})^{-1}$, where N_0/ρ is a scaled version of the noise.

Essentially, CoMP eliminates the interference from what would normally be the most strongly interfering BSs, consequently enhancing SINR and thus throughput at each MS.

A common problem in the CoMP concept is that the gain strongly relies on ideal "user centric" clustering and assignment, meaning all MSs should be served by their individual set of strongest cells. In a large network with irregularly distributed MSs in the cells, the difficulty is to find sufficient users with these sets being identical, since due to shadowing the signal strength of BSs might be widely distributed. Thus, the typical penetration rate of MSs that might be served by such a single specific set of cells is often very small, causing many users to be denied service. In [12] partial CoMP, where shifted coverage areas increase the probability of MSs sharing a set of strongest cells, is suggested to alleviate this problem. While the performance gains of par-

LSMASs are extended MIMO systems with tens or hundreds of radiating elements, each one consuming a small amount of energy, which may be collocated at a BS, spread out on the face of a building, or geographically distributed over an entire network.

Method	Objective	Complexity	Signaling	Optimality	Spec. Eff.	Energy Eff.
FFR	Interference mitigation	++	++	– –	– –	– –
TPC	Uplink fairness	++	+	– –	– –	+
DFR	Interference management	+	+	–	–	– –
SIC	Interference cancellation	–	– –	–	+	–
ULIP	Interference mitigation	+	–	+	+	++
Sys. opt.	Objective optimization	– –	– –	++	++	++
LSMAS	Efficient transmission	–	–	+	+	++
CoMP	Cooperative transmission	–	–	+	++	++
POS	Interference management	+	–	–	+	++

Table 1. *Technique performance comparison: Scores ranging over {– –, –, +, ++} are given in each performance area. '– –' indicates terrible performance, and '++' indicates superior performance (i.e., '++' in complexity means very low complexity, whereas in spec. eff. it signifies very high spectral efficiency). The general trend is evident that the more optimal and efficient the system is, the more complex it becomes.*

tial CoMP reach similar levels as those achievable via full CoMP, the additional signaling and complexity must be compensated by limiting transmission of reference signals by the MSs. However, it is clear that large SINR and throughput gains are achieved.

While CoMP does not achieve the performance of system optimization (i.e., albeit at significantly reduced complexity), it is able to provide large gains over user-based cooperation schemes. In general, a substantial increase in signaling over traditional intercell cooperation techniques is necessary to facilitate CoMP. However, this is true for most of the recent multi-cell cooperation techniques, as the "simple" solutions are no longer sufficient to provide the large throughput demands of today's networks. Furthermore, it is evident that joint processing and cooperative transmission techniques are necessary for future networks, implying that user clustering techniques will also need to be optimized to improve system performance. In general, however, it is clear that multicellular cooperation, and specifically LSMASs, will be paramount to the performance and development of future wireless communications networks. A performance comparison of the discussed cooperation techniques is shown in Table 1. Pareto optimal scheduling (POS), which combines aspects from user-based cooperation and CoMP, is described next.

Pareto-Based Optimal Transmission Scheduling

In our final section, we would like to introduce a recently proposed OFDMA resource allocation and power control method [13] based on Pareto optimal power control (POPC) [14], which combines some of the aspects of user-based cooperation and CoMP. In POPC, given a feasible link allocation (i.e., a group of users in neighboring cells allocated the same resource block [RB](s)), a transmit power vector **P** can be found such that all users achieve their SINR requirements with minimal power. This is, of course, a highly desirable result that, depending on the location and service requirements of the interfering MSs, is clearly not always possible. Hence, by scheduling users in such a manner to maximize the number of feasible user groupings (in principle, there can be as many user groups as there are [orthogonal] RBs in the system), the system spectral efficiency can be maximized.

However, scheduling users in such a manner is not a trivial task. In POPC, the desired and interfering path gains of users in neighboring cells allocated the same frequency resource(s) are combined in a matrix **F**, where

$$F_{ij} = \begin{cases} 0, & \text{if } i = j \\ \dfrac{\gamma_i^* G_{j,i}}{G_{i,i}}, & \text{if } i \neq j \end{cases},$$

the SINR target of user i is γ_i^*, and $G_{j,i}$ is the path gain from MS j in cell j to the BS in cell i. From this, the optimum transmit power vector $\mathbf{P}^* = (\mathbf{I} - \mathbf{F})^{-1}\mathbf{u}$, where $\mathbf{u}$ signifies the SNR of each user, can be calculated if the largest eigenvalue of **F** is shown to be less than 1. Effectively, this constraint establishes the stability (i.e., feasibility) of the POPC system of equations. The largest eigenvalue is found via an eigenvalue decomposition, which can be highly computationally complex even for small matrices. Thus, in order for a group of users in neighboring cells to be purposely scheduled onto the same RB, a simple expression for the allocatability of such a group must be developed.

In [13] the authors, through analytic derivation, have found that for three neighboring cells (for any number of cells greater the complexity increases exponentially) that if three MSs i, j, and k, one in each cell, satisfy the condition

$$F_{ij}F_{ji} + F_{ik}F_{ki} + F_{jk}F_{kj} + F_{ij}F_{jk}F_{ki} + F_{ik}F_{ji}F_{kj} < 1, \quad (1)$$

they may be allocated the same resources in each cell, and POPC can be applied. This is clearly dependent on the individual desired and interfering path gains, along with the SINR targets of the users. Therefore, a scheduler might make use of this condition to schedule users such that the number of feasible groups of MSs is maximized, hence also maximizing the spectral efficiency of these three cells; this is called POS. Of course, in a large network with hundreds of cells, this is not so useful. As indicated in Fig. 4, three neighboring cells may be grouped in the network. By grouping three cells with coinciding beam patterns (see shaded cells in Fig. 4), POS can be applied to these three cells, and the cluster will be relatively shielded from neighboring sectors interference due to the nature of the beam patterns. This clustering is then tessellated over the network, such that POS can be applied separately in each of these clusters without overly excessive CCI from the surrounding cells, allowing the MSs to achieve their transmission requirements. Thus, this form of POS combines the user-based cooperation methods of individual service requirements with the grouping of users that is so essential in CoMP.

In Fig. 5, a performance comparison to maximum power transmission (utilized in downlink transmission) and standard TPC (utilized in system uplink), in terms of spectral and energy efficiency is shown, displaying the potential of POS for future wireless communication networks. In general, POS has a performance advantage over power control over all SINR targets (except $\gamma_i^* =$ 20 dB), whereas also substantial gains over 13% are seen over max. power transmission in the mid-SINR (typically the operational) range. The POS performance begins to suffer for higher SINRs as not enough feasible MS groupings can be found. However, on average the POS spectral efficiency is equivalent to that of max. power transmission.

This becomes even more significant when considered together with the energy efficiency (calculated as data rate per unit of transmitted power [b/s/W ≡ (bits/J)]) results shown in Fig. 5. As expected, maximum power transmission is the least energy efficient of the three considered techniques. POS, on the other hand, provides massive energy efficiency benefits for the system, even when compared to power control, with gains of up to 5 dB for higher SINRs. Hence, it is quite clear that POS drastically reduces the transmit power consumption in a macrocellular network. Considering this together with the spectral efficiency results, it is evident that POS provides overall superior system performance over the standard resource and power allocation techniques.

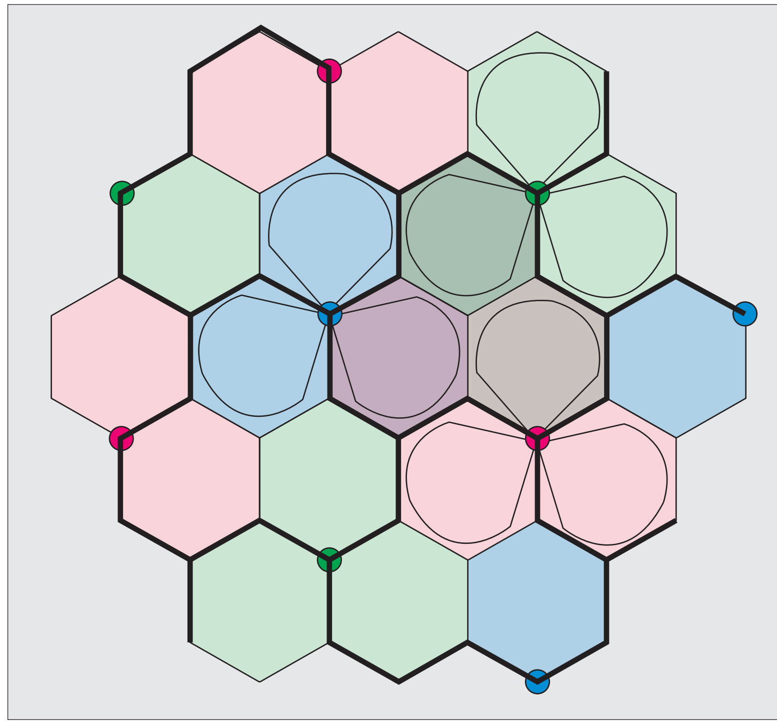


Figure 4. *Extension of three-cell POS to larger multicellular networks [13]. Universal frequency reuse is applied, and the differing coloration of the cells simply demarks which BS is serving them.*

Conclusions

Throughout this article, we have discussed the evolution of multicell cooperation from more simple user-requirements-driven cooperation techniques, over system-wide optimization-based methods, to coordination in LSMASs and cooperative transmission. Clearly, techniques optimizing the resource and power allocation for individual users do not always strike system optimum decisions, and hence from this research moved on to optimization methods. While, if solvable, these techniques can provide system-optimal resource and power allocations by coordinating over all cells in a network, it was explained that these methods are highly intractable, and in most situations must be broken down into multiple subproblems, yielding suboptimal solutions. Finally, the most recent research has considered the opportunity of LSMASs and cooperative transmission from multiple BSs in a network. CoMP provides large throughput gains for users in the network, but at the cost of additional complexity and rather unreliable scheduling. However, it is clear that for future networks to deliver the increasing service demands [1], such cooperation over multiple BSs is more than necessary.

We have concluded our article by introducing a recently proposed technique that combines some of the aspects of user-based cooperation and CoMP scheduling. In POS, users in neighboring cells are grouped based on their desired and interfering path gains, at which point POPC can be applied, and large energy savings can be achieved without sacrificing spectral efficiency. And although a large

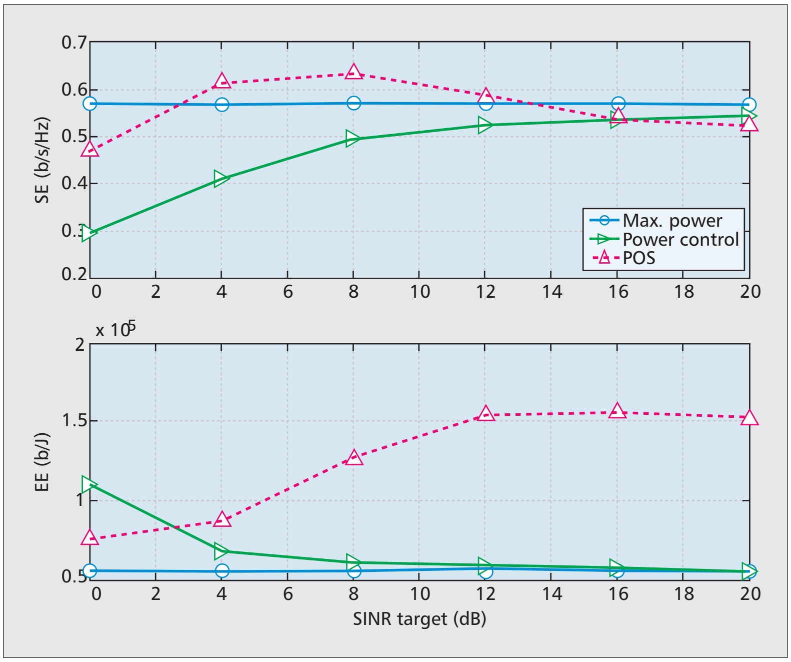


Figure 5. *System spectral efficiency (SE) and energy efficiency (EE) results for the various power control techniques in a multi-cellular network. In modern cellular systems, full frequency reuse is applied, utilizing maximum power transmission in the forward link, and power control in the reverse link.*

amount of information is needed to determine the feasibility of a group of MSs, since this procedure is performed on a three-cell basis, BSs need only the information from their immediate neighbors, limiting the additional complexity and signaling implied. The POS results, combined with the given comprehensive research review, thoroughly show the immense importance of multi-cell cooperation in today's, and more significantly for future, wireless networks.

References

[1] Cisco Visual Networking Index, "Global Mobile Data Traffic Forecast Update, 2011-2016," white paper, Feb. 2012.

[2] A. Molisch, *Multiple Access and the Cellular Principle*, Wiley-IEEE Press, 2011.

[3] S. Ali and V. Leung, "Dynamic Frequency Allocation in Fractional Frequency Reused OFDMA Networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, Aug. 2009, pp. 4286–95.

[4] M. Wang and T. Ji, "Dynamic Resource Allocation for Interference Management in Orthogonal Frequency Division Multiple Access Cellular Communications," *IET Commun.*, vol. 4, no. 6, Apr. 2010, pp. 675–82.

[5] J. G. Andrews, "Interference Cancellation for Cellular Systems: A Contemporary Overview," *IEEE Wireless Commun.*, vol. 12, no. 2, Apr. 2005, pp. 19–29.

[6] H. Burchardt *et al.*, "Uplink Interference Protection and Scheduling for Energy Efficient OFDMA Networks," *EURASIP J. Wireless Commun. and Net.*, May 2012.

[7] Y. Ma and D. Kim, "Rate-Maximization Scheduling Schemes for Uplink OFDMA," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, Jun. 2009, pp. 3193–205.

[8] D. Kivanc, G. Li, and H. Liu, "Computationally Efficient Bandwidth Allocation and Power Control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, Nov. 2003, pp. 1150–58.

[9] G. Wright, "GreenTouch Initiative: Large Scale Antenna Systems," Spring Meeting, Seoul, South Korea, 2011: http://www.greentouch.org/uploads/documents/pres_gregory wright.pdf.

[10] M. Sawahashi *et al.*, "Coordinated Multipoint Transmission/Reception Techniques for LTE-Advanced [Coordinated and Distributed MIMO]," *IEEE Wireless Commun.*, vol. 17, no. 3, June 2010, pp. 26–34.

[11] J.-S. Sheu and C.-H. Hsieh, "Joint Preprocessing Techniques for Downlink CoMP Transmission in Multipath Fading Channels," *IEEE 75th VTC-Spring*, May 2012, pp. 1–5.

[12] W. Mennerich and W. Zirwas, "User Centric Coordinated Multi Point Transmission," *IEEE 72th VTC-Fall*, Sept. 2010, pp. 1–5.

[13] H. Burchardt et al., "Pareto Optimal Power Control Scheduling for OFDMA Networks," *IEEE 76th VTC-Fall*, Sept. 2012.

[14] A. Goldsmith, *Wireless Communications*, Cambridge Univ. Press, 2005.

Biographies

HARALD BURCHARDT (h.burchardt@ed.ac.uk) received his B.Sc. in electrical engineering and computer science in 2007 from Jacobs University Bremen, and his M.Sc. in communications engineering in 2009 from the Technical University of Munich. He is currently working toward his Ph.D. degree at the Institute for Digital Communications at the University of Edinburgh.

HARALD HAAS holds the Chair of Mobile Communications at the University of Edinburgh. He is also part-time chief technical officer (CTO) of a university spin-out company, pureVLC Ltd. His main research interests are interference management in wireless networks, multiple antenna concepts, and optical wireless communication. He holds 23 patents, and has published more than 230 conference and journal papers. He was an invited speaker at TED Global 2011. He currently holds a prestigious Established Career Fellowship of the Engineering and Physical Sciences Research Council (EPSRC) in the United Kingdom.

MULTICELL COOPERATION FOR LTE-ADVANCED HETEROGENEOUS NETWORK SCENARIOS

BEATRIZ SORET AND HUA WANG, AALBORG UNIVERSITY
KLAUS I. PEDERSEN AND CLAUDIO ROSA, NOKIA SIEMENS NETWORKS

ABSTRACT

In this article we present two promising practical use cases for simple multicell cooperation for LTE-Advanced heterogeneous network scenarios with macro and small cells. For co-channel deployment cases, we recommend the use of enhanced interference coordination, or eICIC, to mitigate cross-tier interference and ensure sufficient offload of users from macro to small cells. It is shown how the eICIC benefit is maximized by using a distributed inter-base station control framework for dynamic adjustment of essential parameters. Second, for scenarios where macro and small cells are deployed at different carriers an efficient use of the fragmented spectrum can be achieved by using collaborative inter-site carrier aggregation. In addition to distributed coordination/collaboration between base station nodes, the importance of explicit terminal assistance is highlighted. Comprehensive system-level simulation results illustrate the performance benefits of the presented techniques.

INTRODUCTION

The potential performance benefits of applying multicell cooperation for mobile broadband cellular systems depends on numerous factors such as the network architecture, deployment topology, and the radio standard. As an example, the network architecture for wideband code-division multiple access (WCDMA) included a centralized radio network controller (RNC), hosting major radio resource management (RRM) algorithms such as admission control, mobility control, and packet scheduling for multiple cells. Nevertheless, it seems that most of today's WCDMA implementations only utilize multicell RRM in the form of soft handover. With the introduction of high-speed packet access (HSPA) overlay for WCDMA, the architecture moved toward a distributed functionality by moving RRM algorithms such as packet scheduling to the base station nodes [1]. Thus, faster access to radio channel measurements was achieved (e.g., for packet scheduling and link adaption), but at the cost of reduced possibilities for multicell cooperation. With the introduction of Long Term Evolution (LTE), the architecture evolved further toward a fully distributed paradigm by removing the RNC, but at the same time introducing a new inter-base-station interface (called X2) to allow some degree of multicell coordination and collaboration. Since its first introduction, the LTE standard has evolved significantly toward LTE-Advanced, where numerous spectral efficiency and peak data rate improvements have been introduced. Among those, it is worth mentioning improved multiple-input multiple-output (MIMO) techniques, carrier aggregation (CA), and enhanced intercell interference management [2].

Despite the many enhancements included in LTE-Advanced, it seems that macro-only networks may not be able to carry the predicted future broadband traffic. Therefore, the next big leap in cellular system performance improvement is estimated to come from the introduction of additional small cells to complement traditional macro site installations (i.e., migrating toward heterogeneous network [HetNet] topologies). This raises the questions on how small cells are most spectrally efficiently introduced, and how to integrate them with the macro layer to maximize the performance benefits while keeping capital expenditures (CAPEX) and operational expenditure (OPEX) at tolerable levels. In this study we further address such aspects for cases where small cells are in the form of either standalone picos or remote radio heads (RRHs). As it will be demonstrated, benefits of LTE-Advanced small cells can be improved via tight integration and cooperation between the macro layer and the small cells in a multicell RRM basis. For this purpose, two state-of-the-art LTE-Advanced HetNet integration schemes are examined: interference management between macro and small cells, and cooperative inter-site CA between layers. In both use cases we highlight the benefit of multicell cooperation as well as the importance of explicit terminal involvement to push the system performance to its maximum. As compared to published studies (e.g., [3–5]), we present performance results for highly irregular HetNet scenarios, where clusters of small cells are only present in some macrocells, calling for additional

1536-1284/13/$25.00 © 2013 IEEE

The next big leap in cellular system performance improvement is estimated to come from the introduction of additional small cells to complement traditional macro site installations (i.e., migrating toward heterogeneous network topologies).

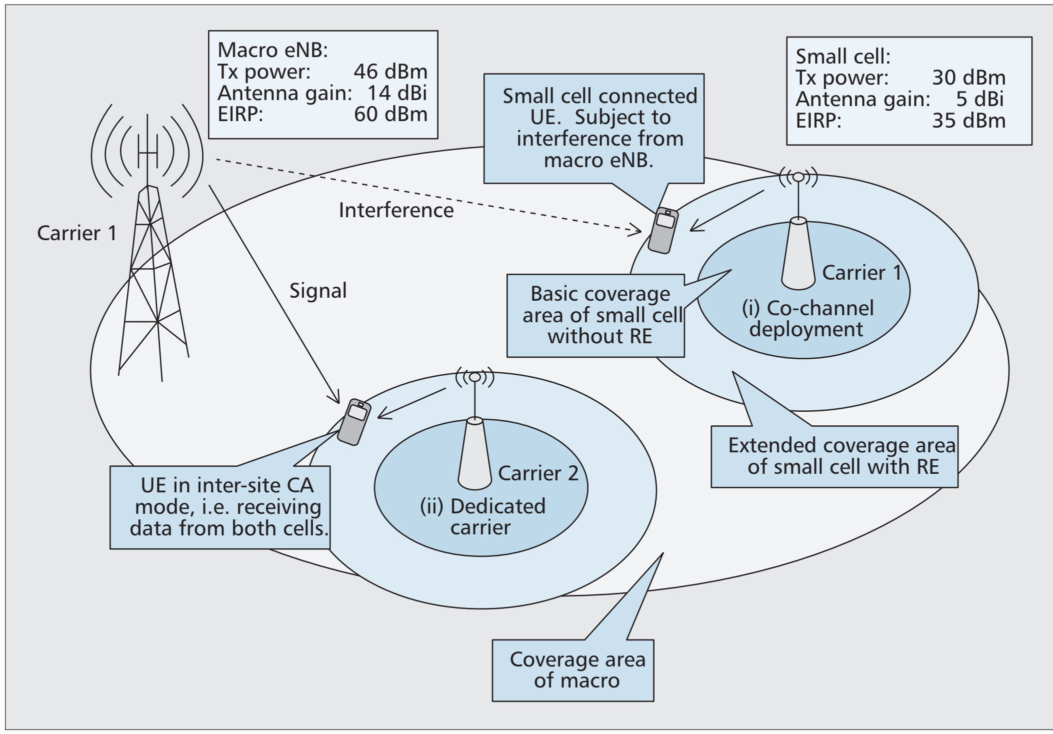


Figure 1. *Sketch of the considered LTE-Advanced HetNet scenario.*

macro to macro coordination. A recent survey on Third Generation Partnership Project (3GPP) LTE-Advanced HetNet features can be found in [3], while more information on the role of terminals in HetNet scenarios is available in [6]. In addition, there is ongoing research toward advanced coordinated multipoint (CoMP) transmission and reception techniques within 3GPP, as reported in [2]. However, advanced CoMP schemes are outside the scope of this article, as we focus here on simpler downlink LTE-Advanced multicell cooperation schemes for HetNet scenarios.

The rest of the article is organized as follows: We first outline the system model and addressed problem. Second, we analyze enhanced intercell interference coordination (eICIC) for co-channel cases, emphasizing the latest research ideas. Afterward, cases with dedicated carriers for macro and small cells are studied, where cooperative inter-site CA aims at exploiting the fragmented bandwidth. In both cases, system-level performance results are presented to demonstrate the benefits of multicell cooperation in terms of end-user-experienced throughput. The article is closed with some concluding remarks.

System Model and Problem Statement

Figure 1 illustrates the considered LTE-Advanced HetNet scenario with macro base stations (called eNBs) and small cells. The macro eNBs are assumed to transmit with 46 dBm power per sector and have 14 dBi antenna gain (including feeder loss), which results in an equivalent isotropic radiated power (EIRP) of 60 dBm. Small cells only have an EIRP of 35 dBm in the example in Fig. 1. In traditional homogeneous networks, the eNB offering the highest signal strength indicator is selected as the serving eNB for the user equipment (UE). In LTE, either the reference signal received power (RSRP) or the reference signal received quality (RSRQ) is used as the comparative metric in the cell-selection procedure. While RSRP only reflects the received power from each cell, RSRQ is defined as RSRP divided by the total received power, capturing the channel quality of the respective resources. Applying the RSRP or RSRQ cell selection criterion in HetNets leads to a downlink imbalance problem: The coverage of the macrocell is much larger than that of the small cell, resulting in a small number of UE devices being served by the small eNB. The coverage area of the small cell is not only limited by its transmition power, but to a large extent by the potential interference experienced from the macro eNB. This imbalance can be corrected by expanding the range of the small cell. Thus, a positive bias denoted as range extension (RE) offset is added to the RSRP or RSRQ measured from small eNBs, pushing more UE devices to this layer. The latter is illustrated in Fig. 1, where the basic coverage area of small cells is plotted with darker blue, while the range extended small cell coverage is depicted with lighter blue.

The performance and the need for multicell cooperation techniques for such a HetNet scenario depend heavily on the carrier deployment strategy of macro eNBs and small cells. Here we consider two distinct cases:
- Co-channel deployment, where macro and small cells occupy the same bandwidth
- Dedicated carrier deployment where macro and small cells use different carriers

The co-channel case has the advantage of being able to exploit the full system bandwidth at both macro and small layers, but the disadvantage of

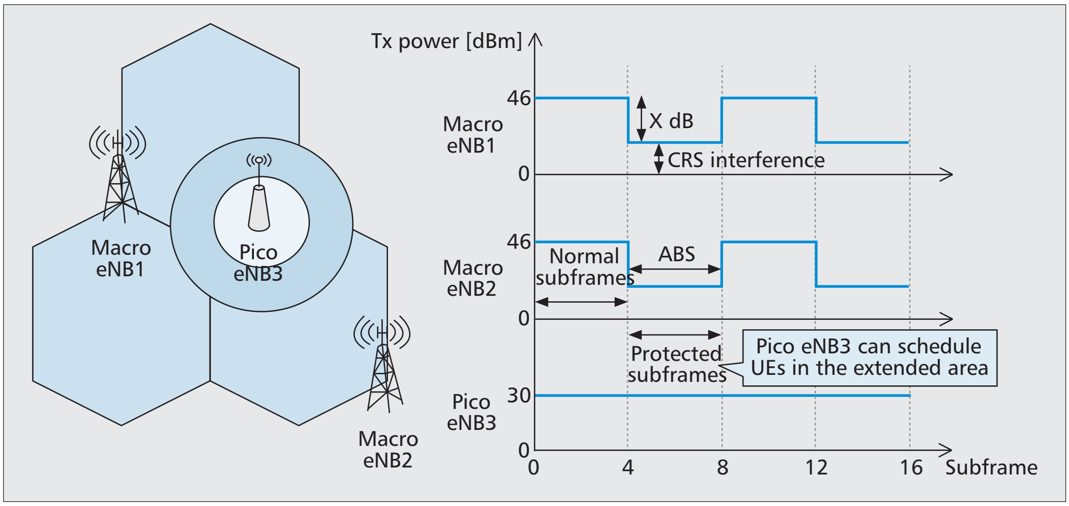


Figure 2. *Basic principle of eICIC.*

eNBs can exchange and request information on ABS patterns, and dynamically change the configuration, taking into consideration not only the overall system performance, but also the individual QoS requirements of individual users.

having interference between the two base station layers. The dedicated carrier case has the benefit of no inter-layer interference, but the disadvantage of not using the full system bandwidth at both macro and small cells, resulting in fragmented spectrum between the two layers. In general, co-channel deployment is preferred for operators with moderate LTE spectrum (e.g., 10 MHz), whereas dedicated carrier deployments are attractive in situations with larger availability of bandwidth (20 MHz or more) and high penetration of small cells. As the characteristics for these two deployment cases are clearly different, they also call for different multicell cooperation techniques.

For the co-channel case we investigate multicell cooperation in the form of inter-layer interference management. More specifically, we study the applicability of time-domain eICIC between macro and small cells [3]. As compared to previous eICIC studies we focus on the inter-site coordination signaling required to have such solutions working in practice, as well as recently considered eICIC enhancements on top of the basic functionality defined for LTE-Advanced Release 10 (Rel-10). New state-of-the-art eICIC performance results are presented for realistic scenarios. For the dedicated carrier case there is no inter-layer interference, and therefore no need to apply eICIC. However, the performance can be further improved by using collaborative inter-site CA between macro and small cells, such that the UE devices will essentially be served on a larger bandwidth simultaneously from multiple cells. For the inter-site CA case we assume that the small cells are RRHs, being connected to macro eNBs via high-bandwidth low-latency fibers. Thus, all baseband processing for the small cells (RRHs) could be placed in the macro eNB, offering further opportunities for various optimizations such as joint multicell packet scheduling. Referring to the 3GPP terminology, the dedicated carrier scenario with macro and RRHs is denoted CA scenario number four [7]. As discussed in detail in the following sections, successful application of eICIC and inter-site CA collaboration techniques requires explicit UE assistance and involvement [6].

Co-Channel Interference Coordination

In the co-channel deployment example in Fig. 1 we assume small cells in the form of outdoor picos, but all the techniques presented here are also valid for other types of small cells. Here, only users in the very close vicinity of the pico-eNB will be served by it, offering limited offload gains. This is due to the high interference received from the macro layer, which essentially prevents picos from using larger RE to increase the offload of users. eICIC techniques (introduced in 3GPP Rel-10 specifications) alleviate this interference problem and enable application of moderate to high values of RE. The basic principle is to establish coordinated resource partitioning between the macro and pico layer with the aim of providing some resources free of macro interference to pico UE devices. This is enabled by preventing the macro eNB from transmitting on certain subframes, the so-called almost blank subframes (ABS), as illustrated in Fig. 2. In this way, victim pico UE devices in the extended area may be scheduled within the subframes overlapping the muted frames of the macrocell, mitigating cross-tier interference. The muting ratio determines the periodicity of the ABS on a time-domain basis (4 over 8 subframes in the example in Fig. 2). The concept relies on accurate time and phase synchronization between all base station nodes. The name "almost" blank subframes refers to the fact that the aggressor cell still transmits the most essential information in order to provide support to legacy UE devices. Thus, ABS does include transmission of common reference signals (CRS) and, depending on the subframe number, also channels such as the physical broadcast channel (PBCH), and the primary and secondary synchronization signals (PSS/SSS). The interference from PBCH and PSS/SSS can be handled by not configuring subframes where those channels appear as ABS, as well as by using time shifts between the macro and pico layers to avoid, or minimize, the probability of experiencing collisions of those channels between layers. The interference from CRS does, however, appear in every ABS. Ultimately, the macro transmission

To fully benefit from macro muting, pico UE may apply CRS IC during subframes where the macro uses ABS, by estimating the CRS interference from neighboring cells and subtracting it from the received signal.

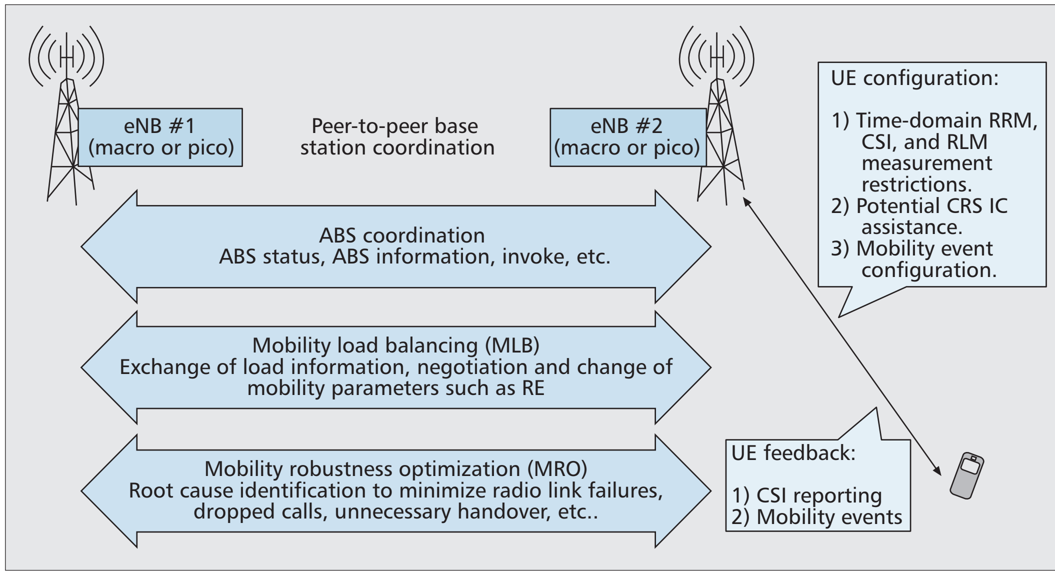


Figure 3. *Multicell and eNB-UE coordination for co-channel deployments.*

power in ABS is not zero, but on average reduced by X dB compared to normal transmission.

Numerous studies in the literature (e.g., [4]) have confirmed that eICIC performance depends heavily on numerous factors such as:
- ABS muting patterns at the different macrocells
- RE settings for the picocells
- UE traffic distributions

3GPP specifications support several mechanisms for distributed configuration of eICIC related parameters by means of coordination among eNBs, as summarized in Fig. 3. Thus, eNBs can exchange and request information on ABS patterns, and dynamically change the configuration, taking into consideration not only the overall system performance, but also the individual QoS requirements of individual users. Other self-organizing network (SON) features like mobility load balancing (MLB) and mobility robustness optimization (MRO) facilitate coordinated adjustment of mobility parameters such as RE [8]. Specifically, MLB allows tuning the handover and cell selection bias for cell pairs such as macro and picocells to balance the load between layers, and MRO monitors failed handovers to fine-tune mobility parameters and minimize the probability of radio link failures (RLFs). Thus, MRO essentially limits settings of RE that would otherwise result in risk of RLF. More details on mobility aspects in HetNet scenarios can be found in [9]. The X2 mechanisms for ABS configuration, and the MLB and MRO procedures form a simple but effective inter-eNB coordination protocol that facilitates autonomous network self-adjustment to maximize the benefits of eICIC. Notice that the inter-eNB coordination mechanisms summarized in Fig. 3 are applicable for macro to macro, macro to pico, and pico to pico. Besides the coordination among eNBs, the explicit involvement of the UE in the process is also important. When eICIC is enabled, highly time-variant interference fluctuations are likely to happen, which basically means that UE measurement reports should be time-restricted to have value for the network. For example, UE channel state information (CSI) measurements for pico UE devices are recommended to be time restricted so that the eNB receives CSI feedback corresponding to normal and protected subframes of the macro. Finally, measurement restrictions for radio link monitoring (RLM) and mobility measurements shall also be configured for terminals to maximize the gain from eICIC [3, 6].

The described coordination mechanisms are available in 3GPP Rel-10 specifications. Further Rel-11 eICIC (FeICIC) work item aims at strengthening the eICIC concept and boosting the overall performance. Particularly, two main novelties are envisioned in the new release: CRS interference cancellation (IC) and low-power ABS (LP-ABS).

UE CRS Interference Cancellation

One of the main UE receiver enhancements in 3GPP Rel-11 is the application of CRS IC to remedy the residual CRS interference during ABS. When the macro is muting a subframe, some resource elements have zero transmission power while resource elements with CRS have normal transmission. Assuming 2×2 MIMO, the CRS overhead during ABS is approximately 9 percent, corresponding to an average transmission power of ABS, which is roughly 10 dB below that of normal subframe transmissions. To fully benefit from macro muting, pico UE may apply CRS IC during subframes where the macro uses ABS, by estimating the CRS interference from neighboring cells and subtracting it from the received signal. In the ideal case, all pico UE devices would suppress the interference from all macro cells with no residual interference. However, a more realistic non-ideal CRS IC assumption entails terminals cancelling only the most relevant interferers with an interference cancellation error [5]. Deciding the suitable number of cancelled interferers is the key aspect here, and it depends on the UE capabilities but also on the user location. Thus, pico UE devices in the extended area are in the best condition for estimating the CRS from the dominant interfering macros and, at the same time, those that mainly require and benefit from CRS IC. The network can provide additional assistance for the UE to further ease the CRS IC, as also indicated in Fig. 3.

Low-Power ABS

During protected subframes the macro layer may, instead of stopping data transmission, reduce the transmission power, scheduling strictly users in the vicinity of the macro eNB. This option is referred to as low-power ABS, in contrast to (zero-) ABS. LP-ABS brings additional flexibility for optimizing the system under different conditions, enabling second order optimizations. Nonetheless, reducing the macro transmission power also has a cost in terms of additional complexity. As happens with ABS, essential information (CRS and other control channels) required by the system has to be transmitted unchanged during the reduced power subframes for backward compatibility. Thus, CRS IC is still relevant for LP-ABS operation. The introduction of LP-ABS mode can be fully exploited by means of extra coordination between layers. Therefore, additional information exchange could include the exchange of suggested (configured) LP-ABS power reduction between pico and macro layers (and vice versa). In order to respect the currently defined eNB transmitter radio frequency (RF) requirements, power reductions of 3 dB and 6 dB are typically considered for LP-ABS.

eICIC Performance

A network layout as defined in [10] is simulated. The network topology consists of hexagonal three-sector macrocells with an inter-site distance of 500 m and four randomly placed hotspots per macrocell area. According to 3GPP simulation guidelines, one picocell is deployed in each hotspot. The system-level simulator follows the LTE specifications, including detailed modeling of major RRM functionalities such as packet scheduling, hybrid automatic repeat request (HARQ), link adaptation, 2×2 closed loop MIMO with precoding and rank adaptation. The available bandwidth is 10 MHz, and Proportional Fair (PF) scheduling is applied. In Fig. 4 we show the performance gain of eICIC under different assumptions. The performance gain in 5th percentile and 50th percentile end-user throughput (macro and pico users) with full buffer traffic is shown, compared to the scenario with no eICIC and no RE. For the case without eICIC and no RE, only 40 percent of the users are offloaded to the picocells, meaning that the majority of users (60 percent) are served by the macro layer. However, by enabling eICIC with 4/8 of subframes as ABS and using 12 dB of RE, the offload increases to 80 percent. Thus, only 20 percent of the users are served by the macro layer, resulting in more resources per macro user even if it is muting 50 percent of the subframes. As a result, a significant eICIC gain for 12 dB is observed in Fig. 4, assuming RSRP-based cell selection. Comparing the three sets of performance results for 12 dB RE clearly shows the importance of having explicit UE support in the form of CRS IC. The cases with ideal and non-ideal CRS IC are fairly close, while there is noticeable performance degradation when CRS IC is not applied (non-ideal UE CRS IC referring to cases where only a few interferers are partially suppressed [5, 6]). Decreasing the RE offset from 12 dB to 9 dB means that a smaller number of users are offloaded to picocells. Therefore, a smaller ABS muting ratio should be applied in order to keep the trade-off between macro loss and pico gain. In this case, the maximum eICIC gain is achieved when reducing the muting ratio to 2/8. Applying LP-ABS with 6 dB power reduction in 4/8 subframes is another variant configuration of eICIC. Here we use ideal CRS IC and only 6 dB RE in coherence with the power reduction at the macro layer. Using 6 dB RE means, of course, that fewer users are offloaded, but this is partly compensated by the fact that for LP-ABS, the macro is still able to schedule users in every subframe. Furthermore, the possibility of scheduling macro users during LP-ABS leads to a higher optimal muting ratio (4/8) even for such small RE offset. The last set of results reported in Fig. 4 is for the case where RSRQ is used for cell selection (i.e., handover). For such cases, the UE devices are configured to only measure on picocells during subframes where the macros apply ABS. This essentially means that only low values of the RE are needed, as the RSRQ metric includes implicit knowledge of the eICIC interference reduction. Comparing the cases with RSRP and RSRQ (both with ideal CRS IC) seems to indicate that the picocell footprint is slightly better controlled by using RSRQ.

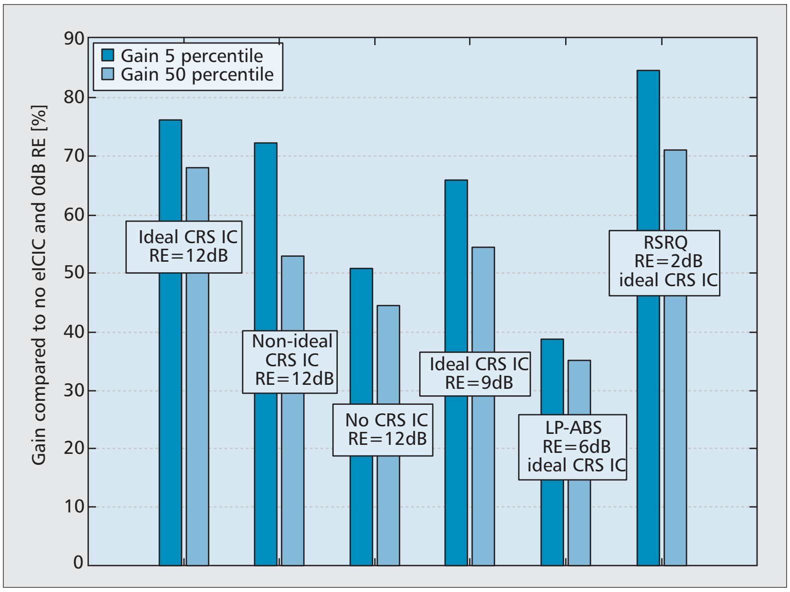


Figure 4. *FeICIC performance gain under different conditions.*

Uniform vs. Non-Uniform Topologies

All results in Fig. 4 assume uniform HetNet cases, that is, with the same number of picos in all macrocells. We next evaluate the effectiveness of eICIC techniques in more realistic non-

The inter-macro coordination described here can be dynamically configured through the X2 interface. As mentioned before, the addition of the power reduction to the existing exchanged messages between eNBs could facilitate the LP-ABS operation.

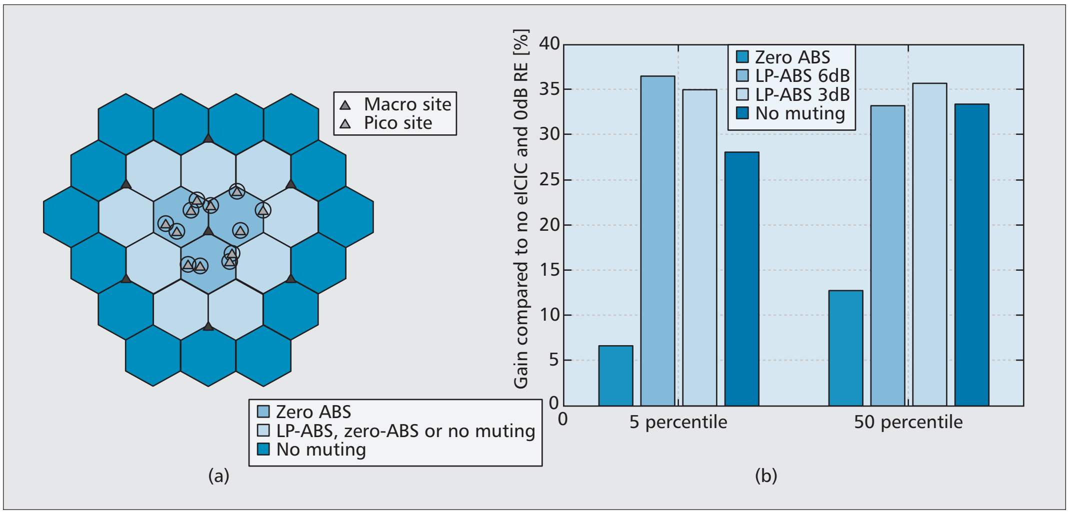


Figure 5. *Intercell coordination for non-uniform topologies: a) scenario; b) performance gain.*

uniform HetNet scenarios. In particular, we consider the scenario pictured in Fig. 5a, where picocells are only present in the three center macrocells, while small cells are not present in all other macrocells. Such deployment cases are envisaged to be typical for initial LTE-A HetNet rollouts. These topologies can also benefit from efficient cooperation among macro eNBs, taking into account that only some cells (center cells in Fig. 5) can gain from eICIC. Particularly, the joint use of LP-ABS and ABS can significantly improve the performance of pico UE in the extended area. Let us assume an RE of 12 dB for the picos and center macrocells using ABS in 4/8 subframes. Due to the deployment of picocells in the center macrocells, neighboring macrocells may be configured with ABS or LP-ABS in order to offer sufficient cumulative macrocell interference reduction to the picos. By applying LP-ABS in the first tier of interfering cells, the interference suffered by victim pico UE is mitigated, and at the same time, the performance of macro UE in the neighboring cells will not degrade much (the experienced signal-to-interference-plus-ratio [SINR] will not be significantly affected since the center macros are assumed to use ABS). Figure 5b shows the eICIC performance gain in this scenario. Notice that the performance gain is computed accounting for the users in the three center cells and those in the surrounding cells. Thus, this includes both effects of gains from center macrocells being able to offload more users to the picos and the potential degradation in surrounding macros from using LP-ABS. Two different values of the LP-ABS power reduction in the neighbor cells are simulated (3 dB and 6 dB), as well as the cases with no muting and zero-power ABS. Here the eICIC performance gain is on the order of approximately 30–40 percent, illustrating a promising use case for applying LP-ABS in macrocells without picos.

The inter-macro coordination described here can be dynamically configured through the X2 interface. As mentioned before, the addition of power reduction to the existing exchanged messages between eNBs could facilitate the LP-ABS operation. Thus, a macro with picos would get the ABS information from all surrounding macrocells, and on detection of one macro not applying ABS it would report its own ABS configuration and suggest reducing the power in order to minimize the interference caused to pico UE.

Collaborative Inter-Site CA

The second considered HetNet scenario is the case where macro and small cells are deployed at different carriers (Fig. 1, case (ii)). In this case there is obviously no need for eICIC as there is no interference between macro and small cells. Here, we therefore recommend exploiting collaborative inter-site CA in order to fully utilize the fragmented radio resources. Particularly, we propose applying CA between the macro and the most dominant small cell whenever a UE device is in a position where it can hear that cell. To be able to apply such inter-site CA techniques, we assume that the small cells are implemented as RRHs. Thus, the macro and RRHs are connected via a high-bandwidth low-latency fiber with the majority of baseband processing hosted in the macro sites. Referring to the established LTE-Advanced CA terminology [11, 12], we propose that UE devices always have the macrocell configured as their primary cell (PCell), whereas RRHs are configured as secondary cells (SCells) whenever possible. UE devices configured to operate with inter-site CA naturally have access to higher transmission bandwidth, and therefore the opportunity to be served at higher data rates. The fact that UE devices always have a macro as PCell results in robust mobility performance as RLF monitoring is only performed for PCell, which typically is found to be good for macro-only carriers. RSRQ measurement reports from terminals are used to decide if UE is configured on both macro and small cells, or whether it is only connected to one layer. Compared to RSRP-based measures, the RSRQ here has the advantage of implicit load information as the interference on the two considered carriers is likely to be significantly different. Aggressive settings for configuring a small cell as an SCell can be used, as the network can choose to deactivate the SCell if it is not being used for data transmission to reduce UE power consumption [11]. As the packet scheduler for the small cells (RRH) is physically located in the macro eNB, joint multicell packet scheduling for those UE devices configured with inter-site CA is feasible. In order to illustrate the application of such techniques, let us consider a simple extension of standard PF scheduling to such cases. Joint PF scheduling across multiple cells basically refers to how the PF metric is computed per cell. In a joint PF scheduler, the denominator of the PF metric is updated as the sum of the past average scheduled throughput over all cells where the UE has been scheduled in the past [13]. It simply requires information exchange on the average scheduled throughput between the scheduler for macro and small cells. In that way, the joint scheduler essentially offers fast and efficient inter-site load balancing between macro and small cells, thereby allowing for more equitable distribution of radio resources among UE.

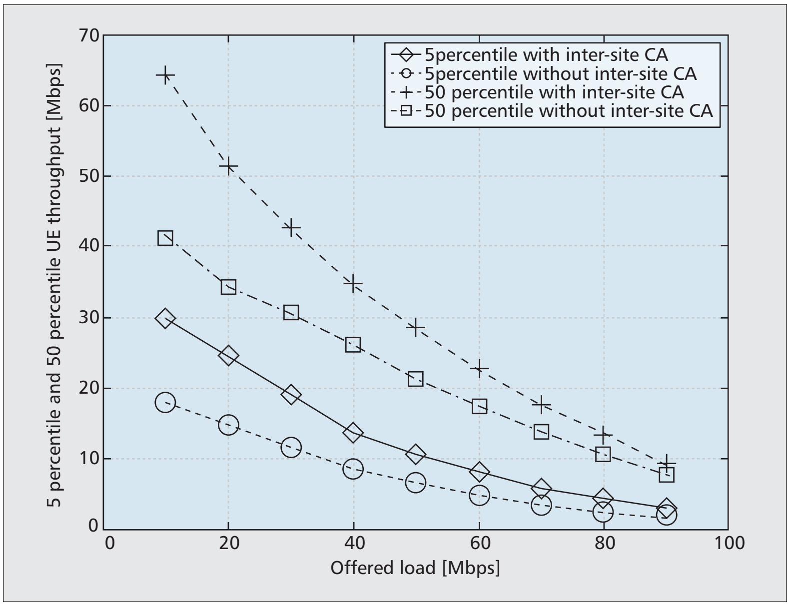


Figure 6. *5 percentile and 50 percentile UE throughput with/without inter-site CA, bursty traffic model.*

Performance of Collaborative Inter-Site CA

In order to further illustrate the benefits of collaborative inter-site CA for the downlink, we present performance results from extensive system-level simulations. These simulations are based on the same assumptions as in the previous section for eICIC evaluation (in line with 3GPP HetNet simulation assumptions in [10]). However, here we assume that macro and small cells are deployed at two different 10 MHz carriers. The intra-site CA case with the two carriers deployed at the macrocell can be found in [11]. In the simulations, we assume that all UE devices are configured with inter-site CA between the macro and the strongest RRH, while load balancing is managed by the joint PF scheduler. In order to illustrate how the inter-site CA performance varies vs. the offered system load, we apply a simple birth-death traffic model with fixed payload size per call. The call arrival follows a homogeneous Poisson process, while the payload size per call is fixed at 10 Mb. The average offered load per macrocell area is calculated as the product of the equivalent user arrival rate (per macrocell area) and the payload size. Figure 6 shows the 5th percentile and 50th percentile end-user throughput vs. the offered load. These results are for the case with four RRHs per macrocell area and with/without inter-site CA. For the cases without inter-site CA, the UE devices are only served by one cell at a time, determined by the RSRQ measurement report from the terminal. It is observed that both the 5th and 50th percentile user throughput with inter-site CA are significantly higher than without inter-site CA. The end-user-experienced gain of inter-site CA is up to 60 percent under low load conditions. However, the gain brought by inter-site CA decreases as the load increases. At high load, the performance with inter-site CA is almost the same as without inter-site CA. This behavior can be explained as follows: At low offered load (small number of UE devices), UE using inter-site CA benefits from larger transmission bandwidth and increased multi-user diversity. When the offered load is high (large number of UE devices in each cell), it does not really matter whether a UE device is assigned on one cell only or assigned on both cells using inter-site CA since the system is saturated, and the scheduler tries to allocate the available resources among all UE in a fair manner. However, at high load, the use of inter-site CA still offers advantages in terms of enhanced HetNet mobility robustness and faster inter-layer load balancing via the use of joint packet scheduling for both macro and RRHs. It is worth mentioning that the highest gain with inter-site CA (at low load) is lower than the theoretical 100 percent gain achievable by doubling the transmission bandwidth since the SINR experienced at the macro and small cell layers is typically quite different.

Conclusions

In this article we have demonstrated the benefits of using tight collaboration between LTE-Advanced macro and small cells, including the importance of having explicit terminal involvement and assistance. Two practical feasible schemes have been studied. For co-channel deployment cases, the advantages of using eICIC were demonstrated. eICIC allows higher off-load of users to the small cells, resulting in overall improvements in end-user-experienced throughput. One of the enablers for successful application of eICIC is the use of distributed inter-site muting pattern adjustment, mobility load balancing, and robustness optimization. For dedicated carrier deployment cases, the benefit of inter-site CA between macro and small cells has been shown. Using such techniques obviously offers higher peak data rate due to larger bandwidth accessibility for the terminals, but also related benefits in terms of faster load balancing and easier mobility man-

At high load, the use of inter-site CA still offers advantages in terms of enhanced HetNet mobility robustness and faster inter-layer load balancing via use of joint packet scheduling for both macro and RRHs.

agement as there is no interference between the network layers. For both of the considered multicell RRM collaboration techniques, explicit terminal support is needed to harvest the full benefit. UE CRS IC and mechanisms for configuring measurement restrictions are required for eICIC cases, as well as CA capability for the multicarrier scenario. In conclusion, we have essentially shown that the introduction of small cells benefits from being tightly integrated with the macro layer in order to unleash the full performance potential.

References

[1] H. Holma and A. Toskala, Eds., *LTE for UMTS, OFDMA and SC-FDMA Based Radio Access*, 2nd ed., Wiley, 2011.

[2] P. Bhat *et al.*, "LTE-Advanced: An Operator Perspective," *IEEE Commun. Mag.*, vol. 50, no. 2, Feb. 2012, pp. 104–14.

[3] A. Damnjanonic *et al.*, "A Survey on 3GPP Heterogeneous Networks," *IEEE Wireless Commun.*, vol. 18, no. 3, June 2011, pp. 10–21.

[4] Y. Wang and K. I. Pedersen, "Performance Analysis of Enhanced Inter-cell Interference Coordination in LTE-Advanced Heterogeneous Networks," *Proc. IEEE VTC-Spring 2012*, May 2012.

[5] B. Soret, Y. Wang, and K. I. Pedersen, "CRS Interference Cancellation in Heterogeneous Networks for LTE-Advanced Downlink," *Proc. IEEE ICC 2012, Int'l. Wksp. Small Cell Wireless Networks*, June 2012.

[6] A. Damnjanovic *et al.*, "UE's Role in LTE Advanced Heterogeneous Networks," *IEEE Commun. Mag.*, vol. 50, no. 2, Feb. 2012, pp. 164–76.

[7] M. Iwamura *et al.*, "Carrier Aggregation Framework in 3GPP LTE-Advanced," *IEEE Commun. Mag.*, vol. 48, no. 8, Aug. 2010, pp. 60–67.

[8] S. Hämäläinen, H. Sanneck, and C. Sartori, Eds., *LTE Self-Organising Networks (SON)*, 1st ed., Wiley, 2012.

[9] S. Barbera *et al.*, "Mobility Performance of LTE Co-Channel Deployment of Macro and Pico Cells," *Proc. IEEE WCNC 2012*, Apr. 2012, , pp. 2863–68

[10] 3GPP Technical Report 36.814, "Further Advancements for E-UTRA Physical Layer Aspects," v. 9.0.0, Mar. 2010, www.3gpp.org.

[11] K. I. Pedersen *et al.*, "Carrier Aggregation for LTE-Advanced: Functionality and Performance Aspects," *IEEE Commun. Mag.*, vol. 49, no. 6, June 2011, pp. 89–95.

[12] S. Zukang *et al.*, "Overview of 3GPP LTE-Advanced Carrier Aggregation for 4G Wireless Communications," *IEEE Commun. Mag.*, vol. 50, no. 2, Feb. 2012, pp. 122–30.

[13] Y. Wang *et al.*, "Carrier Load Balancing and Packet Scheduling for Multicarrier Systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, May 2010, pp. 1780–89.

Biographies

Beatriz Soret received her M.Sc. and Ph.D. degrees in telecommunication engineering from the University of Málaga, Spain, in 2002 and 2010. From 2001 to 2003 she was a research assistant in the same university. From 2003 to 2005 she worked in Cetecom Spain developing protocol test systems for Bluetooth and GSM/UMTS. From 2005 to 2011 she was an assistant professor at the University of Málaga. She is currently a postdoctoral researcher at Aalborg University in collaboration with Nokia Siemens Networks, Denmark. Her current research work is focused on radio resource management for heterogeneous networks. Other main research interests include quality of service and cross-layer design for wireless communications and queueing systems.

Hua Wang received his Bachelor's degree from Zhejiang University, China, in 2003, and his M.Sc. and Ph.D. degrees from Technical University of Denmark in 2005 and 2009, respectively. From 2005 to 2006 he worked as a radio engineer at Danske Telecom, Denmark. From 2008 to 2009 he was a visiting scholar at the University of Texas at Austin. He is currently a postdoctoral researcher at Aalborg University in collaboration with Nokia Siemens Networks, Denmark. His main research interests are in the area of radio resource management and performance analysis for small cell enhancements in heterogeneous networks.

Klaus Ingemann Pedersen received his M.Sc. in electrical enginerring and Ph.D. degrees in 1996 and 2000 from Aalborg University, Denmark. He is currently with Nokia Siemens Networks in Aalborg, Denmark, where he is working as a senior wireless network specialist. His current work is related to radio resource management and 3GPP standardization of LTE and LTE-Advanced. The latter includes work on carrier aggregation, heterogeneous networks, mobility, interference management, and system performance assessment. He is the author/co-author of approximately 100 technical publications on a wide range of topics, as well as an inventor on several patents. He has also contributed to books covering topics such as WCDMA, HSPA, LTE, and LTE-Advanced.

Claudio Rosa received his M.Sc. and Ph.D. degrees from Aalborg University in 2000 and 2005, respectively. He also received an M.Sc. degree in telecommunication engineering from Politecnico di Milano, Italy, in 2003. Curently he works as a wireless network specialist at Nokia Siemens Networks in Aalborg. His work is principally related to the standardization of LTE-Advanced, with particular focus on carrier aggregation evolution and small cell enhancements in future releases of the 3GPP specifications. Since 2003 he has been authoring/co-authoring several publications primarily in the field of uplink radio resource management, and has edited books on HSPA, LTE, and LTE-advanced.

THE PRACTICAL CHALLENGES OF INTERFERENCE ALIGNMENT

OMAR EL AYACH, THE UNIVERSITY OF TEXAS AT AUSTIN
STEVEN W. PETERS, KUMA SIGNALS, LLC
ROBERT W. HEATH JR., UNIVERSITY OF TEXAS AT AUSTIN

The authors at The University of Texas at Austin acknowledge the support of the Office of Naval Research (ONR) (Grant N000141010337), the DARPA IT-MANET program (Grant W911NF-07-1-0028), and the Army Research Labs (Grant W911NF-10-1-0420).

ABSTRACT

Interference alignment is a revolutionary wireless transmission strategy that reduces the impact of interference. The idea of interference alignment is to coordinate multiple transmitters so that their mutual interference aligns at the receivers, facilitating simple interference cancellation techniques. Since interference alignment's inception, researchers have investigated its performance and proposed several improvements. Research efforts have been primarily focused on verifying interference alignment's ability to achieve the maximum degrees of freedom (an approximation of sum capacity), developing algorithms for determining alignment solutions, and designing transmission strategies that relax the need for perfect alignment but yield better performance. This article provides an overview of the concept of interference alignment as well as an assessment of practical issues including performance in realistic propagation environments, the role of channel state information at the transmitter, and the practicality of interference alignment in large networks.

INTRODUCTION

Interference is a major impairment to successful communication in commercial and military wireless systems. In cellular systems, interference is created when different base stations share the same carrier frequency due to frequency reuse. Intercell interference reduces data rates throughout the cells and causes outages at the cell edges. In local area networks, interference is created when different access points share the same channel. The medium access control protocol attempts to deal with interference by avoiding packet collisions (overlapping transmissions). This conservative approach leads to underutilization of system bandwidth. Similarly, neighboring nodes in a dense mobile ad hoc network interfere if they share the same time and frequency resources. The medium access control protocol again limits the number of simultaneous conversations and consequently the system performance. Interference is thus a critical impairment in most wireless systems.

Communication in the presence of interference is often analyzed using an abstraction known as the interference channel. In the example interference channel of Fig. 1, three different transmitters wish to communicate with three receivers. Each transmitter has a message only for its paired receiver. Assuming the transmitters share the same time and frequency resources, each transmission creates interference at the unintended receivers. There may be other sources of interference not illustrated, such as jamming in military networks or self-interference created from nonlinearities in the radio frequency components; those sources are not captured in the basic interference channel.

Traditional methods for dealing with interference often revolve around giving each user exclusive access to a fraction of the communication resources. In frequency-division multiple access (FDMA), the system bandwidth is divided among the transmitters; for example, in Fig. 1 each transmitter would be given a third of the total bandwidth. In time-division multiple access (TDMA), users take turns transmitting on a periodic set of transmission intervals. In random access systems (e.g., carrier sense multiple access), transmitters listen to see if the channel is available and then transmit if it is. Random access protocols are typically much less efficient than preassigned orthogonal access like FDMA or TDMA since the spectrum may not be fully utilized and collisions may occur. Regardless of the access protocol, the unifying concept remains avoiding interference by limiting the number of overlapping transmissions. If simultaneous transmissions are allowed, the resulting interference is often treated as noise, and a loss in data rate ensues.

Recently, a new concept for communication in interference channels, called interference alignment (IA), was proposed in [1, 2]. IA is a cooperative interference management strategy

1536-1284/13/$25.00

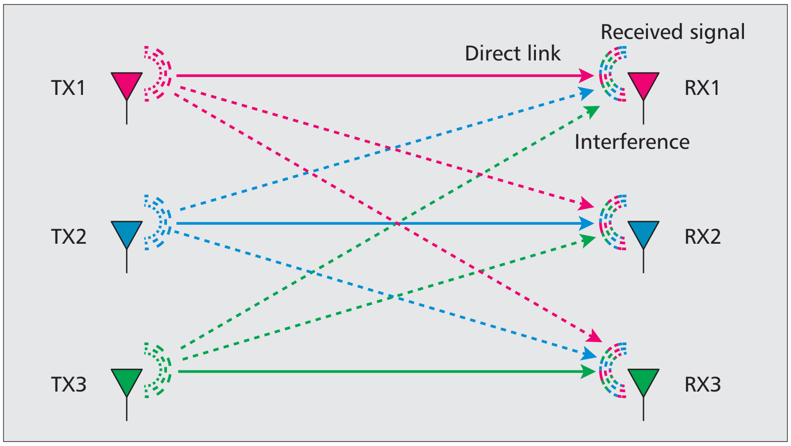


Figure 1. *Illustration of a three transmit/receive pair interference channel. Each transmitter creates a signal that is interpreted as interference by its unintended receiver.*

that exploits the availability of multiple signaling dimensions provided by multiple time slots, frequency blocks, or antennas. The transmitters jointly design their transmitted signals in the multidimensional space such that the interference observed at the receivers occupies only a portion of the full signaling space. An amazing result from [2] is that alignment may allow the network's sum data rate to grow linearly, and without bound, with the network's size. This is in sharp contrast with orthogonal access strategies like FDMA or TDMA, where sum rate is more or less constant since, regardless of network size, only one pair of users can communicate in a given time/frequency block. Since the development of IA, there has been work on a variety of topics to understand its theory and potential applications. A good summary of key results is given in [3].

The objective of this article is to review IA with a focus on making it practical. First, we describe IA in more detail, summarizing relevant practical issues that we believe are critical for its successful deployment. Then we review different techniques for computing alignment solutions and more general interference management solutions. Because computing these solutions relies heavily on channel state information (CSI), we describe the two competing CSI acquisition techniques, reciprocity and feedback, and focus on their respective benefits and limitations. We highlight the fact that the dimensions needed for alignment and the overhead of CSI acquisition both rapidly increase with network size. This places a limit on the gains achievable via IA in large networks. For this reason we summarize work on partial connectivity and user clustering, which leverages network topology information to reduce the requirements of alignment. We conclude with a discussion of the research challenges that remain in realizing IA in practice.

We note that the objective of this article is to provide a high-level introduction to the linear precoding type of IA, which exploits CSI for all users. Deeper discussions of other forms of IA, including blind alignment, are found in [3].

Linear Interference Alignment: Concept

Interference alignment in its simplest form is a precoding technique for the interference channel. It is a transmission strategy that linearly encodes signals over multiple dimensions such as time slots, frequency blocks, and antennas. By coding over multiple dimensions, transmissions are designed to consolidate (i.e., align) the interfering signals observed at each receiver into a low-dimensional subspace. By doing so, IA maximizes the number of non-interfering symbols that can be simultaneously communicated over the interference channel, otherwise known as the multiplexing gain. Interestingly, achieving the channel's maximum multiplexing gain, also known as degrees of freedom, implies that the sum rates provided by IA can approach sum capacity at high signal-to-noise ratio (SNR).

To illustrate the IA concept, consider the four-user system of Fig. 2, where real valued signals are coded over three dimensions and communicated over real valued channels. In this four-user system, a receiver observes a total of three interference signals, each of which is represented as a vector in the real 3D space. Without careful structure, the three interference signals will occupy all three signal dimensions at the receiver; the signals do not fortuitously align. IA allows users to cooperatively precode their transmissions such that each *three* interference signals are fully contained in a *two*-dimensional space. Alignment thus leaves one dimension in which users can decode their messages free from interference by projecting the received signal onto the subspace orthogonal to the interference.

The IA visualization in Fig. 2 translates into an equally intuitive mathematical representation. Consider a K-user interference channel in which each user i transmits a set of S_i information symbols encoded in the $S_i \times 1$ vector $\mathbf{s}_i$. Symbols $\mathbf{s}_i$ are precoded using a matrix $\mathbf{F}_i$ and observed at each receiver ℓ after propagating over the matrix channel $\mathbf{H}_{\ell,i}$; the dimensions of $\mathbf{F}_i$ and $\mathbf{H}_{\ell,i}$ will be made clear shortly. For such a system, the received signal at user i is

$$\mathbf{y}_i = \mathbf{H}_{i,i}\mathbf{F}_i\mathbf{s}_i + \sum_{\ell \neq i} \mathbf{H}_{i,\ell}\mathbf{F}_\ell\mathbf{s}_\ell + \mathbf{v}_i,$$

where the vector $\mathbf{v}_i$ is the noise observed at receiver i. When only the time or frequency dimension is used for precoding (e.g., when coding over a bandwidth of B subcarriers), $\mathbf{H}_{\ell,i}$ is a $B \times B$ diagonal matrix; diagonal since subcarriers are orthogonal. In a multiple-input multiple-output (MIMO) system with N_T transmit antennas and N_R receive antennas, $\mathbf{H}_{\ell,i}$ is $N_\mathrm{R} \times N_\mathrm{T}$ with no special structure.

Regardless of the dimension used to align interference, IA can be summarized as calculating a set of precoders $\mathbf{F}_\ell$ such that any given user, even using a simple linear receiver $\mathbf{W}_i$, can cancel the interference it observes from all other users (i.e., $\mathbf{W}_i^*\mathbf{H}_{i,\ell}\mathbf{F}_\ell = 0\ \forall \ell \uparrow i$) without nulling or destroying its desired signal $\mathbf{W}_i^*\mathbf{H}_{i,i}\mathbf{F}_i$. Early work on IA showed that a system's ability to find such IA precoders is directly related to the number of signal dimensions over which it can code. Intuitively, the more time slots, frequency blocks, or antennas available for precoding, the more flexibility a system has in aligning interference. The problem of characterizing the number of coding dimensions needed to ensure the feasibility of IA has been studied extensively [4].

Linear Interference Alignment: Challenges

Interference alignment relies on some assumptions that must be relaxed before it is adopted in practical wireless systems. In this section, we briefly discuss some of the most pressing challenges facing the transition of IA from theory to practice.

Dimensionality and Scattering

IA is achieved by coding interference over multiple dimensions. Intuitively, the more interfering signals that need to be aligned, the larger the number of dimensions needed to align them.

When using the frequency domain for alignment, prior work has shown that the number of signaling dimensions needed to achieve good IA performance grows *faster than exponentially* with the number of IA users. Properly aligning even as few as four users requires a potentially unreasonable number of subcarriers and a correspondingly large bandwidth [3]. This dimensionality requirement poses a major challenge for IA in practical systems. The dimensionality requirement is relatively milder in the case of MIMO IA where more users can be supported as long as the number of antennas grows linearly with network size [4]. For this reason, IA seems most likely to be implemented in MIMO systems. We thus place a special focus on MIMO IA.

Signal-to-Noise Ratio

IA is often degrees of freedom optimal, meaning that the sum rates it achieves approach the channel's sum capacity at very high SNR. At *moderate SNR levels*, however, the sum rates resulting from IA may fall short of the theoretical maximum. As a result, IA may be of limited use to systems with moderate SNR unless IA algorithms are further improved. Examples of such algorithms are discussed later.

Channel Estimation and Feedback

Channel state information, be it at the transmitter or receiver, is central to calculating IA precoders. As a result, sufficient resources must be allocated to pilot transmission, and in some cases to CSI feedback, to ensure the availability of accurate CSI. Since IA precoders must be recalculated when the channel changes appreciably, the overhead of CSI acquisition in high-mobility fast-fading systems can limit the gains of IA. For this reason, low-overhead signaling strategies must be devised to properly trade off CSI quality with CSI acquisition overhead.

Synchronization

IA via linear precoding is a transmission strategy for the *coherent interference channel*. Thus, IA

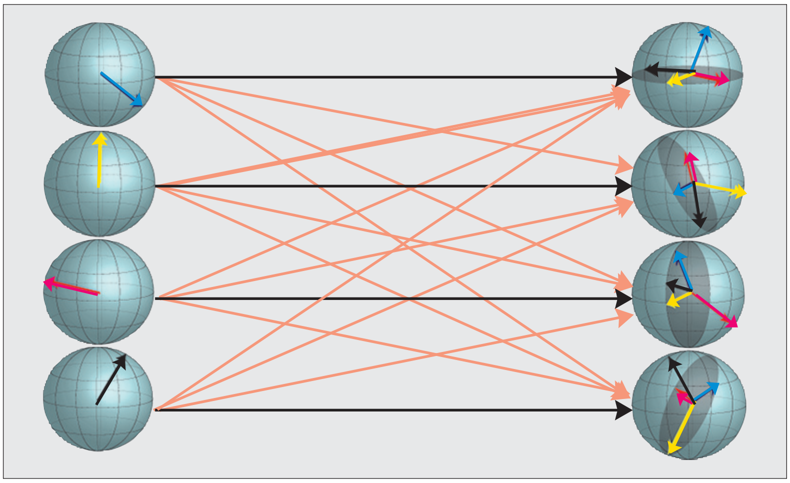


Figure 2. *A diagram illustrating interference alignment. At each receiver, three interferers collapse to appear as two. This enables interference-free decoding in a desired signal subspace.*

requires tight synchronization to remove any timing and carrier frequency offsets between cooperating nodes. In the absence of sufficient synchronization, additional interference terms are introduced to the signal model, rendering the IA solution ineffective. Synchronization strategies that leverage GPS satellite signals could help fulfill this requirement.

Network Organization

Nodes cooperating via IA must not only synchronize, but also negotiate physical layer parameters, share CSI, and potentially self-organize into small alignment clusters if full network alignment proves to be too costly. In the absence of centralized control, distributed network protocols must be redesigned around this more complex and cooperative physical layer.

Further Description of Challenges

The following three sections are devoted to surveying the research efforts invested in overcoming some of the challenges discussed in this section. We present algorithmic improvements that primarily target IA's low SNR performance, we explore methods of obtaining the CSI required by IA, and we address the challenge of applying IA to larger networks.

Computing Interference Alignment Solutions

There are some simple formulas for calculating IA precoders in some special cases [2]. To enable IA in general network settings, however, researchers have relied on developing iterative IA algorithms. Since then, the algorithmic focus has largely been on MIMO IA.

The earliest method for finding MIMO IA precoders was the distributed solution presented in [5]. The idea for the algorithm is that at each iteration, users try to minimize the extent to which their signal leaks into the other users' desired signal subspaces. After the algorithm converges, the IA condition $\mathbf{W}_i^* \mathbf{H}_{i,\ell} \mathbf{F}_\ell = 0$ should

Calculating IA precoders requires accurate knowledge of the interference generated by each transmitter. The premise is that if a transmitter knows the "geometry" of the interference it creates, it can conceivably shape it to mitigate its effect. Two methods of obtaining this knowledge are reciprocity and feedback.

ideally be satisfied and the desired signal spaces be free of interference. While the algorithm performs well at high SNR, it can be far from optimal in badly conditioned channels or at low SNR. The reason is that while this algorithm properly aligns interference, it is oblivious to what happens to the desired signal power during the process.

The first attempt to improve low SNR performance is the Max-SINR algorithm of [5]. Instead of minimizing interference power at each iteration, the improved algorithm accounts for the desired channel by iteratively maximizing the per-stream signal-to-interference-plus-noise ratio (SINR). By relaxing the need for perfect alignment, Max-SINR outperforms IA at low SNR and matches its performance at high SNR. Other algorithms have also relaxed the perfect alignment requirement and adopted more direct objectives such as maximizing network sum rate. For example, [6] highlights the equivalence between maximizing sum rate and minimizing the signal's sum mean square error and uses the properties of the equivalent minimum mean square error (MMSE) problem to give a simple iterative algorithm in which precoders can be updated in closed form or via traditional convex optimization.

Suboptimal low SNR performance is admittedly not the only limitation of IA, and many algorithms have been developed to address various shortcomings. Most early work on IA, for example, focused on the case where all interfering users are able to cooperate. The ability to cooperate, however, is fundamentally limited by constraints such as the number of antennas and cooperation overhead. As a result, large networks will inevitably have uncoordinated interference, or colored noise, which motivates the improved IA algorithms in [7]. Another limitation of precoding for the MIMO interference channel is that it often requires sharing entire channel matrices, potentially incurring a large overhead penalty, as we discuss later. The work in [8] proposes a distributed transmission strategy inspired by game theory in which both precoders and transmit powers are iteratively adjusted to maximize sum rate by sharing *scalar quantities* known as "interference prices." The strategy in [8] thus replaces channel *matrix* feedback with several iterations of *scalar* feedback. Such a strategy could potentially reduce IA's overhead.

While editorial constraints have limited the discussion to the solutions of [2, 5–8], a large number of noteworthy solutions have been developed to improve on IA performance. The interested reader is encouraged to see the extensive list of IA-inspired algorithms and results in http://www.profheath.org/research/interferencealignment/.

Obtaining CSI in the Interference Channel

Calculating IA precoders requires accurate knowledge of the interference generated by each transmitter. The premise is that if a transmitter knows the "geometry" of the interference it creates, it can conceivably shape it to mitigate its effect. Two methods of obtaining this knowledge are reciprocity and feedback.

Interference Alignment via Reciprocity

In time-division duplexed systems where transmissions on the forward and reverse links overlap in frequency and are minimally separated in time, propagation in both directions will be identical. In such systems, channels are said to be reciprocal. Reciprocity enables IA by allowing transmitters to infer the structure of the interference they *cause* by observing the interference they *receive*.

Consider, for example, the IA strategy proposed in [5] in which the precoders $\mathbf{F}_i$ and the combiners $\mathbf{W}_i$ are updated iteratively to reduce the power of uncanceled interference $\mathbf{W}_i^*\mathbf{H}_{i,\ell}\mathbf{F}_\ell$ to zero. To do so, transmitters begin by sending pilot data using an initial set of precoders. Receivers then estimate their interference covariance matrix and construct combiners that select the receive subspaces with least interference. When the roles of transmitter and receiver switch, thanks to reciprocity, that same *receive subspace* that carried the least interference becomes the *transmit direction* which causes the least interference on the reverse link. Iterating this subspace selection on the forward and reverse links results in a set of precoders satisfying the IA conditions.

While [5] at each iteration chooses the subspace with least interference, a variety of more sophisticated objectives may be considered. For example, [9] considers an update rule that minimizes the signal's mean square error, resulting in improved sum rate performance. Regardless of the subspace selection rule, the general framework for precoding with reciprocity is shown in Fig. 3a and proceeds as follows:

- *Forward link training*: Transmitters send precoded pilot symbols using a set of initial precoders. Receivers estimate forward channel parameters and compute combiners that optimize a predefined objective.
- *Reverse link training*: Receivers send precoded pilot symbols using the combiners from the above step as transmit precoders. Transmitters in turn optimize their combiners/precoders and initiate a second training phase with the updated precoders.
- *Iterations*: Communicating pairs iterate the previous steps until convergence.
- *Data transmission*: Payload data is then communicated.

Relying on reciprocity for precoding over the interference channel has a number of potential drawbacks. First, iterating over the air incurs a non-negligible overhead due to the recurring pilot transmissions. While the results in [9] consider pilot overhead, more work is needed to settle the viability of reciprocity. Second, reciprocity may not suffice for all IA-based algorithms.

For example, one of the algorithms in [7] attempts to improve IA by considering sources of uncoordinated interference. Since the uncoordinated interference observed by the transmitters and receivers is not "reciprocal," reciprocity cannot be used. Third, reciprocity does not hold

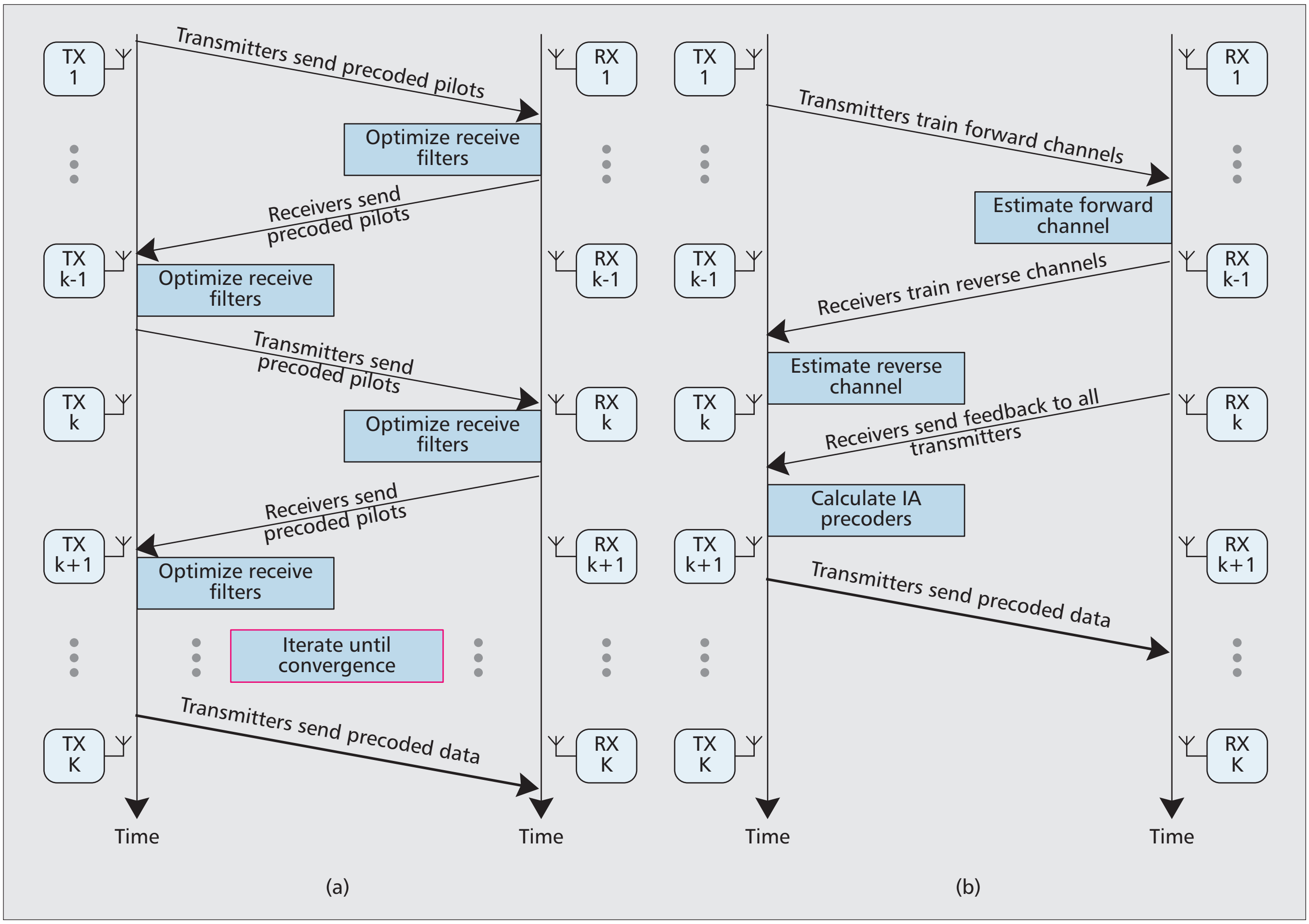


Figure 3. *The two CSI acquisition paradigms considered in IA literature: a) reciprocity-based strategies that infer the forward channel from reverse link pilots; b) feedback strategies that rely on explicitly communicating forward channel information to the transmitters.*

in frequency duplexed systems, and ensuring reciprocity with time duplexing requires tightly calibrated radio frequency (RF) devices.

Interference Alignment with Feedback

Several IA results have also considered systems with CSI feedback. A general feedback framework is shown in Fig. 3b. In such a system, the transmitters first send training sequences with which receivers estimate the forward channels. Receivers then feed back information about the estimated forward channels, potentially after training the reverse link. After feedback, the transmitters have the information needed to calculate IA precoders. Feedback, however, invariably introduces distortion to the CSI at the transmitters and incurs a non-negligible overhead penalty. Therefore, the difficulty lies in designing low-overhead low-distortion feedback mechanisms for IA.

An established method to provide high-quality feedback with low overhead is limited feedback (i.e., channel state quantization). Limited feedback was first considered in [10] for single-antenna systems where alignment is achieved by coding over orthogonal frequency-division multplexed (OFDM) subcarriers. The feedback strategy in [10] leverages the fact that IA solutions remain unchanged if channels are scaled or rotated, which allows efficient quantization via what is known as Grassmannian codebooks. Unfortunately, maintaining proper alignment requires that the accuracy of quantized CSI improve with SNR, which in turn implies that the quantization codebook size must scale exponentially with SNR. The complexity of quantized feedback, however, increases with codebook size and large Grassmannian codebooks are difficult to design and encode. Furthermore, this strategy relies on the CSI's Grassmannian structure for efficient quantization. As a result, it cannot be applied to systems where the CSI exhibits no special structure. This alienates a main case of interest, namely IA in multiple antenna systems where the CSI to be fed back is the set of channel matrices.

To support MIMO IA and overcome the problem of scaling complexity, IA with analog feedback was considered in [11]. Instead of quantizing the CSI, analog feedback directly transmits the channel coefficients as uncoded quadrature and amplitude modulated symbols. Since no quantization is performed, the only source of distortion is the thermal noise introduced during training and feedback. As a result, the CSI quality naturally increases with the SNR on the reverse link. This implies that IA's multiplexing gain is preserved as long as the forward and reverse link SNRs scale together.

Feedback, however, poses a number of chal-

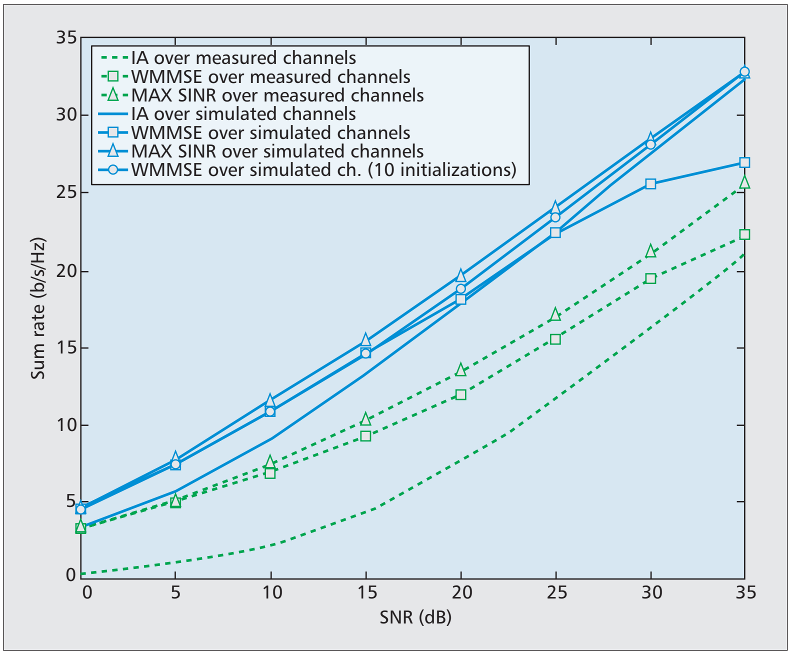


Figure 4. *A performance comparison for three precoding solutions for the interference channel over both measured and simulated channels.*

lenges. Not only does the required quality of CSI scale with SNR, but the required quantity also scales super-linearly with network size. If feedback is inefficient, IA's overhead could dwarf its promised theoretical gains. Second, IA relies on sharing CSI with interfering transmitters: multicell systems that share transmitter CSI over a backhaul may suffer from queuing delays that render CSI obsolete on arrival. While work on addressing these issues is underway, most research still considers systems with a limited number of users. In large-scale networks, however, practical challenges are amplified manifold.

Interference Alignment in Large Scale Networks

Consider applying IA, or any other interference channel precoding strategy, to a large-scale network, such as a regional cellular network. In this setting, the need for signaling dimensions grows, the overhead of CSI acquisition explodes, and the task of synchronization becomes daunting. This grim outlook, however, is a byproduct of a naive overgeneralization of the basic interference channel. In large-scale networks not all interfering links are significant, and thus not all of them should be treated equally.

To examine the feasibility of IA in large networks, [12] considered a large network wherein each user receives interference from a finite subset of users, such as the first tier of interferers in a grid-like cellular system. Using this model, [12] showed that the number of antennas needed for network-wide alignment is a linear function of the size of the interfering subset rather than the size of the entire network. This implies that perfect alignment in an *infinitely* large network is theoretically possible with a *finite* number of antennas. This partially connected model, however, is ultimately a simplification of network interference. While some interferers may indeed be too weak to merit alignment, that threshold remains unclear. Moreover, neglecting interferers that fall below that threshold, as is implicitly done in the partially connected model, may not be optimal. Exploring the optimal transmission strategy in this gray area between full connectivity and true partial connectivity is the focus of [13].

The work in [13] considers a wireless network and assigns a finite channel coherence time and different path loss constants to each link. This setup allows [13] to gauge the gain from aligning a set of users vs. the CSI acquistion overhead involved. On one hand, it is noticed that the large coherence time in static channels dilutes the cost of CSI and enables alignment over the whole network. On the other hand, fast fading channels do not allow enough time to acquire the CSI needed for IA. In this latter case, systems perform better with lower overhead strategies such as TDMA. In between the two extremes, it is noticed that systems benefit most from a hybrid IA/TDMA strategy wherein smaller subsets of users cooperate via alignment, and TDMA is applied across groups. The objective then becomes finding an optimal user grouping, or network partitioning, for which [13] provides various algorithms based on geographic partitioning, approximate sum rate maximization, and sum rate maximization with fairness constraints. In all grouping solutions only long-term path loss information is used, which further reduces the overhead of IA by avoiding frequent regrouping.

The IA research community has made successful initial attempts at demonstrating that the reach of IA can extend well beyond the information theoretic interference channel through work such as [7, 12] and many others. Perhaps the most crucial next step toward realizing IA gains is to rethink the upper layers above IA in the communication stack (e.g., the medium access layer). These protocols have traditionally been built around point-to-point physical layers and must be redesigned to enable more cooperative multi-user physical layers instead. Preliminary work on this front is already in progress through the prototyping work in [14], which extends the earlier prototyping efforts by the same authors.

A Practical Performance Evaluation

In this section we provide numerical results on the performance of IA with a special focus on the concepts discussed throughout the article. Figure 4 compares the performance of the algorithms presented earlier in a three-user system with two antennas per node. For this assessment both simulated channels and measured MIMO-OFDM channels from the testbed in [15] are considered. Simulated channels are assumed to be Rayleigh fading with equal average received power on all links. Simulated channels are further assumed independent across all users and antennas. This constitutes a baseline of highly scattered channels, considered ideal for IA. To demonstrate the effect of limited scattering, we report measurements from [15] corresponding to a scenario in which antennas are spaced half a wavelength apart.

The results in Fig. 4 verify that IA, in this case, achieves a multiplexing gain of three as predicted in theory. Multiplexing gain can be calculated by examining the slope of the sum rate curves. The same results, however, also demonstrate IA's suboptimality at low SNR, where precoding algorithms such as the WMMSE algorithm in [6] or the MAX-SINR algorithm in [7] consistently outperform. Finally, Fig. 4 shows that the correlation present in measured channels can have significant effects on the performance of IA. Interestingly, in such measured channels, the algorithms in [6, 7] provide large gains over IA even at a significant SNR of 35dB.

In Fig. 5 we relax the assumption of perfect channel knowledge and consider a five user system that acquires CSI through training and analog feedback, thus suffering from both CSI distortion and CSI acquisition overhead. The results indicate that, at pedestrian speeds, the slow channel variation enables acquiring accurate CSI at a low overhead cost, which translates into substantial gains over non-cooperative transmission strategies such as TDMA. The same results indicate that in very high mobility settings, where CSI acquisition is costly, one may elect to either coordinate a smaller user group or altogether adopt a lower overhead strategy such as TDMA. In a large wireless network, grouping users based on overhead constitutes selecting from a continuum of network structures. For example, Fig. 6 demonstrates how even a simple six-user system should be partitioned differently as the overhead cost of full IA varies. The throughput maximizing strategy transitions from full six-user IA in static channels to basic TDMA for fast fading channels, passing through a range of "hybrid" network structures in which users cluster into small cooperating groups.

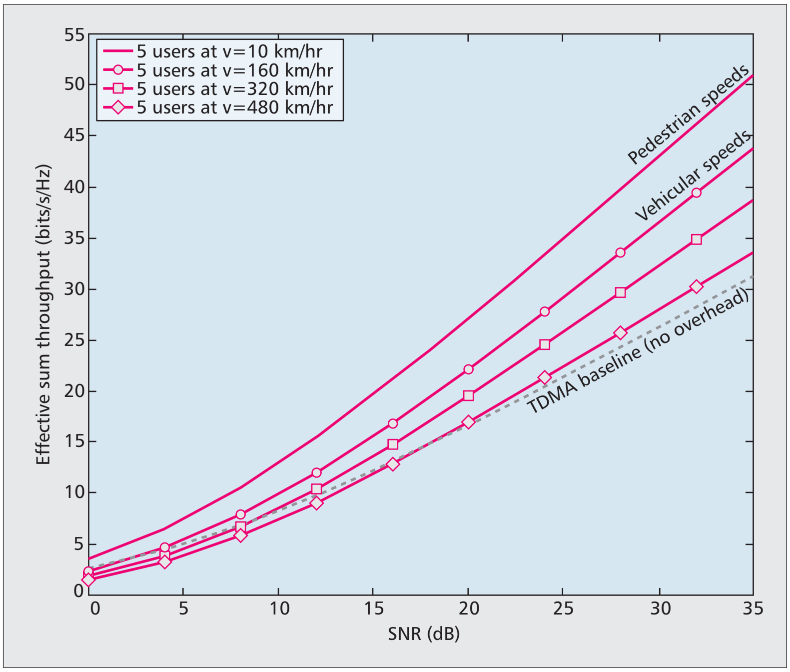


Figure 5. *Effective throughput vs. SNR for a five user system at varying normalized Doppler spreads. Each transmitter is equipped with 3 antennas, and every user communicates via one spatial stream. The figure demonstrates how as Doppler increases, so does overhead and CSI error, resulting in a drop in effective sum rate. Channels are assumed to be Rayleigh and user speeds are calculated for a system with a coherence bandwidth of 300 kHz at a carrier frequency of 2 GHz.*

Future Research Directions

In this section we present some active areas of interest and general topics for further research on interference alignment.

Algorithms: Algorithms remain a hot topic for IA research as a number of "extra features" can be incorporated or further improved such as complexity, low SNR performance, distributed computation, and robustness to CSI imperfections.

Feedback: Temporal correlation can be leveraged to both improve the accuracy of CSI as well as reduce the overhead of feedback. Limited feedback strategies such as [10] can be generalized to better support MIMO IA. The development of limited feedback strategies can benefit from the lessons learned in the MIMO single-user and broadcast channels where feedback can be significantly compressed by exploiting the mathematical properties of the quantized CSI. Finally, additional work is needed to better understand the effects of feedback overhead and feedback delay.

IA in multihop networks: Preliminary work on multihop IA suggests that relays can greatly reduce the coding dimensions needed to achieve a network's degrees of freedom, and otherwise simplify the optimal transmission strategies [3]. The importance of practical precoding algorithms for the relay-aided interference channel is further amplified by the standards community's growing interest in deploying relays in future wireless systems.

IA in modern cellular networks: To make IA a viable multicell cooperation technique, IA research must characterize the effect of cellular system complexities such as scheduling, resource allocation, and backhaul signaling delay. IA must also be re-evaluated using models that more accurately resemble modern cellular systems, which could include heterogeneous infrastructure such as macrocells, picocells, small cells, relays, and distributed antennas.

Conclusion

Interference alignment is a transmission strategy that promises to improve throughput in wireless networks. This article reviews the key concept of linear interference alignment, surveys recent results on the topic, and focuses on bringing the concept closer to implementation by discussing some of its main limitations. After the key hurdles in areas such as low SNR performance, CSI acquisition overhead, synchronization, and distributed network organization can be overcome, interference alignment will be ready for practical implementation.

References

[1] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X Channels: Interference Alignment, Decomposition, and Performance Analysis," *IEEE*

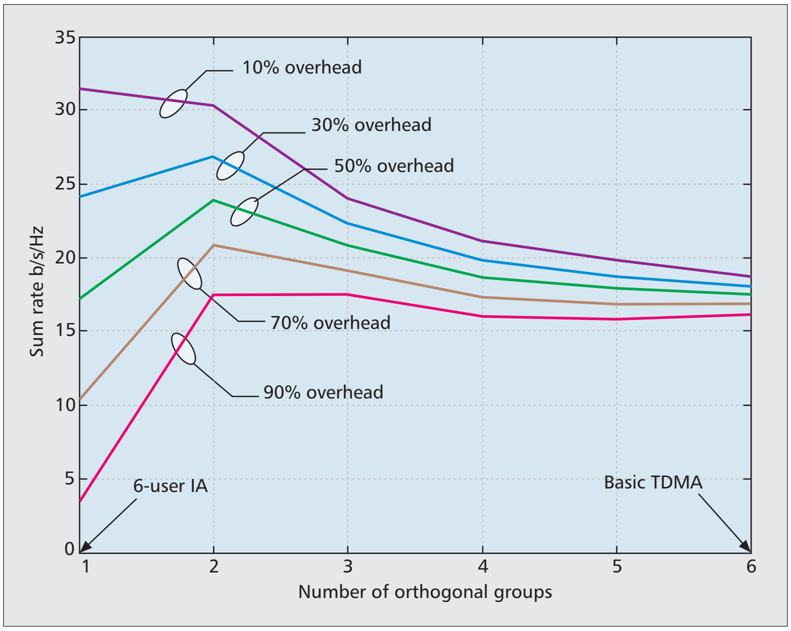


Figure 6. *The effective sum rate achieved by full network IA, TDMA, or a partitioned network in which smaller groups coordinate via IA. As overhead increases, the plot indicates that it is often optimal to allow fewer users to communicate simultaneously.*

Trans. Info. Theory, vol. 54, no. 8, 2008, pp. 3457–70.

[2] V. Cadambe and S. Jafar, "Interference Alignment and Degrees of Freedom of the *K*-User Interference Channel," *IEEE Trans. Info. Theory*, vol. 54, no. 8, Aug. 2008, pp. 3425–41.

[3] S. Jafar, *Interference Alignment — A New Look at Signal Dimensions in a Communication Network*, Now Publications, 2011.

[4] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the Degrees of Freedom Achievable through Interference Alignment in A MIMO Interference Channel," *IEEE Trans. Signal Proc.*, vol. 60, no. 2, Feb. 2012, pp. 812–21.

[5] K. Gomadam, V. Cadambe, and S. Jafar, "A Distributed Numerical Approach to Interference Alignment and Applications to Wireless Interference Networks," *IEEE Trans. Info. Theory*, vol. 57, no. 6, June 2011, pp. 3309–22.

[6] Q. Shi *et al.*, "An Iteratively Weighted MMSE Approach to Distributed Sum — Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Trans. Signal Proc.*, vol. 59, no. 9, Sept. 2011, pp. 4331–40.

[7] S. W. Peters and R. W. Heath, Jr., "Cooperative Algorithms for MIMO Interference Channels," *IEEE Trans. Vehic. Tech.*, vol. 60, no. 1, Jan. 2011, pp. 206–218.

[8] C. Shi *et al.*, "Local Interference Pricing for Distributed Beamforming in MIMO Networks," *Proc. IEEE MILCOM*, Oct. 2009.

[9] C. Shi, R. A. Berry, and M. L. Honig, "Adaptive Beamforming in Interference Networks via Bi-Directional Training," ArXiv preprint arXiv:1003.4764, Mar. 2010, http://arxiv.org/abs/1003.4764

[10] H. Bölcskei and I. Thukral, "Interference Alignment with Limited Feedback," *Proc. IEEE Int'l. Symp. Info. Theory*, July 2009, pp. 1759–63.

[11] O. El Ayach and R. W. Heath, Jr., "Interference Alignment with Analog Channel State Feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, Feb. 2012, pp. 626–36.

[12] M. Guillaud and D. Gesbert, "Interference Alignment in the Partially Connected K-User MIMO Interference Channel," *Proc. Euro. Signal Proc. Conf.*, Barcelona, Spain, Sept. 2011, pp. 1–5.

[13] S. W. Peters and R. W. Heath, Jr., "User Partitioning for Less Overhead in MIMO Interference Channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, 2012, pp. 592–603.

[14] K. Lin, S. Gollakota, and D. Katabi, "Random Access Heterogeneous MIMO Networks," *Proc. ACM SIGCOMM Conf. 2011*, pp. 146–57.

[15] O. El Ayach, S. Peters, and R. W. Heath, Jr., "The Feasibility of Interference Alignment over Measured MIMO-OFDM Channels," *IEEE Trans. Vehic. Tech.*, vol. 59, no. 9, Nov. 2010, pp. 4309–21.

Biographies

Omar El Ayach [S'08] (owarayach@utexas.edu) received his B.E. degree in computer and communications engineering from the American University of Beirut, Lebanon, in 2008. He completed his M.S. in electrical engineering degree in 2010 at the University of Texas at Austin (UT Austin) where he is now a Ph.D. student. He was an intern at the University of California, Berkeley, in the summer of 2007 and an intern at Samsung Electronics in the summers of 2011 and 2012.

Steven W. Peters [S'03] (steven.peters@kumasignals.com) is co-founder and chief executive officer of Kuma Signals, LLC in Austin, Texas. He received B.S. degrees in electrical engineering and computer engineering from the Illinois Institute of Technology in Chicago in 2005, and an M.S.E. degree in electrical engineering from UT Austin in 2007, where he is also working toward a Ph.D. in electrical engineering.

Robert W. Heath Jr. [S'96, M'01, SM'06, F'11] (rheath@utexas.edu) is a professor in the Electrical and Communications Engineering Department at UT Austin and the director of the Wireless Networking and Communications Group. He received his B.S. and M.S. in electrical engineering from the University of Virginia and his Ph.D. in electrical engineering from Stanford University. He is the president and CEO of MIMO Wireless Inc. and chief innovation officer at Kuma Signals LLC. He is the recipient of the David and Doris Lybarger Endowed Faculty Fellowship in Engineering and is a registered Professional Engineer in Texas.

CSI SHARING STRATEGIES FOR TRANSMITTER COOPERATION IN WIRELESS NETWORKS

PAUL DE KERRET AND DAVID GESBERT, EURECOM

ABSTRACT

Multiple-antenna "based" transmitter cooperation has been established as a promising tool toward avoiding, aligning, or shaping the interference resulting from aggressive spectral reuse. The price paid in the form of feedback and exchanging channel state information between cooperating devices in most existing methods is often underestimated, though. In reality, feedback and information overhead threatens the practicality and scalability of TX cooperation approaches in dense networks. Hereby we addresses a "Who needs to know what?" problem when it comes to CSI at cooperating transmitters. A comprehensive answer to this question remains beyond our reach and the scope of this article. Nevertheless, recent results in this area suggest that CSI overhead can be contained for even large networks provided the allocation of feedback to TXs is made non-uniform and to properly depend on the network's topology. This article provides a few hints toward solving the problem.

INTRODUCTION

Wireless communication has become essential to our lives in many ways, through a variety of services as well as devices ranging from pocket phones to laptops, tablets, sensors, and controllers. The advent of multimedia dominated traffic poses extraordinary constraints on data rates, latency, and, above all, spectral efficiency. In order to deal with the expected saturation of available resources in currently used bands, new wireless systems are designed based on:

- Greater densification of infrastructure equipment (small cells)
- Very aggressive spatial frequency reuse, which in turn results in severe *interference* conditions for cell-edge terminals

The role played by multiple antenna combining in mitigating interference by means of zero forcing (ZF) (or related criteria) beamforming is well established. Over the last few years, the combination of multiple-antenna approaches together with the concept of cooperation among interfering wireless devices was explored, showing strong promise (see [1] and references therein). In particular, transmitter (TX)-based cooperation allows for avoidance of interference before it even takes place (e.g., multicell multiple-input multiple-output [MIMO] or "joint processing coordinated multipoint [CoMP]"), or helps to shape it in a way which makes it easier for the receivers (RX) to suppress it (e.g., alignment). TX cooperation methods can be categorized depending on whether the data messages intended at the users must be known at several TXs simultaneously or not. For systems not allowing such an exchange (e.g., due to privacy regulations or low backhaul capabilities), interference alignment (IA) has been shown to be instrumental [2]. In contrast, when user data message exchange is made possible by a specific backhaul routing architecture, multicell, a.k.a. "network" MIMO, methods offer the best theoretical benefits [3].

A distinct advantage of TX cooperation over conventional approaches relying on egoistic interference rejection lies in the reduced number of antennas needed at each RX to ZF residual interference. This gain is further amplified when user data messages exchange among TXs is made possible. For instance, in the case of three interfering two-antenna TXs, relying on RX-based interference rejection alone requires three antennas to ZF the interference at each RX, while just two are needed when coordination is enabled via IA [4]. Furthermore, if the three user messages are exchanged among the TXs, thus enabling network MIMO precoding, just one antenna per TX and RX is sufficient to preserve interference-free transmission.

However, the benefits of multiple antenna transmit cooperation are gained at the expense of requiring channel state information (CSI) at the TXs. Indeed, whether one considers cooperation with or without user data sharing, the TXs should in principle acquire the complete CSI pertaining to every TX and RX pair in the network. This is also the case for distributed schemes (e.g., [5]) where the computation of precoders typically relies on iterative techniques where each iteration involves the acquisition of local feedback. However, as local feedback is updated over the iterations, this approach implicitly allows each TX to collect information about the precoders and channels of other TXs, hence amounting to an iterative global CSI acquisition at all TXs.

 1536-1284/13/$25.00 © 2013 IEEE

A distinct advantage of TX cooperation over conventional approaches relying on egoistic interference rejection lies in the reduced number of antennas needed at each RX to ZF residual interference. This gain is further amplified when user data messages exchange among TXs is made possible.

At first glance, CSI feedback and sharing requirements grow unbounded with network size. Since over the air feedback and backhaul exchange links are always rate and latency limited, the practical application of TX cooperation in dense large networks is difficult.

In this article, we challenge the common view that interfering TXs engaging in a cooperative scheme *can* or *should* share global (network-wide) CSI. Instead, we formulate the problem of a suitable CSI *dissemination* (or allocation) policy across transmitting devices while maintaining performance close to the full CSIT sharing scenario. We report a couple of findings revealing how the need for CSIT sharing can be alleviated by exploiting specific antenna configurations or decay property of signal strength vs. distance, hence making TX cooperation distributed and scalable. We use interference alignment and network MIMO as our driving scenarios.

More specifically, for the cooperation scenario without user data message sharing where alignment of interference is sought, we show how perfect alignment is possible in certain antenna topologies without knowledge of all the channel elements at some TXs. For the network MIMO scenario, this is not the case, and we illustrate instead how power decay vs. distance can be exploited to substantially reduce the CSI sharing requirements while fulfilling optimal asymptotic rate performance conditions. A common trait behind the findings is that different cooperating TXs can (and often must) live with their own individual partial version of the global CSIT. Hence, CSIT representation quality is bound to be non-uniform across TXs. Consequently, we discuss briefly the problem of multiple-antenna precoding with *TX-dependent* CSI.

Brief Notions in Multiple-Antenna Transmitter Cooperation

We consider fast fading multiple-antenna wireless networks where the transmission can be mathematically represented by writing $\mathbf{y}_i$, the received signal at RX i, as

$$y_i = \mathrm{H}_{ii}\mathrm{x}_i + \sum_{j\neq i}\mathrm{H}_{ij}\mathrm{x}_j \tag{1}$$

where $\mathbf{x}_i$ is the signal emitted by TX i and $\mathbf{H}_{ij}$ is a matrix containing the channel elements between TX j and RX i. The transmitted symbols $\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_K]^\mathbf{T}$ are then obtained from the user's data symbols $\mathbf{s} = [\mathbf{s}_1, \ldots, \mathbf{s}_K]^\mathbf{T}$ by multiplication with a precoder $\mathbf{T}$ (i.e., $\mathbf{x} = \mathbf{Ts}$). If the user's data symbols are not shared between the TXs, the precoder $\mathbf{T}$ is restricted to a particular block-diagonal structure, while it can otherwise take any form. The received filter $\boldsymbol{g}_i^\mathbf{H}$ is then applied to the received signal $\mathbf{y}_i$ to obtain an estimate of the transmitted data symbol.

Here, we briefly discuss the leading techniques for MIMO-based cooperation with or without user data message exchange. We point out commonly made assumptions in terms of CSIT sharing and feedback design.

Interference Alignment for Interference Channels

When the user's data symbols are not shared between the TXs, the setting is referred to as an interference channel (IC) in the communication theoretic literature. In MIMO ICs, a method called interference alignment (IA) has recently been developed and shown to achieve the maximal number of degrees of freedom (DoF), or pre-log factor, in many cases [2, 4]. As a consequence, IA has attracted a lot of interest in the community. In this work, we take the DoF as our key performance metric, such that we focus on IA schemes.

IA is said to be *feasible* if the antenna configuration (i.e., the distribution of antenna elements at the TXs and RXs) yields enough optimization variables to allow for interference-free transmission of all user data symbols, which means fulfilling [4]

$$\forall i, \forall j \uparrow i, \quad \boldsymbol{g}_i^\mathbf{H}\mathbf{H}_{ij}t_j = 0. \tag{2}$$

Intuitively, IA consists of letting the TXs coordinate among themselves to beamform their signals such that the interferences received at each of the RXs are confined in a subspace of reduced dimensions, which can then be suppressed by linear filtering at the RXs with a smaller number of antennas.

Precoding in the Network MIMO

When the user's data symbols are shared between the TXs, the TXs form a *distributed antenna array*, and a joint precoder can be applied at the transmit side [3]. Consequently, this setting becomes similar to the single TX multi-user MIMO downlink channel, and the interference between the TXs can be completely canceled, for example, by applying a global ZF precoder $\mathbf{T} \propto \mathbf{H}^{-1}$.

Limited Feedback vs. Limited Sharing

The limited feedback capabilities have been recognized as a major obstacle for the practical use of the precoding schemes described above. Consequently, a large body of literature has focused on this problem, and both efficient feedback schemes and robust transmission schemes have been derived for network MIMO [6, 7] and IA [8, 9].

However, all these works assume that the imperfect channel estimates obtained via limited feedback are *perfectly* shared between all the TX antennas. This is a meaningful assumption when the TXs are collocated but less realistic otherwise, as we shall now see.

CSIT Sharing Issues — One obstacle to the sharing of global CSIT follows from the fact that the amount of CSI that has to be exchanged increases very quickly with the number of TXs. In fact, each TX needs to obtain the CSI relative to the full multi-user channel, which consists of $(NK)^2$ scalars in a K-user setting with N antennas at each node.

In addition, acquiring the CSI at a particular TX can be realized by either direct broadcast of the CSI to all listening TXs or an

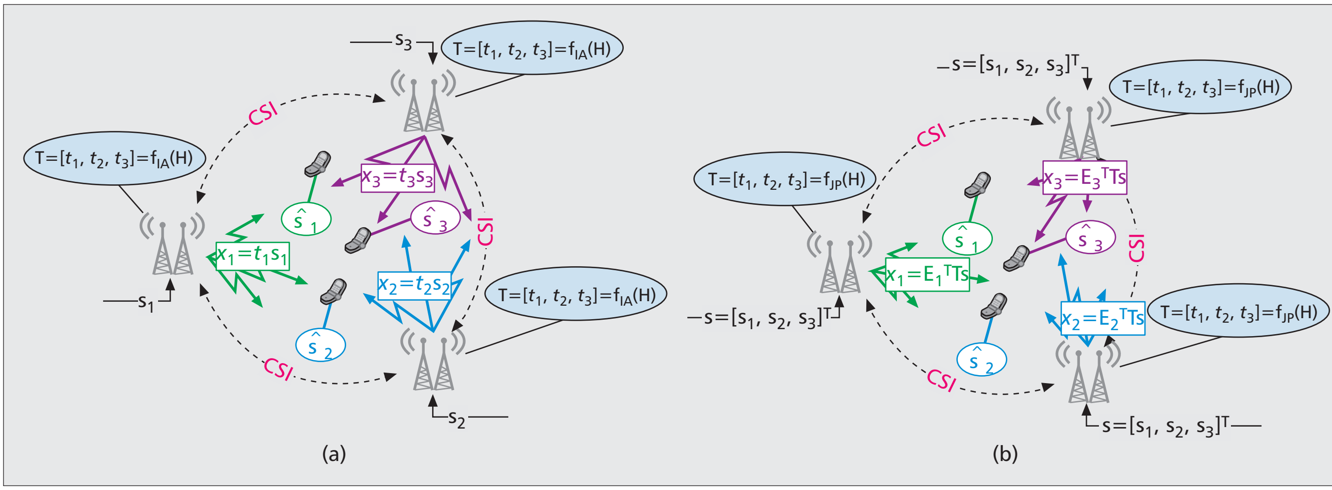


Figure 1. *The description of a distributed IA algorithm is done in (a) while (b) represents the distributed precoding in a Network MIMO with user's data sharing. The matrix* $\mathbf{E}_i$ *is a matrix which selects the rows of the multi-user precoder T corresponding to the antennas at TX* i.

over-the-air feedback to the *home* base station alone, followed by an exchange of the local CSIs over the backhaul, as it is currently advocated by Third Generation Partnership Project (3GPP) Long Term Evolution — Advanced (LTE-A) standards [10]. See the illustrations for such scenarios in Fig. 2. Note that exchange over the backhaul can involve further quantization loss and may lead to a different CSI aging at each TX due to protocol latency. In either case, the channel estimates available at the various TXs will not be exactly the same. This leads to a form of CSI discrepancy that is inherent in the cooperation among non-collocated TXs.

In order to capture multiple-antenna precoding scenarios whereby different TXs obtain an imperfect *and* imperfectly shared estimate of the overall multi-user channel, we denote by $\mathbf{H}^{(j)}$ the network-wide channel matrix estimate available at TX j. Consequently, the precoding schemes in Fig. 1 have to be modified to take into account that each TX will compute its precoder based on its own channel estimate. Thus, TX j transmits $\mathbf{x}_j = e_j^{\mathrm{T}}\mathbf{T}^{(j)}\boldsymbol{s}$ based on the knowledge of $\mathbf{H}^{(j)}$ only.

The fundamental questions that arise are:

- How complete and accurate should the estimate $\mathbf{H}^{(j)}$ be for each j while operating under reasonable CSI overhead constraints?
- How should precoders be designed given the likely discrepancies between various channel estimates?

Although these questions prove to be difficult and to a large extent remain open, we shed some light on the problem for two key scenarios in the following sections.

Aligning Interference with Incomplete CSIT

Let us first consider an IC without a user's data sharing. Feasibility studies for IA are typically carried out under the assumption of full CSIT. However, one can show that IA feasibility and the CSIT model are in fact tightly coupled notions. Assume, for instance, that all the RXs were given a generous number of antennas equaling or exceeding the number of TXs; it is well known that the interference could be suppressed at the RXs alone, and no precoding, and hence no CSIT, is necessary. This example suggests the existence of a trade-off between the number of antennas and the CSI sharing requirements. Thus, it is possible to design IA algorithms using less CSIT than conventionally thought, without performance degradations by exploiting the availability of extra antennas at a subset of devices. More specifically, the problem of finding the minimal CSIT allocation that preserves IA feasibility can be formulated. The minimality refers to the size of a CSIT allocation, defined as the total number of scalars sent through the multi-user feedback channel.

We differentiate between antenna configurations where IA is feasible, and the number of antennas at the TXs and the RXs provide just enough optimization variables to satisfy alignment conditions, denoted as *tightly feasible*, and the ones where extra antennas are available, denoted as *super feasible*. Furthermore, we call a CSIT allocation *strictly incomplete* if at least one TX does not have the complete multi-user CSI. With such concepts in place, the following lesson can be drawn.

Tightly Feasible ICs

A strictly incomplete CSIT allocation implies that some TXs compute their precoders in order to fulfill IA inside a smaller IC formed by a subset of RXs and a subset of TXs. Most of the time, this creates additional constraints for the optimization of the other precoders, which makes IA infeasible. However, it can be shown that IA feasibility can be preserved under the following condition [11].

Lesson 1 *— In a tightly feasible IC, there exists a strictly incomplete CSIT allocation preserving IA feasibility if there exists a tightly feasible sub-IC strictly included in the full IC.*

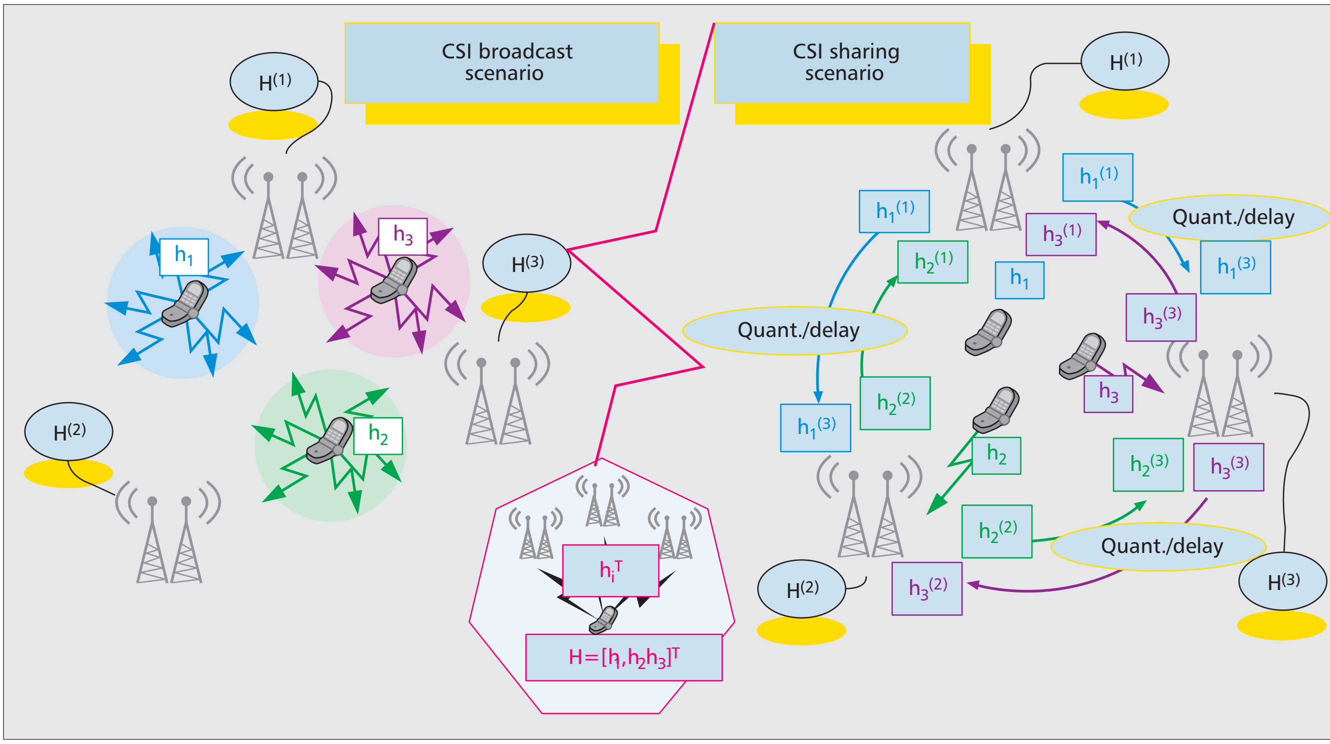


Figure 2. *Possible scenarios for the feedback of the CSI.*

Exploiting this result, a CSIT allocation algorithm is derived in [11] along with an algorithm that achieves IA based on this incomplete CSIT allocation. In a few words, it consists of giving to each TX the CSI relative to the smallest tightly feasible sub-IC to which it belongs. The precoders are then designed to align interference inside this smaller sub-IC, thus requiring only a part of the total CSIT, while IA feasibility is preserved. We will see in the simulations results presented in the following that the reduction in CSIT size is significant. In fact, the reduction of the CSIT allocation feeds on the heterogeneity of the antenna configuration such that the more heterogeneous the antenna configuration is, the larger is the saving brought by using the minimal CSIT allocation. This is particularly appealing in regard to future networks, where mobile units and base stations from different generations with different numbers of antennas are likely to coexist.

Super-Feasible ICs

In super-feasible settings, the additional antennas can be used to reduce the size of the minimal CSIT allocation. But how to optimally exploit these additional antennas to reduce the feedback size is a very intricate problem. Still, a low-complexity heuristic CSIT allocation can be derived [11]. The main idea behind the algorithm is to let some TXs or RXs ZF less interference dimensions such that small tightly-feasible settings are formed inside the original setting.

The effective CSIT reduction is illustrated in Fig. 3 for a three-user IC. The results are averaged over 1000 random distributions of the antennas across the TXs and RXs. If 12 antennas are distributed between the TXs and the RXs, the setting is tightly feasible, and the previous CSIT allocation policy for tightly feasible settings is used. With more than 12 antennas, the algorithm exploits every additional antenna to reduce the size of the CSIT allocation.

When the setting is tightly feasible, the reduction in feedback size requires neither a DoF reduction nor any additional antenna and comes in fact "for free": It simply results from exploiting the heterogeneity in the antenna numbers at the TXs and RXs.

A CSIT Allocation Policy for Network MIMO

When the user's data symbols are jointly precoded at the TXs, complete CSIT allocation, in the sense defined above, is needed in all practically relevant scenarios. So in this case, a different notion of reduced CSIT sharing must be advocated. The essential ingredient of this approach is the classical intuition that a TX should have a more accurate estimate for channels creating the strongest interference (i.e., originating from devices in the close neighborhood). This means the fact that interference decays with path loss can be exploited in principle to reduce the CSIT sharing requirements. This concept was recently introduced in [12]. A mathematical tool known as generalized DoF comes in handy to capture the effect of path loss on the multiplexing gain of cooperating networks with partially shared CSIT. Additionally, a simplified model referred to as a Wyner model is used in this context to aid analytical tractability and provide first insights into this problem.

Generalized Degrees of Freedom

TX cooperation methods are often evaluated through the prism of DoF performance. Unfortunately, the DoF is essentially path-loss-independent, such that a DoF analysis fails to properly capture the behavior of a large (extended) network MIMO. An extension of the notion of DoF, introduced in [13] as the *generalized DoF*, offers a much better grip on the problem as it can better take path loss models into account. Upon defining the *interference level* γ as $\gamma = \log(\text{INR})/\log(\text{SNR})$ with SNR denoting the signal-to-noise ratio and INR the interference-to-noise ratio, it is possible to define the generalized DoF as the DoF obtained when the SNR and INR both tend to infinity for *a given interference level* γ.

For ease of exposition, we consider scenarios where all the TXs and RXs have single antennas. The CSI is distributed, meaning that each TX has its own channel estimate based on which it computes its transmit coefficient without further exchange of information with the other TXs. We denote the estimate at TX j by $\mathbf{H}^{(j)}$ and its ith row, which corresponds to the channel from all TXs to RX i, by $\boldsymbol{h}_i^{(j)}$. We consider a digital quantization with a number of bits quantization $\boldsymbol{h}_i^{(j)}$ equal to $B_i^{(j)}$. Therefore, TX j computes its own version of the precoding matrix $\mathbf{T}^{(j)}$ based on its own estimate $\mathbf{H}^{(j)}$. It then transmits $\mathbf{x}_j = e_j^{\mathrm{T}}\mathbf{T}^{(j)}\boldsymbol{s}$.

The One-Dimensional Wyner Model

In the simple 1D *Wyner model* [14], the TXs are regularly placed along a line, and only the direct neighboring TXs emit non-zero interference. The channel is thus represented by a tridiagonal matrix. Furthermore, we assume that the interference from the direct neighbors is attenuated with a coefficient $\mu = P\gamma^{-1}$, according to the generalized DoF model. The transmission in the Wyner model is schematically presented in Fig. 4.

Our objective is to evaluate how small $B_i^{(j)}$ can be while guaranteeing the same DoF as a system with perfect CSIT. Obviously, sharing the most accurate CSIT to all the TXs is a possible solution, but the size of the CSI required *at each TX* then grows unbounded with the number of users K, making this solution both inefficient and unpractical. In contrast, a much more efficient CSI sharing policy achieving the maximal generalized DoF, denoted as *distance-based*, is summarized below [12].

Distance-Based CSIT Allocation

We are interested in a CSIT allocation strategy, referred to below as distance-based, whereby each TX receives a number of CSI scalars that remains *bounded* as the number of users K increases.

The distance-based CSIT allocation is obtained by setting for all i, j,

$$B_i^{(j)\text{Dist}} = \lceil([1 + (\gamma - 1)|i-j|]^+ + 2[\gamma + (\gamma-1)|i-j|]^+)\log_2(P)\rceil \quad (3)$$

where $[\bullet]^+$ is equal to zero if the argument is negative and to the identity function otherwise, and $\lceil\bullet\rceil$ is the ceiling operator. It can be shown that the CSIT allocation $\{B_i^{(j)\text{Dist}}\}_{i,j}$ allows the maximal generalized DoF to be achieved (i.e., those achieved in a system with perfect feedback) [12]. The proof is based on the off-diagonal exponential decay of the inverse of the tridiagonal channel matrix.

Setting $\gamma = 1$ (no significant path loss attenuation) in the previous equation, a conventional CSIT allocation is obtained where all channels are described with the same number of bits at all TXs (uniform allocation). For $\gamma < 1$, however, the number of bits allocated decreases with the distance $|i-j|$ between the considered TX and the index of the quantized channel until no bit at all is used for quantizing the channel if the distance $|i-j|$ between the RX and the TX is larger than $\lceil 1/(1-\gamma)\rceil$. Crucially, this solution allocates to each TX a total number of bits that no longer grows with K as only the CSI relative to a *neighborhood* is shared at each TX.

The different CSIT allocations are compared in Fig. 5 for $K = 15$ users and $\gamma = 0.5$. The conventional CSIT allocation consists in providing the best quality to all the TXs, while the other strategies have the same size as the distance-based CSIT allocation, but the feedback bits are shared uniformly and according to a conventional clustering of size 3. It can be seen from Eq. 3 that the ratio between the sizes of the distance-based CSIT allocation and the conventional CSIT allocation is independent of the SNR. Here, the distance-based CSIT allocation represents only 10 percent of the size of the conventional CSIT allocation.

Hence, the distance-based CSIT allocation achieves the maximal number of generalized DoF with only a small share of the total CSIT, and outperforms the other schemes of comparison. Additionally, the user's data sharing can also be reduced to a neighborhood without loss of performance. Consequently, this scheme can be seen as an alternative to clustering in which

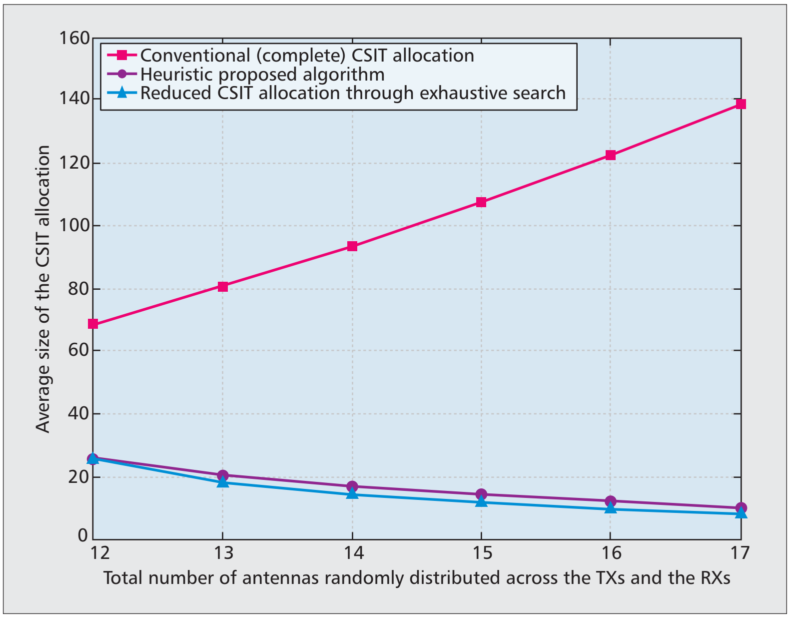


Figure 3. *Average CSIT allocation size in terms of the number of antennas randomly distributed across the TXs and RXs for* $K = 3$ *users.*

the hard boundaries of the clusters are replaced by a smooth decrease of the level of cooperation.

Precoding in the Network MIMO Channel with Distributed CSIT

As becomes clear from the previous section, an efficient CSI dissemination policy naturally leads to a significant reduction of the CSI sharing requirements. As a consequence, the CSIT is represented non-uniformly across the TXs. Since non-uniform sharing is the best strategy in order to maximize performance under a given feedback overhead constraint, it is a natural consequence that some users' channels will be coarsely described at certain TXs and more accurately at others. Interestingly, the problem of designing precoders that can accommodate such a peculiar CSIT scenario is by and large open. In particular, new robust precoding schemes should be developed, as conventional precoders are designed under the assumption that the same imperfect CSIT is shared *perfectly* among the TXs.

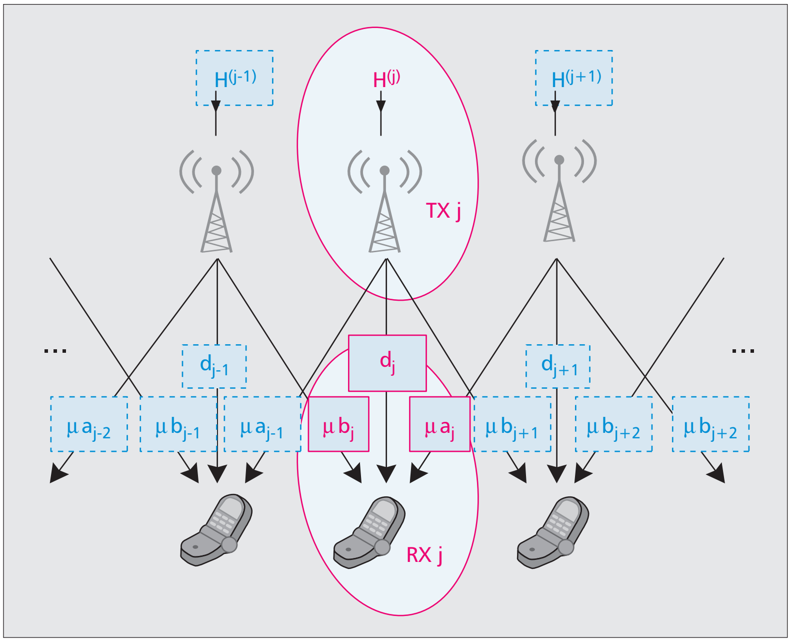


Figure 4. *Schematic representation for the Wyner model considered.*

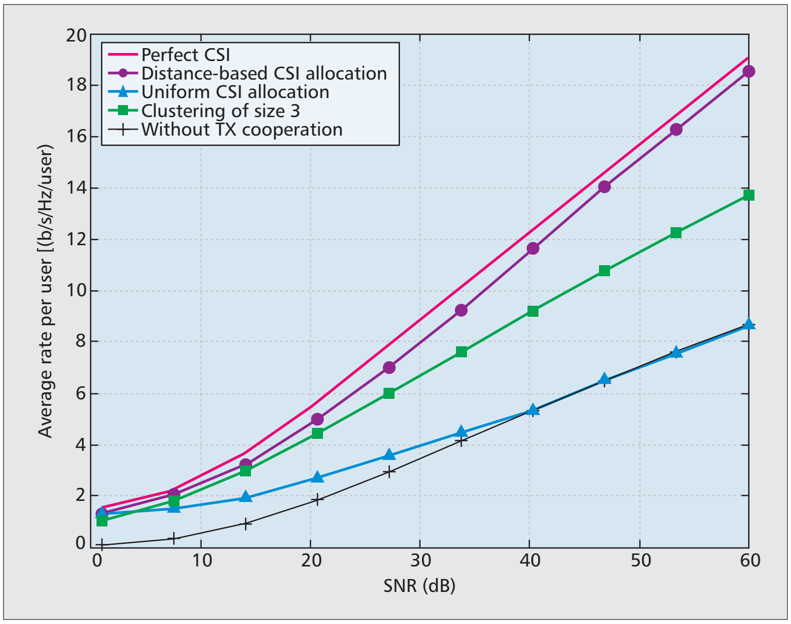


Figure 5. *Average rate per user in terms of the SNR. The distance-based CSIT allocation, the uniform allocation, and the clustering one have all a size equal to 10 percent of the size of the conventional CSIT allocation.*

Let us consider the model of distributed CSI described earlier where TX j receives its own channel estimate $\mathbf{H}^{(j)}$ with the ith row, denoted by $\boldsymbol{h}_i^{(j)}$, obtained using $B_i^{(j)}$ quantizing bits. Assume a network where each TX has roughly the same average path loss to each RX. The DoF, which can be achieved with limited feedback, is studied in [7] for the single TX MIMO downlink channel. We can extend this to the setting of non-uniform CSI to gain insight into the design of efficient precoders. In this case, CSI scaling coefficients $\alpha_i^{(j)}$ are introduced and defined by the limit of $B_i^{(j)}/((K-1)\log_2(P))$ when P goes to infinity.

ZF is widely used and well known to achieve the maximal DoF in the MIMO downlink channel with perfect CSIT [7]. One may wonder how conventional ZF performs in the presence of CSI discrepancies brought by imperfect sharing. The answer is strikingly pessimistic: The sum DoF achieved can be shown to be equal to just $K \min_{i,j} \alpha_i^{(j)}$ [15]. Intuitively, this can be restated as follows.

***Lesson 2** — In the network MIMO with distributed CSI, the worst channel estimate across the TXs and the users limits the DoF achieved by each user using ZF precoding.*

This is in strong contrast to the single TX case studied in [7] where the quality of the feedback of user i relative to $\boldsymbol{h}_i$ only impacts the DoF of user i. It shows clearly the disastrous impact of CSI non-uniformity, since one inaccurate estimate at one TX degrades the performance of *all* the users. One may also wonder whether conventional-type robust precoders [6] can offer a better response. The answer is negative, unfortunately. Instead, a novel precoder design is needed that is tailored to the non-uniform CSIT sharing model.

Preliminary results to this end [15] suggest that it is possible to dramatically improve the DoF in certain scenarios. For instance, in the two-TX network, a scheme referred to as active-passive (AP) ZF, consisting of letting the TX with degraded CSIT arbitrarily fix its transmit coefficient while the other TX compensates to zero-out the interference, can be shown to recover the optimal DoF.

The average rate achieved with conventional ZF, AP ZF, and ZF with perfect CSIT are compared in Fig. 6. In that case, the sum rate of conventional ZF saturates at high SNR, while AP ZF is more robust and achieves a better DoF.

Open Problems

New concepts for CSIT sharing in wireless networks have been derived, and their potential to reduce signaling overhead has been shown.

We have presented some insights into a new problem that presents serious challenges, but also opportunities for the future. This leads to new intriguing open questions. First, IA algorithms with incomplete CSIT are based on a DoF-preserving criterion only (i.e., on the performance at asymptotically high SNR). The impact of the incomplete CSIT on the performance at finite SNR should then be investigated to obtain practical solutions. Similarly, robust precoding schemes for the MIMO network with distributed CSI have so far considered DoF only as a metric. By and large, precoding over network MIMO with distributed CSI remains a challenging problem. Regarding the optimization of the CSIT allocation in a network MIMO channel, we believe that the distance-based CSIT allocation has the potential to impact the design of cellular systems in practical settings. However, it has been examined for simplified channel models only and should be adapted to more realistic scenarios.

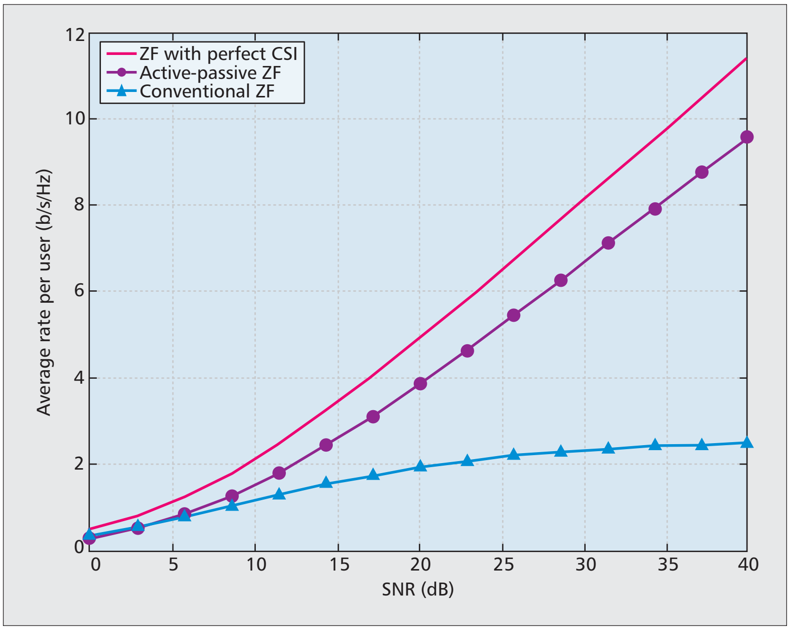


Figure 6. *Average rate per user in terms of the SNR for* $\alpha_1^{(1)} = 1$, $\alpha_2^{(1)} = 0$, $\alpha_2^{(1)} = 0.5$, $\alpha_2^{(2)} = 0.7$.

Conclusion

The uniform allocation of CSI resources toward all TXs does not lead to efficient use of the feedback and backhaul resources: The CSIT dissemination should be designed to allocate to each TX the right CSI, in terms of which elements to share and accuracy. For both network MIMO and IA, the CSIT sharing requirements have been significantly lowered by designing appropriate CSIT dissemination policies, thus making TX cooperation more practical and thereby paving the way for large performance improvement. Additionally, the problem of robust precoding schemes taking into account the inconsistency between the CSIs at the TXs has been tackled and shown to be a key element.

Acknowledgment

This work has been performed in the framework of the European research project HARP, which is partly funded by the European Union under its FP7 ICT Objective 1.1 — The Network of the Future.

References

[1] D. Gesbert *et al.*, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," *IEEE JSAC*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
[2] V. R. Cadambe and S. A. Jafar, "Interference Alignment and Degrees of Freedom of the *K*-User Interference Channel," *IEEE Trans. Info. Theory*, vol. 54, no. 8, Aug. 2008, pp. 3425–41.
[3] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network Coordination for Spectrally Efficient Communications in Cellular Systems," *IEEE Wireless Commun.*, vol. 13, no. 4, Aug. 2006, pp. 56–61.
[4] C. Yetis et al., "On feasibility of interference Alignment in MIMO Interference Networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, Sept. 2010, pp. 4771–82.
[5] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A Distributed Numerical Approach to Interference Alignment and Applications to Wireless Interference Networks," *IEEE Trans. Info. Theory*, vol. 57, no. 6, June 2011, pp. 3309–22.
[6] M. B. Shenouda and T. N. Davidson, "Robust Linear Precoding for Uncertain MISO Broadcast Channels," *Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing*, 2006.
[7] N. Jindal, "MIMO Broadcast Channels with Finite-Rate Feedback," *IEEE Trans. Info. Theory*, vol. 52, no. 11, Nov. 2006, pp. 5045–60.
[8] H. Bolcskei and I. J. Thukral, "Interference Alignment with Limited Feedback," *Proc. IEEE Int'l. Symp. Info. Theory*, 2009.
[9] O. E. Ayach and R. W. Heath, "Interference Alignment with Analog Channel State Feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, Feb. 2012, pp. 626–36.
[10] S. Sesia, I. Toufik, and M. Baker, *LTE — The UMTS Long-Term Evolution: From Theory to Practice*, 2nd ed., Wiley, 2011.
[11] P. de Kerret and D. Gesbert, "Interference Alignment with Incomplete CSIT Sharing," submitted to *IEEE Trans. Wireless Commun.*, Nov. 2012.
[12] —, "CSI Feedback Allocation in Multicell MIMO Channels," *Proc. IEEE ICC*, 2012.
[13] R. Etkin, D. Tse, and H. Wang, "Gaussian Interference Channel Capacity to Within One Bit," *IEEE Trans. Info. Theory*, vol. 54, no. 12, Dec. 2008, pp. 5534–62.
[14] A. D. Wyner, "Shannon-Theoretic Approach to a Gaussian Cellular Multiple-Access Channel," *IEEE Trans. Info. Theory*, vol. 40, no. 6, Nov. 1994, pp. 1713–27.
[15] P. de Kerret and D. Gesbert, "Degrees of Freedom of the Network MIMO Channel with Distributed CSI," *IEEE Trans. Info. Theory*, vol. 58, no. 11, Nov. 2012, pp. 6806–24.

Biographies

Paul de Kerret [S] (dekerret@eurecom.fr) graduated from Ecole Nationale Superieure des Telecommunications de Bretagne, France, in 2009 and obtained a diploma degree in electrical engineering from Munich University of Technology (TUM), Germany. He also earned a four-year degree in mathematics at the Universite de Bretagne Occidentale, France, in 2008. From January to September 2010 he was a research assistant at the Institute for Theoretical Information Technology, RWTH Aachen University, Germany. Since October 2010 he has been working toward a Ph.D. degree in the Mobile Communications Department at EURECOM, France.

David Gesbert [F] (gesbert@eurecom.fr) is a professor and head of the Mobile Communications Department, EURECOM, France. He received his Ph.D. in 1997 from Telecom Paris Tech. From 1997 to 1999 he was with the Information Systems Laboratory, Stanford University. In 1999 he was a founding engineer of Iospan Wireless Inc, San Jose, California, a startup company pioneering MIMO-OFDM (now Intel). He has guest edited six special issues, and published about 200 papers and several patents in the area of wireless networks, among which three won international awards. He co-authored the book *Space Time Wireless Communications: From Parameter Estimation to MIMO Systems* (Cambridge Press, 2006).

MULTICELL COOPERATIVE SYSTEMS WITH MULTIPLE RECEIVE ANTENNAS

INSOO HWANG, UNIVERSITY OF TEXAS AT AUSTIN
CHAN-BYOUNG CHAE, YONSEI UNIVERSITY
JUNGWON LEE, SAMSUNG ELECTRONICS U.S. R&D CENTER
ROBERT W. HEATH, JR., UNIVERSITY OF TEXAS AT AUSTIN

The work of C.-B. Chae (corresponding author) was in part supported by the KCC (Korea Communications Commission), Korea, under the R&D program supervised by the KCA (Korea Communications Agency) (KCA-2012-911-01-106) and the Ministry of Knowledge Economy under the IT Consilience Creative Program (NIPA-2012-H0201-12-1001).

ABSTRACT

Multicell cooperation may play an important role in achieving high performance in cellular systems. Multicell cooperation with a single receive antenna at the mobile station has been widely investigated. Applying cooperation with multiple receive antennas, allowed in several emerging wireless standards, has met with some challenges. This is primarily because the transmit precoding/beamforming vector and receive processing have to be jointly optimized in multiple cells to combat out-of-cell interference. In this article, we first discuss the role of the receive antennas in a multicell environment, and then review recently proposed multicell cooperative algorithms and receive antenna techniques for different interference statistics. Finally, we highlight recent work on the fundamental limits of cooperation and the possibility of overcoming such limits by using multiple receive antennas in multicell cooperative networks.

INTRODUCTION

In wireless cellular systems, a geographical region is partitioned into small cells. Within each cell, a transmitter is dedicated to serving receivers within that cell. Within this cell structure, all voice and data traffic that the receivers request and generate will be transmitted from and to the transmitter dedicated to serving them. One of the key performance indicators that decides the maximum rate at which the transceivers communicate with each other is the received signal-to-interference-plus-noise ratio (SINR) — the higher the SINR, the higher the maximum rate. Transmissions from neighboring co-channel transmitters, however, result in out-of-cell interference, degrading the SINR, especially to the receivers located at the cell edge. To mitigate the effects of out-of-cell interference, multicell cooperative processing between the transmitters has been actively investigated [1–5].

Multicell cooperative processing may also be an important technology for small cell networks. To deal with the increasing data traffic due to smart phones, wireless cellular systems such as Long Term Evolution (LTE) networks will offload portions of their traffic onto smaller cells such as picocells, femtocells, and relays, which together with the currently existing macrocells form a heterogeneous network. As these cells get smaller, and more cells are packed into the same amount of space, the receiver may see more diverse and stronger out-of-cell interference; thus, a novel way of handling out-of-cell interference is required. Therefore, in emerging wireless cellular standards such as Third Generation Partnership (3GPP) LTE-Advanced, coordinated multipoint (CoMP) transmission and reception techniques, which use the core concept of multicell cooperative processing, are being actively discussed [6].

In multicell cooperative processing, the transmitters cooperate to jointly combat interference from adjacent cells and fading to receivers. This cooperation requires exchanging some information such as downlink channel information and may require multiple antennas. The barriers preventing such cooperation have become less costly; the wired communication between transmitters (e.g., X2 interface in the 3GPP LTE-Advanced networks) enable the transmitters to exchange information quite reliably with only marginal delay. The decrease in the price of radio frequency (RF) electronic components enable the transmitters and receivers to use multiple-input multiple-output (MIMO) technologies. By combining beamforming/precoding techniques enabled by MIMO with multicell cooperative processing, the post-processing SINR per receiver can be increased, thus further improving the system throughput.

If the number of data streams designated to a receiver is smaller than the number of receive antennas, the receiver can utilize the receive antennas to combat the other-cell interference. Additional degrees of freedom at each receiver enable such operation, even with the use of linear receive filters [2, 4, 5]. When the number of receive antennas is not greater than the number of streams to decode, or transceiver coordination is not feasible due to inaccurate feedback,

 1536-1284/13/$25.00

more advanced receiver algorithms to suppress the other-cell interference can also be used. Advanced receiver algorithms such as the joint maximum likelihood (ML) detector may give better error rate performance through decoding the desired signal and interference signal *jointly* rather than treating the interference as noise [7].

Another potential benefit of using multiple receive antennas in multicell cooperative networks is the measurement of network utility performance. The capacity region of the MIMO interference channel with multiple receive antennas at each receiver—the system model of interest in this article—is still unknown for most cases. The degrees of freedom of the MIMO interference channel has recently been well studied [8], and a better measure of the achievability that takes interference signal strength into account has been proposed in [9]. Despite its importance, the role of the receive antenna in multicell cooperative coordinated processing has yet to be fully understood.

To achieve this diverse set of objectives, in this article, we first introduce basic CBF systems, each of which is mainly categorized by the type of information sharing. Next, we investigate the role of multiple receive antennas in CBF (CBF) systems and introduce advanced receiver processing techniques to efficiently mitigate the other-cell interference. Then we introduce recent theoretical findings on the fundamental limits of cooperation and discuss their expansion to more general multicell scenarios. Finally, we review multicell CBF in emerging standards, mainly focusing on 3GPP LTE-Advanced systems, and then follow with some conclusions.

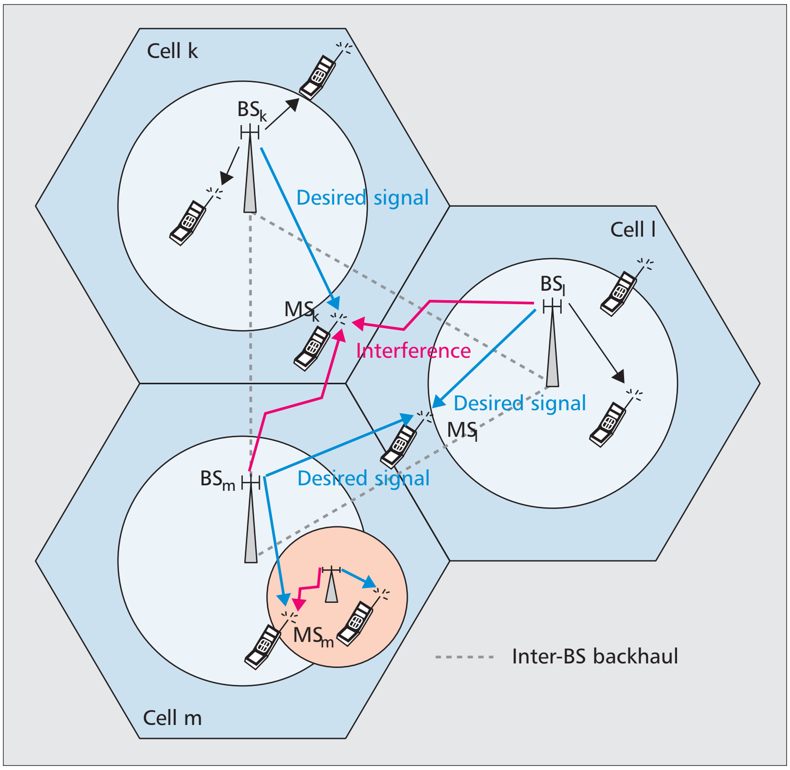


Figure 1. *Coordinated multicell processing system model (a three-cell scenario). Receiver MS_k is in a coordinated beamforming scenario, receiver MS_l is under a joint processing scenario, and receiver MS_m is in a spatial sensing scenario.*

Coordinated Beamforming Systems

Prior work on multicell cooperative processing can be categorized as either:

- Coordinated beamforming (CBF)
- Joint processing (JP)
- Spatial sensing (SS)

In this section, we present the key concepts of the three different multicell cooperative processing techniques.

Cellular Networks: Coordinated Beamforming with No Data Sharing

In CBF, the transmitters exchange only the downlink channel information, and each receiver receives a stream of data from only one transmitter. The receiver in CBF is illustrated as MS_k in Fig. 1. In this case, the multicell MIMO channel is the MIMO interference channel, and the goal of CBF is to jointly design optimized transmit and receive beamformers to minimize the effect of out-of-cell interference. A number of CBF algorithms have been proposed in the literature [4, 8], with interference alignment being one of the technical enablers.

In CBF, data for a receiver is only available at and transmitted from one cooperating base station for a time-frequency resource. User scheduling and beamforming decisions, however, are made with coordination among transmitters corresponding to the cooperating set. The transmit beamforming and receive combining vectors are generated to reduce the unnecessary interference to other receivers scheduled within the coordinated cell. The scheduler is targeted to maximize the SINR or minimize the leakage power to other receivers while maximizing the signal power to the desired receiver. The latter is often called Max-SLNR (signal-to-leakage-plus-noise ratio) scheduler and often leads to better sum rates than the Max-SINR scheduler. Through scheduling, the desired set of receivers is selected so that the transmit and/or receive beamforming vectors are chosen to reduce the interference to/from other neighboring users, while increasing the serving cell's signal strength. The cell edge receiver's SINR can be improved by CBF, resulting in a cell-edge throughput improvement.

When the receivers are equipped with multiple receive antennas, transceiver algorithms at the cooperating transmitters and receivers need to be jointly optimized to improve the sum rate. In a two-cell system with multiple receive antennas, non-iterative CBF algorithms called interference-aware CBF (IA-CBF) were proposed in [4]. It was proven that the proposed algorithms are optimal in terms of the degrees of freedom of the two-cell MIMO channel, where the mobile stations have two receive antennas. An asymptotic expression of the achievable sum rate of the proposed system with respect to the number of receive antennas was also derived. Interestingly, as the number of receive antennas increases, the proposed algorithm achieves not only full degrees of freedom but also the sum rate of the point-to-point upper bound.

Recent demand for higher data rate support and the use of multiple receive antennas at the receiver changes the aspect of joint processing from classical soft-handoff or macro diversity to joint transmit and receive beamformer/precoder design in multicell cooperative processing.

Generalization of IA-CBF algorithms to more than two-cell scenarios is also presented in the same paper [4]. The authors proposed a novel (physical) beam-switching mechanism that intentionally creates a beam conflict; a conflict that in the conventional system was, as much as possible, avoided. A beam sequence optimization problem is NP-hard; simple algorithms are discussed by prioritizing the transmitter to decide the beam index first based on its own channel information. Next the transmitter determines its own beam index that creates a beam conflict with the top priority transmitter. The philosophy used here is that instead of mitigating other-cell interference and treating them as background noise, the transmitters can create a strong interference link and use it to further minimize the background interference. This can be done by additional receiver processing, so the importance of multiple receive antennas in CoMP has become more obvious.

Cellular Networks: Joint Processing with Perfect Data Sharing

In joint processing, each user receives a data stream from multiple transmitters. To enable it, downlink data streams as well as downlink channel information must be exchanged via intertransmitter cooperation. The joint processing transforms the MIMO interference channel (in the coordinated case) into the MIMO broadcast channel in a case of ideal cooperation because multiple coordinated transmitters effectively form a distributed super base station. The capacity region of the MIMO broadcast channel has been studied and analyzed [10]. The receiver in joint processing is illustrated as MS_l in Fig. 1.

Similar to spatial division multiple access (SDMA) technologies in the MIMO broadcast channel, in joint processing, transmit beamforming is performed to eliminate the other-cell interference at the transmitter while the receivers with single receive antenna performs simple decoding. Typically the transmitter computes a transmit precoding vector using the aggregated channel knowledge. Linear precoding is one of the simplest solutions, although zero-forcing channel inversion linear precoder suffers from excessive power penalty. Regularized channel inversion precoder is hard to find the exact regularization factor. Nonlinear pre-processing such as vector perturbation (VP) or Tomlinson-Harashima precoding (THP) can be applied on top of the linear precoder, showing a better sum rate performance. The role of multiple receive antennas at the receiver side, however, has not been well addressed.

The network CBF (N-CBF) algorithm proposed in [11] approaches the joint processing system in different ways. Instead of multiple transmit antennas with a single receive antenna, the paper focuses on the multicell downlink channel where each transmitter is equipped with one transmit antenna and each receiver is equipped with multiple receive antennas. Linear and nonlinear CBF algorithms are proposed for both clustered broadcast channels and full broadcast channels that achieve near multicell capacity. The proposed algorithms minimize the largest eigenvalue of the inverse of the effective matched channel matrix to minimize the transmit power consumption and simultaneously maximize the effective channel gain. This can be achieved through the fact that the effective channel, $\boldsymbol{R} = \boldsymbol{H}^*\boldsymbol{H}$ where $\boldsymbol{H}$ is the channel matrix, goes to the identity matrix as the number of receive antennas grows large; the more receive antennas the system uses, the better throughput can be obtained even with a single transmit antenna. The expansion of this problem with multiple transmit antennas is nontrivial and may give better sum rate performance.

It is known that joint processing provides higher system throughput than CBF, but at the cost of downlink data stream exchange between transmitters. We note that the down side of joint processing is the requirement that the transmitters exchange downlink data streams. The price to be paid for such additional cooperation, however, may be decreasing. For instance, in a 3GPP LTE-Advanced network [6], the transmitters are *already* connected over a wired backhaul line, so data exchange between transmitters is becoming easier. Recent demand for higher data rate support and the use of multiple receive antennas at the receiver changes the aspect of joint processing from classical soft handoff or macro diversity to joint transmit and receive beamformer/precoder design in CoMP.

Cognitive Networks

So far, we have focused on homogeneous cellular networks and addressed the role of the receive antennas. In this section, we consider cognitive networks and also investigate how much gain can be achieved due to multiple receive antennas. As the standardization of heterogeneous networks become more concrete, cognitive multicell network is also drawing lots of attention. The use of femtocells in a heterogeneous network may use the same licensed band as the macro transmitter, but the femtocell may not be allowed to create any noticeable interference to the legacy receivers. When the primary transmitters are active, the secondary receiver which is attached to the secondary cell (e.g., femtocells) suffers interference from the primary transmitters. This type of interference has motivated new research into joint transceiver designs in cognitive multicell cooperative networks. The receiver MSm in Fig. 1 fits into this scenario.

In cognitive multicell cooperative networks, most prior work has focused on the design of precoding matrices to suppress interference to the primary receivers. In conventional cognitive multicell systems, the secondary receiver treats interference from the primary transmitter as an additive noise. It is thus mainly an interference avoidance approach that minimizes the signal leakage to the primary receivers. The interference to the secondary receivers is considered a secondary problem, although the secondary receivers are also desired to communicate reliably. Again, multiple receive antennas at the secondary receivers may help to achieve both objectives simultaneously:

- To prevent interference to the primary receivers

Interference	Scenario	Possible Receiver Algorithm	Required Parameters
White Gaussian	Many interferers without dominant one	MMSE	σ_I^2
Colored Gaussian	A few interferers with dominant one	IW-MMSE or MMSE-IRC	$\boldsymbol{R_I}$
Modulated symbol	Single or a few strong interferers	Joint detection	$\boldsymbol{H_I}$, modulation order

Table 1. *Comparisons of receiver processing. Interference channel, interference plus noise variance, and interference covariance matrix are respectively denoted by $\boldsymbol{H_I}$, σ_I^2, and $\boldsymbol{R_I}$.*

In all three of the proposed SS-CBF algorithms, multiple antennas at the secondary transceiver are used in two ways: to suppress interference to the primary receivers and from the primary transmitters, as well as to maximize the achievable rate of the secondary link.

- To remove the interference, due to primary transmissions, at the secondary receiver

This can be done with the help of multiple receive antennas at the secondary receiver.

The spatial sensing CBF (SS-CBF) algorithms proposed in [5] address this problem. With single-antenna primary terminals and two-antenna cognitive terminals, a linear transceiver design has been introduced under a global channel state information (CSI) assumption. Depending on the required information of the secondary transceiver, the approach and the role of multiple receive antennas may be different. If the secondary transceiver has only the local CSI, which consists of the channel knowledge of primary transmitter-secondary receiver and secondary transceivers, the secondary transmitter constructs the projected-channel singular value decomposition (P-SVD) to enable error-free communication to the secondary receiver. If the channel between primary transceivers is known to the secondary transceiver in addition to the local CSI, the secondary transceivers can adopt a joint transmitter-receiver optimization, which has proved to be optimal under the zero-interference constraint at both the secondary transmitter and receiver. If the local CSI and side information (precoding and postcoding information of the primary transceivers) is known to the secondary transceivers, iterative precoding and decoding is proposed, outperforming the P-SVD approach of the local CSI.

In all three of the proposed SS-CBF algorithms, multiple antennas at the secondary transceiver are used in two ways: to suppress interference to the primary receivers and from the primary transmitters, as well as to maximize the achievable rate of the secondary link. The conventional approach in cognitive multicell cooperative networks is focused solely on constructing a transmitter by treating the interference from the primary transmissions to the secondary receiver as additive noise. The proposed SS-CBF jointly designs the secondary transceivers, resulting in an achievable rate similar to that of equivalent point-to-point MIMO channels. As the number of antennas at the second transceivers increase, the performance of SS-CBF with global CSI converges to that of the point-to-point MIMO channel, while the other two SS-CBF algorithms also show the achievable rates very close to that bound.

Advanced Receiver Algorithms

If the number of receive antennas is insufficient or full coordination between transmitters and receivers is not feasible, the receiver may utilize its receive antennas in different ways. An advanced receiver technique may also be required when a small group of adjacent interferers are not fully synchronized. Thus, the other-cell interference becomes non-Gaussian [12]. Assuming the receiver can estimate both the desired channel and interference channel, advanced interference mitigation algorithms can be used. Depending on the statistics of interference, the appropriate receiver algorithms can differ. Table 1 presents the three possible receiver algorithms under three types of out-of-cell interference, which are minimum mean square error (MMSE), interference whitening with MMSE (IW-MMSE) ,or MMSE with interference rejection combining (MMSE-IRC), and joint detection.

In this section, we introduce the advanced receiver algorithms that can be used in multicell cooperative CBF systems. Note that the use of advanced receivers is also possible on top of the CBF algorithms introduced earlier. Typically, when the channel is slowly varying so that channel knowledge at the transmitter can be accurate, CBF can be effectively utilized. In an agile channel condition, the receiver may apply advanced receiver techniques rather than relying on full coordination. The advanced receive algorithm can thus be a supplementary to CBF algorithms, not a competitive one.

White Gaussian Interference: MMSE Filter

If the out-of-cell interference comes from many non-dominant interferers, and the interference from each of the interferers is uncorrelated, the aggregated interference seen by the receiver can be assumed to be white Gaussian. If the variance of the interference is known or precisely measured by the receiver, the receiver can treat the white Gaussian interference as noise. Since the interference-plus-noise covariance matrix becomes an identity matrix with the same diagonal entry sI2, it can be treated as additive white Gaussian noise, and the MMSE receiver provides the best link quality to the receiver in multicell cooperative systems.

In a multicell coordinated system with white Gaussian interference, the MMSE filter performs differently in different SINR levels. If the

As the signal-to-interference level increases, the packet error rate performance becomes floored. If the interference is highly non-Gaussian, which typically happens for the out-of-cell interference with low modulation order, the IW-MMSE and MMSE-IRC is not a desirable approach.

signal power is noticeably higher than the interference-plus-noise power, the receiver may ignore the interference and perform as a zero-forcing decorrelator. The detector simply divides the received signal by the serving cell channel gain and then slices it to the nearest signal constellation point. If the interference-plus-noise power is relatively higher than the signal power, the MMSE receiver matches with the serving cell channel. In the CBF system, the MMSE filter maximizes the post-processing SINR when the interference is white Gaussian, regardless of the SINR levels.

Colored Gaussian Interference: IW-MMSE or MMSE-IRC

The noise-plus-interference term at the receiver may contain either intercell interference or intracell interference, or a mix of both plus additive Gaussian noise. In general, such interference is not white, especially when the number of interferers is small and there are a few dominant interferers. The noise-plus-interference term is then spatially colored with its covariance matrix having non-zero off-diagonal terms. The colored interference-plus-noise term can cause a serious problem to the detector and decoder, which is generally designed assuming interference and spatially temporally white noise. Assuming the interference-plus-noise covariance matrix is nonsingular, a spatial whitening filter matrix needs to be applied on the receive signal such that the transformed interference becomes white, with a slight sacrifice to the desired signal strength.

The interference whitening filter $\boldsymbol{W}$ transforms the interference-plus-noise term $\boldsymbol{n}$ into a spatially white Gaussian vector. The transformed noise after whitening, $\boldsymbol{Wn}$, has the covariance of an identity matrix, $\mathbb{E}\{\boldsymbol{Wn}(\boldsymbol{Wn})^*\} = \boldsymbol{W}\mathbb{E}\{\boldsymbol{nn}^*\}\boldsymbol{W}^* = \boldsymbol{I}_{N_r}$. This yields $\boldsymbol{WRW}^* = \boldsymbol{I}_{N_r}$, and after taking the eigen-decomposition on the effective channel matrix $\boldsymbol{R}$, $\boldsymbol{R} = \boldsymbol{U}\Sigma\boldsymbol{U}^*$, the whitening matrix can be obtained as $\boldsymbol{W} = \Sigma^{-1/2}\boldsymbol{U}^*$. After the whitening filter is applied, an MMSE detector can be used to eliminate the white Gaussian noise.

An alternative way to detect the symbol with colored interference is to use interference rejection combining (IRC). In the MMSE-IRC filter, the non-diagonal interference covariance matrix $\boldsymbol{R}_I$ is added to the MMSE whitening matrix, enabling the receiver to reject the interference. Note that the MMSE-IRC receiver can be biased or unbiased, depending on the use of a normalization factor. The biased MMSE-IRC receiver maximizes the received SINR, while the unbiased MMSE-IRC filter minimizes the MSE. With a proper scaling, however, the two MMSE filters show identical coded block error rate performance [13]. Thus, we may observe that the two MMSE-IRC filters may result in similar throughput performance.

The IW-MMSE/MMSE-IRC detector is one of the simplest solutions to combat (slightly) non-Gaussian interference. By sacrificing the desired channel gain a bit, the interference plus noise can be approximated as Gaussian, so a conventional detector and decoder can be utilized. The IW-MMSE/MMSE-IRC receiver, however, can perform well only with high modulation order, high code rate in a weak to medium interference region. As the signal-to-interference level increases, the packet error rate performance becomes floored. If the interference is highly non-Gaussian, which typically happens for out-of-cell interference with low modulation order, IW-MMSE and MMSE-IRC are not desirable approaches.

Interference with Known Modulation Order: Joint Detection

In practical multicell cooperative systems, the transmit symbols are modulated, so the receiver needs to detect the transmitted symbol using a proper demodulator. The modulation order can be changed over time to adapt the transmission rate close to the instantaneous channel capacity. If the modulation order of both the desired signal and interference signal are known to the receiver, the receiver can attempt to decode the desired symbols by jointly decoding the desired symbol and interference symbol. In [7], a significant performance gain is shown to be obtained if the detectors explicitly take into account the modulation formats of the desired and interference signals.

Unlike the conventional interference canceler and successive interference canceler, which work well only in noise-limited and interference-limited regimes, respectively, the joint ML detector can cover variety of interference range. Specifically, the joint ML detector incorporates knowledge of the interfering channel and the finite modulation of the interference, turning interference-limited transmission system into a noise-limited one. The joint ML detector detects the desired and interfering signals jointly, then discards the detector output for the interference, which aids the detection of the desired signal. As the computational complexity can easily be double that of the interference-ignorant ML detector, a less complex joint minimum distance (MD) detector can also be used. In a multicell cooperative network, however, the joint MD detector is slightly worse than the joint ML detector in a low SNR region.

The joint detection approach is beneficial when the channels are fast fading. In such a condition, channel knowledge at the transmitter is easily outdated; hence, rigorous coordinated processing is somewhat infeasible. The joint detector, however, performs well in such an agile channel condition as it turns an interference-limited channel into a noise-limited channel. It is also shown that the full diversity gains can be achieved with the joint detector even in the presence of interference, so the transmit diversity scheme is advantageous. The joint detector is thus quite complementary to multicell cooperative CBF systems with multiple receive antennas.

Multicell Cooperative Processing in Emerging Standards

The next generation wireless cellular systems will use coordinated multicell joint processing. In the 3GPP LTE-Advanced systems, CoMP is considered one of the strongest and best proven tools to improve cell edge throughput. It is also proven that CoMP gives a higher gain than the traditional approaches of cellular network planning such as frequency reuse, sectoring, and spread spectrum. Despite the fact that the standardization of CoMP is still in progress and scheduled to be finalized in the first half of 2013, it is meaningful to review the practical scenarios and limitations. It is noteworthy that, based on the proposals submitted to the Release 12 (Rel-12) LTE-Advanced workshop, CoMP will be further evolved and enhanced to meet the requirements for so-called fifth generation (5G) systems. In this section, we briefly introduce the scenarios of interest and key challenges when implemented in real-world multicell cellular systems.

CoMP Algorithms and Scenarios

The scenarios and supported CoMP algorithms in Rel-11 3GPP LTE-Advanced are introduced in [6]. In the standard, both CBF and JP are considered as the main CoMP algorithms, and JP is further decomposed into joint transmission and dynamic point selection.

In coordinated scheduling/CBF (CS/CB), which contains both coordinated scheduling and CBF, the desired data stream is transmitted only from a transmitter in the CoMP cooperating set. The resource block is assigned to the receiver with CBF by scheduling of the serving cell after coordination among multiple coordinated cells. The transmit beamforming and receive combining vectors are generated to reduce the unnecessary interference to or from other receiver(s) scheduled within the coordinated cell. The best serving set of users will be selected so that the transceiver beamforming vectors are chosen to reduce the interference to/from other neighboring receivers, while increasing the serving cell signal strength. The other-cell interference can be mitigated, and the cell edge receiver throughput can be improved.

In joint transmission (JT), a data stream is simultaneously transmitted from multiple transmitters to a receiver in a given time-frequency resource. Unlike the CBF receiver, the receiver in joint processing tries to *coherently* combine the symbols from multiple transmitters; if the data streams from multiple transmitters are synchronized, the receiver can have a (coherent) combining gain, which additionally gives both SNR and diversity gains. To enable coherent combining at the receiver, a strict time-frequency synchronization should be enforced, and the phase difference from different transmitters needs to be fed back to the transmitter. In this sense, joint transmission may be more demanding and fragile than CS/CBF.

In dynamic point selection (DPS), data is transmitted from one transmitter within the cooperating transmitters in a time-frequency resource, and the designated transmitter is dynamically selected. The other transmitter(s) may mute so the receiver(s) do not see out-of-cell interference. The transmitting point may change in time depending on the feedback indicating the most desirable transmitting point by the receiver. Dynamic point selection also includes dynamic cell selection, which is a CoMP transmission scheme that can be seen as a natural extension of existing scheduler implementations. Unlike joint transmission, dynamic point selection does not need the interpoint information. Dynamic point selection supports CoMP scenarios 3 and 4 (explained later), where low-power remote radio heads (RRHs) are deployed as a heterogeneous network. In situations where the mobile station penetrates across the RRH transmission boundaries, dynamic point selection shows similar gain to joint transmission.

In Table 2, the differences of the three CoMP algorithms are presented. It is expected that joint transmission may give the highest throughput gain at the cost of additional feedback and higher use of inter-base-station communication. Dynamic point selection is simpler to implement but preferable only for the heterogeneous network scenarios in which the mobile station sees more cell edges so that received SINRs are similar to each other. Coordinated scheduling/CBF is probably the simplest and one of the most robust schemes with decent performance gain.

The four CoMP scenarios considered in the LTE-Advanced specifications and supported CoMP algorithms are presented in Table 3. Scenarios 1 and 2 are designed for classical homogeneous deployments, while scenarios 3 and 4 are for the heterogeneous deployment. Scenario 3 targets typical picocell heterogeneous deployments, and scenario 4 represents a distributed antenna system (DAS) where the points distributed within the macrocell can be thought of as RRHs. Scenario 4 differs from scenario 3 in the fact that all points belong to the same cell. The DAS approach using RRHs (scenario makes it easier to handle the out-of-cell interference and to apply joint processing than the picocell (scenario 3), although the orthogonal cell-specific pilot sequence design issue ought to be resolved. Note that in all scenarios, two receive antennas are the default receive antenna configuration; hence, aforementioned CoMP algorithms can be exploited.

Practical Issues in Receiver Processing

The main practical issues of CoMP include feedback channel design and downlink reference signal design as they determine the accuracy of the channel information at the transmitter and receiver, respectively. This receiver processing is also practically important because the receiver needs to learn about the channel with limited reference signals. Although the discussions are still in the consensus building stage, it is meaningful to learn about the working agreements made so far in the 3GPP LTE-Advanced standard.

The main practical issues of CoMP include feedback channel design and downlink reference signal design as they determine the accuracy of the channel information at the transmitter and receiver, respectively.

Feedback Channel Design — To enable CoMP operation, each receiver feeds back the necessary information to the transmitter via reliable feedback channel. The channel quality indicator (CQI) is the main feedback information. With CQI, along with other feedback information such as precoding matrix index and rank indicator, the transmitter may be able to learn about the necessary downlink and interference channel, and to schedule a transmission with an appropriately selected modulation and coding index. As the transmitter solely relies on the

Increased hypotheses for CQI generation per CoMP scenarios and ambiguity on the CQI metric with certain receiver algorithm may cause a serious problem both to the transmitter and the receiver.

	CS/CB	JT	DPS
Performance gain	Medium	High	Medium to high
Feedback overhead	Medium	High	High
Requirements on frequency/time sync	Low	High	High
Backhaul requirement	Medium	High	High
Relative phase information at the transmitter	Not required	Required	Not required
Relative amplitude information	Required	Required	Not required

Table 2. *A summary of the characteristics for each CoMP algorithm.*

Scenarios and descriptions	Equivalent system	CoMP algorithms
1. Homogeneous network with intrasite coordination	Single-cell coordination	CS/CB, JT
2. Homogeneous network with high-power RRHs	Multicell coordination	CS/CB, JT
3. Heterogeneous with low-power RRHs (different cell IDs)	Macro-pico voordination	CS/CB, JT, DPS
4. Heterogeneous with low-power RRHs (same cell ID)	Distributed sntenna dystem (DAS)	CS/CB, JT, DPS

Table 3. *CoMP scenarios and supported algorithms in 3GPP LTE-Advanced.*

information fed back from the receiver, the design of feedback channel is important in CoMP operation.

One of the main issues of the CoMP feedback design is CQI generation because the definitions of CQI vary in different CoMP algorithms. A CQI without considering interference is the maximum eigenvalue of the effective downlink channel, while a CQI with interference is the maximum eigenvalue of the effective channel multiplied by the MMSE-IRC whitening matrix. With the use of an ML detector, CQI computation can become more complicated. The CQI thus depends on the receiver algorithm as well as CoMP algorithms.

Increased hypotheses for CQI generation per CoMP scenarios and ambiguity on the CQI metric with a certain receiver algorithm may cause a serious problem to both the transmitter and the receiver. One typical issue is what interference should be taken into account in CQI computation; the interference from cooperating transmitters may be eliminated after proper coordination, so it may not be considered while computing the CQI. Interference from non-interfering transmitters should be considered as interference in the CQI computation metric. However, as the cooperating sets may change in time, the reported CQI may not represent the true link quality, and the transmitter may not detect this problem since *the CQI computation metric is not transparent to the transmitter*. This makes the feedback design at the receiver harder.

Downlink Reference Signal Design — The downlink reference signal is used for the receiver to estimate the multicell channel. The reference signal needs to be carefully designed to meet the performance requirement. If the density of the reference signal in the time-frequency grid is too high, significant spectral efficiency loss occurs. If the reference signals are too sparse, the multicell channel measurement can become very coarse. In 3GPP LTE-Advanced systems, a cell-specific reference signal with adequate reference signal density is considered for CoMP operation. In such a cell-specific reference signal, the scrambling pattern and cyclic shifting depends on the cell identification (ID) number.

In some CoMP scenarios, especially CoMP scenario 4 where the RRHs use the same cell IDs, the reference signal from multiple transmission points can easily collide. Even for the CoMP scenarios with different cell IDs, as the number of transmitting nodes gets higher, the reference signal collision probability gets higher. To provide additional orthogonality between reference signals from different transmitters, the 3GPP LTE-Advanced systems introduce *virtual* cell IDs. The reference signal sequence is initiated and scrambled, not in the cell-specific manner using the cell ID, but by the node-specific virtual ID. By doing this, additional orthogonality among reference signals can be obtained.

The second issue is the possibility of different granularity of channel estimation intervals in the serving and interfering cells. In general, CoMP operation can be done when the receiver does not move or moves very slowly, so the channel coherence time is strictly longer than the round-trip delay. The issue is that the interference channel does not change significantly in time, and its estimation need not be as accurate as that of the serving cell channel. In most cases, only the covariance of the interference can be utilized. The interference channel estimation period can be longer, so the receiver can save its power and devote more resources to serving cell channel estimation.

Receiver Architecture — In 3GPP LTE-Advanced CoMP, the default receive antenna number is set to two or four. This enables the receiver to apply an interference mitigation process. Specifically, the receiver algorithms considered in [6] are MMSE (mandatory) or MMSE-IRC (recommended). For the MMSE-IRC receivers, both biased and unbiased receivers are considered even though the throughput performance difference may be marginal regardless of the biasedness. The performance of the receiver depends on the accuracy of the covariance matrix estimation and frequency selectivity of the covariance matrix. Joint detection has yet to be considered in CoMP literature, but it can also be utilized if the modulation order of the interference signal is precisely known to the receiver.

System Performance Evaluation Results

To consider the practical issues discussed in the previous sections, we present the system performance evaluation results. We mostly use the simulation parameters presented in [6]. We use a 2×2 antenna configuration with dual polarized antennas and 6 ms feedback delay under full buffer traffic. For intercell interference modeling, we explicitly consider seven configurable intercell interference links to the mobile station, and the remaining links are modeled with Rayleigh distribution. Among the scenarios in Table 3, we simulate a homogeneous network (scenario 1) and a heterogeneous network with low power RRHs with difference cell IDs (scenario 3).

We first present a simulation result for rank statistics with different transmit modulation orders. The rank statistic is an indicator that a mobile station is capable of receiving more than one stream using multiple receive antennas. The benefit of using additional antennas at the receiver can be interpreted as the probability to choose rank 2 under a good channel condition. The simulation results with the use of the IW receiver is presented in Table 4. As shown in the table, the mobile station tends to choose more than one stream in a homogeneous scenario with high probability. The receiver processing with multiple receive antennas can benefit in such a good channel. For a heterogeneous network, however, the UE sees more diverse sources of interference. It tends to reject the interference using part of the receive antennas; hence, the chances that the mobile station is scheduled with more than one stream are limited. The simulation results show the benefit of using multiple receive antennas in both cases.

Simulation results with different transmitter and receiver techniques discussed in this article are also presented in Table 5. We compare the cell average throughput and cell edge throughput (5 percent user throughput) using the baseline IW receiver with JD and CoMP. By doing so, we show how to handle the way interference affects the system performance. In Table 5, we observe that the use of an advanced receive antenna and coordination technique improves the cell edge throughput the most. In a homogeneous scenario, the cell edge throughput is improved by 52 percent when CoMP with joint detection is applied. In a heterogeneous scenario, CoMP is the best technique to improve cell edge user throughput, while CoMP with JD is best for average cell throughput. This demonstrates that both multicell cooperative processing and the advanced receive technique are good methods to overcome other-cell interference in practical systems.

	Macro UE (homogeneous)	Macro UE (heterogeneous)	Pico UE (heterogeneous)
Rank 1	54.90%	92.4%	84.2%
Rank 2	45.10%	7.6%	15.8%

Table 4. *Rank statistics: homogeneous (left column) and heterogeneous (right two columns) scenarios.*

Fundamental Limits of Cooperation

Multicell cooperation is promising as it promises to convert the multicell network from interference-limited to noise-limited. Coordination between transmitters and receivers or the use of an advanced receive algorithm enables the transformation. The use of multiple receive antennas is helpful to alleviate other-cell interference. The natural question is whether such benefits are still fruitful in a large multicell network. When the cooperating transmitters form a cluster, a recent article [14] derives fundamental limits of cooperation in a general uplink setting. Based on their analysis, the spectral efficiency gets saturated as the transmit power increases in clustered large-scale multicell networks.

Surprisingly, in [14], it is shown that saturation of the spectral efficiency at sufficiently high transmit powers is unavoidable in large networks. A key insight used in that paper is that not all channels can be estimated due to time variation; thus, there is residual interference that limits the performance. Since the interference power is proportional to the transmit power, increasing the transmit power will not solve the problem. This results in the collapse of the multiplexing gain in the higher SNR region. The paper also claims that the fundamental limitations cannot be overcome through any other technological advances. Based on the insights from [14], the discrepancy in terms of throughput gains between theory [4, 5, 11] and practice [6] in multicell cooperative processing comes from this fundamental limit, mainly from the latency of the backhaul link or limited feedback of the practical cellular systems.

Reference [14] provides intuition on how large multicell networks are likely to perform. It would be interesting to study how multiple receive antennas impact the theory derived in [14], for example, to show if they can extend the region of multicell cooperation where capacity scales with SNR. The results in small-scale mul-

	Cell avg. throughput	Cell edge throughput	Cell avg. throughput	Cell edge throughput
IW	1.13 (Mb/s)	0.016	9.62	0.054
JD	1.27	0.023	9.89	0.062
CoMP	1.14	0.017	9.58	0.067
CoMP with JD	1.34	0.024	10.09	0.066

Table 5. *Cell throughput: homogeneous (left two columns) and heterogeneous (right two columns) scenarios.*

ticell cooperative systems [4, 5, 11] show that additional receive antennas help achieve a better spectral efficiency. An investigation of the tradeoffs between numbers of antennas at a single transmitter and coordination of many transmitters is an interesting topic for the purpose of any fundamental preference for centralized (e.g., massive MIMO) or distributed (e.g., distributed antenna systems) architectures. This may give some insights on the evolution of the multicell cooperative network for the next generation cellular standards and the role of multiple receive antennas therein.

Conclusions

This article introduces multicell cooperative systems with multiple receive antennas and advanced receiver techniques. In particular, coordinated beamforming, joint processing, and spatial sensing techniques are introduced, explaining their potential use of multiple receive antennas. Advanced receiver algorithms, which include MMSE, interference rejection combining (IRC), interference whitening, and joint detection in different interference statistics were also introduced. CoMP, as envisioned by emerging wireless standards, is reviewed. Finally, the fundamental limits of cooperation are discussed, and the potential role multiple receive antennas may play is reviewed. Multiple receive antennas and receiver processing have become more important in multicell cooperative networks, both theoretically and practically.

References

[1] D. Gesbert *et al.*, "Multicell MIMO Cooperative Networks: A New Look at Interference," *IEEE JSAC*, vol. 28, no. 9, Dec. 20, pp. 1380–4810.

[2] P. Marsch and G. P. Fettweis, *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*, Cambridge Univ. Press, 2011.

[3] H. Dahrouj and W. Yu, "Coordinated Beamforming for the Multicell Multi-Antenna Wireless Systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, May 2010, pp. 1748–95.

[4] C.-B. Chae *et al.*, "Interference-Aware Coordinated Beamforming in a Multi-Cell System," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, Oct. 2012, pp. 3692–3703.

[5] K. Lee *et al.*, "MIMO Transceiver Designs for Spatial Sensing in Cognitive Radio Networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, Nov. 2011, pp. 3570–76.

[6] 3GPP Long Term Evolution, "Coordinated Multi-Point Operation for LTE Physical Layer Aspects," tech. rep. TR36.819, 2011.

[7] J. Lee, D. Toumpakaris, and W. Yu, "Interference Mitigation via Joint Detection," *IEEE JSAC*, vol. 29, no. 6, June 2011, pp. 1172–84.

[8] T. Gou and S. A. Jafar, "Degrees of Freedom of the K User $M \times N$ MIMO Interference Channel," *IEEE Trans. Info. Theory*, vol. 56, no. 12, Dec. 2010, pp. 6040–57.

[9] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian Interference Channel Capacity to Within One Bit," *IEEE Trans. Info. Theory*, vol. 54, no. 12, Dec. 2008, pp. 5534–62.

[10] H. Weingarten, Y. Steinberg, and S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," *IEEE Trans. Info. Theory*, vol. 52, no. 9, Sept. 2006, pp. 3936–64.

[11] C.-B. Chae, S. Kim, and R. W. Heath, Jr., "Network Coordinated Beamforming for Cell-Boundary Users: Linear and Non-Linear Approaches," *IEEE J. Select. Topics in Sig. Proc.*, vol. 3, no. 6, Dec. 2009, pp. 1094–105.

[12] C. Seol and K. Cheun, "A Statistical Inter-Cell Interference Model for Downlink Cellular OFDMA Networks Under Log-Normal Shadowing and Multipath Rayleigh Fading," *IEEE Trans. Commun.*, vol. 57, no. 10, Oct. 2009, pp. 3069–77.

[13] Z. Bai *et al.*, "On the Equivalence of MMSE and IRC Receiver in MU-MIMO Systems," *IEEE Commun. Lett.*, vol. 15, no. 12, Dec. 2011, pp. 1288–90.

[14] A. Lozano, R. W. Heath, Jr., and J. G. Andrews, "Fundamental Limits of Cooperation," submitted for publication, Mar. 2012; http://arxiv.org/abs/1204.0011.

Biographies

INSOO HWANG (ihwang@utexas.edu) is a Ph. D candidate in the Electrical and Communications Engineering Department at the University of Texas, Austin (UTA)under the supervision of Prof. Robert W. Heath Jr. He received his M.S. in electrical engineering from the University of Southern California and his B.S. from POSTECH, Korea. He has more than five years of experience in wireless systems research at Qualcomm Research (San Diego), Samsung Electronics US R&D Center (San Diego), Telecommunications R&D Center (Korea), and Samsung Advanced Institute of Technology (Korea). He is currently working on next generation cellular systems and standards, focusing on investigation of the limits and challenges in large-scale heterogeneous and small cell networks.

CHAN-BYOUNG CHAE [SM'12] (cbchae@yonsei.ac.kr) is an assistant professor in the School of Integrated Technology, Yonsei University, Korea. Before joining Yonsei University, he was with Bell Labs, Alcatel-Lucent, Murray Hill, New Jersey, as a member of technical staff, and Harvard University, Cambridge, Massachusetts, as a post-doctoral research fellow. He received his Ph.D. degree in electrical and communications engineering from UTA. He was the recipient of the IEEE ComSoc AP Outstanding Young Researcher Award in 2012 and IEEE VTS Dan. E. Noble Fellowship Award in 2008. He currently serves as an Editor for *IEEE Transactions on Wireless Communications*, and his research interest includes resource management in energy-efficient networks and nano networks.

JUNGWON LEE (jungwon@stanfordalumni.org) is a senior director at Samsung Mobile Solutions Lab in San Diego, California. He received his Ph.D. degree in electrical engineering from Stanford University. He has co-authored more than 50 papers as well as numerous standards contributions for various standards bodies, and he holds over 60 U.S. patents. He currently serves as an Editor for *IEEE Communications Letters*, and his main research interests lie in wireless and wireline communication theory.

ROBERT W. HEATH, JR. [F'11] (rheath@ece.utexas.edu) is a professor in the Electrical and Communications Engineering Department at UTA and director of the Wireless Networking and Communications Group. He received his B.S. and M.S. in electrical engineering from the University of Virginia, and his Ph.D. in electrical engineering from Stanford University. He is the president and CEO of MIMO Wireless Inc. and chief innovation officer at Kuma Signals LLC. He is the recipient of the David and Doris Lybarger Endowed Faculty Fellowship in Engineering and is a registered Professional Engineer in Texas.

INTERFERENCE MANAGEMENT THROUGH COMP IN 3GPP LTE-ADVANCED NETWORKS

SHAOHUI SUN, PEKING UNIVERSITY AND STATE KEY LABORATORY OF WIRELESS MOBILE COMMUNICATIONS (CATT)
QIUBIN GAO, YING PENG, YINGMIN WANG, STATE KEY LABORATORY OF WIRELESS MOBILE COMMUNICATIONS (CATT)
LINGYANG SONG, PEKING UNIVERSITY

ABSTRACT

Intercell interference management has become a critical issue for future cellular mobile systems. Coordinated multipoint transmission/reception, or CoMP, is an effective way of managing intercell interference, and has been regarded as a key technology of LTE-Advanced. This article first provides an overview of downlink CoMP techniques specified in 3GPP LTE Rel-11, which mainly focuses on transmission schemes, channel state information reporting, interference measurement, and reference signal design. Then uplink CoMP is discussed in brief as most of the coordination gain can be achieved by implementation with little standardization support. Evaluation results are provided to show the efficiency of CoMP. The challenges as well as possible solutions for future CoMP standardization are also discussed.

INTRODUCTION

During the standardization process of Third Generation Partnership Project (3GPP) Long Term Evolution (LTE)-Advanced [1], a number of techniques have been proposed to boost the peak data rates, such as carrier aggregation and downlink multiple-input multiple-output (MIMO) up to eight layers. By taking advantages of these techniques, peak rates up to 100 Mb/s for high mobility and 1 Gb/s for low mobility are envisaged. However, intercell interference is preventing ubiquitous user experience of such high data rate (i.e., the data rate achieved by cell edge users is a small fraction of the peak data rate). Nonuniform signal-to-interference-plus-noise ratio (SINR) distribution of a typical homogeneous network with 500 m inter-site distance is depicted in Fig. 1a. Over 20 dB SINR is achieved in the cell center; however, it can only reach 5 dB or below in most areas due to intercell interference.

The heterogeneous network is an emerging network deployment. In a heterogeneous network the layer of a planned homogeneous macro network is overlaid by layers of low-power nodes (LPNs), such as picocells and remote radio heads (RRHs). The macrocell provides wide-area coverage, and LPNs provide coverage for hotspots and offload traffic from the macrocell. The gain of offloading traffic is limited due to imbalanced transmission power (i.e., LPN coverage is shrunk by the high-power macrocell). To extend the LPN coverage, an offset to the received power measurement is introduced in cell selection, allowing more user equipment (UE) to be attached to the LPN. The problem of this approach is that it dramatically increases the interference the macrocell imposes on UE attached to the LPN by coverage expansion. Figures 1b and 1c show SINR distribution of a heterogeneous network deployment without and with coverage expansion, respectively. Two LPNs are placed randomly in the coverage of each macrocell. It can be seen that the expanded region suffers very low SINR, which is depicted by deep color in the figure.

From information theory it is already known that intercell interference can be seen as an opportunity if cells process signals cooperatively. Coordination techniques can be classified as semi-static coordination and dynamic coordination according to the requirement of information sharing across cells. Semi-static coordination techniques, including fractional frequency reuse (FFR) [2], soft frequency reuse (SFR) [3], and enhanced intercell interference coordination (eICIC) [4], require little time-insensitive information exchange across cells. The basic idea of FFR is to create partitions between cell edge and cell center UE based on SINR. Resources of cell center UE in every cell are reused, , thus removing adjacent cell interference to cell edge UE. In the case of SFR, edge UE devices are restricted to those allocated orthogonal resources by coordination of neighboring cells as in FFR. Higher per-cell

 1536-1284/13/$25.00

resource utilization is achieved by reusing all resources in every cell but with a lower transmit power given to center UE to mitigate the increased interference. Both FFR and SFR are supported since LTE Release 8 (Rel-8).

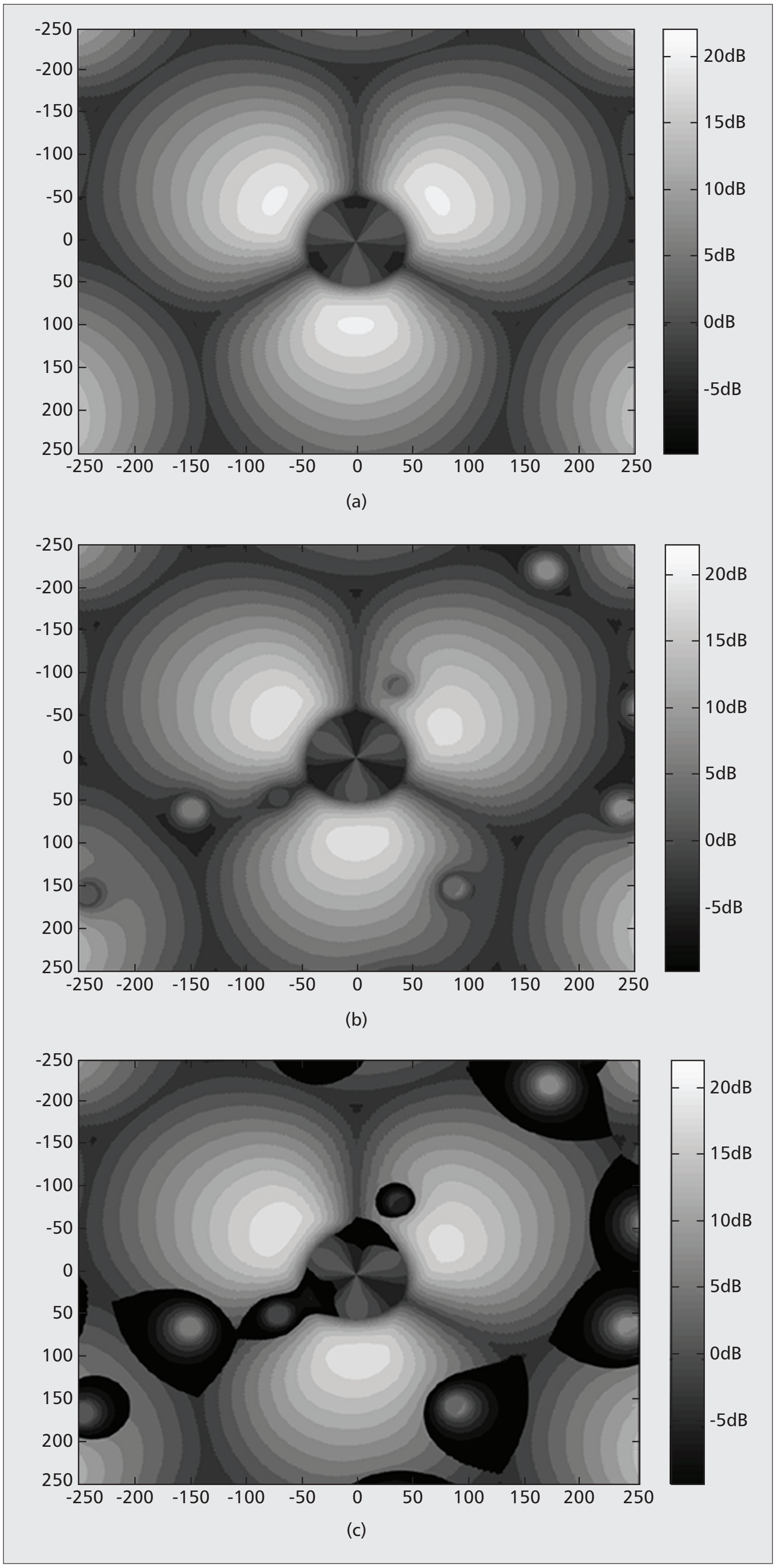


Figure 1. *SINR distribution: a) homogeneous macro network; b) heterogeneous network without coverage expansion; c) heterogeneous network with coverage expansion.*

In LTE Rel-10, the time-division multiplexing (TDM) eICIC principle was exploited to mitigate interference from macrocell to picocells in heterogeneous network. The macrocell transmits an almost blank subframe (ABS) or multimedia broadcast over single-frequency network (MBSFN) subframe with a certain periodicity, but the picocells transmit normal subframes all the time. The macrocell is not allowed to schedule any UE in an ABS or MBSFN subframe. The UE connected to picocells in the expanded region can be scheduled during transmission of an ABS or MBSFN from the macrocell, while UE close to picocells may be scheduled in any subframe.

Semi-static coordination techniques developed in LTE Rel-8/9/10 can provide benefits in managing intercell interference with relatively little backhaul overhead and low implementation complexity. However, the achievable benefit is limited as it is not able to follow the dynamic interference condition caused by dynamic scheduling and beamforming. Moreover, semistatically preserved resources result in low efficiency of resource utilization.

Dynamically coordinating the transmission and reception of signals at multiple cells could lead to further performance improvement. Existing dynamic multi-cell coordinating schemes include network MIMO [5], distributed MIMO [6], multicell MIMO [7], and so on. These schemes are conceptually the same, providing promising improvements of spectral efficiency and more uniform throughput distribution. They are unified into coordinated multipoint (CoMP) transmission/reception in 3GPP discussion as a key technology of LTE-Advanced. CoMP was initially brought forward in the study item of LTE-Advanced [1] to meet the requirements of IMT-Advanced [8] in 2008. However, there was no consensus on the gain of CoMP under realistic conditions, and it was not included in LTE Rel-10. Instead, a study item on CoMP was proposed to continue to study the performance benefits and specification impacts. After further study, performance gain was identified, and a work item was approved to specify CoMP in LTE Rel-11.

In the following sections, standardized CoMP in 3GPP LTE-Advanced is described. Downlink and uplink CoMP are introduced. Evaluation results are presented to show performance gains of CoMP. Conclusions are drawn, and future challenges are also discussed.

CoMP in 3GPP LTE-Advanced

Both downlink and uplink CoMP are standardized to support four identified CoMP scenarios. Downlink CoMP involves channel state information (CSI) reporting from UE, interference measurement, and reference signal design, which require much standardization work. Uplink CoMP, on the other hand, can be implemented with little standardization support. To optimize performance of uplink CoMP, uplink reference signal and uplink power control are enhanced in a CoMP-favorable manner.

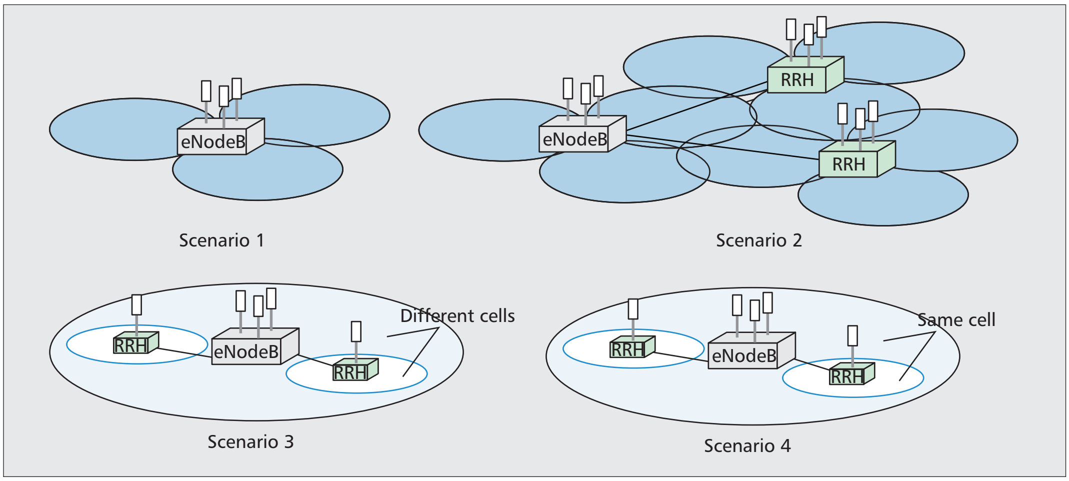


Figure 2. *CoMP scenarios in 3GPP LTE-Advanced.*

In the framework of 3GPP, downlink CoMP transmission is categorized as dynamic point selection (DPS), dynamic point blanking (DPB), joint transmission (JT), and coordinated scheduling/beamforming (CS/CB).

SCENARIOS

Four scenarios were defined in the 3GPP LTE-Advanced CoMP study [9]. The defined scenarios are deployed in numerical evaluations, and standardization is aimed at providing support for all four scenarios. It is worth noting that 3GPP standardization focuses on scenarios that are with backhaul of low latency and sufficient capacity in LTE Rel-11. The four defined scenarios are illustrated in Fig. 2.

Scenario 1: Homogeneous network with intra-site CoMP. The coordination takes place between three sectors of the same site in a homogeneous macro network. This is the most prevalent deployment scenario at the current stage, and CoMP can be applied in this scenario without any additional investment in network infrastructure.

Scenario 2: Homogeneous network with high-transmission-power RRHs. RRHs with equal transmission power to that of the eNodeB are employed to construct a homogeneous network. The coordination area consists of RRHs controlled by the same eNodeB. More CoMP gain could be achieved than with scenario 1 due to a possibly larger coordination area.

Scenario 3: Heterogeneous network with low-power RRHs within macrocell coverage. A number of low-power RRHs connected to an eNodeB are deployed in macrocell coverage. Each RRH forms a small cell with physical cell identity (PCI) independent from the macrocell. As explained in the introduction, interference imposed by a macrocell on UE associated with RRH severely degrades system performance. The interference problem could be resolved by applying CoMP in this scenario.

Scenario 4: Heterogeneous network with low-power RRHs within macrocell coverage where the transmission/reception points created by the RRHs have the same PCI as the macrocell. The main difference from scenario 3 is that no new cell is formed by RRH in this scenario, which results in distinct control channel design and mobility management. However, from aspect of data transmission, there is no difference between the two scenarios since the network topology is the same.

In scenario 1, the coordinated cells are collocated, and it does not suffer from the restrictions of underlying backhaul. In scenarios 2–4 the RRHs are connected to the eNodeB via optical fiber, and it is assumed that information is exchanged without latency and capacity constraints. For the sake of simplicity, a macrocell or an RRH is called a transmission point (TP) in downlink and receiving point (RP) in uplink.

DOWNLINK COMP

For downlink CoMP techniques, specification efforts mainly focus on transmission schemes, CSI reporting, interference measurement, reference signal design, and control signaling [9]. Transmission schemes are not mentioned explicitly in the specification; however, all the specification items are to support CoMP transmission schemes. CSI reporting facilitates link adaptation in the network, while interference measurement and reference signal are essential in deriving a CSI report. Besides common standardization support for frequency-division duplex (FDD) and time-division duplex (TDD) systems, special attention was paid to TDD to exploit channel reciprocity.

General Specification Support for FDD and TDD Transmission Schemes — In the framework of 3GPP, downlink CoMP transmission is categorized as dynamic point selection (DPS), dynamic point blanking (DPB), joint transmission (JT), and coordinated scheduling/beamforming (CS/CB). The four categories of transmission schemes are shown in Fig. 3.

With DPS, a TP is dynamically selected according to the instantaneous channel condition on per-subframe basis. The TP with the best link quality is selected to exploit channel variations opportunistically (i.e., selection diversity gain is achieved). In general, the gain of selection diversity in addition to the already achieved frequency and spatial diversity is marginal. To get further gain, DPB can be used in conjunction with DPS. By DPB, the dominant interferer(s) to UE in the coordination area is identified and muted dynamically (i.e., no signal is transmitted from the identified TP). By muting the dominant interferer, SINR of the UE may be enhanced

With CS/CB, the scheduling and beamforming across different TPs are aligned to reduce interference. Coordinating scheduling decisions (i.e., which specific UE is selected for transmission) can help alleviate strong interference conditions.

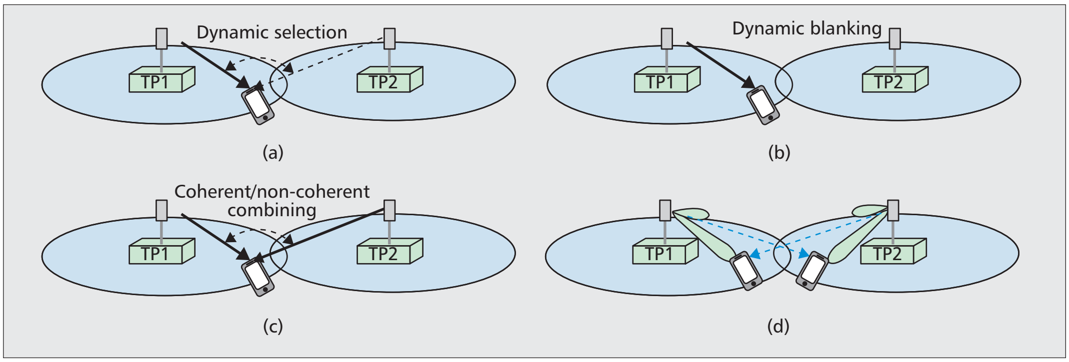


Figure 3. *Illustration of CoMP transmission schemes: a) dynamic point selection; b) dynamic point blanking; c) joint transmission; d) CS/CB.*

significantly as a few dominant interferers may contribute an absolute majority of the total interference. A TP is muted only when there is gain on the utility of the whole coordination area. Muting the dominant interferer causes performance loss of the particular TP. If performance improvements of those beneficiaries outweigh the loss, the TP can be muted; otherwise, it should remain active to transmit data. It seems unfair to UE connected to the muted TP; however, since scheduling is dynamically performed, those UE devices may also benefit from muting other TPs in subsequent subframes. On the whole, it is beneficial to all UE if the scheduling algorithm is designed carefully. DPS/DPB is especially useful in CoMP scenarios 3 and 4 as a muted macrocell benefits when multiple RRHs reside in the macrocell. It is very likely that the boosted RRH performance could outweigh the loss of the macrocell.

JT represents the case where two or more TPs are transmitting signal to a single UE device on the same time-frequency resource. JT is categorized into coherent and non-coherent JT, depending on whether coherent combining of signals from multiple TPs is targeted. Coherent JT offers the best performance among all considered CoMP schemes, since it allows multiple TPs to completely nullify interference toward co-scheduled UE. As evidenced in [10], coherent JT is especially sensitive to imperfections of CSI, so it would be premature to include it in the standard considering the current state of the art. Therefore, there is no explicit support for coherent JT in LTE Rel-11; nevertheless, it can be implemented in TDD by channel reciprocity. Non-coherent JT is not able to nullify interference across multiple TPs due to the lack of relative phase information across TPs. The SINR at UE could be boosted by non-coherent JT, but with area-splitting-gain loss, as multiple TPs could instead schedule multiple separate UE devices. Only UE devices located on the coverage edge of two TPs are able to get benefit from non-coherent JT, for the area-splitting-gain loss could be compensated by the boosted throughput in low SINR regions. The gain of non-coherent JT is more prominent in a network with low traffic load, as vacant resources of neighboring TPs could be borrowed to perform non-coherent JT.

With CS/CB, the scheduling and beamforming across different TPs are aligned to reduce interference. Coordinating scheduling decisions (i.e., which specific UE is selected for transmission) can help alleviate strong interference conditions. In addition, a wise selection of beamforming weights may further contribute to reduce interference. Ideally, scheduling and beam selection should be carried out jointly to optimize performance. However, it is impractical to make joint decisions on UE and beams across multiple TPs. A simple and prevalent way to implement CS/CB is an iterative scheduling procedure. Each cell revisits the choice of UE and underlying transmit beam(s) based on scheduling decisions and beams decided by other cells in the previous iteration. An updated scheduling decision not only accounts for the utility of the scheduled UE but also for the utility of the victim UE that has been tentatively scheduled by other cells in the previous iteration. The goal of each updating is to maximize a global utility (e.g., the total weighted throughput with the coordination area). Only a few numbers of iterations are needed to reach a near-optimal solution.

In practice, there is no clear boundary between transmission schemes; that is, combination of transmission schemes is possible. For example, an RRH is selected for data transmission for UE, and the scheduling and beams decisions are coordinated across multiple RRHs while the macrocell is muted. In LTE Rel-11, all of the above transmission schemes can be supported in a transparent manner from the specification perspective.

CSI Reporting — Channel state information at the transmitter is necessary to take advantage of multiple antenna techniques. LTE Rel-8 employs implicit feedback; that is, the reported CSI is a set of transmission parameters recommended by UE corresponding to a transmission hypothesis. The parameters are derived under the transmission hypothesis. CSI includes the precoding matrix indicator (PMI), rank indicator (RI), and channel quality indicator (CQI). The RI is the transmission rank or number of independent data streams that can be supported by the channel in spatial multiplexing transmission. The PMI is an index to a codeword in a predefined codebook comprising precoding matrices. The indexed codeword can be used as transmit beamforming weight for UE directly or used in deriving transmit beamforming weight. The CQI reflects the quality of the target link, which is a reference for modulation and coding scheme (MCS) selection.

A transmission hypothesis in CoMP comprises two parts: signal hypothesis and interference hypothesis. The signal hypothesis specifies TP(s) from which the packet data is assumed to be transmitted, and the interference hypothesis stands for interference suffered during the assumed data transmission. For DPS, CSI of transmission hypotheses with signal hypothesis from each candidate TP is needed at the network to facilitate dynamic selection of TPs. In addition, for DPB, CSI of transmission hypotheses with the same signal hypothesis but different interference hypotheses (e.g., whether an interfering TP is muted or not) are necessary for the network to make a decision on whether or not to mute the interfering TP.

To maximize flexibility in scheduling in the network, CSI of all possible transmission hypotheses should be reported by UE. However, the feedback overhead and UE implementation complexity is proportional to the number of hypotheses. To limit feedback overhead while retaining flexibility, Rel-11 UE can be configured to report CSI corresponding to one or more transmission hypotheses. CSI corresponding to one transmission hypothesis is defined as a *CSI process*. A CSI process is determined by the association of a signal hypothesis and an interference hypothesis, where the signal hypothesis and interference hypothesis are measured through CSI-RS and interference measurement resource (IMR), respectively. Note that there is no transmission hypothesis corresponding to JT and CS/CB in Rel-11 feedback, since that CSI can be calculated from the CSI of the DPS/DPB transmission hypothesis in the network without significant performance loss.

Assuming there are two TPs involved in CSI reporting, the possible CSI processes are illustrated in Table 1. In each transmission hypothesis, the signal can be transmitted from either TP1 (TP1 on, TP2 off) or TP2 (TP2 on, TP1 off), and the interference from the interfering TP can be either muted (off) or not (on). There are four possible CSI processes in the considered example. For CSI_1 in Table 1, the signal from TP1 is "on" and the signal from TP2 is "off," which means that the signal is transmitted from TP1 only. Regarding the interference hypothesis, the interference from TP1 is "off" and that from TP2 is "on," which means that the assumed transmission experiences interference from TP2. The corresponding transmission hypothesis is DPS from TP1. Unlike CSI_1, interference from TP2 is "off" for CSI_2, and thus CSI_2 can be used directly for DPB transmission from TP1 while muting TP2. CSI with JT and CS/CB transmission hypotheses are not supported in LTE Rel-11, and the required CSI could be inferred from CSI_2 and CSI_4. It is worth noting that by the association of signal hypothesis and interference hypothesis, CSI processes besides the listed ones could also be configured.

CSI process	Transmission hypothesis			
	Signal hypothesis		Interference hypothesis	
	TP1	TP2	TP1	TP2
CSI_1	on	off	off	on
CSI_2	on	off	off	off
CSI_3	off	on	on	off
CSI_4	off	on	off	off

Table 1. *Example CSI process configuration for two TPs.*

Interference Measurement — To derive the CSI of each CSI process, appropriate interference corresponding to the interference hypothesis should be measured by UE. In addition, accuracy is another requirement since CoMP performance is sensitive to accuracy of interference measurement. To facilitate CSI reporting, LTE Rel-11 introduces a UE-specific interference measurement resource (IMR) for UE to measure interference. Each CSI process is linked with a configured IMR, and the interference hypothesis of the CSI process is derived on the IMR.

Each IMR occupies a subset of resource elements (REs) that are muted intentionally on certain TPs. No signal is transmitted from those TPs on IMR occupied REs. UE simply receives the signal on those REs to get interference estimation. The network should coordinate the signal transmission on configured IMR for TPs in the coordination area to ensure that the signal on the IMR matches the interference hypothesis. To elaborate more clearly, Fig. 4 depicts several IMR configurations. On IMR1, both TP1 and TP2 are muted; the received signal on those REs represents interference from TPs other than TP1 and TP2. Similarly, UE could estimate interference out of {TP1} and {TP2} on IMR2 and IMR3, respectively.

Reference Signal Design — The design philosophy of LTE-Advanced reference signal (RS) is to separate RS for CSI reporting (CSI-RS) and RS for demodulation (DM-RS). For CSI reporting, RS sparse in time and frequency domain is sufficient. The sparsity significantly reduces system overhead. For demodulation, higher RS density is necessary to meet demodulation requirements. DM-RS only needs to be transmitted in the resources scheduled for data transmission. It is transmitted in the same way as data, and thus the transmitted data can be demodulated by referring to DM-RS without knowing the actual CoMP transmission scheme. This gives the network sufficient flexibility in choosing a CoMP transmission scheme to optimize performance.

A CSI-RS-resource refers to some predefined time/frequency/code resources used for transmission of CSI-RS from a certain TP. To support CoMP transmission, multiple CSI-RS-resources need to be configured for UE. The configured set of CSI-RS-resources is defined as a CoMP

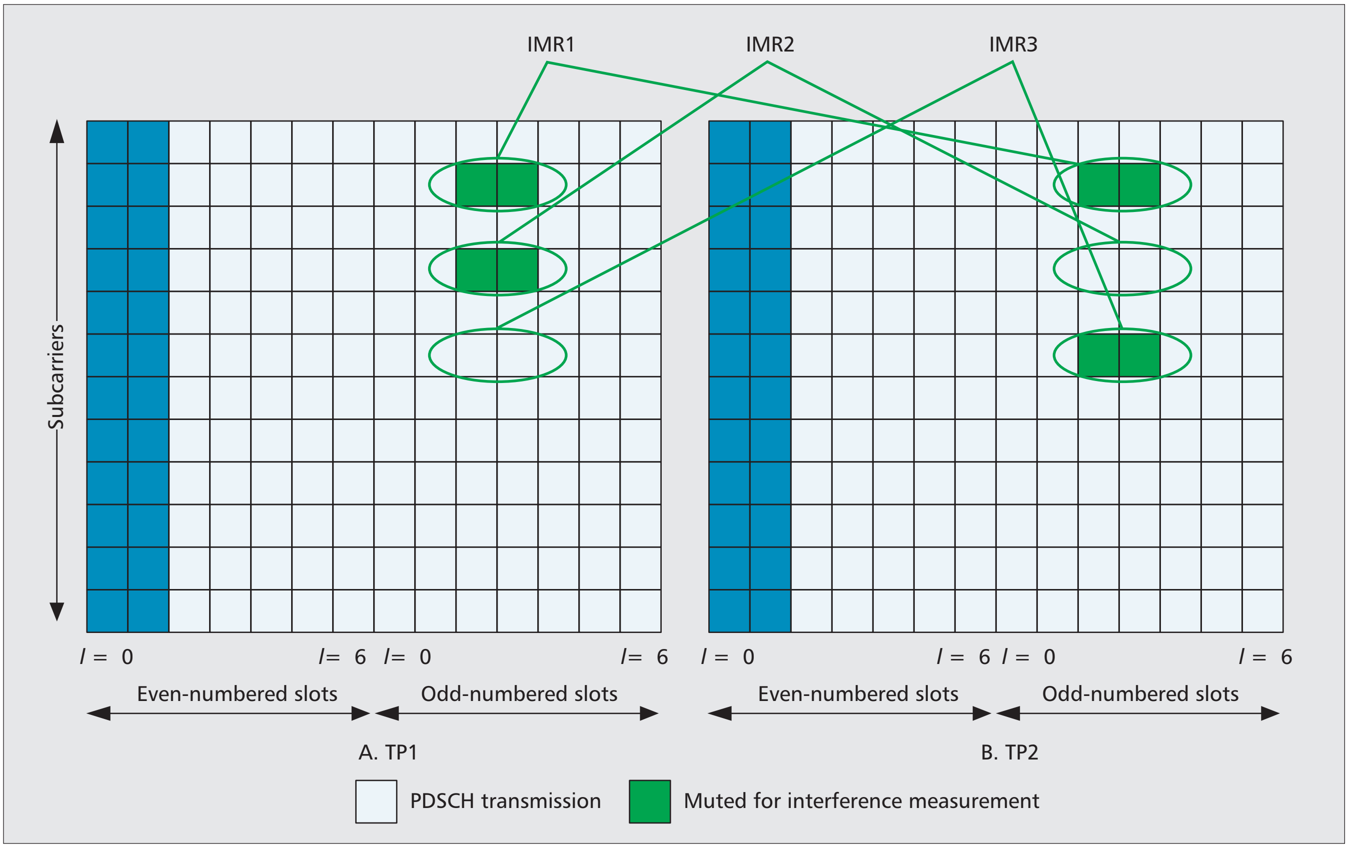


Figure 4. *Interference measurement resource (IMR) configuration examples.*

measurement set. Each CSI-RS-resource in a CoMP measurement set is used to track the channel of one TP. A CSI process is associated with a CSI-RS-resource in the CoMP measurement set to derive the signal hypothesis of the CSI process.

Special Considerations for TDD — In a TDD system, channel reciprocity can be exploited to reduce feedback overhead and improve CSI reporting accuracy. CSI obtained by channel reciprocity is free of quantization and feedback compression errors, and thus more accurate than that reported by UE. CQI reporting from UE is still needed to deliver information about downlink interference, which is different from interference measured on an uplink channel. In principle, reported CQI along with CSI by channel reciprocity enables all kinds of transmission schemes. Even with CQI reporting, the feedback overhead is still greatly reduced by the absence of PMI/RI reporting.

Although coherent JT is not supported explicitly in Rel-11, it can be implemented in a TDD system to achieve significant performance gain. A prerequisite of coherent JT is that channel reciprocity holds over multiple TPs. Antenna calibration across multiple TPs should be carried out to ensure overall channel reciprocity.

Channel-reciprocity-based schemes rely heavily on the accurate reception of an uplink sounding reference signal (SRS). Coordination of SRS transmission among multiple TPs is beneficial in improving channel estimation accuracy. By such coordination, the SRS transmission in one cell (TP) is orthogonal to other cells in the coordination area, reducing interference imposed on SRS transmission.

The targets of SRS transmission may be located at separate positions, and the distances from UE to those targets are different. The transmission power should be set to guarantee that the furthest TP can receive the SRS correctly. Enhanced transmission power is also beneficial to uplink CoMP in deriving uplink CSI at multiple RPs.

Uplink CoMP

For uplink CoMP, uplink CSI is available in the network without resource-consuming transmission on uplink for CSI reporting. Terminals need almost no modification to support uplink CoMP transmission. It is easier to implement CoMP in uplink than in downlink from the UE perspective. Cooperation schemes are implementation-specific in the network, except that possible backhaul transmission needs to be standardized in future. To optimize uplink CoMP, uplink reference signal and uplink power control are enhanced in a CoMP-favorable manner.

Uplink CoMP includes joint reception (JR) of the received signal at multiple RPs and/or coordinated scheduling (CS) decisions among RPs to control interference and improve coverage.

Joint reception (JR): The basic concept behind JR is to utilize antennas at different sites. The signals received by the RPs are jointly processed to produce the final output. Received data needs to be transferred between the RPs for them to operate. The amount of exchanged information between cooperating RPs depends on to what extent the received signals are pre-processed before information exchange. Quantized baseband samples convey the most complete information for joint processing. It achieves the best performance, while the load on backhaul is also the highest. Further processing at RPs can lower the load of backhaul, but with less CoMP gain. There is clearly a trade-off between CoMP gain and backhaul overhead.

Coordinated scheduling (CS): UE scheduling and precoding selection decisions are aligned among RPs in a coordination area to minimize interference. A much reduced load is achieved in the backhaul network because only CSI and resource allocation information needs to be shared among cooperating RPs.

In a heterogeneous network, the transmission power of various nodes may differ dramatically. For example, transmission power of a macro node may be 10~20 dB higher than that of a lower-power RRH. As a result, the cell boundary between a macro node and an RRH node (similar received signal level between nodes) is actually very close to the RRH node. It is very likely that cell edge UE connected to a macro node is closer to an RRH. In this case, the RRH received signal strength of the UE uplink transmission is stronger than that of the macro node. That is, the uplink transmission of such UE can target the RRH instead of the macro node. The transmission power can thus be lowered to reduce interference and extend battery life. The transmission of uplink and downlink are decoupled in this scenario. Note that this can be viewed as a special type of joint reception.

It is worth noting that uplink CoMP cooperation schemes are not limited to those mentioned in this article. Network is free to choose any schemes that benefit uplink transmission.

Evaluations

Extensive evaluation work has been carried out during the study phase of CoMP in 3GPP [9]. By the collected results in [9], performance benefits of both downlink and uplink CoMP are identified. In this section, evaluation results are given to verify the performance of downlink CoMP according to the latest progress in 3GPP. The simulation is implemented via Microsoft Visual C++ 2008.

Evaluation Results for Scenarios 1 and 2

Coherent JT is evaluated in scenarios 1 and 2. CSI is obtained by channel reciprocity in the network. CQI for each cell is reported by UE. Each cell first performs single cell scheduling to select at most two UE devices without considering intercell coordination. The transmit beamforming of each selected UE is then jointly calculated to suppress interference to co-scheduled UE in the coordination area. Figure 5a shows that coherent JT provides gain for cell center UE as well as cell edge UE because spatial reuse is allowed by multiple-user transmission. If scheduling decisions are made jointly over the coordination area, further gain could be expected. Other detailed evaluation assumptions can be found in [9].

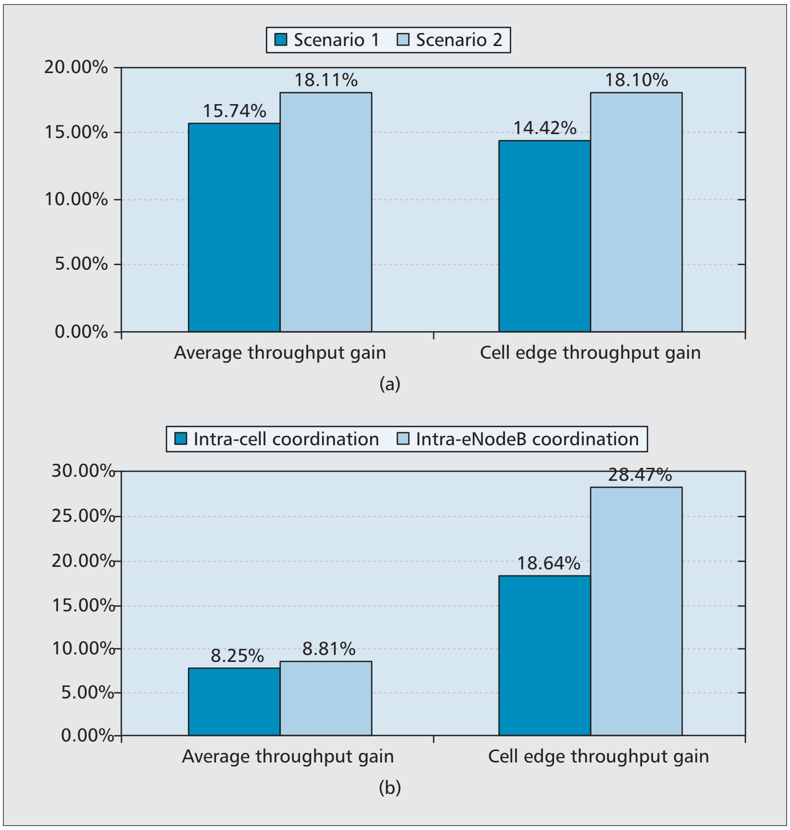


Figure 5. *Evaluation results of CoMP schemes: a) scenarios 1 and 2; b) scenarios 3 and 4.*

Evaluation Results for Scenarios 3 and 4

The combined scheme of DPS/DPB and non-coherent JT is evaluated in scenarios 3 and 4. TPs within a CoMP coordination area are jointly scheduled, and each CoMP coordination area is scheduled independently. Within each coordination area, exhaustive search is utilized to schedule the UE group and the corresponding transmission schemes (DPS, DPB, and non-coherent JT) that achieve the highest total weighted estimated throughput. UE reports all of the four CSI processes (CSI_1–CSI_4) to the network as shown in Table 1. For DPS and DPB, the reported CSI is used directly, while for non-coherent JT, the network figures out required CSI from reported CSI. Both intra-sector and intra-eNodeB cooperation are evaluated. The coordination area includes one macrocell and four RRHs within the macrocell (sector) in intra-sector cooperation. For intra-eNodeB cooperation, coordination is carried out between three collocated macrocells (sectors) and 12 RRHs connected to those macrocells (sectors). From a data transmission perspective, scenarios 3 and 4 are identical; thus, only one set of results is given. As seen in Fig. 5b, DPS/DPB and non-coherent JT provide throughput gain mainly for UE with median to low rates. Furthermore, the larger coordination area provides additional gain. Other detailed evaluation assumptions are aligned with [9].

DPS/DPB and non-coherent JT provide throughput gain mainly for UE with median to low rates. Furthermore, the larger coordination area provides additional gain.

Conclusions and Future Work

This article has presented an overview of the standardization work on CoMP in 3GPP LTE-Advanced. Due to the limited timeframe of LTE Rel-11 standardization, only basic functions of CoMP are supported. The CoMP technique is far from mature, and many problems remain to be solved.

1) Backhaul-efficient CoMP. Standardized CoMP in 3GPP assumes that an ideal backhaul connection exists. However, the available backhaul types are diverse in practice, such as T1/E1 and microwave. On these practical backhaul types, assumption of infinite capacity and zero latency are no longer valid. Smart and backhaul-efficient cooperation techniques are needed.

2) Efficient CSI feedback. CoMP performance gain, especially coherent JT, relies heavily on the accuracy of CSI in the network. In practical systems, the accuracy of CSI in the network is limited by the capacity of uplink transmission. How to design efficient feedback over the limited uplink to support aggressive CoMP schemes such as coherent JT poses big challenges on research and standardization.

3) Multiple-site antenna calibration for TDD. As seen from evaluation results, coherent JT exploiting channel reciprocity provides significant gain for TDD. However, channel reciprocity does not hold strictly. To ensure channel reciprocity and hence maximize the capability of coherent JT in TDD, antennas across multiple TPs need to be jointly calibrated. Self-calibration is difficult to implement, especially for scenarios 2–4. Over-the-air (OTA) calibration seems necessary in these scenarios. With the constraint of feedback overhead, the content of information reported by UE to network and the calibration algorithm need further study.

Acknowledgment

The authors acknowledge the support from the National Science and Technology Major Project of China under Grant 2011ZX03003-001-01 and the National Basic Research Program of China under Grant 2012CB724103.

References

[1] 3GPP TR 36.814 v9.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Further Advancements for E-UTRA Physical Layer Aspects (Release 9)," TSG RAN.

[2] G. Boudreau *et al.*, "Interference Coordination and Cancellation for 4G Networks," *IEEE Commun. Mag.*, vol. 47, no. 4, Apr. 2009, pp. 74–81.

[3] J. Li, N. Shroff, and E. Chong, "A Reduced-Power Channel Reuse Scheme for Wireless Packet Cellular Networks," *IEEE/ACM Trans. Net.*, vol. 7, no. 6, Dec. 1999, pp. 818–32.

[4] 3GPP R2-106897, "Introduction of Enhanced ICIC," 3GPP TSG RAN WG2 meeting #72.

[5] H. Huang *et al.*, "Increasing Downlink Cellular Throughput with Limited Network MIMO Coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, June 2009, pp. 2983–89.

[6] A. Saleh, A. Rustako, and R. Roman, "Distributed Antennas for Indoor Radio Communication," *IEEE Trans. Commun.*, vol. 35, no. 12, Dec. 1987, pp. 1245–51.

[7] D. Gesbert *et al.*, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," *IEEE JSAC*, vol. 28, no. 9, Dec. 2010, pp. 1380–408.

[8] ITU-R Rep. M.2134, "Requirements Related to Technical Performance for IMT-Advanced Radio Interface(s)," 2008.

[9] 3GPP TR 36.819 v11.1.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Coordinated Multi-Point Operation for LTE Physical Layer Aspects (Release 11)," TSG RAN.

[10] S. Annapureddy *et al.*, "Coordinated Joint Transmission in WWAN," *IEEE Commun. Theory Wksp.*, May 2010.

Biographies

Shaohui Sun (sunshaohui@catt.cn) received a B.S. in auto control engineering and an M.S. in computer engineering from Xidian University in 1994 and 1999, respectively, and a Ph.D. in communication and information systems from Xidian University in 2003. He has been deeply involved in the development and standardization of LTE/LTE-Advanced since 2005. His research area of interest includes multiple-antenna technology, heterogeneous wireless networks, and relay.

Qiubin Gao (gaoqiubin@catt.cn) received his B.S. and Ph.D. in control science and engineering from Tsinghua University, China. He is currently a senior research engineer at the Datang Wireless Mobile Innovation Center of the China Academy of Telecommunication Technology (CATT). His current research interests include physical layer design for mobile communication, multiple-antenna technology, CoMP, and system performance evaluation. He is inventor/co-inventor of more than 100 patents in wireless communications, and author/coauthor of a number of journal and conference papers.

Ying Peng (pengying@catt.cn) received her M.Sc. and Ph.D. in electrical and electronic engineering from the University of Bristol, United Kingdom. She is presently a senior standardization researcher in the Wireless Mobile Innovation Center of CATT. Her research interests include 3GPP/ITU standardization, LTE/LTE-A physical layer design, CoMP, heterogeneous networks, and cognitive radio. She received First Prize in "4G TD-LTE-Advanced" of the Science and Technology Prize from CCSA in 2011.

Yingmin Wang [M] (wangyingmin@catt.cn) received his Ph.D. degree in signal and information systems from Xidian University, Xi'an, China. He joined R&D of TD-SCDMA systems at the Datang Mobile Communications Equipment Co., Ltd., Beijing, China, in 2000. He has served as chief scientist since 2010 in the Datang Telecom Technology and Industry Group. He is currently leading the Datang Wireless Mobile Innovation Center. His current research interests include digital communication signal processing and wireless networks. He has applied for more than 150 patents in wireless communications and published a number of papers.

Lingyang Song [SM] (lingyang.song@pku.edu.cn) received his Ph.D. from the University of York, United Kingdom, in 2007, where he received the K. M. Stott Prize for excellent research. He worked as a postdoctoral research fellow at the University of Oslo, Norway, and Harvard University, Cambridge, Massachusetts, until rejoining Philips Research UK in March 2008. In May 2009, he joined the School of Electronics Engineering and Computer Science, Peking University, China, as a full professor. He is a co-inventor of a number of patents (standards contributions), and author or co-author of over 100 journal and conference papers. He has received best paper awards at three international conferences.

HOW DO WE DESIGN COMP TO ACHIEVE ITS PROMISED POTENTIAL?

CHENYANG YANG, SHENGQIAN HAN, XUEYING HOU, BEIHANG UNIVERSITY
ANDREAS F. MOLISCH, UNIVERSITY OF SOUTHERN CALIFORNIA

ABSTRACT

Coordinated multipoint, or CoMP, transmission has been recognized as a spectrally efficient technique for full frequency reuse cellular systems, in which base stations cooperate to reduce or eliminate intercell interference. However, there are still many obstacles before it can be put into practical use. In this article, we first discuss the features of CoMP systems and channels that are distinct from single-cell multi-antenna systems. We then give an overview of state-of-the-art approaches for coping with the factors that limit the potential of CoMP. A major issue is the acquisition of channel state information, which creates different challenges for TDD and FDD systems. Another set of challenges arises from the limited capacity available on the backhaul connections between the cooperating base stations. Both the fundamentals of possible solutions and their relations to cellular standards are discussed.

INTRODUCTION

Multiple-input multiple-output (MIMO) transmission can greatly increase capacity when the signal-to-interference-plus-noise ratio (SINR) is high. In many practical systems (e.g., full frequency reuse cellular systems), however, SINR is low, especially near the cell edge. In these scenarios, increasing the number of antenna elements might not yield significant performance improvement. For this reason, spectral efficiency of cellular systems with MIMO is far below the promised value.

Consider as a starting point a single-cell MIMO system. For ease of exposition, from here on we assume that each mobile station (MS) has a single antenna, while the base station (BS) has multiple antenna elements. In the downlink, the BS can now form a beam, such that the desired signal is concentrated at the location of an MS, while minimizing (interfering) energy to the MSs that do not want this signal. We note that beamforming is usually done based on the instantaneous channel realizations, not just on the average directional characteristics of the channel. The BS can form multiple beams simultaneously, and thus provide multiple data streams to the MSs at the same time. This is the principle of multi-user MIMO (MU-MIMO). However, MU-MIMO does not, by itself, alleviate the problem of intercell interference (ICI). Without coordination among different cells, the beams formed by one BS might "point" toward an MS in a neighboring cell, and thus provide concentrated interference energy to an MS of interest (Fig. 1a).

To avoid the ICI, multiple BSs can be connected via backhaul links so that they can collaborate and thus act as a single huge MIMO system. This effectively eliminates ICI and furthermore enhances the desired signal (similar to macro diversity in soft handover), thus ensuring high SINR even at the cell edge. This concept was (to our knowledge) first elaborated in [1] and has since then attracted broad interest in the cellular industry, because it can improve both cell average and cell edge throughput. Currently, it is mostly known as coordinated multipoint (CoMP) transmission, although BS cooperation and network MIMO [2] are often used synonymously. In 2008, a study item was started for CoMP in Third Generation Partnership Project (3GPP) Long Term Evolution (LTE)-Advanced; it is also a part of Advanced WiMAX.We only consider *inter-BS* CoMP, where multiple geographically separated BSs cooperate. While *intra-BS* CoMP (which exploits coordination between sectors covered by the same BS) has many aspects in common, its particular features are beyond the scope of this article.

CoMP is generally divided into two categories: CoMP joint processing (CoMP-JP) and CoMP coordinated beamforming (CoMP-CB), depending on what kinds of information are shared among BSs. An illustration of CoMP-JP and CoMP-CB is shown in Figs. 1b and 1c.

For CoMP-JP, both data and channel state information (CSI) need to be shared via backhaul links between each BS and a central unit (CU). The CU could be a separate entity, possibly collocated with one of the BSs. Alternatively, each BS can serve as a CU — a concept called decentralized cooperation in the literature. This approach allows the cooperating BSs to behave like a single large multi-antenna BS with distributed antenna elements. The distributed array forms beams toward all the users

 1536-1284/13/$25.00

in its coverage area (i.e., the cells covered by all the cooperating BSs) simultaneously, employing all available BS antennas for each beam. Consequently, ICI is turned into desired signals. On the downside, this approach places high demands on the backhaul links and requires signal-level synchronization as well as data-level synchronization among BSs.

In contrast, CoMP-CB only needs the coordinated BSs to share CSI and the scheduling information. CoMP-CB, also called spatial ICI coordination (ICIC), retains the concept of cells: in each cell, the BS forms beams toward the users in such a way that it not only increases the desired signal strength towards the desired user in its own cell, but also reduces interference towards the users in the adjacent cells. In this approach, each BS needs only the data for the users in its own cell, as well as the CSI and scheduling information of the adjacent cells. This significantly reduces (compared to CoMP-JP) the demands for backhaul links because sharing CSI among BSs needs much lower capacity than sharing data [3]. However, CoMP-CB only passively avoids ICI rather than proactively exploiting it.

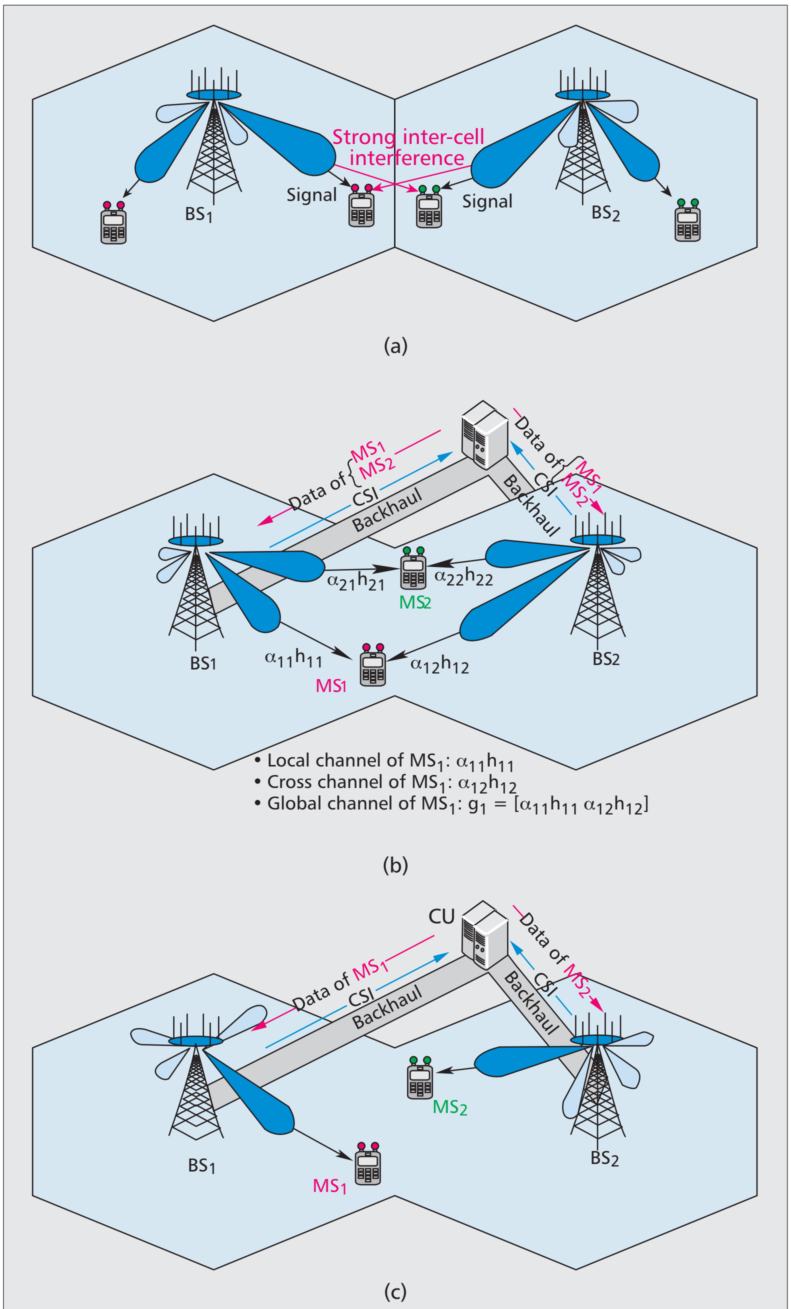


Figure 1. *Illustration of non-CoMP, CoMP-JP and CoMP-CB. For the CoMP systems, BS_1 is the master BS of MS_1, from which the average channel gain is stronger: a) non-CoMP: no coordination between BSs, and strong interference might be caused to adjacent cell edge users; b) CoMP-JP: BSs cooperate by jointly serving multiple users in their covered area; c) CoMP-CB: BSs cooperate by avoiding interference to adjacent cell edge users.*

While CoMP is useful — and challenging — both for the uplink and downlink of cellular systems, we focus in this article on the downlink, due to the fact that limitations of downlink transmission speed are currently considered the more important bottleneck of cellular communications. Since the number of antennas available at the (coordinated) large BS is much larger than that on the (individual) MSs, CoMP-JP should be used in conjunction with MU-MIMO transmission to fully exploit the abundant spatial resources provided by the BS cooperative system.

Since the above description pointed out great similarities of downlink CoMP-JP to single-cell MU-MIMO, one might wonder why the manifold and well explored techniques for implementing MU-MIMO cannot be applied in a straightforward manner. As a matter of fact, there are three important obstacles to such a simplistic approach:

- Single-cell MIMO differs in some subtle but important aspects from a true CoMP-JP setup, thus requiring changes in the transmission strategies.
- The acquisition of CSI is more difficult in CoMP systems.
- There are restrictions on the sharing of information between the cooperating BSs due to the limitations of the backhaul network.

In the following three sections, we deal with those issues one by one.

Single-Cell MIMO vs. CoMP-JP

CoMP-JP differs from single-cell MIMO even if the BSs are synchronized and connected with perfect backhaul, which comes from the distributed BSs and practical limitations.

PBPC

It has been widely recognized that CoMP is subject to a per-BS power constraint (PBPC), while most of the optimization of beamforming, scheduling, and power allocation for single-cell MIMO is subject to a sum power constraint (SPC). Since power cannot be shared among BSs, directly applying the transmit strategies optimized under SPC will lead to optimistic results, especially for heterogeneous networks including macro, micro, and pico BSs with very different transmit powers. In homogeneous networks where multiple BSs have the same transmit power, the transmission schemes designed for maximizing the sum rate under PBPC perform close to those under the SPC.

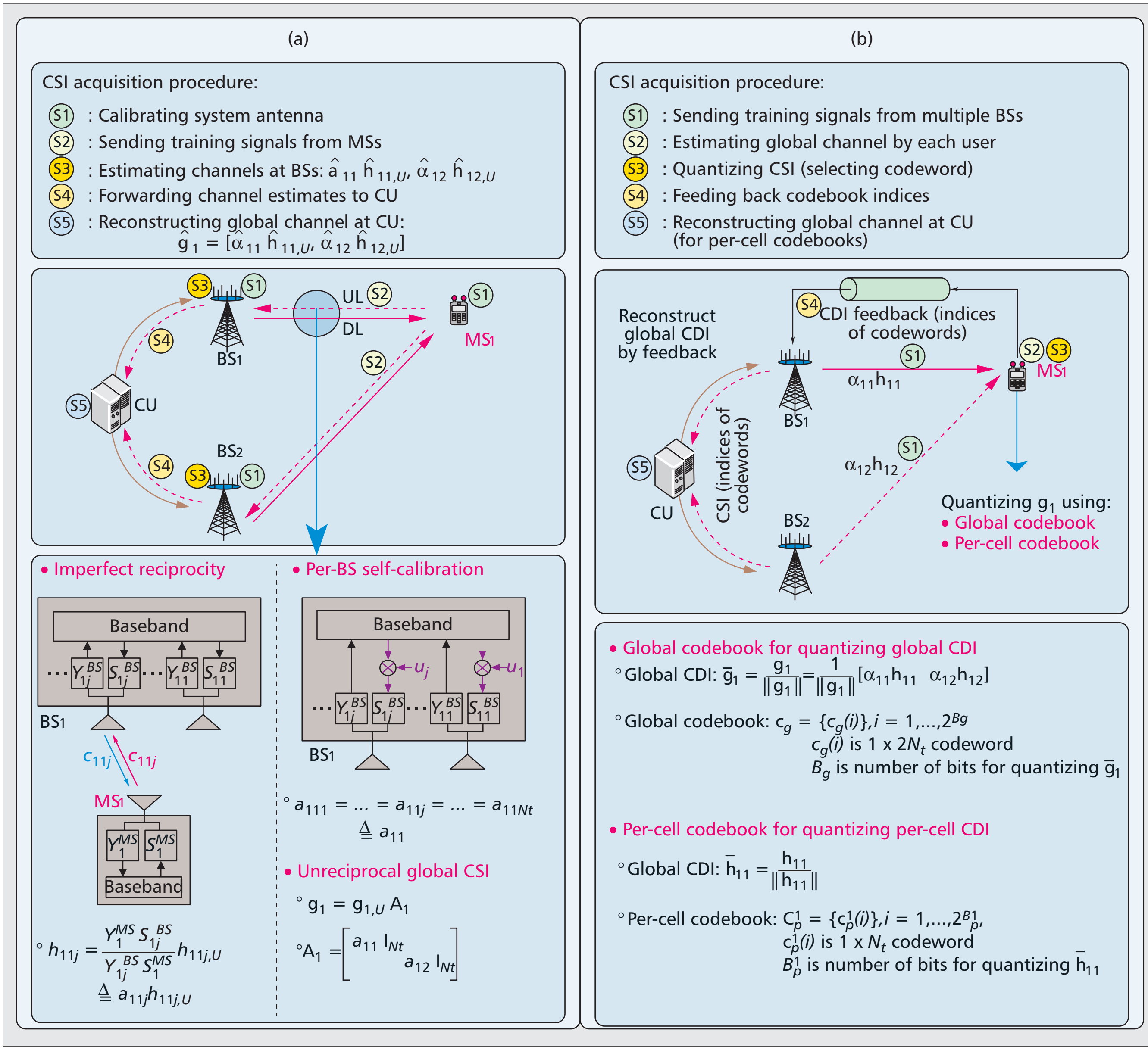


Figure 2. *CSI acquisition procedures in a) TDD and b) FDD CoMP systems, where the specific steps from S1 to S5 are respectively shown in each procedure.*

Asynchronous Interference

Since BSs are not collocated, interference from different BSs received at each MS are asynchronous [4]. This cannot be compensated by the timing-advance technique conventionally applied in cellular systems, because the degree of freedom of the timing advance is used up by ensuring that the signals from the BSs arrive synchronously at a *desired* MS. For orthogonal frequency-division multiple access (OFDMA) systems, this problem can simply be solved by prolonging the cyclic prefix. Considering that cell size is continually reduced and CoMP is more desirable for high-density networks, even such a prolonged cyclic prefix does not need to be very long.

Dynamic Clustering

Due to the prohibitive complexity and overhead, it is not possible to allow all BSs (which might span a whole city) to cooperate. Moreover, a CoMP system with many BSs in a network exhibits negligible performance gain compared to one where only a few BSs are cooperating [5, Chapt. 7]. As a pragmatic trade-off, cooperative clusters can be formed, within which several adjacent BSs jointly transmit. Dynamic clustering outperforms fixed clustering, but it leads to a dynamic overall number of transmit antennas. Considering the flexibility and scalability, channel training and feedback mechanisms need to be redesigned for such a cooperative network.

Non-i.i.d. Global Channels

The global channel for each MS in a CoMP-JP system is a stacking of multiple single-cell channel vectors, as shown in Fig. 1b. The distributed antennas yield a special non-independent and identically distributed (i.i.d.) global channel, where the average channel energies of the links between multiple BSs and each MS differ. As a result, the statistics of the global channel depend on the MS location (i.e., path loss and shadowing). An ongoing research goal is to investigate how such non-i.i.d. channel statistics and the stacked channel structure can be exploited to reduce the overhead to gather CSI for downlink multiuser precoding and scheduling.

Challenges Related to Channel Information Acquisition

CSI at the transmitter (CSIT) is essential for all kinds of CoMP transmission to obtain their full benefits. It is well known that the performance gain of MU-MIMO is largely dependent on the CSIT quality, and the same holds true for CoMP. To facilitate downlink spatial precoding and scheduling, the CU needs to gather CSIT from all coordinated BSs to all MSs in their serving cells. In time-division duplexing (TDD) systems, the CSI is estimated at each BS by uplink training via exploiting channel reciprocity. In frequency-division duplexing (FDD) systems, since uplink and downlink operate in different frequency bands, the CSI is estimated at each MS by downlink training, and then is fed back to the MS's master BS via uplink channels. The channel acquisition procedures in TDD and FDD systems are illustrated in Fig. 2.

In this section, we assume perfect backhaul, that is, the backhaul links connecting the CU and multiple BSs within each cluster have unlimited capacity and zero latency. With such an assumption, CoMP-JP could realize the full potential of BS cooperation transmission if the global channels of multiple MSs were perfectly available at the CU. We focus on various issues associated with obtaining the global CSI in TDD and FDD systems.

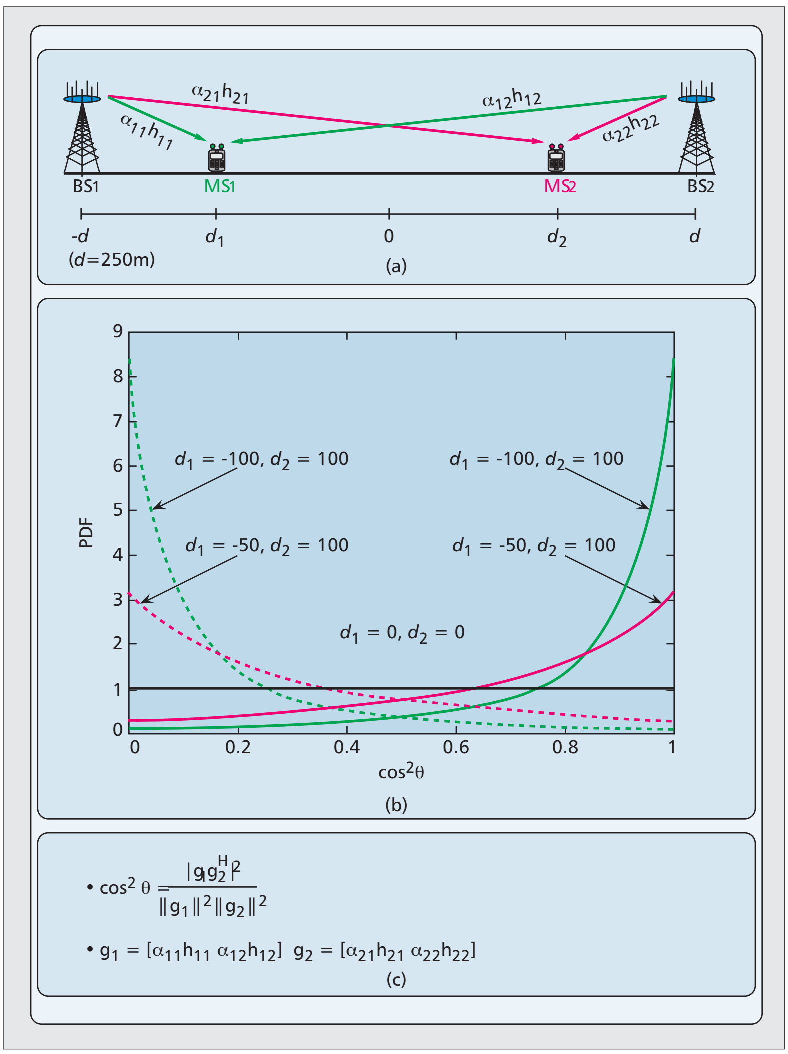


Figure 3. *The probability density function (PDF) of $\cos^2\theta$; θ is the angle between the channels of two MSs located in different places d_1 and d_2. The global channels between two MSs in different cells (e.g., $d_1 = -100$ m and $d_2 = 100$ m) are orthogonal in high probability since $\cos^2\theta \to 0$. By contrast, the channels between two MSs in the same cells (e.g., $d_1 = 100$ m and $d_2 = 100$ m) have the same direction in high probability since $\cos_2\theta \to 1$. When both MSs are located at the cell edge ($d_1 = d_2 = 0$), their channel directions are distributed uniformly within (0, 2π).*

TDD Systems

Imperfect Channel Reciprocity — TDD was expected to be more desirable for CoMP because it can obtain CSIT by exploiting the reciprocity between uplink and downlink channels. Unfortunately, only uplink and downlink *propagation channels* are reciprocal. The baseband equivalent channel is no longer reciprocal due to the different characteristics of radio frequency (RF) chains used in reception and transmission in practical systems.

Self-calibration is a popular antenna calibration method in single cell systems. It adjusts all antennas at one BS to achieve the same RF analog gain as that of a reference antenna, and hence ensures a constant scalar ambiguity between the uplink and downlink channels for all antennas. As shown in Fig. 2a, the equivalent downlink and uplink channels between BS_1 and MS_1 are related by $\mathbf{h}_{11} = a_{11}\mathbf{h}_{11,U}$, where a_{11} is a complex ambiguity factor. Such a scalar ambiguity does not affect the performance of single-cell single-user systems. However, when self-calibration is employed at each BS in CoMP systems, it will lead to N_b ambiguity factors between the uplink and downlink channels at different coordinated BSs (i.e., α_{11} and α_{12} for the two BSs CoMP illustrated in Fig. 2a). Then the global equivalent uplink channel $\mathbf{g}_{1,U}$ and downlink channel $\mathbf{g}_1$ of MS_1 is related by $\mathbf{g}_1 = \mathbf{g}_{1,U}\mathbf{A}_1$, where $\mathbf{A}_1$ is the diagonal ambiguity matrix as shown in Fig. 2a.

The multiple ambiguity factors in the global equivalent channels lead to imperfect downlink global CSIT even if the uplink channel estimation is perfect. Such an ambiguity is more detrimental than channel estimation errors. Because it is a kind of multiplicative noise rather than additive noise, it will hinder co-phasing of coherent CoMP transmission. Under some circumstances, sharing all data among BSs might not be advantageous anymore, given such undesirable multiplicative noise [6].

One possible way to avoid this dilemma is over-the-air calibration. It has been applied in single-cell systems to achieve the same goal as self-calibration for one BS. In CoMP it can be employed to calibrate the antennas among the BSs, where the channel is estimated simultaneously through two different approaches:

reciprocity of propagation channels, and measurement on the downlink and feedback; knowledge of those two channel estimates allows multiple ambiguity factors to be estimated. The performance of this method depends on the accuracy of the channel estimation. To improve the performance, the calibration needs to be obtained from the measurements of multiple uplink and downlink frames or multiple users.

In practice, the interference experienced at the BSs and MSs are also not reciprocal. As a result, the actual SINR to assist downlink modulation and coding selection (MCS) should be estimated at the MS, then fed back to the BSs.

Training Overhead — To estimate the global channel for downlink cooperative transmission, the uplink training overhead is proportional to the number of MSs and transmit antennas at each MS, which is roughly N_b times the overhead of single-cell systems, where N_b is the number of the coordinated BSs. To ensure that the downlink spectral efficiency gain over non-CoMP is not "eaten up" by the uplink training overhead, it is paramount to reduce the training overhead, especially for relatively fast fading channels.

The trade-off between the performance gain of CoMP-JP and the required overhead for channel estimation is nontrivial. For a given coherence time, the training length as well as the number of cooperative BSs can be optimized such that they maximize the net throughput (excluding the uplink overhead). This trade-off depends on the MS location and SNR as analyzed in [7]. Under which circumstances the ultimate performance of CoMP-JP — with optimized training overhead — will outperform non-CoMP systems is still an open problem.

A simplified form of this parameter optimization is a switching between CoMP and non-CoMP transmission modes. For users that require asymptotically zero training overhead (e.g., static users), CoMP-JP always outperforms non-CoMP transmission. For users requiring high training overhead, the net data rate of some cell-center MSs under CoMP transmission may be even lower than non-CoMP. This suggests developing transmission mode selection either from a system perspective or independently from each MS's perspective.

We can also reduce the required CSI by differentiating what we use it for. For instance, channel direction information (CDI) is essential for MU-MIMO precoding, and the channel norms and channel angles among MSs are essential for multi-user scheduling. If using full CSIT (i.e., perfect instantaneous CSIT), the spatial scheduling needs enormous training overhead even in single-cell systems. Fortunately, in CoMP systems, the channel orthogonality between MSs is largely dependent on their locations, as shown in a numerical result in Fig. 3. Intuitively, the global channels of two cell-center users are more orthogonal, since they are separated geographically. This indicates that for MSs in different cells their average channel gains can be exploited for scheduling [8] and possibly even for precoding. Since average channel gains vary slowly, they can be obtained at longer intervals, so training overhead can be reduced accordingly.

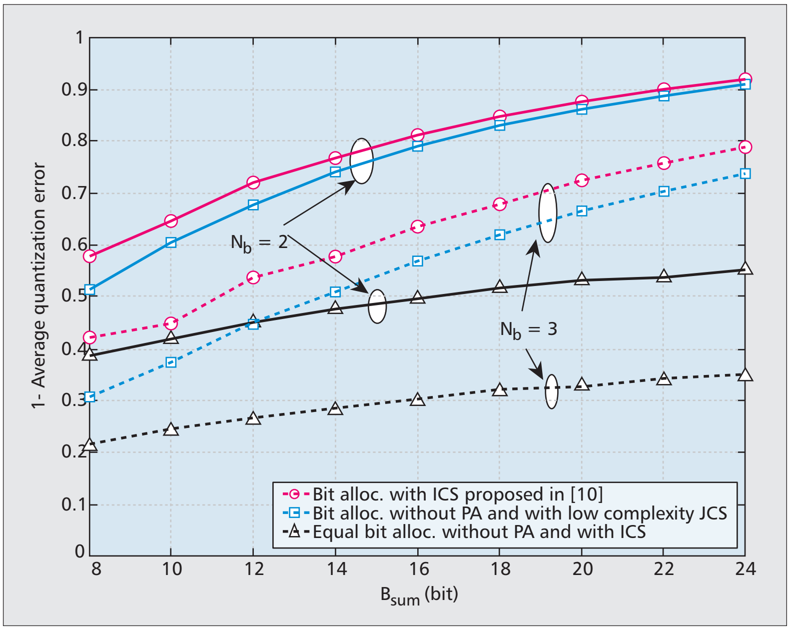


Figure 4. *Average quantization accuracy of the global CDI, E* $\{\cos^2(\angle\bar{\boldsymbol{h}}, \hat{\boldsymbol{h}})\}$, *v.s the overall number of bits for each user,* B_{sum}, *which is reproduced from Fig. 6 in [10]. The number of coordinated cells* N_b *= 2 or 3, each BS has four antennas. An optimal bit allocation among the per-cell CDIs and the phase ambiguity (PA) with independent codeword selection (ICS) proposed in [10] is compared with other two feedback schemes: (i) optimal bit allocation with a low complexity joint codeword selection (JCS) method, where only per-cell CDIs are quantized and no bits are allocated for quantizing the PA; (ii) the overall bits are equally allocated for the per-cell CDI quantization and no bits are allocated for PA quantization, where ICS is used. "Without PA" in the legend means without PA feedback.*

Non-Orthogonal Training — In CoMP systems, training signals for all MSs in multiple cells should be mutually orthogonal to obtain optimal channel estimation. However, in systems complying with the LTE standard, the training sequences of MSs in the same cell are orthogonal, but those for the MSs in different cells are not orthogonal (and are not identical). Orthogonal training for the MSs in different cells demands intercell signaling and protocols to coordinate the training sequences among cells. From the viewpoint of system compatibility and complexity, the training sequences of MSs in different cells are preferred to be non-orthogonal.

In fact, due to the non-i.i.d. feature of CoMP channels, orthogonal training for intercell global CSI may not be necessary. When the global channel is estimated under a minimum mean square error (MMSE) criterion, non-orthogonal training for the MSs in different cells leads to acceptable performance degradation for downlink precoding [9].

FDD Systems

In order to save uplink resources for channel feedback, the feedback is done using a quantization codebook known at both the BS and MS. Such limited feedback techniques have been investigated extensively in the context of single-cell MIMO.

In the following we consider the feedback of CDI, which is essential for precoding. We do not address the feedback of channel quality indica-

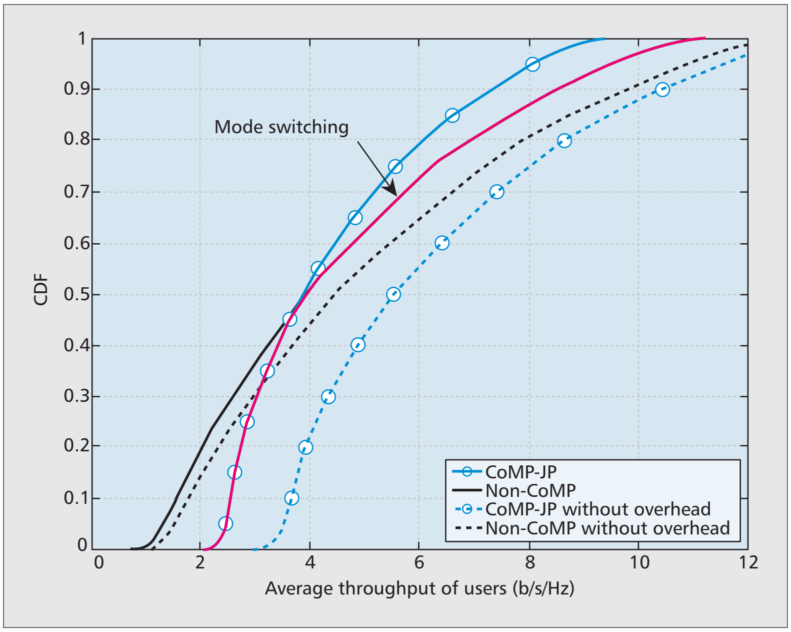


Figure 5. *Cumulative distributed function (CDF) of the net throughput of users averaged over small-scale fading channels. We consider the urban macrocell scenario of 3GPP. The path loss factor is 3.76, and the cell radius* R *= 250 m. The cell edge SNR (i.e., the received SNR of the user located at distance* R *from the BS) is set as 5 dB, where the inter-cluster interference is regarded as white noise.* N_b *= 3, each BS has four antennas, and 10 users are randomly placed in each cell. The downlink training overhead includes both the common and dedicated pilots. The overhead of CoMP-JP and non-CoMP systems occupy 28.1% and 12.6% of the overall downlink resources, respectively, which are typical values in 3GPP. The systems where all users are served by CoMP-JP or non-CoMP are also simulated, either with or without considering the overhead. The performance of a mode switching method with a distance threshold is shown. As addressed earlier, when the training overhead is considered, it is shown that with CoMP-JP the net throughputs of the cell center users become inferior to those with non-CoMP.*

tion (CQI) facilitating user scheduling and MCS, which largely depends on the employed beamforming.

The CDI quantization is chosen from a codebook with unit norm vectors of size 2^B, where B is the number of feedback bits. Each MS quantizes its CDI, $\bar{\mathbf{h}}$, for example, by choosing the closest codeword to the CDI as measured by the inner product (which reflects the quantization accuracy), $|\bar{\mathbf{h}}^H \hat{\mathbf{h}}| = \cos(\angle\bar{\mathbf{h}}, \hat{\mathbf{h}})$, where $\hat{\mathbf{h}}$ is the quantized CDI. Then each MS feeds back B bits to indicate the index of this codeword in the codebook. It has been shown that for MU-MIMO, quantization errors lead to severe throughput limits at high SNR levels. The feedback overhead increases with SNR to ensure a constant rate loss compared to the ideal case.

In CoMP systems, the dimension of the global CDI may vary dynamically, and the statistics of the CDI depends on each MS's location. This implies that every MS needs a unique codebook. This is unrealistic due to the prohibitive complexity of generating codebooks as well as selecting the codewords. Considering network scalability and compatibility, it is highly desirable to design a per-cell codebook-based feedback strategy, where the existing codebooks can be reused to quantize each single-cell channel as shown in Fig. 2b. Although such a structured codebook is suboptimal, its performance can be enhanced by representing the global CDI in an optimal manner.

Feedback Overhead — To increase the network spectral efficiency, a fundamental question is: how much uplink overhead is required to achieve the downlink performance gain of CoMP-JP over single-cell MU-MIMO?

To reduce the feedback overhead, the non-i.i.d. feature of CoMP channels should be exploited. This can be realized by optimizing the sizes of the per-cell codebooks. In [10], a scaling law of the feedback overhead for each user in CoMP-JP systems was provided. The analysis showed that the feedback overhead of CoMP-JP is about N_b times of that of single-cell MIMO, which depends on user location. When a user has equal average per-cell channel gains, its overhead is largest. When a user is in the cell center, its overhead is less because only the channels to one of the BSs is strong enough to need a fine quantization (large size codebook).

Alternatively, we can switch the transmission modes between CoMP-JP and non-CoMP, or design spatial scheduling exploiting the channel statistics, as in the TDD case. We can also employ selective feedback [11], where the CSI of weak links of each user are not fed back.

How to Represent the Global CDI

Codeword selection: Considering that a global channel is a stacking of multiple per-cell channels, each MS can select codewords either jointly to minimize the quantization error of global CDI or independently to minimize the quantization error of each per-cell CDI.

Independent codeword selection leads to severe performance degradation for global CDI quantization due to phase ambiguities among per-cell CDIs, especially for cell edge MSs [12]. Such a phase ambiguity does not affect the performance of a single-cell limited feedback MIMO system. However, it introduces multiplicative noise to the downlink global CDI analogous to the imperfect channel reciprocity in a TDD CoMP system. The phase ambiguities can be either compensated by feedback or mitigated by joint codeword selection.

On the other hand, simply selecting multiple codewords jointly does not necessarily lead to minimal quantization error of the global CDI. In order to achieve this, the joint codeword selection should consider the *structure* of the global CDI reconstruction, which affects the overall quantization accuracy [12]. In addition, to enjoy the optimality of the judiciously developed joint codeword selection, its prohibitive complexity needs to be reduced.

Codebook bit allocation: When we quantize the global CDI with per-cell codebooks, the features of the CoMP channel give rise to new parameters that can be optimized. Since different per-cell CDIs have different contributions to the global CDI, we can allocate bits to different per-cell codebooks. The optimal bit allocation that minimizes the average global CDI quantization error (i.e., $1 - \mathrm{E}\{\cos^2(\angle\bar{\mathbf{h}}, \hat{\mathbf{h}})\}$) under an overall constraint on the number of feedback bits turns out to be similar to water filling: more bits will be allocated to stronger channels [10]. Moreover, since different users have very different SINRs, we can also allocate the total number of bits among the users to maximize the weighted sum rate under an overall uplink feedback constraints in the cluster.

In Fig. 4, we provide simulation results for the average global CDI quantization accuracy of several feedback schemes. It shows the impact of the bit allocation among the per-cell CDIs and the phase ambiguity as well as the impact of the codeword selection.

Common Pilot Optimization — To facilitate channel feedback and data detection at the MS, training signals are transmitted for channel estimation, which consume downlink resources for data transmission. We can further distinguish between:

- "Common pilots" for channel feedback, which are transmitted from a particular transmit antenna and allow the MS to estimate the channel to this particular antenna
- "Dedicated pilots" for data detection, which are transmitted for each data stream with the same precoding settings as the actual user data

For common pilots, the induced overhead grows in proportion to the overall number of cooperating BSs, while for dedicated pilots, the overhead is in proportion to the overall number of data streams intended for all MSs, and thus does not necessarily increase when CoMP is used.

Analogous to the uplink pilot optimization in a TDD system, downlink pilot optimization that maximizes net throughput is also a nontrivial trade-off between the performance gain achieved by CoMP and the reduced downlink spectral efficiency resulting from spending additional time-frequency resources on non-payload transmissions.

Imperfect Backhaul

In existing cellular systems, the backhaul links among BSs are not perfect as assumed. On one hand, a limited-capacity backhaul does not allow BSs to share a large amount of data. On the other hand, the CSI shared among BSs may have quantization error and severe latency if the existing X2 interface (interface for communication between BSs) in LTE systems is used. The backhaul imperfection is even more severe in parts of heterogeneous networks, such as femto-cells, which hinders the application of CoMP. While backhaul links can be upgraded by high-speed optical fiber without technical challenges, this creates very high costs to the operators and therefore may not be realized in the near future.

The impact of limited-rate backhaul is largest for uplink CoMP, since in that case (quantized) analog signals need to be conveyed on the backhaul. Although not as stringent as in the uplink case, downlink CoMP-JP might also have to reduce its throughput to accommodate the capacity limitation of the backhaul links [5, Ch. 12.2]. In that case, it cannot achieve its full performance potential.

For both TDD and FDD systems, outdated channel information can considerably decrease the effectiveness of CoMP. The latency of the X2 interface (which arises from the IP-based protocols in the backhaul networks), which creates delays in the distribution of the CSI to different cooperating BSs, may reach 10 ms and more. This is more significant than the turnaround time of a TDD system (or the feedback latency of CSI in an FDD system), which creates delays between the measurement of CSI and its availability at one particular BS. Recent investigations have shown that CoMP-JP can benefit more from channel prediction than non-CoMP due to the non-i.i.d. channel feature [13]. If predicted channels instead of estimated channels are used for precoding, the performance degradation of CoMP-JP will be largely alleviated.

In the following, we discuss several possible approaches to mitigate the effect.

Switching between Different Transmission Modes

CoMP-CB needs much less backhaul capacity than CoMP-JP, since it only shares the CSIT, not the user data [3]. Considering that both of these schemes are able to eliminate interference, CoMP-CB may outperform CoMP-JP under stringent backhaul capacity constraints [14]. Consequently, mode switching between these two CoMP transmission modes — adaptive to location and number of MSs — will provide better overall throughput.

Another natural way is to switch between CoMP-JP and non-CoMP. Since cell center users experience lower ICI, they will not benefit as much from CoMP as cell edge users. Intuitively, we can simply divide the users in each cell with a threshold based on their average channel gains. Since only the users to be served by CoMP need to share their data among the BSs, the backhaul load can be controlled by judiciously selecting the threshold. In Fig. 5, we provide simulation results of mode switching, where the results for pure CoMP-JP and pure non-CoMP schemes are shown as a baseline, either with or without considering the training overhead. In the considered simulation setting (as shown in the caption of the figure), the distance threshold to divide the users into the CoMP-JP and non-CoMP users is 90 m, and 62 percent of the users prefer to be served by CoMP. After mode switching, the backhaul load is reduced by 27 percent.

Except for these "hard" mode switching methods where all data of some cell edge users are shared among the backhaul, partial cooperation among BSs is possible where partial data of all users are shared. For example, "soft" transmit mode switching can be operated with rate splitting, where the downlink data of each user is split into common and private parts, and only the common data is shared among the cooperating BSs [15].

Interference Coordination

If the backhaul capacity is too limited, such that no data are able to be shared among BSs, we are faced with an interference channel problem in information theoretic terminology. When only CSIT is shared, each BS serves multiple MSs in

For the common pilots, the induced overhead grows in proportion to the overall number of cooperating BSs, while for dedicated pilots, the overhead is in proportion to the overall number of data streams intended for all MSs, and thus does not necessarily increase when CoMP is used.

Although various innovative approaches have been devised in the literature, the downlink spectral efficiency gain of CoMP is still obtained at the price of a high uplink and downlink overhead. To provide high spectral efficiency from the network viewpoint, many challenging issues remain.

its own cell and coordinates with other BSs. In such a scenario, PBPC is no longer a performance limiting factor, and inter-BS calibration in TDD systems is unnecessary.

CoMP-CB is one of the popular ways to coordinate ICI. When full CSIT for MSs, both within the cell and in adjacent cells, is available at each BS, various criteria can be used to design the precoding, such as zero forcing (ZF), MMSE, and maximal signal-to-leakage-plus-noise ratio (SLNR). To implement centralized beamforming, scheduling, or power allocation, these channels need to be forwarded to a CU via backhaul links. The disadvantage of CoMP-CB is its limited performance gain. When MU-MIMO precoding is applied, the maximal multiplexing gain of CoMP-CB equals the number of transmit antennas at each BS, which is the same as single-cell MU-MIMO systems.

To reduce the impact of outdated CSIT caused by the X2 interface latency, statistical CSIT (long-term averaged CSIT), such as the angular power spectrum of the MSs in other cells, or scheduling results can be shared among BSs. It can also be combined with instantaneous local CSIT for ICI avoidance. However, this only performs well when each per-cell channel is highly spatially correlated. Other CoMP-CB schemes that take latency on the X2 interface into account are described in [5, Sec. 5.3.2].

Conclusion

We have addressed fundamental limiting factors that prevent CoMP from achieving its full potential, in particular the training or feedback overhead to gather channel information, and the backhaul constraints. We have briefly summarized critical issues in channel acquisition and typical ways to tackle the technical challenges in TDD as well as FDD systems. Considering the scenarios with imperfect backhaul, we have discussed possible ways to coordinate interference. Although various innovative approaches have been devised in the literature, the downlink spectral efficiency gain of CoMP is still obtained at the price of high uplink and downlink overhead. To provide high spectral efficiency from the network viewpoint, many challenging issues remain to be solved, especially for large-scale systems, including cooperative clustering selection, various ways of overhead reduction, transmit strategy optimization with imperfect backhaul links, and decentralized and distributed realization.

References

[1] A. F. Molisch, Internal Memo, AT&T Labs-Research, 2001; see also A. F. Molisch *et al.*, "Method and Apparatus for Reducing Interference in Multiple-Input-Multiple-Output (MIMO) Systems," U.S. Patent 7,912,014 (filed 2002, granted 2011).

[2] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network Coordination for Spectrally Efficient Communications in Cellular Systems," *IEEE Wireless Commun.*, vol. 13, no. 4, Aug. 2006, pp. 56–61.

[3] D. Samardzija and H. Huang, "Determining Backhaul Bandwidth Requirements for Network MIMO," *EUSIPCO*, 2009.

[4] H. Zhang *et al.*, "Asynchronous Interference Mitigation in Cooperative Base Station Systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, Jan. 2008, pp. 155–65.

[5] P. Marsch and G. P. Fettweis, *Coordinated Multi-Point in Wireless Communications: From Theory to Practice*, Cambridge Univ. Press.

[6] S. Han *et al.*, "Coordinated Multi-Point Transmission with Non-Ideal Channel Reciprocity," *Proc. IEEE WCNC*, 2011.

[7] J. Hoydis, M. Kobayashi, and M. Debbah, "Optimal Channel Training in Uplink Network MIMO Systems," *IEEE Trans. Signal Proc.*, vol. 59, no. 6, June 2011, pp. 2824–33.

[8] S. Han *et al.*, "Channel Norm-based User Scheduling in Multicell Cooperative Transmission Systems," *Proc. IEEE GLOBECOM*, 2009.

[9] X. Hou, C. Yang, and B. K. Lau, "Impact of Non-Orthogonal Training on Terformance of Downlink Base Station Cooperative Transmission," *IEEE Trans. Vehic. Tech.*, vol. 60, no. 9, Nov. 2011, pp. 4633–39.

[10] F. Yuan and C. Yang, "Bit Allocation Between Per-Cell Codebook and Phase Ambiguity Quantization for Limited Feedback Coordinated Multipoint Transmission Systems," *IEEE Trans. Commun.*, vol. 60, no. 9, Sep. 2012, pp. 2546–59.

[11] A. Papadogiannis *et al.*, "Efficient Selective Feedback Design for Multicell Cooperative Networks," *IEEE Trans. Vehic. Tech.*, vol. 60, no. 1, Jan. 2011, pp. 196–205.

[12] D. Su, X. Hou and C. Yang, "Quantization based on Per-Cell Codebook in Cooperative Multi-Cell Systems," *Proc. IEEE WCNC*, 2011.

[13] L. Su, C. Yang, and S. Han, "The Value of Channel Prediction in CoMP Systems with Large Backhaul Latency," *Proc. IEEE WCNC*, 2012.

[14] N. Seifi *et al.*, "Coordinated Single-Cell vs. Multi-Cell Transmission with Limited-Capacity Backhaul," *Proc. IEEE ACSSC*, Nov. 2010.

[15] R. Zakhour and D. Gesbert, "Optimized Data Sharing in Multicell MIMO with Finite Backhaul Capacity," *IEEE Trans. Signal Proc.*, vol. 59, no. 12, Dec. 2011, pp. 6102–11.

Biographies

CHENYANG YANG [SM'08] is a full professor at Beihang University, China. She was supported by the 1st Teaching and Research Award Program for Outstanding Young Teachers of Higher Education Institutions of the Chinese Ministry of Education during 1999–2004. She is an Associate Editor for *IEEE Transactions on Wireless Communications* and was Chair of the Beijing chapter of IEEE Communications Society. Her recent research interests include network MIMO, energy-efficient transmission, and interference management.

SHENGQIAN HAN [S'05, M'12] received his B.S. degree in communication engineering and Ph.D. degree in signal and information processing from Beihang University in 2004 and 2010, respectively. From 2010 to 2012 he held a postdoctoral research position with the School of Electronics and Information Engineering, Beihang University. He is currently a lecturer ay the same school. His research interests are multiple antenna techniques, cooperative communication, and energy-efficient transmission in the areas of wireless communications and signal processing.

XUEYING HOU received her B.S. degree from Beihang University in 2007. She is currently working toward her Ph.D. degree in signal and information processing in the School of Electronics and Information Engineering, Beihang University. From September 2009 to June 2010 she was a visiting student with the School of Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden. Her research interests include cooperative communication, channel estimation, and limited feedback techniques in wireless systems.

ANDREAS F. MOLISCH [S'89, M'95, SM'00, F'05] (molisch@usc.edu) is a professor of electrical engineering at the University of Southern California. His research interests include the measurement and modeling of mobile radio channels, ultra-wideband communications and localization, cooperative communications, multiple-input multiple-output systems, and wireless systems for health care. He is a Fellow of the AAAS and IET, an IEEE Distinguished Lecturer, a member of the Austrian Academy of Sciences, as well as the recipient of numerous awards.

ON THE IMPACT OF NETWORK GEOMETRIC MODELS ON MULTICELL COOPERATIVE COMMUNICATION SYSTEMS

ANVAR TUKMANOV, ZHIGUO DING, AND SAID BOUSSAKTA, NEWCASTLE UNIVERSITY
ABBAS JAMALIPOUR, UNIVERSITY OF SYDNEY

ABSTRACT

Recognized as an efficient technology allowing enhanced high-data-rate coverage, multicell cooperation builds on collaborative operation of distributed network elements aiming to reduce the impact of intercell interference, or even to exploit it as a source of additional information. However, performance gains arising from this collaboration come at a price of increased complexity and coordination among involved network elements. An immediate question is how to adequately model a multicell cooperative network, while striking the balance between analytic complexity and precision of design decisions. In this article we highlight two categories of multicell network models, Wyner models and random network models, and their impact on the design and analysis of MCC systems. Recent examples of realistic yet tractable multicell network models are discussed.

THE STRUCTURE OF A MULTICELL COOPERATIVE NETWORK

The requirement for increased high-data-rate coverage has led to intensive research activities in academia and industry aiming to extend limits of current communication networks with the constraint of scarce bandwidth resource. Departing from the conventional cellular network operation paradigm, where interaction between cells was limited mainly to handover procedures, multicell cooperation (MCC) resembles a more proactive approach by utilizing and rethinking the opportunities for networks with multiple user terminals (UTs) and base stations (BSs). The key idea in MCC is to exploit the resources of multiple distributed BSs and UTs in order to more efficiently combat fading and interference, which are the main limiting factors for current communication systems.

BS COOPERATION

Base station cooperation has been one of the active research topics in the area of MCC for the past decade and is currently being considered for inclusion in future releases of Long Term Evolution — Advanced (LTE-A) in the form of cooperative multipoint (CoMP) technology [1]. In BS cooperation, the resources of distributed BSs are utilized to mitigate or even exploit intercell interference (ICI) through coordinated beamforming/coordinated scheduling (CB/CS) or joint processing and transmission (JPT), respectively. By tuning the transmitting BS's radiation pattern and signaling, CB/CS aims to strike a balance between maintaining optimal performance of UTs and minimizing the level of unintended interference caused to other mobiles. While CB/CS is primarily a downlink technique with a single BS performing the coordinated transmission, JPT involves multiple BSs sharing, cooperatively processing, transmitting, and receiving uplink and downlink streams.

Promising performance gains from state-of-the-art BS cooperation techniques do not come for free: there are certain requirements to enable any cooperative communication scheme, which result in increased coordination and processing complexity. In general, to enable cooperation it may be necessary to:

- Collect system information (e.g., channel state information [CSI] or even raw signal samples) [2] from other nodes
- Process this information to make a decision on how to act cooperatively
- Deliver this decision with individual instructions to all involved parties

Clearly, it is desirable to completely eliminate some of the above steps, or at least minimize associated overhead. Relaxation of these requirements in BS cooperation typically involves the selection of a subset of collaborating BSs [1], which in turn can be based on intelligent selection of cooperating nodes (e.g., incentive-based recruitment) [3].

Given such a nontrivial task — to efficiently coordinate multiple distributed transceivers — it becomes crucially important to adequately model the environment where this communication takes place to ensure that the way we account for the key influencing parameters (e.g., interfer-

 1536-1284/13/$25.00

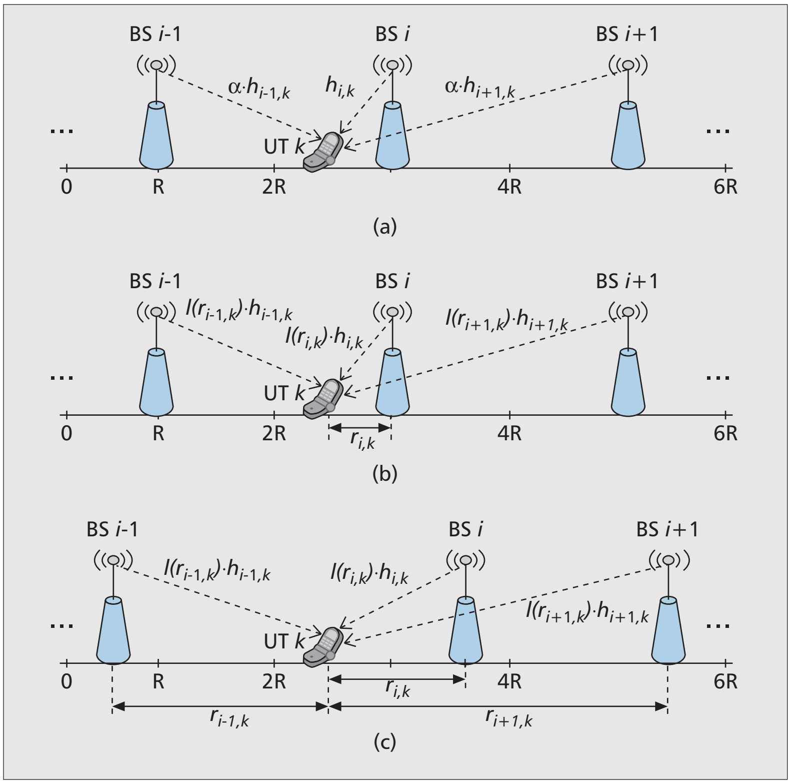


Figure 1. *Multicell network topologies in 1D: a) linear Wyner network model with fading; b) linear network model with fixed BSs and random UT location and account for fading and path loss; c) linear network model with random BSs and UT locations with fading and path loss.*

ence and fading) allows us to predict the actual system performance.

Multicell System Models

In developing MCC techniques, researchers take different approaches to modeling multicell networks. Current multicell cooperative network models generally are variations of the Wyner model, where network elements have deterministic locations, while the cellular network itself has a simple regular macroscopic structure [2]. While it is well understood that this is a simplistic model, the insights obtained through this approach still provide valuable guidance for the practical design of current MCC communication systems. However, the main drawback of the Wyner model is that obtained performance metrics, conclusions, and design guidelines are usually applicable only to the specific studied system setup. In other words, the extent to which such results are valid for different setups is not known. The impact of this issue is that large-scale simulations are required for hypothesis testing in systems designed using the Wyner model.

This was the key motivating factor for introducing an alternative approach to system modeling, which would capture the random nature of communication networks and the impact of network topologies on various performance metrics. The first efforts to capture such macroscopic properties for both wired and wireless communication networks appeared in the late 1990s [4]. Specifically, in [4] the point process theory and stochastic geometry were used to obtain network performance metrics that to a greater extent accounted for the natural randomness of wireless networks. Recently these methods found applications in the area of MCC [5–7], where network topology has been shown to have particular importance (e.g., due to introduced propagation delay and propagation path loss).

The main contribution of this article is in comparing existing and emerging approaches to modeling of MCC systems, and highlighting associated advantages and limitations. Our aim is to attract attention to more recent advances that, in conjunction with expertise in traditional methods, may ultimately support the design of future communication systems. We first review the Wyner model, which is used in the majority of published works on MCC, is described. Following that, heoretical tools that allow a departure from such a deterministic approach are briefly introduced to facilitate the discussion on random network topologies for MCC. Remarks then conclude the article.

Wyner-Type Network Topologies

Introduced in the early 1990s, a general Wyner model (WM) consists of M identical, regularly placed cells with K fixed users per cell. All network nodes are asumed to be equipped with a single antenna. In a linear or one-dimensional (1D) WM, the BSs are aligned along a line such that a cell i has two adjacent cells with BSs indexed $i - 1$ and $i + 1$, respectively. In a more advanced 2D setup, a *planar* WM is represented by a regular hexagonal structure, where the central cell is surrounded by six adjacent cells. While the latter seems to be the more realistic of these two setups, the majority of research results on achievable rates and outage probabilities of MCC techniques exist primarily for the linear WM, mainly due to its analytic tractability [2]. Figure 1a depicts three BSs for the 1D setup, where each cell spans over a segment of length $2R$, and each BS is located at the center of its corresponding cell. The position of the reference user terminal (UT) k is not explicitly used in performance evaluation.

Recall that the reason for introduction of multicell models is to study coexistence of multiple BSs in terms of, for example, ICI. In the original WM, the contribution of each cell to ICI at the UT k in downlink is assumed to be limited to $L = 2$ adjacent cells, where parameter L is called *interference span*. In particular, the average channel gain between cells i and j is characterized by a single, globally available, and deterministic parameter $\alpha_{i,j} \in [0,1]$, which is the same for all UTs in cell i. For the serving BS such homogeneous channel gain $\alpha_{i,i} = 1$. If these channel gains α are also symmetric, such a setup with no randomness is referred to as *Gaussian WM*, and allows system performance to be studied through a single design parameter α. In contrast, a *fading* WM in addition to the static parameter α includes a random coefficient $h_{i,k}$ to account for the fading channel between BS i and UT k. In this way, while the network topology remains static, channel uncertainty is introduced through the fading coefficient $h_{i,k}$, which is independent for each communicating/interfering pair of network elements. Introduction of fading into the model allows the multipath effects on top of the static interference picture to be accounted for. Other extensions and variations of the linear WM include non-symmetric channel gains, extended interference span $L > 2$, for performance results; the interested reader is referred to [2].

In the following, *outage probabilit*y will be used as a performance metric, which corresponds to the probability of the event that the signal-to-interference-plus-noise ratio (SINR) at the receiver falls below some predefined threshold. Note that for the Gaussian WM with static parameter α, SINR is a constant parameter and admits no notion of outage probability; whereas for versions of WM with randomness, and of course for random models to be discussed further, outage probability is a tool to capture the effect of randomness of SINR.

The key advantage of WM is its simplicity, which allows tractable models for various signal processing methods and communication protocols for MCC. Currently, with the aid of this model, capacity results are available for CB/CS, JPT, including the account for limited inter-BS backhaul, limited feedback, relaying, and various optimization methods for MCC. However, there remains a question on the extent to which WM captures performance characteristics for realistic network geometries. This issue was first systematically addressed in [8], where it was found that a WM can adequately represent systems with large amounts of averaging (e.g., when the number of users is large). However, such results are only valid for uplink, whereas if downlink communication is considered, the 1D WM was reported to fail in capturing the impact of random user locations in terms of outage probability and average throughput. In order to better describe the random nature of wireless networks, researchers have developed alternative approaches for modeling networks, which are discussed in the following section.

A Stochastic Geometry Approach to Random Wireless Networks

To elaborate on the impact of random network modeling on the design of MCC systems, in this section, we first provide an overview of instruments used to describe random multicell network topologies and to account for randomness in node locations in wireless networks in general.

Motivation

Real-world patterns of deployed BSs may have little in common with regular structures used in WMs due to, for example, inhomogeneous signal propagation conditions, availability of BS deployment sites, or regulatory requirements. To illustrate the impact of different assumptions on network performance, let us consider a simple illustrative example where assumptions based on the linear WM will be gradually relaxed toward random placement of both the BSs and UTs.

We are interested in the outage probability P of a user in cell i due to interference from adjacent BSs $i - 1$ and $i + 1$. Consider the linear WM with fading in Fig. 1a as a baseline scenario. Specifically, fixed BSs are located in centers of cells covering segments of $2R$ each, while interference from adjacent cells is characterized by a symmetric design parameter α for both cells. Recall from earlier that the design parameters α are homogeneous for all users in a given cell; therefore, the specific location of the studied user in this scenario plays no role in performance analysis. Channels between UT k and BSs undergo independent Rayleigh-distributed fading, and additive white Gaussian noise (AWGN) power is assumed to be equal and normalized at 1.

In the second scenario in Fig. 1b, a grid model with randomly located reference UTs and static BSs is considered: the reference user location is now a uniformly distributed random variable, characterized by the distance $r_i \sim \mathcal{U}(0,R)$ to the serving BS. In addition to fading, channel gains between BSs and UTs now include *bounded* path loss function $l(r_i) = (1+r_i^{\beta})^{-1}$, where β is the path loss exponent for the propagation environment.[1] Distances and values of associated path loss functions between UT k and adjacent BSs can be explicitly found once r_i is known, since BSs are located in cell centers.

In the third scenario random BSs and UT locations are assumed: BSs are now uniformly distributed on the same total coverage segment of length $6R$, such that cell dimensions become (dependent) random variables, and BSs are not necessarily located in cell centers (Fig. 1c). Specifically, coverage segments for BSs $i - 1$, i, and $i + 1$ are now segments of random lengths X_{i-1}, X_i, and X_{i+1}, respectively, subject to the condition $X_i - 1 + X_i + X_{i+1} = 6R$. The distance from the reference UT k to its home BS i is uniformly distributed conditioned on the size of the home cell, that is, $r_i \sim \mathcal{U}(0, X_i | b_i)$, where b_i is the location of BS i.

Outage performance for the above three scenarios is illustrated in Fig. 2. Specifically, 10^6 network realizations were simulated to capture the impact of network topologies on outage. The power of interfering BSs was set to 1, and the design parameter α for the WM has been chosen as an average of the channel gains from either of the adjacent BSs to a uniformly placed UT as in [5]. Simulation results show that although the slopes for outage probability curves match for all three scenarios, there is a significant difference in performance of all three models caused by the assumptions and factors involved in network modeling. Specifically, the WM provides a 10 dB optimistic picture compared to the case when the UT's location becomes random and path loss is accounted for explicitly. In addition, outage performance for the scenario with random BSs is approximately 1.5 dB worse compared to the random UTs and fixed BSs scenario.

While a more realistic system model is desirable, in simulation or analysis of large-scale systems it can be impractical to explicitly account for individual effects of multiple factors; but one can attempt to capture significant properties and the impact of network topology via certain probabilistic law. Clearly, such a probabilistic model would also address only a limited number of

While a more realistic system model is desirable, in simulation or analysis of large-scale systems it can be impractical to explicitly account for individual effects of multiple factors; but one can attempt to capture significant properties and the impact of network topology via certain probabilistic law.

[1] *Different path loss models can be used depending on the application. For example, the mentioned bounded model is one method to avoid ambiguities of a less complicated* unbounded *model* $l(r_i) = r_i^{-\beta}$, *which are associated with the increase in the received power for* $r_i < 1$ *and with a singularity at* $r_i = 0$. *Another model* $l(r_i) = C(r_i/r_{ref})^{-\beta}$ *is used in applications, where an estimate* C *of the signal power is available for certain reference distance* $r_{ref} < r_i$ *in the antenna far-field.*

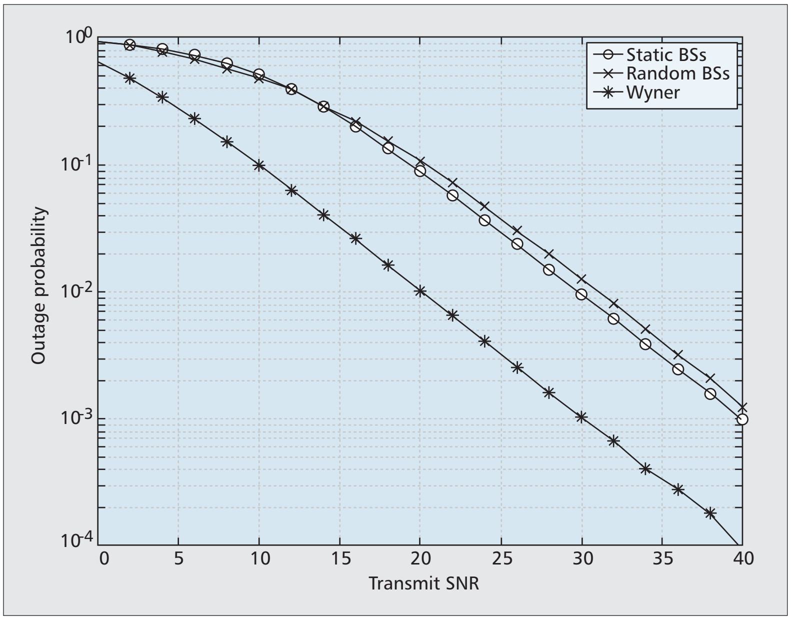


Figure 2. *Outage probability as a function of transmit SNR in the presence of interference from adjacent cells for WM and random UT, and random BS models.*

aspects of the real-world process; however, network topology and space are factors that are believed to be key for the design of future generation networks. Fortunately, these types of interactions can be described using point process theory and stochastic geometry, where network elements can be modeled as point patterns [4, 9].

Basics of Point Processes

The core element in moving from the static WM to a more general setup is the point process. A point process (PP) Φ in d-dimensional space $\mathbb{R}^d$ is a pattern of points distributed in $\mathbb{R}^d$ with respect to some probabilistic law, where point $x \in \Phi$ can represent a location of a BS or a user. Here d is the number of space dimensions, so $d = 1$ corresponds to a PP on line and $d = 2$ corresponds to a PP on a Euclidean plane. Such probabilistic law for any PP can be mathematically described using finite-dimensional (fi-di) distributions of the form

$$\Pr(\Phi(B_1) = n_1, \ldots, \Phi(B_k) = n_k), \qquad (1)$$

where the notation $\Phi(B_i) = n_i$ states that there are exactly n points of some PP Φ in a subset B of space $\mathbb{R}^d$. These fi-di distributions, in turn, allow distributions of random distances between points of the PP to be obtained, which are of particular importance for MCC processing, since they are associated with variable signal arrival times and statistical internode channel characterization.

Different types of PPs exist, capturing various levels of interactions between member points. Examples of such interactions could be minimal internode distance, node clustering, and repelling or attraction conditions, which in practice could translate into minimal inter-BS distance or grouping of users around business and shopping centers. The ultimate desired result of applying PP theory to modeling MCC topologies is the same as for Wyner-type models: obtaining an accurate prediction of the performance of some communication technique for a *typical user*.[2] And, as with a WM, PP-based models require certain assumptions to be made for analytical tractability.

Point Process Models for MCC and Their Limitations

The simplest and most commonly used PPs for modeling nodes in wireless networks are the Poisson point process (PPP) and the binomial PPs (BPP). The reason for such common application of PPPs is that they allow for analytically tractable description of statistical properties of networks, where nodes are independently distributed (e.g., ad hoc networks) [4]. Specifically, the counts of points of the PPP in any number of disjoint subsets B_i are independent Poisson distributed random variables (see [10] for distribution expressions). In general, the level of captured correlation between points is the main limiting factor for any PP-based model because of associated analytical complexity. In the case of relatively simple PPPs, points are assumed to be absolutely uncorrelated (i.e., the number n_1 of points of a PPP in a subset B_1 conveys no information about the number n_2 of points in another disjoint subset B_2. While this is a clearly unrealistic assumption for MCC due to dependence in BS placements, it works well for ad hoc networks [9], and even such simple PPs can still be used to obtain performance bounds (rather than precise results) for such metrics as mutual information, outage probability, coverage and level of interference in MCC systems.

Point process theory is gradually moving toward capturing general types of interactions between points of a PP [10]. One useful example is the BPP, where the numbers of points in disjoint subsets B_i become correlated through the total number of points

$$N = \sum_i n_i$$

so that any n_i is now binomial-distributed. Accounting for this type of correlation adds precision to the network model; however, this also increases the complexity of performance analysis [11]. High-precision results for stronger point correlations (e.g., minimal internode distance and clustering) are possible through more sophisticated PPs, which still use PPPs and BPPs as building blocks.

For instance, hard-core PP (HCPP) models point patterns where elements lie no closer than a certain distance h [10, p. 162]. HCPP has been recently applied to modeling of spatial correlation among active transmitters in networks with carrier-sense multiple access [12], and can be of particular importance for modeling MCC networks. The presence of a BS at a certain location in practice implies that neighboring BSs must be geographically separated by some distance. This real-world inter-BS distance depends on such factors as the type of terrain and population density; hence, the HCPP provides a sim-

[2] *In the literature on stochastic geometry and random networks, the term "typical user" refers to a "randomly chosen user," which itself is formally defined using Palm theory [4, 10]. To see the need for a formal definition, consider, for example, a possible ambiguity when a realization of a PP contains no points, so there is nothing from which to choose.*

plified version of BS placement correlation that allows avoiding highly unlikely BS locations. Recall that pure PPP-based models lack any mechanism to capture such dependence, or to prevent impractical placement of BS masts in the network model. Consider models of cellular networks in Fig. 3 and Fig. 4 for example. In the PPP-based model in Fig. 3 some BSs are placed very closely, such as three BSs at approximately (4, –6). In contrast, Fig. 4 illustrates the case when BSs are restricted to lie closer than $h = 1$, which leads to a more realistic network realization. It is important to stress that the fixed minimal distance h is still a simplification of the real-world correlation, and in actual deployed networks h may depend on various parameters. To extend the flexibility in modeling of BS placement correlations further, soft-core PP (SCPP) [10, p. 164] can be used to account for inhomogeneous terrain properties through variable distance h.

In overall, PP-based models naturally produce results for a *typical user's* performance, which is one of the benchmarks for practical communication system designs [4]. The key tasks involved with PP-based network modeling are to:

- Build a PP model to capture interactions or correlations between network elements
- Calculate desired metrics based on this model

Therefore, the main limitation for PP-based approaches is in analytical complexity stemming from accounting for correlations between network elements, such as inter-BS distance. This is in contrast to Wyner-type models, where an inherent drawback is in large-scale simulations, necessary to generalize and verify results obtained for a static setup. Table 1 compares key features of PP-based random models and Wyner-type models.

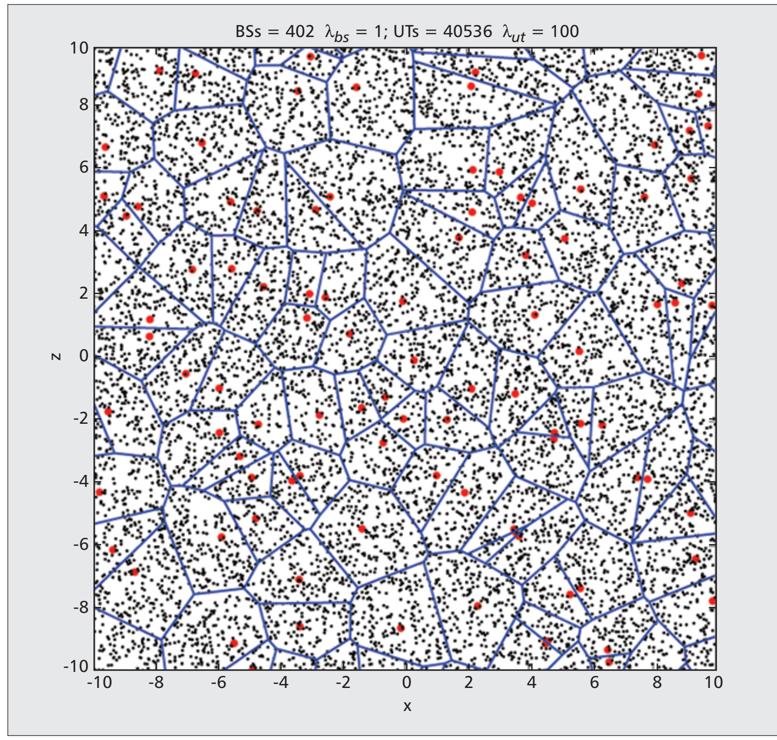


Figure 3. *Completely randomly distributed BSs and UTs on a plane. Each UT is associated with the nearest BS, forming Voronoi tessellations. Blue lines represent cell borders, red points are randomly placed BSs, and small black points are the UTs. Illustration is zoomed from 20 × 20 scale.*

Random Network Topologies for Multicell Cooperation

Different to the case with a WM for fixed setups, there is no generally accepted network model accounting for random network topology for MCC. Rather, there is a common set of tools that can be applied to modeling certain network properties. In the following, we discuss recent examples of application of such instruments to multicell scenarios.

Fixed BSs and Random User Locations

Joint processing of uplink transmission from uniformly distributed users by two fixed BSs was considered in [7] for 1D topology. Specifically, two BSs were fixed at certain positions on a line to form two identical cells with the same radius R. Each BS was equipped with N antennas, and could share both CSI and signal data through unlimited backhaul links in order to jointly decode user signals with another BS. There were K uniformly distributed users in each cell. Although this setup resembles the WM (Fig. 1a), the key difference is that the performance results are averaged over random user locations using the density function of user distribution.

The main advantage of such an approach over WMs with static locations is that it has been demonstrated to provide insights into the performance of a typical user, avoiding computationally intensive simulations. Specifically, with the results for such typical user performance, one can obtain optimal placement of the BSs in terms of maximizing mutual information through varying system parameters in asymptotic expressions for the mutual information.

Distributed Antenna System with Randomness

Although the account for random user distribution has allowed for efficient design and optimization in the *linear* setup with two cooperating cells in [7], more advanced models are required to describe planar scenarios. One of the first attempts to address the performance of a random multicell planar network was presented in [5], where the impact of placement of multiple distributed antenna systems (DAS) on the interference experienced by a user was addressed.

A cluster with multiple single antenna elements distributed randomly around a BS was considered in [5] as a building block, so the multicell network consisted of multiple such DAS clusters. Antenna elements (AEs) in each cluster were modeled as a BPP with N AEs per cell, while users were represented as points of a homogeneous PPP with density λ. For such a setup, outage probability performance of a typical user in a single cell was reported to be close to the case with fixed placement of the AEs, which indicates that even random placement of AEs can deliver the same performance gains. The multicell scenario, however, was studied in [5] with the assumption that AEs and user posi-

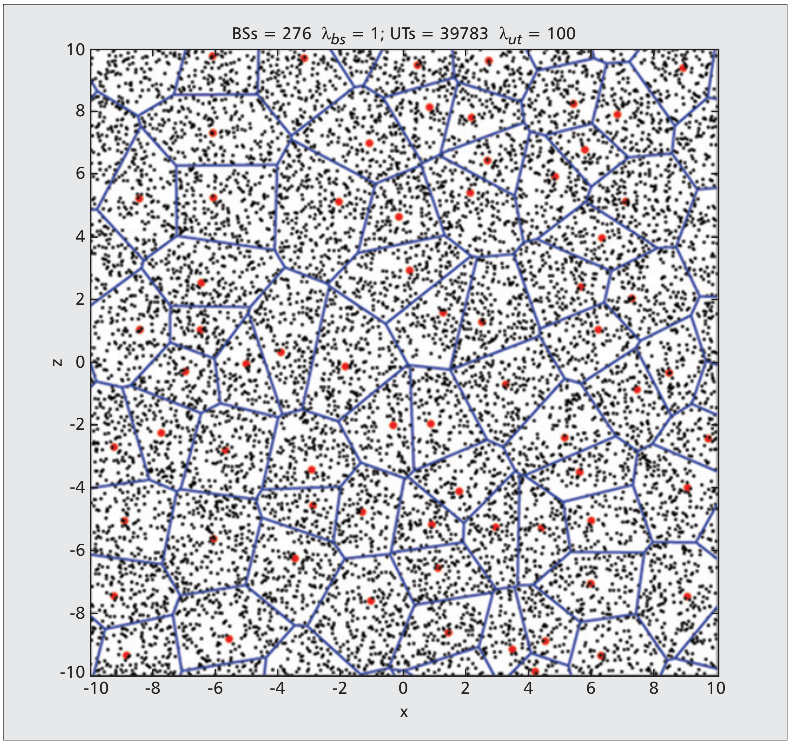


Figure 4. *BSs modeled as a realization of a hard-core point process with* h = *1, so the minimal inter-BS distance is* h = *1.*

tions are fixed, for which case it has been demonstrated that selecting a single AE for transmission is preferable for a multicell environment over blanket transmission, due to reduced ICI of the former approach. It is important to stress that in [5] network topology exhibits randomness only for derivation of outage performance for a single cell.

Random BS and User Locations

Greater flexibility in multicell topology was considered recently in [6]. Specifically, interference coordination between multiple randomly distributed BSs serving multiple randomly distributed users on a plane was studied. Locations of BSs and UTs were modeled as independent homogeneous PPPs, and BSs were clustered into cooperating groups with respect to an overlaid regular hexagonal lattice. Each user was associated with the nearest BS, such that each cell in the network has irregular shape and size, depending on proximity of other BSs. Figure 3 illustrates the setup.

The presented cellular network topology is different from regular hexagonal structures not only visually, but also through distribution of distances to interfering BSs. For example, the distance to the nth nearest BSs in the setup in Fig. 1c is a random variable, the probability distribution function (PDF) of which can be found for PPPs in closed form [10]. Although placement of BSs can be only approximately described using PPPs, by using such a PDF we can get a rough estimate of the distance to the nearest (and probably the strongest) interfering BS. Note that this interfering BS will have index $n = 2$, because the BS with $n = 1$ would be the serving BS. Therefore, such statistics can play an important role in MCC models by providing estimates for the distances.

In the network in Fig. 3, BSs share CSI to direct their radiation patterns toward the intended users, avoiding transmission toward unintended stations. Clearly, the decisions made by a BS will depend on the realization of node locations, which is visibly different from regular placement. For such a setup, shapes of outage probabilities for a *typical* mobile station and for the one located close to the center of cooperating cluster were compared analytically in [11]. Intuitively, the chance of outage for the nodes located closer to the cell center should be lower than that for a typical user; however, the analysis in [6] based on PP theory allowed quantification of the approximate rate of outage probability decay.

Conclusion and Future Directions

The aim of this article was to highlight current approaches to network modeling with applications to multicell cooperation, motivated by the

Feature	Wyner models	Random models
BS locations	Deterministic	Points of a point process
UT locations	Not used	Points of a point process
Cell shape	Line segment (1D); Generally hexagon or circle (2D)	Irregular, typically Voronoi tesselations
Path loss	Implicitly included in design parameter a	Explicit path loss model $l(r_i)$
Fading	Explicit (except Gaussian WM)	Explicit
Advantages	Simplicity; analytical tractability; account for essential channel characteristics	Model real processes (e.g. interference); account for randomness, direct results for typical user
Disadvantages	Replace real processes (e.g. interference); static topology; hypothesis testing and generalization by simulations required, hence computational complexity to implement/verify/generalize results	Analytical complexity to create accurate point process model for systems with correlated nodes; analytical complexity to obtain performance metrics; model revision/new model required for different communication strategies

Table 1. *Comparison of key features of Wyner models and random models based on point processes.*

fact that associated assumptions drive such performance-determining parameters as signal propagation delay, attenuation, and interference picture. These approaches can generally be classified into:

- Regular models, represented by the Wyner model family, where network topology is static and channel gains are typically described using a single parameter to capture channel attenuation and interference
- Random network models, which use point processes to model placements and interactions between network elements

To understand future research directions in MCC network modeling, it is important to once again recall the ultimate aim of such modeling. A network model is required to provide a version of the environment or a setup where developed communication techniques will operate. Ideally, one would want such a version to replicate all important features of the real environment; however, usually detailed modeling is associated with high complexity. Given the growing number of mobile terminals and evolution of communication techniques, it is a question whether the growth in computational power will be sufficient to continue delivering acceptable results for techniques developed based on Wyner-type constructions. More important, not taking network topology into account may limit our understanding of the operation of modern communication systems. On the other hand, the limitations of PP-based models originate mainly from analytical complexity, although surprisingly simple closed form expressions are already available for systems with little dependence between network elements.

Therefore, further research into analytically feasible yet accurate PP-based models is necessary to capture or approximate the dependencies in BS placement and mobile user clustering. Ideally, such models could not only reduce or eliminate the need for large-scale simulations, but also deliver more accurate results and advance our understanding of MCC systems. If closed form solutions are intractable, asymptotic results could still provide valuable insights into operation and design of modern MCC networks.

References

[1] E. Hossain, D. I. Kim, and V. K. Bhargava, *Cooperative Cellular Wireless Networks*, Cambridge Univ. Press, 2011.

[2] D. Gesbert *et al.*, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," *IEEE JSAC*, vol. 28, no. 9, Dec. 2010, pp. 1380–408.

[3] Z. Hasan, A. Jamalipour, and V. Bhargava, "Cooperative Communication and Relay Selection Under Asymmetric Information," *Proc. IEEE WCNC '12*, Apr. 2012, pp. 2373–78.

[4] F. Baccelli *et al.*, "Stochastic Geometry and Architecture of Communication Networks," *Telecommun. Sys.*, vol. 7, no. 1–3, 1997, pp. 209–27.

[5] J. Zhang and J. Andrews, "Distributed Antenna Systems with Randomness," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, Sept. 2008, pp. 3636–46.

[6] K. Huang and J. Andrews, "A Stochastic-Geometry Approach to Coverage in Cellular Networks With Multi-Cell Cooperation," *Proc. IEEE GLOBECOM '11*, Dec. 2011.

[7] J. Hoydis *et al.*, "Analysis of Multicell Cooperation with Random User Locations via Deterministic Equivalents," *Proc. 10th Int'l. Symp. Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, May 2012, pp. 374–79.

[8] J. Xu, J. Zhang, and J. Andrews, "On the Accuracy of the Wyner Model in Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, Sept. 2011, pp. 3098–3109.

[9] M. Haenggi *et al.*, "Stochastic Geometry and Random Graphs for the Analysis and Design of Wireless Networks," *IEEE JSAC*, vol. 27, no. 7, 2009, pp. 1029–46.

[10] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 2nd ed. Wiley, 2008.

[11] A. Tukmanov *et al.*, "On the Broadcast Latency in Finite Cooperative Wireless Networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, Apr. 2012, pp. 1307–13.

[12] H. ElSawy and E. Hossain, "Modeling Random CSMA Wireless Networks in General Fading Environments," *Proc. IEEE ICC '12*, 10–15 June 2012, pp. 6522–6989.

Biographies

ANVAR TUKMANOV [S'09] (anvar.tukmanov@ncl.ac.uk) received his B.Eng. degree in mobile communications from Kazan State Technical University in 2007 and his M.Sc. degree in communications and signal processing from Newcastle University in 2009, all with distinction. Since 2010 he has been working toward his Ph.D. degree at Newcastle University. From 2008 to 2010 he was an IP/MPLS network engineer with Tattelecom, Kazan. His research interests are in cooperative diversity, stochastic geometry, and information theory.

ZHIGUO DING [S'03, M'05] received his B.Eng. in electrical engineering from the Beijing University of Posts and Telecommunications in 2000, and his Ph.D. degree in electrical engineering from Imperial College London in 2005. From July 2005 to June 2010, he worked with Queen's University Belfast, Imperial College, and Lancaster University. Since October 2010 he has been with Newcastle University, currently as a reader. His research interests are cross-layer optimization, cooperative diversity, statistical signal processing, and information theory.

SAID BOUSSAKTA [S'89, M'90, SM'04] received his Ingenieur d'Etat degree in electronic engineering from the National Polytechnic Institute of Algiers (ENPA), Algeria, in 1985 and his Ph.D. degree in electrical engineering (signal and image processing) from the University of Newcastle upon Tyne, United Kingdom, in 1990. From 1990 to 1996 he was with the University of Newcastle upon Tyne as a senior research associate in digital signal and image processing. From 1996 to 2000 he was with the University of Teesside, United Kingdom, as a senior lecturer in communication engineering. From 2000 to 2006 he was with the University of Leeds as a reader in digital communications and signal processing. He is currently a professor of communications and signal processing at the School of Electrical, Electronic and Computer Engineering, University of Newcastle upon Tyne, where he is lecturing on advanced communication networks and signal processing subjects. His research interests are in the areas of fast DSP algorithms, digital communications, communications networks systems, cryptography and security, and digital signal/image processing. He has authored and co-authored more than 250 publications. He is a Fellow of the IET.

ABBAS JAMALIPOUR [S'86, M'91, SM'00, F'07] received his Ph.D. degree from Nagoya University, Japan. He is the Professor of Ubiquitous Mobile Networking with the School of Electrical and Information Engineering, University of Sydney, Sydney, NSW, Australia. He is a Fellow of the Institute of Electrical, Information, and Communication Engineers (IEICE) and the Institution of Engineers Australia, an IEEE Distinguished Lecturer, and a Technical Editor of several scholarly journals. He has been a Chair of several international conferences, including IEEE ICC and IEEE GLOBECOM, General Chair of IEEE WCNC '10, as well as technical program chair of IEEE PIMRC '12 and IEEE ICC '14. He is Vice President — Conferences and a member of the Board of Governors of the IEEE Communications Society (ComSoc). He is the recipient of several prestigious awards, including the 2010 IEEE ComSoc Harold Sobol Award for Exemplary Service to Meetings and Conferences, the 2006 IEEE ComSoc Distinguished Contribution to Satellite Communications Award, and the 2006 IEEE ComSoc Best Tutorial Paper Award.

To understand future research directions in MCC network modeling, it is important to once again recall the ultimate aim of such modeling. A network model is required to provide a version of the environment or a setup where developed communication techniques will operate.

ON GREENING CELLULAR NETWORKS VIA MULTICELL COOPERATION

TAO HAN AND NIRWAN ANSARI, NEW JERSEY INSTITUTE OF TECHNOLOGY

ABSTRACT

Recently, green communications has received much attention. Cellular networks are among the major energy hoggers of communication networks, and their contributions to the global energy consumption increase fast. Therefore, greening cellular networks is crucial to reducing the carbon footprint of information and communications technology. In this article, we overview the multicell cooperation solutions for improving the energy efficiency of cellular networks. First, we introduce traffic-intensity-aware multicell cooperation, which adapts the network layout of cellular networks according to user traffic demands in order to reduce the number of active base stations. Then we discuss energy-aware multicell cooperation, which offloads traffic from on-grid base stations to off-grid base stations powered by renewable energy, thereby reducing the on-grid power consumption. In addition, we investigate improving the energy efficiency of cellular networks by exploiting coordinated multipoint transmissions. Finally, we discuss the characteristics of future cellular networks, and the challenges in achieving energy-efficient multicell cooperation in future cellular networks.

INTRODUCTION

Greening is not merely a trendy concept, but is becoming a necessity to bolster social, environmental, and economic sustainability. Naturally, green communications has received much attention recently. As cellular network infrastructures and mobile devices proliferate, an increasing number of users rely on cellular networks in their daily lives. As a result, the energy consumption of cellular networks keeps increasing. Therefore, greening cellular networks is attracting tremendous research efforts in both academia and industry [1]. Meanwhile, it has been shown that, with the aid of multicell cooperation, the performance of a cellular network in terms of throughput and coverage can be enhanced significantly. However, the potential of multicell cooperation on improving the energy efficiency of cellular networks remains to be unlocked. This article overviews the multicell cooperation solutions for improving the energy efficiency of cellular networks.

The energy consumption of a cellular network is mainly drawn from base stations (BSs), which account for more than 50 percent of the energy consumption of the network. Thus, improving the energy efficiency of BSs is crucial to green cellular networks. Taking advantage of multicell cooperation, the energy efficiency of cellular networks can be improved from three aspects. The first one is to reduce the number of active BSs required to serve users in an area [2]. The solutions are to adapt the network layout according to traffic demands. The idea is to switch off BSs when their traffic loads are below a certain threshold for a certain period of time. When some BSs are switched off, radio coverage and service provisioning are taken care of by their neighboring cells. The second aspect is to associate users with green BSs powered by renewable energy. Through multicell cooperation, off-grid BSs enlarge their service area while on-grid BSs shrink their service area. Zhou *et al.* [3] proposed the handover parameter tuning algorithm and power control algorithm to guide mobile users to associate with BSs powered by renewable energy, thus reducing on-grid power expenses. Han and Ansari [4] proposed an energy-aware cell size adaptation algorithm named ICE. This algorithm balances the energy consumption among BSs, enables more users to be served with green energy, and therefore reduces the on-grid energy consumption. Envisioning future BSs to be powered by multiple types of energy sources (e.g., the grid, solar energy, and wind energy), Han and Ansari [5] proposed to optimize the utilization of green energy for cellular networks by cell size optimization. The proposed algorithm achieves significant main grid energy savings by scheduling the green energy consumption along the time domain for individual BSs, and balancing the green energy consumption among BSs for the cellular network.

The third aspect is to exploit coordinated multipoint (CoMP) transmissions to improve the energy efficiency of cellular networks [6]. On one hand, with the aid of multicell cooperation, the energy efficiency of BSs on serving cell edge users is increased. On the other hand, the coverage area of BSs can be expanded by adopting multicell cooperation, thus further reducing the number of active BSs required to cover a certain area. In addition to discussing the multicell

[1] *The static power consumption of a BS refers to the power consumption of the BS when there are no active users in the coverage of the BS.*

 1536-1284/13/$25.00

cooperation solutions, we investigate the challenges for multicell cooperation in future cellular networks.

Traffic Intensity Aware Multicell Cooperation

There are two reasons traffic demands of cellular networks experience fluctuations. The first one is the typical day-night behavior of users. Mobile users are usually more active in terms of cell phone usage during the day than during the night, and therefore, traffic demands during the day are higher than those at night. The second reason is the mobility of users. Users tend to move to their office districts during working hours and come back to their residential areas after work. This results in the need for large capacity in both areas at peak usage hours, but reduced requirements during off-peak hours. However, cellular networks are usually dimensioned for peak hour traffic, and thus, most of the BSs work at low work load during off-peak hours. Due to the high static power consumption,[1] BSs usually experience poor energy efficiency when they are operating under a low work load. In addition, cellular networks are typically optimized for the purpose of providing coverage rather than operating at full load. Therefore, even during peak hours, the utilization of BSs may be inefficient regarding energy usage. Adapting the network layout of cellular networks according to traffic demands has been proposed to improve the energy efficiency of cellular networks. The network layout adaptation is achieved by dynamically switching BSs on/off. Figure 1 shows several scenarios of network layout adaptations. Figure 1a shows the original network layout, in which each BS has three sectors. For the green cell, if most of the traffic demands on it are coming from sector three, and the traffic demands in sectors one and two are lower than a predefined threshold, the green cell could switch off sectors one and two to save energy, and the users in the sectors that are switched off will be served by the neighboring cells. In this case, the network layout after the adaptation is shown in Fig. 1b. If traffic demands from sector three of the green cell also decrease below the threshold, the entire green cell is switched off, and the network layout is adapted, as shown in Fig. 1c. Under this scenario, the radio coverage and service provisioning in the green cell are taken care of by its active neighboring cells. When a BS is switched off, the energy consumed by its radio transceivers, processing circuits, and air conditioners can be saved. Therefore, adapting the network layout of cellular networks according to traffic demands can save a significant amount of energy consumed by cellular networks.

While network layout adaptation can potentially reduce the energy consumption of cellular networks, it must meet two service requirements:

- The minimum coverage requirement
- The minimum quality of service (QoS) requirements of all mobile users

Therefore, in carrying out network layout adaptation, multicell cooperation is needed to guar-

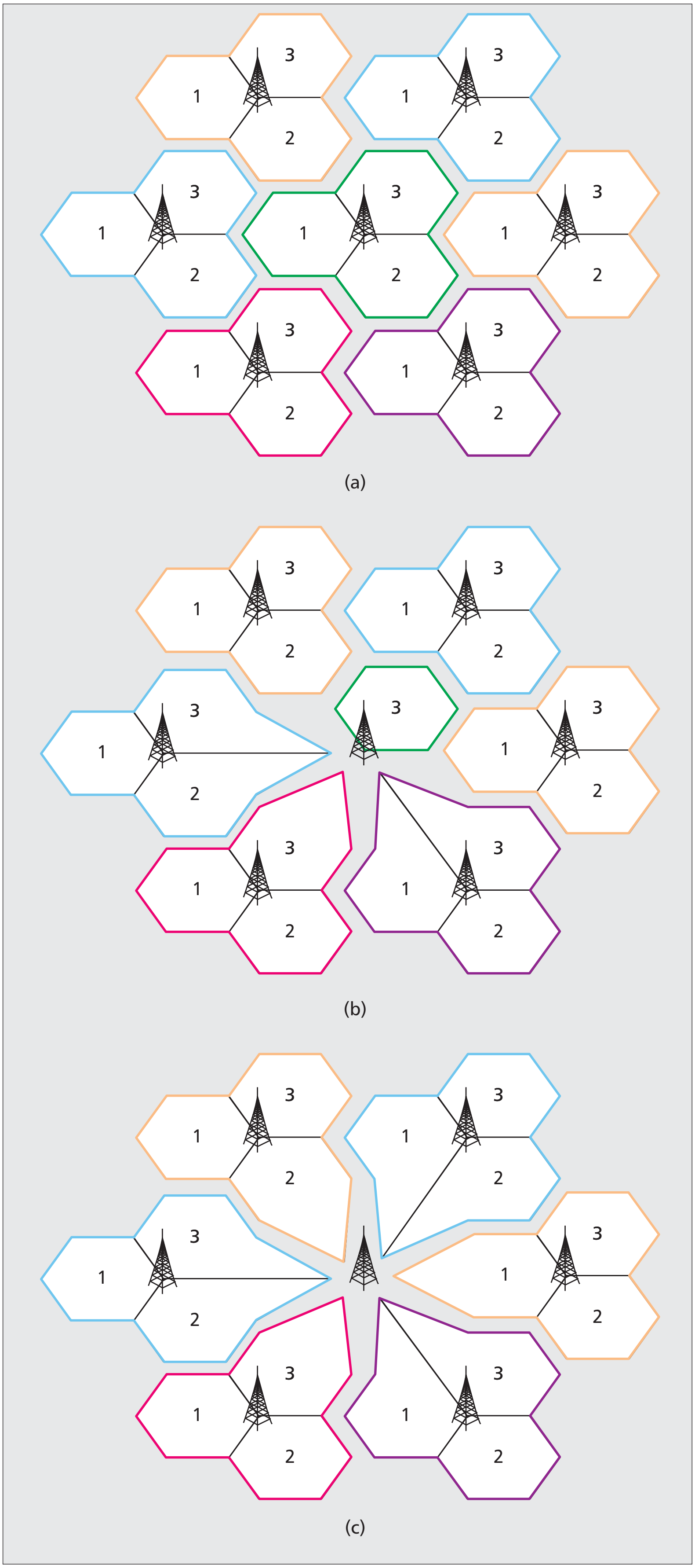


Figure 1. *Illustrations of network layout adaptations for macro cellular sites: a) original network layout; b) BS partially switched off; c) BS entirely switched off.*

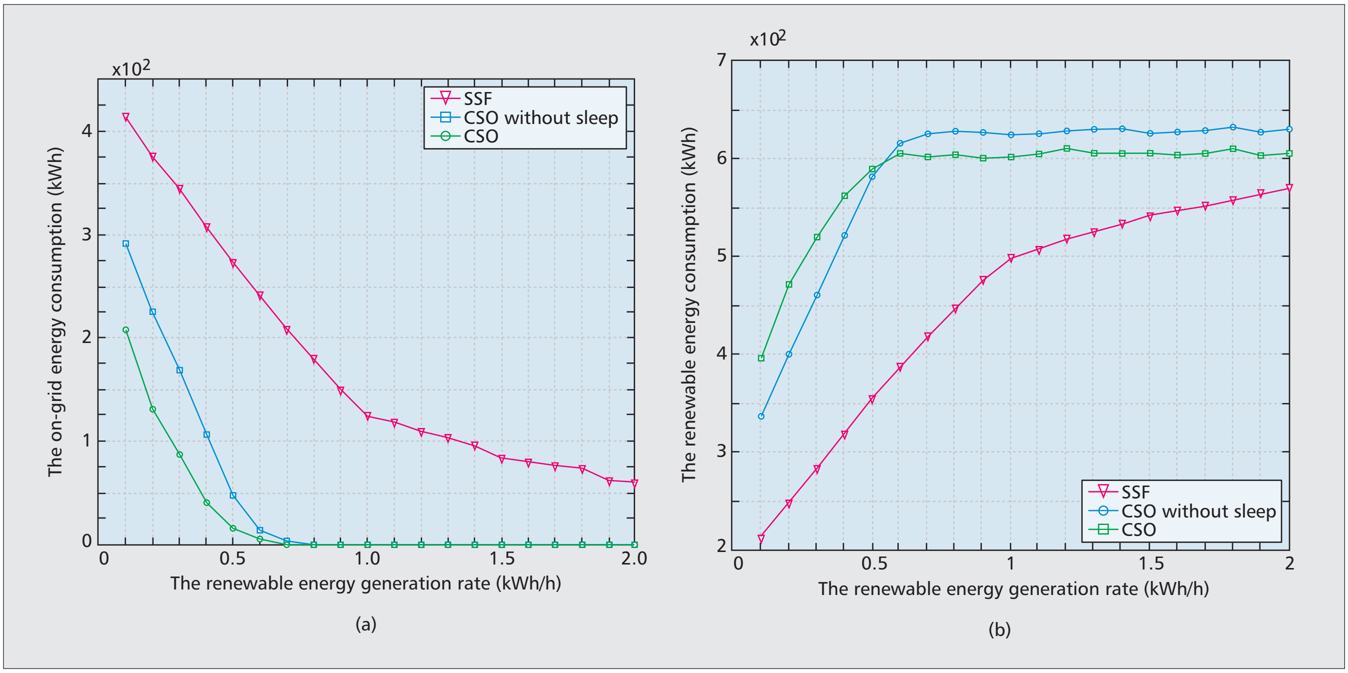


Figure 2. *Energy utilization of the cellular network: a) on-grid energy consumption; b) renewable energy consumption.*

antee service requirements. Otherwise, it will result in a high call blocking probability and severe QoS degradation. For example, two adjacent BSs both experience low traffic demands. However, only one BS can be switched off to save energy, and the other BS should be active to sustain the service provisioning in both coverage areas. In this case, if both BSs are switched off according to their own traffic demands, their subscribers will lose connections. Therefore, cooperation among BSs is essential to enable traffic-intensity-aware network layout adaptation.

Cooperation to Estimate Traffic Demands

Network layout adaptation is based on the estimated traffic demands at individual cells. The traffic demand estimation at individual cells requires the cooperation of their neighboring cells. To avoid frequently switching BSs on/off, they are switched off only when their traffic demands are less than a predefined threshold for a minimum period, *T*. Therefore, the estimated traffic demands should represent the traffic demands at individual cells for at least a time period of length *T*. Hence, traffic demands at a BS consist of three parts: traffic demands from users who are currently attached to the BS, traffic demands from users who will hand over to other BSs, and traffic demands from users who will hand over to the BS from its neighboring BSs. While the first two components can be measured and estimated by individual BSs, the estimation of the third component requires cooperation from neighboring cells. Two factors contribute to the handover traffic from the neighboring cells. The first one is the user mobility. A user's motion, including the direction and velocity, can be measured by various signal processing methods. Therefore, individual BSs can predict:

- How many users will hand over to other cells in the near future
- To which cells these users are highly likely to hand over

Such information is important for their neighboring BSs to estimate their traffic demands. The second reason for handover traffic is the switching off of neighboring BSs. If one of the neighboring cells is switched off, the users under its coverage will be handed over to its neighboring cells, thus increasing traffic demands of the neighboring cells. Therefore, cooperation among radio cells is important for the traffic demand estimation at individual cells.

Cooperation to Optimize the Switching Off Strategy

With the traffic demand estimation, cellular networks optimize the switching on/off strategies to maximize the energy saving while guaranteeing users' minimal service requirements. Currently, most existing switching on/off strategies are centralized algorithms, which assume that there is a central controller that collects the operation information of all BSs and optimizes network layout adaptations. Three methods have been proposed to determine which BSs are to be switched off. The first one is randomly switching off BSs with low traffic loads. This method mainly applies to BSs in the night zone, where few users are active. The method randomly switches off some BSs to save energy, and the remaining BSs provide coverage for the area. The second method is a greedy algorithm, which enforces BSs with higher traffic loads to serve more users and switches off BSs with no traffic load. The third method is based on the user–BS distance. The required transmission power of BSs for serving users depends on the distance between users and BSs. The

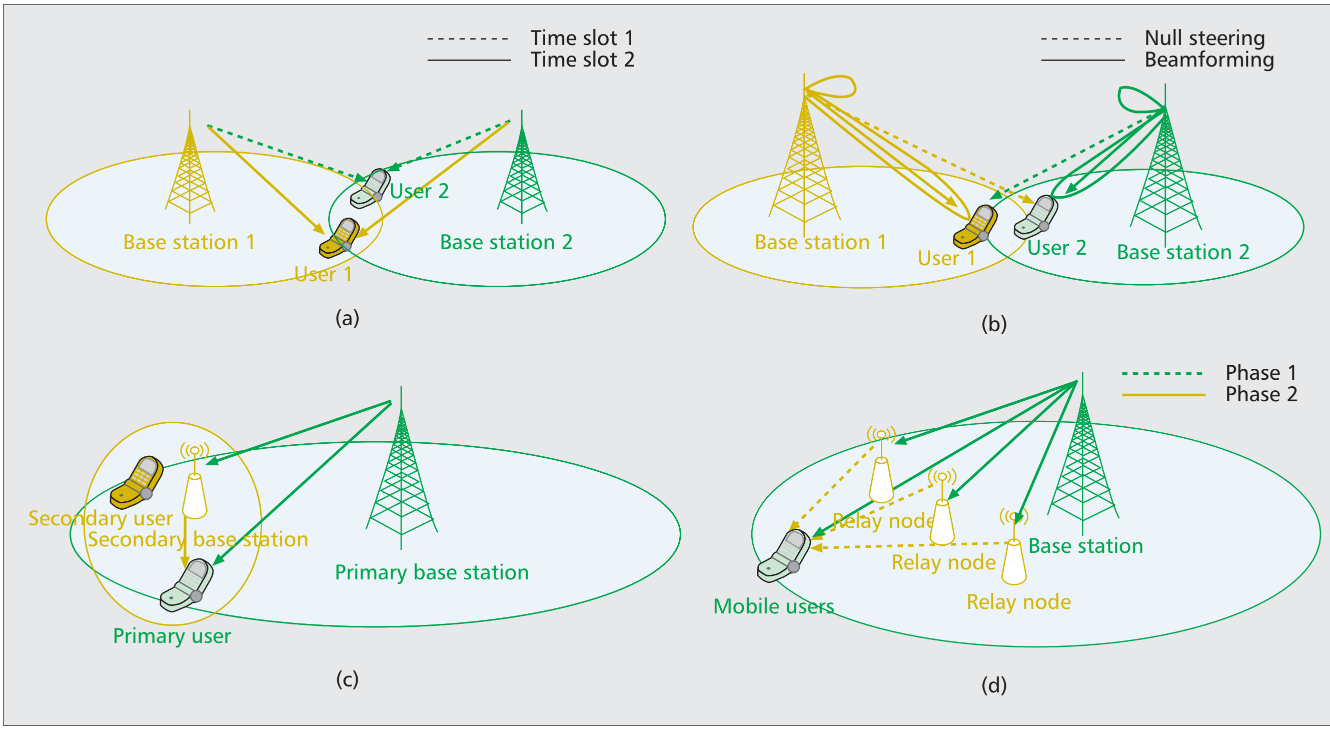


Figure 3. *Energy utilization of the cellular network: a) joint transmission; b) cooperative beamforming; c) cooperative relaying; d) distributed space-time coding.*

longer the distance, the greater the transmission power required in order to meet users' minimal service requirements. Therefore, the user–BS distance can be an indicator of the energy efficiency of cellular networks: the shorter the average user–BS distance, the higher the energy efficiency. Hence, the algorithm tends to switch off BSs with the longest user–BS distance to improve the energy efficiency of cellular networks.

Centralized algorithms, however, require the channel state information and traffic load information of every cell. Collecting this information centrally may impose tremendous communication overheads, and thus reduce the effectiveness of centralized algorithms in improving energy efficiency. Therefore, distributed algorithms are favored, especially for heterogeneous networks that consist of various types of cells such as macrocells, microcells, picocells, and femtocells. To enable distributed algorithms, individual cells may cooperatively form coalitions, and share the channel state information and traffic load information. Based on the shared information, individual BSs optimize their operation strategies to minimize the total energy consumption of the BS coalition.

Energy Aware Multicell Cooperation

With the worldwide penetration of distributed electricity generation at medium and low voltages, it is proved that renewable energy such as sustainable biofuels, and solar and wind energy can be effectively utilized to substantially reduce carbon emission. Therefore, researchers have proposed to power BSs with renewable energy to save the on-grid energy consumption and reduce the carbon footprint. Nokia Siemens Networks [7] has developed off-grid BSs that rely on a combination of solar and wind power supported by fuel cell and deep cycle battery technologies. Compared to on-grid BSs, off-grid BSs achieve zero carbon emission and save a significant amount of on-grid energy. However, due to the dynamic nature of renewable resources, renewable energy generation highly depends on environmental factors such as temperature, light intensity, and wind velocity. In addition, because of the limited battery capacity, generated energy should be utilized well to avoid energy overflow. Therefore, how to optimize the utilization of renewable energy in cellular networks is not a trivial problem. One intuitive solution for this problem is the energy-aware cell breathing method. Here, "energy-aware" pertains to two perspectives. The first one is the awareness of energy sources such that users are guided to draw energy from off-grid BSs rather than on-grid BSs. This can be achieved by enlarging the cell sizes of off-grid BSs and shrinking those of on-grid BSs [3]. The second perspective is the awareness of the amount of energy storage such that off-grid BSs with higher amounts of energy storage are forced to serve a larger area. In this way, energy overflow can be avoided, thus enabling the renewable energy generation system to harvest more energy from renewable sources [4]. The energy-aware cell breathing technique is a step forward to further traffic-intensity-aware network layout adaptations. In addition to the traffic load information and channel state information, energy-aware cell breathing requires multiple cells to cooperatively share the energy informa-

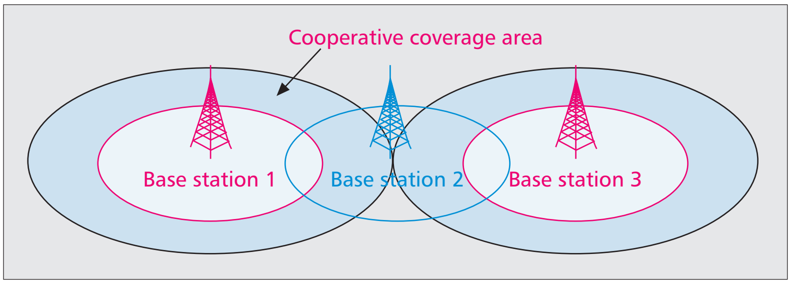


Figure 4. *Cooperation to increase coverage area.*

tion, including the estimated amount of energy arrival and the amount of energy storage, and to coordinate to minimize the energy consumption of cellular networks.

A Case Study

Envisioning future BSs powered by multiple types of energy sources (e.g., on-grid energy and renewable energy), we investigate the optimal energy utilization strategies in cellular networks with both on-grid and off-grid BSs. In such cellular networks, BSs are powered by renewable energy if they have enough renewable energy storage; otherwise, the BSs switch to on-grid energy to serve mobile users. To optimize energy utilization, BSs cooperatively determine their cell size through energy-aware cell breathing based on the proposed cell size optimization (CSO) algorithm [5]. By cooperative cell breathing, the on-grid energy consumption of the cellular network over a period of time consisting of several cell size adaptations is minimized. The optimal energy utilization strategy includes two parts: the multi-stage energy allocation policy and the single-stage energy consumption minimization. The multi-stage energy allocation policy determines the amount of renewable energy allocated at individual off-grid BSs during each cell size update. The single-stage energy consumption minimization involves two steps. The first step is to minimize the number of active on-grid BSs by offloading traffic cell breathing. On-grid BSs shrink their cell sizes to enforce the associated users to hand over to off-grid BSs, and the off-grid BSs with higher amounts of energy storage enlarge their cell sizes to absorb more users. The second step is to minimize the number of active BSs (both on-grid and off-grid) by applying the sleep mode checking algorithm. When traffic demands on individual BSs are lower than a threshold, the BSs seek to put themselves into sleep mode by cooperating with their neighboring cells. If the BS is in sleep mode, its associated users will be handed over to its neighboring cells. In this case, if the neighboring BSs have enough energy storage to serve the handed over users, and the total energy consumption of the cellular network is reduced, the BS is switched into sleep mode to save energy.

Figure 2 shows the comparisons between the CSO algorithm and the strongest-signal first method, which always associates a user with the BS with the strongest received signal strength. Figure 2a shows the on-grid energy consumption at different renewable energy generation rates. In this simulation, there are 400 mobile users in the mobile network. As the renewable energy generation rate increases, the on-grid energy consumption of the mobile network is reduced. When the renewable energy generation rate is larger, more electricity is generated from renewable energy. Therefore, more BSs can serve mobile users using electricity generated by renewable energy instead of consuming on-grid energy. When the renewable energy generation rate is larger than 0.6 kWh/h, the CSO algorithm achieves zero on-grid energy consumption by optimizing the cell size of each BS, while the SSF algorithm still consumes a significant amount of on-grid energy. When the renewable energy generation rate is low, the CSO algorithm saves more than 200 kWh electricity than the SSF algorithm. As the generation rate increases, the gap between the energy consumption of CSO and that of SSF narrows because the CSO algorithm has already achieved zero energy consumption, and the energy consumption of SSF reduces due to increased availability of renewable energy. Figure 2b shows the renewable energy consumption at different energy generation rates. When the energy generation rate is less than 0.7 kWh/h, the renewable energy consumptions of all the simulated algorithms increase as the renewable energy generation rate increases. This is because the larger the renewable energy generation rate, the more renewable energy can be used to serve mobile users. In addition, the CSO algorithm consumes more renewable energy than the SSF algorithm does because the CSO algorithm optimizes the cell sizes of BSs and maximizes the utilization of renewable energy in order to reduce the on-grid energy consumption. When the renewable energy generation rate is larger than 0.7 kWh/h, the renewable energy consumption of the CSO algorithm does not change much, while that of SSF keeps increasing. This is because when the renewable energy generation rate is larger than 0.7 kWh/h, the on-grid energy consumption of the CSO algorithm is zero, as shown in Fig. 2a. This indicates that all the users are served by renewable energy. Therefore, the renewable energy consumption of the CSO algorithm reflects the total energy consumption of the network, which is similar under different energy generation rates in the simulation.

Energy Efficient CoMP Transmission

Coordinated multipoint (CoMP) transmission is a key technology for future mobile communication systems. In CoMP transmission, multiple BSs cooperatively transmit data to mobile users to improve their receiving signal quality. BSs share the required information for cooperation via high-speed wired links. Based on the information, BSs determine joint processing and coordinated scheduling strategies. While CoMP transmission has been proved to be effective in improving the data rate and spectral efficiency of cell edge users [8], its potential for improving the energy efficiency of cellular networks has not been fully exploited. In this section, we discuss the potentials of CoMP transmission in greening cellular networks.

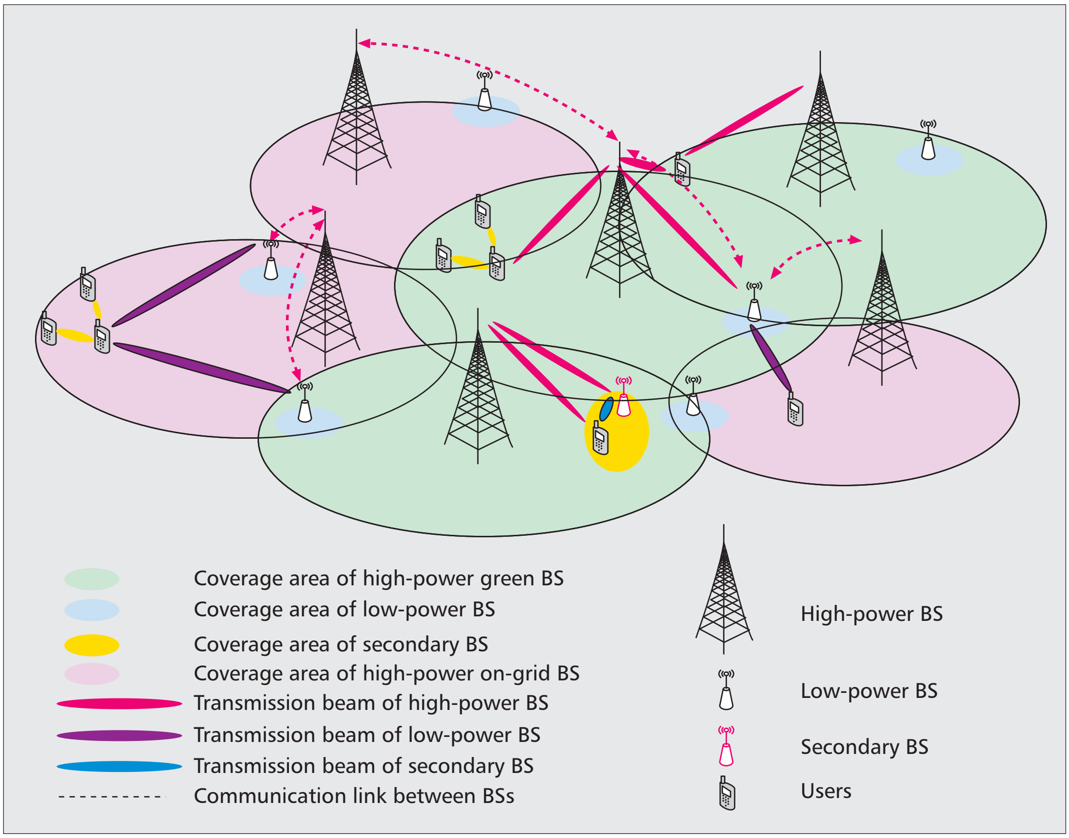


Figure 5. *Future cellular networks with holistic cooperation.*

Owing to joint transmission, cell edge users experience higher receiving signal strength and lower interference, therefore requiring less transmission power from both BSs. Thus, the joint transmission improves energy efficiency of the cell edge communications.

Increase Energy Efficiency for Cell Edge Communications

BSs usually have lower energy efficiency in serving the users at the cell edge than in serving the users within the cell. This is not only because cell edge users are far away from BSs and require more transmission power, but also because cell edge users experience more interference from neighboring cells. Multicell cooperation can improve energy efficiency for cell edge communications [6].

Joint Transmission — Joint transmission exploits the cooperation among neighboring BSs to transmit the same information to individual users located at the cell edge [9]. Figure 3a illustrates the joint transmission scheme. User 2 associates with BS 2 while user 1 associates with BS 1. Instead of serving their associated users independently, BSs 1 and 2 cooperatively transmit information to their users. For example, if both BSs apply time-division multiple access (TDMA) in the first time slot, both BS 1 and 2 transmit the same information to user 2. The user combines the signals from both BSs to decode the information. In the second time slot, the BSs cooperatively serve user 1. Due to joint transmission, cell edge users experience higher receiving signal strength and lower interference, hence requiring less transmission power from both BSs. Thus, joint transmission improves the energy efficiency of the cell edge communications.

Cooperative Beamforming — In cooperative beamforming, a BS forms radio beams to enhance the signal strength of its serving users while forming null steering toward users in its neighboring cells. As shown in Fig. 3b, BS 1 forms the radio beam toward its associated user, user 1, to increase the user's receiving signal strength. On the other hand, in order to reduce the interference to user 2 who is served by a neighboring BS, BS 1 forms the null steering toward user 2. BS 2 executes the same operation as BS 1. Therefore, through cooperative beamforming, the signal-to-noise ratio (SNR) of both user 1 and user 2 is enhanced. Thus, less transmission power is required for BSs to serve the cell edge users, and the energy efficiency of cell edge communications is increased.

Cooperative Relaying — Taking advantage of cognitive radio techniques, cooperation can be exploited between a primary cell and a secondary cell as shown in Fig. 3c. The secondary BS cooperatively relays the signal from the primary BS to the primary user. The primary user combines the signals from both the primary and secondary BSs to decode the information. To stimulate cooperative relaying, the primary BS grants the secondary BS some spectrum that can be utilized for communications between the secondary BS and secondary users. By cooperating with secondary BSs, the transmission power required from primary BSs to serve the cell edge users is reduced. Therefore, the cooperative relay scheme reduces the power consumption of primary BSs [10]. However, due to the time-varying

Taking advantage of the increasing network diversity, BSs have more cooperation opportunities, and thus can achieve additional energy savings. However, realizing optimal multicell cooperation is not trivial.

wireless channels, the scheduling delay involved in the relaying process may increase the outage probability. Fan *et al.* [11] showed that the outage probability caused by the scheduling delay can be alleviated by optimizing the transmission power of the network nodes.

Distributed Space-Time Coding — Distributed space-time coding [12] explores the spatial diversity of both the relay nodes and the mobile users to combat multipath fading, and can be applied to improve the spectral and energy efficiency of cell edge users. As illustrated in Fig. 3d, the distributed space-time coding scheme consists of two phases. In the first phase, the BS broadcasts a data packet to its destination and the potential relays. In the second phase, the potential relays that can decode the data packet cooperatively transmit the decoded data packet to the destination using a suitable space-time code. By exploring spatial diversity, distributed spaced-time coded cooperative transmission improves the diversity gain of cellular networks. However, the multiplexing gain of the wireless system may be sacrificed [13]. Zou *et al.* [13] proposed an opportunistic distributed space-time coding (O-DSTC) scheme to increase the multiplexing gain. Taking advantage of DSTC cooperative transmission, the spectral efficiency is increased. As a result, fewer transmissions are required in order to transmit a given number of data packets. Therefore, the energy efficiency of the cellular networks is enhanced.

ENABLE MORE BSS TO ENTER SLEEP MODE

As discussed in the previous subsection, CoMP transmission (e.g., joint transmission) enhances the receiving SNR of cell edge users, implying that BSs are able to cover a larger area with the same transmission power. Therefore, under low traffic demands, adopting the CoMP transmission can reduce the number of BSs required to cover an area, thus improving the energy efficiency of cellular networks [14].

As presented earlier, traffic-intensity-aware multicell cooperation enables BSs with low traffic demands to go into sleep mode to save energy. The users in the off cells will be served by their active neighboring cells. However, due to the limitation of BSs' maximal transmission power, a few BSs should stay awake to provide coverage in the area. By applying CoMP transmission, the number of required active BSs can be further reduced. As illustrated in Fig. 4, the inner circle of BS 1 is the original coverage area, while the outer circle indicates the coverage area after cooperation with its neighboring BS, BS 3. The green area is the additional coverage area achieved by multicell cooperation. Under the condition of lower traffic demands, if there is no cooperation among BSs, three BSs have to stay awake to cover the whole area. However, if multicell cooperation is enabled, BS 1 and BS 3 can provide services to the users in the area by applying cooperative transmission; thus, BS 2 can be switched into sleep mode. Therefore, the total energy consumption of cellular networks is reduced.

FUTURE CELLULAR NETWORKS WITH MULTICELL COOPERATION

With advances of cellular networks and communication techniques, future cellular networks will be featured as heterogeneous networks in terms of both network deployments and communication techniques, as shown in Fig. 5.

Regarding network deployments, heterogeneous networks refer to deploying a mix of high-power and low-power BSs in order to satisfy the traffic demands of service areas. High-power BSs (e.g., macro and micro BSs) are deployed to provide coverage to a large area, while low-power BSs are deployed to provide high capacity within a small coverage area. In addition, based on energy supplies, BSs can be further divided into two types: on-grid BSs, and off-grid BSs powered by green energy such as solar power and wind power. Therefore, future heterogeneous networks consist of four types of BSs: high-power on-grid BSs, high-power off-grid BSs, low-power on-grid BSs, and low-power off-grid BSs. In terms of technology diversity, heterogeneous networks consist of a variety of technologies such as multiple input multiple output (MIMO), cooperative networking, and cognitive networking. Taking advantage of the increasing network diversity, BSs have more cooperation opportunities, and thus can achieve additional energy savings. However, realizing optimal multicell cooperation is not trivial.

Coalition Formulation — The problem of coalition formation is to determine the coalition size and members of the coalition. Although the problem of coalition formation is well studied in the field of economics, and some game theory methods have been applied to solve the coalition formation problem in wireless networks, few models can be directly applied to form the coalitions of BSs for the purpose of reducing the energy consumption of cellular networks. This is because the original coalition formation problem is to form a coalition to maximize the benefit of all the members in the coalition, while the problem of forming coalitions of BSs is to maximize total benefits of the coalitions. In addition, due to the diversity of BSs and radio access technologies, a successful coalition may consist of BSs and techniques that are complementary to each other to maximize the energy savings of the coalition. Hence, novel models or methods of coalition forming must be developed to facilitate energy-efficient multicell cooperation.

Green Energy Utilization — For the sake of environmental friendliness, off-grid BSs powered by renewable energy are favored over on-grid BSs. In order to optimize the utilization of off-grid BSs, the fundamental design issue is to effectively utilize the harvested energy to sustain traffic demands of users in the network. The optimal utilization of green energy over a period of time depends on the characteristics of the energy arrival and energy consumption at the current stage as well as in future stages, and on the cooperation strategies of the neighboring cells or cooperating cells.

Incentive Mechanism — Note that in future heterogeneous networks, multicell cooperation may involve cells not owned by the same operator. For example, the coalition may consist of a BS from a different operator, a secondary BS operated via cognitive radio, and a relaying cell formed by mobile users. To incentivize cooperation among these cells, a simple but effective incentive mechanism should be exploited.

The coalition may consist of a BS from a different operator, a secondary BS operated via cognitive radio, and a relaying cell formed by mobile users. To incentivize cooperation among these cells, a simple but effective incentive mechanism should be exploited.

CONCLUSION

This article discusses how to reduce the energy consumption of cellular networks via multicell cooperation. We focus on three multicell cooperation scenarios that enhance energy efficiency of cellular networks. The first one is traffic-intensity-aware multicell cooperation, in which multiple cells cooperatively estimate traffic demands and adapt the network layout based on the estimated traffic demands. Through network layout adaptation, the number of active BSs can be reduced, thus reducing the energy consumption of cellular networks. The second scenario is energy-aware multicell cooperation, in which off-grid BSs powered by green energy are enforced to serve a large area to reduce the on-grid power consumption. The third one is energy-efficient CoMP transmission, in which the overall energy consumption is reduced by improving the energy efficiency of BSs in serving cell edge users. In addition, we discuss the diversity of future cellular networks in terms of radio access technologies and BS characteristics, and the challenges to optimally exploit multicell cooperation by capitalizing on diversities.

REFERENCES

[1] L. Correia *et al.*, "Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks," *IEEE Commun. Mag.*, vol. 48, no. 11, Nov. 2010, pp. 66–72.

[2] Z. Niu, "Tango: Traffic-Aware Network Planning and Green Operation," *IEEE Wireless Commun.,* vol. 18, no. 5, Oct. 2011, pp. 25–29.

[3] J. Zhou *et al.*, "Energy Source Aware Target Cell Selection and Coverage Optimization for Power Saving in Cellular Networks," *Proc. 2010 IEEE/ACM Int'l Conf. Green Computing and Commun. & Int'l Conf. Cyber, Physical and Social Computing*, Hangzhou, China, Dec. 2010.

[4] T. Han and N. Ansari, "ICE: Intelligent Cell Breathing to Optimize the Utilization of Green Energy," *IEEE Commun. Letters*, vol. 16, no. 6, June 2012, pp. 866–69.

[5] T. Han and N. Ansari, "Optimizing Cell Size for Energy Saving in Cellular Networks with Hybrid Energy Supplies," *IEEE GLOBECOM '12*, Dec. 2012, pp. 3–7.

[6] F. Heliot, M. Imran, and R. Tafazolli, "Energy Efficiency Analysis of Idealized Coordinated Multi-Point Communication System," *IEEE VTC-Spring '11,* Budapest, Hungary, May 2011.

[7] "E-Plus, Nokia Siemens Networks Build Germany's First Off-Grid Base Station," http://www.nokiasiemensnetworks.com/news-events/press-room/press-releases/e-plus-nokia-siemens-networks-build-germany-s-first-off-grid-base-station

[8] R. Irmer *et al.*, "Coordinated Multipoint: Concepts, Performance, and Field Trial Results," *IEEE Commun. Mag.*, vol. 49, no. 2, Feb. 2011, pp. 102–11.

[9] M. Sawahashi *et al.*, "Coordinated Multipoint Transmission/Reception Techniques for LTE-Advanced (Coordinated and Distributed MIMO)," *IEEE Wireless Commun.*, vol. 17, no. 3, June 2010, pp. 26–34.

[10] S. Jayaweera, M. Bkassiny, and K. Avery, "Asymmetric Cooperative Communications based Spectrum Leasing Via Auctions in Cognitive Radio Networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, Aug. 2011, pp. 2716–24.

[11] L. Fan *et al.*, "Outage Probability Analysis and Power Allocation for Two-Way Relay Networks with User Selection and Outdated Channel State Information," *IEEE Commun. Letters*, vol. 16, no. 5, May 2012, pp. 638–41.

[12] J. N. Laneman and G. W. Wornell, "Distributed Space-Time-Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks," *IEEE Trans. Info. Theory*, vol. 49, no. 10, Oct. 2003, pp. 2415–25.

[13] Y. Zou, Y.-D. Yao, and B. Zheng, "Opportunistic Distributed Space-Time Coding for Decode-and-Forward Cooperation Systems," *IEEE Trans. Signal Proc.*, vol. 60, no. 4, Apr. 2012, pp. 1766–81.

[14] M. Herlich and H. Karl, "Reducing Power Consumption of Mobile Access Networks with Cooperation," *2nd Int'l. Conf. Energy-Efficient Computing and Networking*, New York City, NY, May 2011.

BIOGRAPHIES

TAO HAN [S'08] received his B.E. in electrical engineering and M.E. in computer engineering from Dalian University of Technology and Beijing University of Posts and Telecommunications, respectively. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering, New Jersey Institute of Technology (NJIT), Newark. His research interests include mobile and cellular networks, network optimization, and energy-efficient networking.

NIRWAN ANSARI [S'78, M'83, SM'94, F'09] received his B.S.E.E. (summa cum laude with a perfect GPA) from NJIT, his M.S.E.E. from the University of Michigan, Ann Arbor, and his Ph.D. from Purdue University, West Lafayette, Indiana. He joined NJIT in 1988, where he is a professor of electrical and computer engineering. He has also assumed various administrative positions at NJIT. He has been a visiting (chair/honorary) professor at several universities. His current research focuses on various aspects of broadband networks and multimedia communications. He has served on the Editorial/Advisory Boards of eight journals. He was elected to serve on the IEEE Communications Society (ComSoc) Board of Governors as a member-at-large (2013–2014) as well as IEEE Region 1 Board of Governors as the IEEE North Jersey Section Chair. He has chaired ComSoc technical committees, and has actively organized numerous IEEE international conferences/symposia/workshops, assuming leadership roles as Chair or TPC Chair for various conferences, symposia, and workshops. He has authored *Computational Intelligence for Optimization* (Springer 1997) with E.S.H. Hou and *Media Access Control and Resource Allocation for Next Generation Passive Optical Networks* (Springer, 2013) with J. Zhang, and edited *Neural Networks in Telecommunications* (Springer 1994) with B. Yuhas. He has also contributed over 400 publications, over one third of which were published in widely cited refereed journals/magazines. He has been granted over 15 U.S. patents. He has also guest edited a number of special issues, covering various emerging topics in communications and networking. He has frequently been selected to deliver keynote addresses, distinguished lectures, and tutorials. Some of his recent recognitions include a couple of best paper awards, several Excellence in Teaching Awards, a Thomas Alva Edison Patent Award (2010), an NJ Inventors Hall of Fame Inventor of the Year Award (2012), and designation as a ComSoc Distinguished Lecturer (2006–2009).

OPTIMAL TRADE-OFF BETWEEN POWER SAVING AND QOS PROVISIONING FOR MULTICELL COOPERATION NETWORKS

XI ZHANG AND PING WANG, TEXAS A&M UNIVERSITY

The authors propose the semi-Markov decision process model-based stochastic optimization scheme for the optimal trade-off between power saving and QoS provisioning over multicell cooperation networks.

This work was supported in part by the U.S. National Science Foundation under Grant CNS-205726 and the U.S. National Science Foundation CAREER Award under Grant ECCS-0348694.

ABSTRACT

Multicell cooperation is emerging as a promising wireless communication technique to significantly improve the performance of cellular networks in terms of the coverage of high-data-rate services and the system throughput for the next generation wireless networks. On the other hand, multicell cooperation also imposes the new challenges associated with the two mutually conflicting constraints between the power/energy efficiency/saving and the QoS provisioning required by mobile users, which differ from the conventional cellular networks. To overcome the above problem, we propose the semi-Markov decision process model-based stochastic optimization scheme for the optimal trade-off between power saving and QoS provisioning over multicell cooperation networks. Our objective is to efficiently minimize the power consumption at base stations while guaranteeing diverse QoS provisioning for mobile users through multicell cooperation power scheduling. To achieve this goal, we apply finite-state Markov chains to model the mobile users' density, mobility, multicell cooperation power-profile scheduling, and QoS provisioning. Using these models, we formulate the SMDP-based stochastic optimization problem and develop an efficient iteration algorithm to implement our SMDP-based scheme. We evaluate our proposed schemes through numerical analyses, which show that our proposed schemes significantly outperform the other schemes in terms of minimizing power consumption while guaranteeing the required QoS provisioning.

INTRODUCTION

Recent years have witnessed the rapid emergence of a variety of multimedia wireless services over cellular networks, such as video conferences, mobile TV, music downloading, online gaming, and so on. Motivated by the concept of ubiquitous computing, these promising wireless services are designed to be available for people anytime and anywhere. Quality of service (QoS) provisioning plays a crucial role in supporting multimedia wireless services in cellular networks. Consequently, the research community has made significant efforts in the development of various advanced and efficient wireless cellular networking architectures to support the diverse QoS requirements. Among them, multicell cooperation networking is receiving significant research attention from both academia and the industry. This is because multicell cooperation enables significant performance improvement of cellular networks in terms of the coverage of high-data-rate services and system throughput. As a result, multicell cooperation is being adopted in various communication standards, including Third Generation Partnership Project (3GPP) [1] Long Term Evolution (LTE)-Advanced , or LTE-A [2], and IEEE 802.16m [3] for the next generation wireless networks.

However, multicell cooperation also imposes the new challenges associated with the two mutually conflicting constraints between the power/energy efficiency/saving and the QoS provisioning required by mobile users, which differ from conventional cellular network systems. On one hand, to ensure the demanded wireless transmission capacity for guaranteeing the stringent QoS provisioning over time-varying channels covering the sizable wireless cells, a large amount of power supply and consumption is desired/needed at the base stations to sustain the required wireless channel qualities [4, 5]. On the other hand, one of the main design goals/criterions of multicell cooperation networks is to minimize the power/energy consumption [1, 6] among all base stations through multicell cooperation and power scheduling, which, however, may significantly affect the QoS provisioning required by mobile users. Some standardization and research efforts have been made to address these problems. For instance, as a promising technique under the multicell cooperation framework in the LTE-A standard [2], the cooperative multipoint or coordinated multipoint (CoMP) [7] approach was proposed to improve the coverage of high data rates and increase the overall mulicell cooperation network throughput while efficiently mitigating intercell interference. While CoMP has the potential to improve QoS provisioning, it does not explicitly address any specific/concrete power scheduling schemes which take into account the trade-off between power saving and QoS provisioning over multicell cooperation networks. In [8] the authors studied a joint power and subcarrier allocation algorithm that aims at minimizing network power consumption with a lower-bounded data rate constraint. The authors of [9] developed a sleeping mode control scheme under the given performance requirements. In [10] the authors derived an optimal sleep mode policy to maximize a multiple-objective function of QoS and the power consumption. However, the above-mentioned research works only consider the simple power scheduling action of either totally switching off the base stations or keeping them in a fully operating state in a deterministic manner [11]. As a result, this type of deterministic switching-on/off mode control scheme is less effective for the important application scenarios where the mobile users' traffic loads are relatively heavy and their variations are highly random, especially when taking the users' mobility/location and dynamic traffic load density/distribution variations into account.

To overcome the aforementioned problems, in this article we propose the semi-Markov decision process (SMDP) model-based stochastic optimization scheme to achieve the optimal trade-off between power saving and QoS provisioning over multicell cooperation networks with the dramatic variations in mobile users' traffic load density and mobility. The objective of our proposed stochastic optimization framework is to efficiently minimize the power consumption at the base stations while guaranteeing the diverse QoS provisioning for mobile users through multicell cooperation power scheduling. To achieve this goal, we apply finite-state Markov chain techniques to develop a set of analytic models to characterize mobile users' density, mobility, and multicell cooperation power-profile scheduling. Based on these models, we formulate the SMDP-based stochastic optimization problem to optimize the trade-off between power saving at base stations (BSs) and QoS provisioning to mobile users over multicell cooperation networks. We also develop an efficient iteration algorithm to implement our proposed SMDP-based optimization scheme. Then we conduct numerical analyses to evaluate our proposed schemes, which show that our proposed SMDP-based stochastic control process converges to the unique optimal solution and significantly outperforms the other schemes in terms of minimizing the power consumption at BSs while guaranteeing the required QoS provisioning to mobile users.

The rest of this article is organized as follows. First, we develop the system models for optimizing the trade-off between power saving and QoS provisioning over our multicell cooperation networks. Then, we formulate the SMDP-based stochastic optimization problem to derive the optimal power scheduling policy to optimize the trade-off between power saving and QoS provisioning through multicell cooperation. Finally, we conduct numerical analyses to evaluate our proposed schemes. The article then concludes.

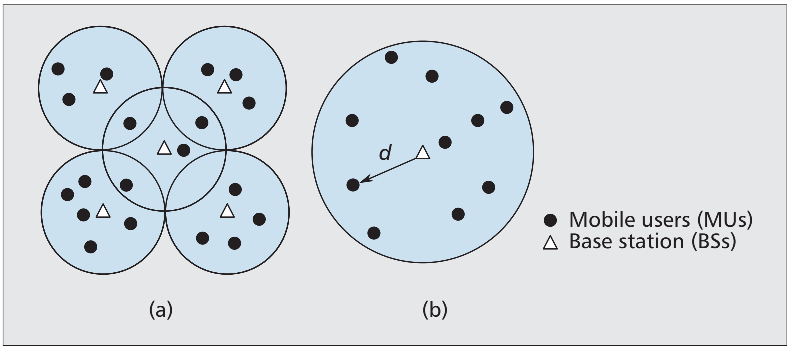


Figure 1. *Wireless cellular network model: a) muticell cooperation network model; b) single cell model with a base station at the center of the cell, where* d *denotes the radius distance from the base station of the cell to the mobile user.*

The System Models

Multicell Cooperation Network Model

We consider cellular networks where multicell cooperation is employed to minimize the power consumption at BSs while guaranteeing QoS provisioning to mobile users (MUs). Our cellular network model is illustrated by an example shown in Fig. 1. It is a cellular network with five cells. As shown in Fig. 1a, one central cell is surrounded by four neighboring cells. BSs are located at the respective center of the cells, illustrated by hollow triangles. As shown in Fig. 1b for each cell's internal architecture, a number of MUs, represented by solid dots, are randomly distributed in the cell with each MU located at the variable radius distance, denoted by d, between the BS and the MU. Clearly, the MUs' traffic load distributions among wireless cells are independent of each other. Within each cell, the BS can perceive or sense the users' information on mobility, dynamic behaviors, and traffic load spatial distributions, which all vary with time. The MUs' traffic load and distributions in each cell are characterized by the number of MUs, which varies with time. Thus, the power control of the BS in each cell needs to be dynamically scheduled and cooperated according to the traffic load variations in order to minimize the total power consumption while guaranteeing the QoS for mobile users.

Traffic Load Model for Mobile Users

For each wireless cell, our MUs' traffic load model consists of two components:
• Mobile users' density model
• Mobile users' mobility model
which are detailed as follows, respectively.

Mobile Users' Density Model — We use the finite-state Markov chain to model the MU density in each wireless cell. Assuming the maximal number of MUs that a cell can support is $(D - 1)$, our MU density Markov chain in each cell can have at most D state levels, with each level representing the number of MUs existing in the cell. Each cell's BS acquires/senses the current state of MU density through the sensing mechanism at the BS. Under the finite-state Markov chain model, the next state of user density in the cell is

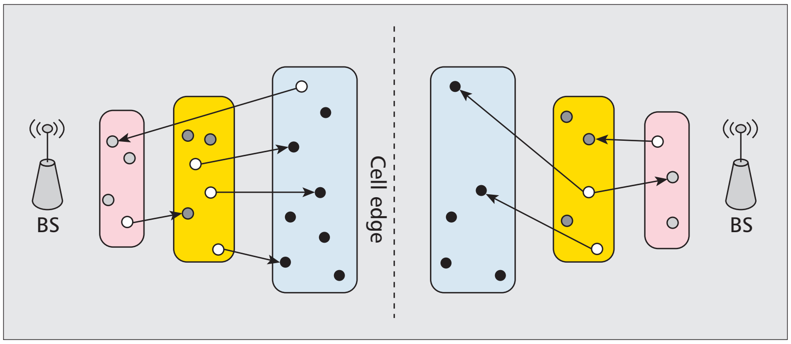


Figure 2. *Mobile user mobility model with different classes (categorized by the different radius distances from the BS to its corresponding MUs) denoted by the three different sizes of boxes.*

determined by the current state of user density and the Markov transition probability. Let $\mathcal{D} = \{0, 1, \ldots, (D-1)\}$ represent the Markov-chain state space of user density in cell i, $N_i(t)$ denote the number of MUs in cell i with $N_i(t) \in \mathcal{D}$, and $\phi_{g_i h_i}(t)$ be the transition probability that the state of user density $N_i(t)$ changes from state g_i to state h_i, where $g_i, h_i \in \mathcal{D}$. Then the state transition probability matrix of user density is defined for cell i: $\Phi_i(t) = [\phi_{g_i h_i}(t)]$, where $\phi_{w_i z_i}(t) = \Pr\{N_i(t+1) = h_i | N_i(t) = g_i\}$, $\forall\, g_i, h_i \in \mathcal{D}$.

Mobile Users' Mobility Model — We use the finite-state Markov chain to model the MU mobility in terms of MUs' spatial distributions in each wireless cell. According to users' different location distributions (e.g., an MU moves from the center location of the BS to the edge of the cell), we can shut down the BS or reduce its power for power/energy saving. To accurately model the impact of user mobility for cell power saving, we apply the k-means clustering algorithm [12, 13] to categorize the MUs into b classes according to the radius distances from the BS to the MUs. As shown in Fig. 2, for example, the MUs in each cell are classified into three classes ($b = 3$) illustrated by three different-sized boxes, respectively. The numbers of users in each class varies as MUs move around. Assuming that there are K ($\leq (D-1)$) users in cell i, the spatial distribution state of k MUs' locations in cell i is denoted by $L_i(t) = (l_{i,1}, \ldots, l_{i,k})$, where $k = \{1, 2, \ldots, K\}$ and $l_{i,k}$ represents the current location (measured by the radius distance from the corresponding BS) of user k in cell i. Since there are b possible locations for each MU, there are totally b^k possible spatial location distributions for any k ($\leq K$) mobile users in cell i. The BS can acquire/sense the current state of MUs' spatial location distribution through a sensing mechanism, and the next state is determined by the current state and the user mobility Markov-chain transition probability. Let $\mathcal{L} = \{L_1, \ldots, L_{bK}\}$ stand for the user mobility Markov chain state space for K users' spatial distributions in a cell, and let $\varphi_{U_iV_i}(t)$ denote the probability that the state of users' location distribution in cell i moves from location state U_i to location state V_i at time t, where $\{U_i, V_i \in \mathcal{L}$. Then we have the $b^K \times b^K$ user mobility Markov chain state transition probability matrix of cell i, which is defined as $\Theta_i(t) = [\varphi_{U_iV_i}(t)]$, where $\varphi_{U_iV_i}(t) = \Pr\{L_i(t+1) = V_i | L_i(t) = U_i\}$, $\forall\, U_i = (u_{i,1}, \ldots, u_{i,k}, \ldots, u_{i,K})$, $V_i = (v_{i,1}, \ldots, v_{i,k}, \ldots, v_{i,K}) \in \mathcal{L}$, where $u_{i,k}$, $v_{i,k}$ are the values that $l_{i,k}$ can take, respectively.

Power Consumption Model for Base Stations

We use the finite-state Markov chain to model the transmit power scheduling in each cell. Assuming our transmit power the Markov chain in each cell has E power-state levels, with each level representing a transmit power level at the BS in a cell. Then we let $\mathcal{E} = \{0, 1, \ldots, (E-1)\}$ represent the power scheduling Markov chain state space in cell i, $J_i(t)$ denote the implemented transmit power of cell i where $J_i(t) \in \mathcal{E}$ at time t, and $\psi_{w_i z_i}(t)$ be the Markov-chain state transition probability that the transmit power state of cell i moves from state w_i to state z_i at time t, where $w_i, z_i \in \mathcal{E}$. The BSs are assumed to be able to transmit at one of E discrete power levels, and each level corresponds to a state in the power scheduling Markov chain. At each decision time, all cells have the associated power profiles denoted $J(t) = [e_1, \ldots, e_i, \ldots, e_N]$, where N denotes the total number of cells in multicell networks, and e_i denotes the transmit power of cell i. Then the transmit power Markov-chain state transition probability matrix of cell i is defined as $\Psi_i(t) = [\psi_{w_i z_i}(t)]$, where $\Psi_{w_i z_i}(t) = \Pr\{E_i(t+1) = z_i | E_i(t) = w_i\ \forall\, w_i\}, z_i \in \mathcal{E}$. For example, we set the BS transmit power to have $E = 4$ different power levels: 0, 5, 10, and 15 W, which are the values we will use in our numerical performance analyses, described later. Also, for the smoothness and convenience of power control, in this article our transmit-power control can only increase or decrease the transmit power across the neighboring transmit power levels. These imply that in this particular case, our power control Markov chain state space in cell i is represented by $\{0, 5, 10, 15\}$, and the Markov chain state transition can only take place across adjacent states.

The total operating power, denoted by $\mathcal{P}$, of a BS to cover a single cell usually includes two parts: a constant part $\mathcal{P}_{const}$, the power that is independent of the number of MUs and their distances, and the dynamic part, which varies proportionately with the traffic intensity denoted by ρ (which is equal to $(D-1)$), supplying the power needed to transmit the signal from the BS to all the mobile users in a cell, and depends on ρ of the BSs. Therefore, the power consumption in cell i can be calculated by $\mathcal{P}_i = \mathcal{P}_{const} + \rho_i \mathcal{P}_{tran}(i)$ [14], where ρ_i is the traffic intensity of cell i, which varies over time, and $\mathcal{P}_{tran}(i)$ is the transmit power dynamically allocated to cell i.

Objective Functions and QoS Constraints for Mobile Users

The proposed optimal power scheduling for multicell cooperation scheme is based on the stochastic control theory. Due to the time-varying traffic distribution of wireless cells, we need to derive the optimal wireless-cell power scheduling policy to efficiently implement multicell cooperation, which aims at minimizing the total power consumption subject to the QoS-threshold constraint of signal-to-interference-plus-noise ratio (SINR). The SINR is calculated for each transmission power allocation and is used to determine the QoS achieved by an MU in the multicell cooperation network. Since the wireless channel quality is a time-varying function of multiple variables and parameters, we can derive its corresponding SINR implemented by MU k as follows:

$$\gamma_k(i) = \frac{\eta \mathcal{P}_{tran}(i)}{[d_k(i)]^{\xi}\left(N_0 + \sum_{j=1, j\neq i}^{N} \mathcal{P}_{tran}(j)\right)} \quad (1)$$

where $\gamma_k(i)$ represents the SINR for mobile user k in cell i, η is the propagation coefficient, ξ is the attenuation coefficient, typically ranging from 2 to 5, $\mathcal{P}_{tran}(i)$ represents the transmit power of cell i, $d_k(i)$ denotes the distance from mobile user k to the BS of cell i, and N_0 denotes the additive white Gaussian noise.

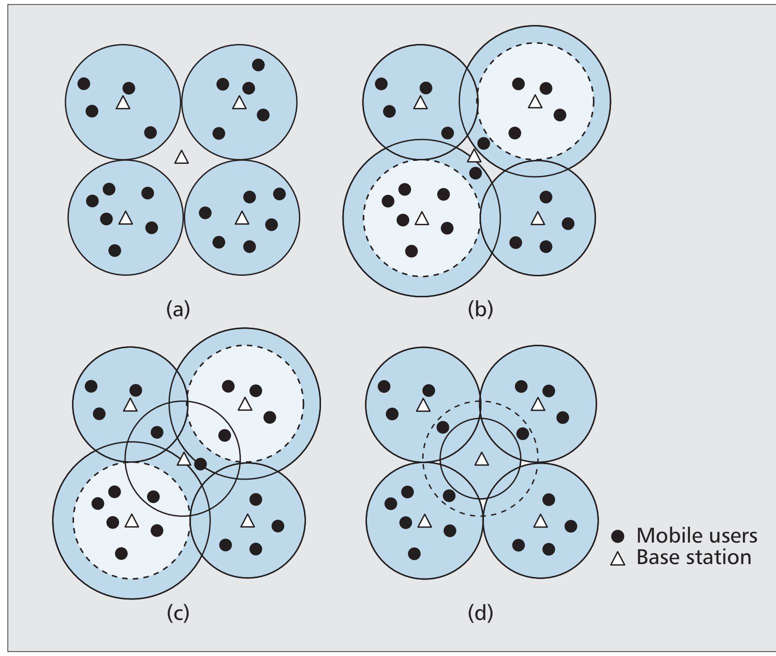


Figure 3. *Multicell cooperation power scheduling with time-varying mobile-users' traffic loads in multicell cooperation networks: a) the central cell is turned off while the transmit power of neighboring cells remain unchanged; b) the central cell is turned off while increasing the transmit power of two neighboring cells through multilevel power control to guarantee the QoS of MUs in the original central cell; c) the central cell remains unchanged while increasing the transmit power of two neighboring cells with multilevel power control to guarantee the QoS provisioning to mobile users in the central cell; d) the central cell decreases its transmit power through multilevel power control due to the reduced mobile users' traffic load while the neighboring cells remain unchanged.*

Stochastic Optimization for Optimal Trade-off between Power Saving and QoS Provisioning

Using the system models of MU density, mobility/spatial distribution, and power-profile scheduling described in "The System Models" section, for multicell cooperation power scheduling, we formulate the stochastic optimization problem by applying the semi-Markov decision process (SMDP) [15] model to achieve the optimal trade-off between power saving and QoS provisioning for multicell cooperation networks. Then we solve the stochastic optimization problem by deriving a set of optimal multilevel power scheduling policies and system reward functions that optimize the trade-off between power saving and QoS provisioning over multicell cooperation networks. Finally, we develop a concrete and efficient iteration algorithm to solve the SMDP-based stochastic optimization, which can yield the optimal decision/action functions to minimize the entire power consumption at BSs while guaranteeing the QoS provisioning required by MUs over multicell cooperation networks.

The SMDP Model and Its State and Action Functions

We observe that the optimization of trade-off between the power saving at BSs and QoS provisioning at MUs over multicell cooperation networks can be efficiently solved by formulating a stochastic optimization problem by using the SMDP model. In particular, we can develop the stochastic reward/utility and constraint functions to maximize the power saving subject to the constraints of QoS provisioning. To apply the SMDP model, we first need to define its state space and action functions. The SMDP state of each cell at time t is characterized by the state of user density, denoted by $N_i(t)$, the state of users' locations $L_i(t)$, and the state of transmit power of each cell, denoted by $E_i(t)$. Then the SMDP state of cell i can be expressed as $s_i(t) = \{N_i(t), L_i(t), E_i(t)\}$, and the state transition probability matrix of the SMDP model is defined as: $\mathbf{P}_i(t) = [\zeta_{s_i s'_i}(t)] = [\Phi_i(t), \Theta_i(t), \Psi_i(t)]$, where $\zeta_{s_i s'_i}(t) \triangleq \phi_{g_i h_i}(t)\varphi_{U_i V_i}(t)\psi_{w_i, z_i}(t)$, which is the SMDP state transition probability that the state of cell i changes from state s_i to state s'_i.

Let $\mathcal{S}$ be the state space of the SMDP model where $\mathcal{S} \triangleq \{\mathcal{D}, \mathcal{L}, \mathcal{E}\}$ and $\mathcal{A}$ are the SMDP model action space, denoted by $\mathcal{A} = \{a_1(t), a_2(t), a_3(t), a_4(t)\}$, respectively, where $a_1(t)$ denotes the action through which the central cell is turned off while the transmit power of neighboring cells remain unchanged; $a_2(t)$ denotes the action threough which the central cell is turned off while increasing (within E levels) the transmit power of two neighboring cells through multilevel power control to guarantee the QoS of MUs in the original central cell; $a_3(t)$ denotes the action through which the central cell remains unchanged while the transmit power of two neighboring cells is increased with multilevel power control to guarantee the QoS provisioning to mobile users in the central cell; and $a_4(t)$ denotes the action through which the central cell decreases (within E levels) its transmit power through multilevel power control due to the reduced MUs' traffic load while the neighboring cells remain unchanged. Given the current SMDP state $s(t) \in \mathcal{S}$, and the selected action $a(t) \in \mathcal{A}$, the SMDP state transition probability function for the next state $s(t + 1)$ is denoted by $\Pr\{s(t + 1)|s(t), a(t)\}$. This function is Marko-

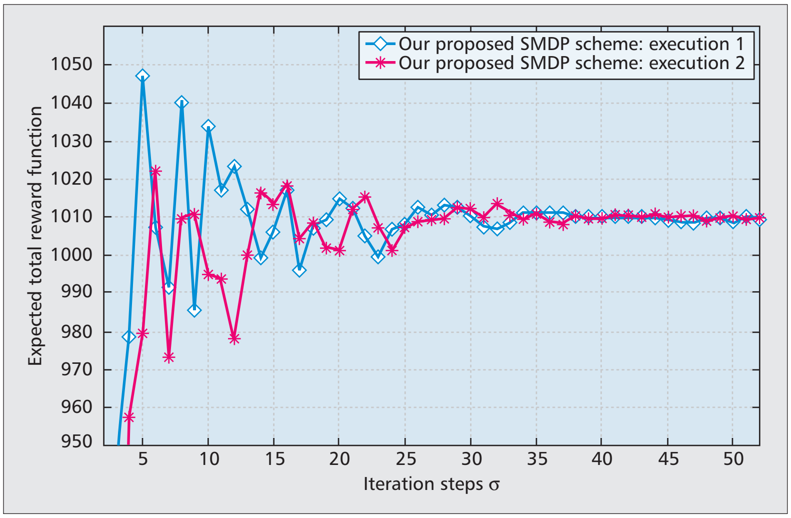


Figure 4. *The expected total reward function $\upsilon_\sigma^\pi(s)$ vs. the SMDP optimization iteration steps σ.*

vian because the state transition probability of the next state is independent of the previous states when given the current state.

In our SMDP model, at each decision epoch, the multicell cooperation power scheduling control has to decide how to allocate the transmit power to each cell according to the changes of cell state in terms of traffic load and power profiles. As indicated in Figs. 3a–d, the multicell cooperation with power scheduling is implemented by the selected action $a(t) \in \mathcal{A}$ at each SMDP model decision epoch.

The SMDP Reward and Policy Functions for Multicell Cooperation Power Scheduling

Since the objective of our proposed scheme is to minimize the total power consumption while guaranteeing QoS provisioning for MUs, we define the system reward function to be the function of power consumption for all cells with QoS provisioning constraints. Thus, we can define the system reward function as follows:

$$R(s(t),a(t)) = \frac{1}{\mathcal{P}(s(t),a(t))} \tag{2}$$

where $\mathcal{P}(s(t), a(t))$ is the total power at all BSs of the multicell cooperation networks as a function of the SMDP state and SMDP action for the optimal trade-off between power saving and QoS guarantee for multicell cooperation power scheduling under the current SMDP state and the action selected in our multicell cooperation networks. Given the current SMDP state and the selected SMDP action, we can derive the power consumption $\mathcal{P}(s(t), a(t))$ for our proposed multicell cooperation power scheduling scheme.

The SMDP decision rules specify a mapping function $\delta(t)$: $\mathcal{S} \to \mathcal{A}$, which selects an action function $a(t)$ for given cells' state at SMDP decision epoch t. Thus, our SMDP policy function, denoted by $\pi \triangleq (\delta(1), \delta(2), \ldots, \delta(t))$, is a sequence of decision rules that are used at all decision epochs $\{1, 2, \ldots, t\}$. Then we can define the expected total reward function of the SMDP model, which is defined as a utility function given in Eq. 3.

The Stochastic Power Saving Optimization and QoS Provisioning Constraints

We need to derive an optimal policy, denoted by π^*, for the optimal trade-off between power saving and QoS provisioning in multicell cooperation networks, which maximizes the expected total reward value under the given QoS-threshold SINR constraint on the kth mobile user's wireless channel SINR $\gamma_k(i)$ within cell i as follows:

$$\begin{cases} \upsilon^{\pi^*}(s(t)) = \max\limits_{a \in \mathcal{A}} \begin{Bmatrix} R(s(t),a(t)) + \sum\limits_{s(t+1)\in\mathcal{S}} \lambda \\ \cdot \Pr\{s(t+1) | s(t),a(t)\} \upsilon^{\pi}(s(t+1)) \end{Bmatrix} \\ \text{subject to}: \gamma_k(i) \ge \gamma_{th} \end{cases} \tag{3}$$

where $\upsilon^{\pi^*}(s(t))$ represents the maximum expected total reward value under the optimal policy π^* for the current state $s(t)$, λ is the discount factor, $\gamma_k(i)$ represents the SINR of user k in cell i (which is defined in Eq. 1), and γ_{th} denotes the targeted QoS-threshold SINR of MUs in a cell.

The Iteration Algorithm to Implement the Stochastic Power-Saving Optimization

We propose to use the iterative algorithm to solve the optimization problem specified by Eq. 3. In particular, we use the iterative algorithm [15] to derive a stationary optimal policy and the corresponding expected total reward function. The iterative algorithm is described as follows, where ε denotes an infinitesimal gap, $\sigma \in \{0, 1, 2, \ldots\}$ is the iteration steps index, and $\upsilon_\sigma^\pi(s(t))$ is defined in Eq. 3.

Step 1: Set $\upsilon_0^\pi(s(t)) = 0$ for each state $s(t)$. Specify $\varepsilon > 0$ and initialize σ to be 0.

Step 2: For each state $s(t)$, compute $\upsilon_{\sigma+1}^\pi(s(t))$.

Step 3: If $|\upsilon_{\sigma+1}^\pi(s(t)) - \upsilon_\sigma^\pi(s(t))| < \varepsilon\{(1-\lambda)/(2\lambda)$, then go to **Step 4**;

Else, increase σ by 1, then go to **Step 2**.

Step 4: For each $s(t) \in \mathcal{S}$, compute the stationary optimal policy π^*.

The computational complexity of the iteration algorithm is $O(\|\mathcal{A}\| \|\mathcal{S}\|^2)$ [15], where $\|\cdot\|$ is the cardinality of a set. While the system state space can be large, in the real network implementation the computational and time complexity can be significantly reduced because the optimal transmit-power control policy searching through our proposed iteration algorithm is implemented by an offline approach. Once the optimal policy is obtained corresponding to a given space state through the offline approach, it is stored in an optimal-policy lookup table. Each entry of the lookup table specifies the optimal action policy for any given system state, consisting of the MU density state, MU location state, and BS transmit power state of each cell. For the online operation, at each decision epoch, the multicell cooperation power scheduling controller can quickly look up the table to obtain the optimal action policy corresponding to the current system state and then execute the optimal decision or the optimal action policy. The lookup table can readily be generated offline by using separate and independent high-speed computer systems with the super processing capability and power supply.

Performance Evaluations

We conduct numerical analyses to evaluate the performance of our proposed schemes based on the SMDP model for optimal trade-off between power saving at base stations and QoS provisioning to mobile users over multicell cooperation networks. The multicell cooperation in our numerical analyses follows the SMDP model-based stochastic power scheduling by using the action functions as described in Figs. 3a–d. Using the SMDP model-based power scheduling optimization and QoS provisioning constraints and solving for the stochastic optimization problem by the iteration algorithm described in "The Iteration Algorithm to Implement the stochastic Power-Saving Optimization" section, we obtain a set of numerical solution results detailed as follows. Our performance evaluation model consists of the following parameters: the constant transmit power of each cell is 500 W, the propagation coefficient η is –31.54 dB, the discount factor λ is 0.8, and the attenuation coefficient ξ is 3.

Figure 4 plots the convergence process of the two executions of the expected total reward function $\upsilon^{\pi}(s(t))$ specified in Eq. 3. We can obtain the following observations from Fig. 4. First, Fig. 4 shows that our SMDP-based stochastic control process for optimal multicell cooperation power scheduling is stable as it quickly converges to its equilibrium state from any initial state. Second, Fig. 4 confirms that there is always the optimal solution for our SMDP-based stochastic power scheduling. Finally, Fig. 4 shows that the multiple executions (we only plot two of them in Fig. 4 for lack of space) of the iteration algorithm for solving SMDP-based stochastic optimization converge to the unique optimal solution regardless of the initial states.[1] The above observations imply that our proposed SMDP-based stochastic optimization and its corresponding iteration algorithm are feasible and valid.

Figure 5 plots the power consumption functions against the MUs' traffic load under three different multicell cooperation power scheduling schemes, including our proposed SMDP scheme, the random scheme, and the static scheme, respectively. As shown in Fig. 5, although the average power consumption of these three schemes all increase as MUs' traffic load increases, our proposed SMDP scheme significantly outperforms the other two schemes, especially when the traffic load is heavy. This is because our proposed SMDP scheme consumes much less power than the other two schemes due to multicell cooperation power scheduling. In contrast, the static scheme statically allocates the maximum power specified by cell capacity, and the random scheme's average power allocation is also much larger than that of our proposed SMDP scheme as multicell cooperation is not employed.

Under the given power consumption dynamics and conditions obtained in Fig. 5, Fig. 6 plots the implemented average SINR functions against MUs' traffic load also for the same three power scheduling schemes. We observe that as the MUs' traffic load and power consumption increase, the average SINR implemented by our proposed SMDP-based scheme increases monotonically and is always much larger than the required QoS-threshold SINR, as shown in Fig. 6, which implies that our proposed SMDP scheme can guarantee much better QoS provisioning with the least power consumption, as shown in the corresponding Fig. 5 in multicell cooperation networks. Figure 6 also indicates that the implemented average SINR by our proposed SMDP scheme is much higher than that by the static scheme, and thus, our SMDP scheme can provide much better QoS provisioning than the static scheme. Figure 6 shows that the SINR implemented by our SMDP scheme and the static scheme can both guarantee the required QoS provisioning, because they are both larger than QoS threshold SINR as shown in Fig. 6, but the power consumed by the static scheme is much larger than that consumed by our proposed SMDP scheme, as shown in the corresponding Fig. 5. On the other hand, Fig. 6 also shows that the average SINR implemented by the random scheme cannot guarantee the QoS provisioning requirement because its implemented average SINR is often lower than the QoS-threshold SINR, as shown in Fig. 6.

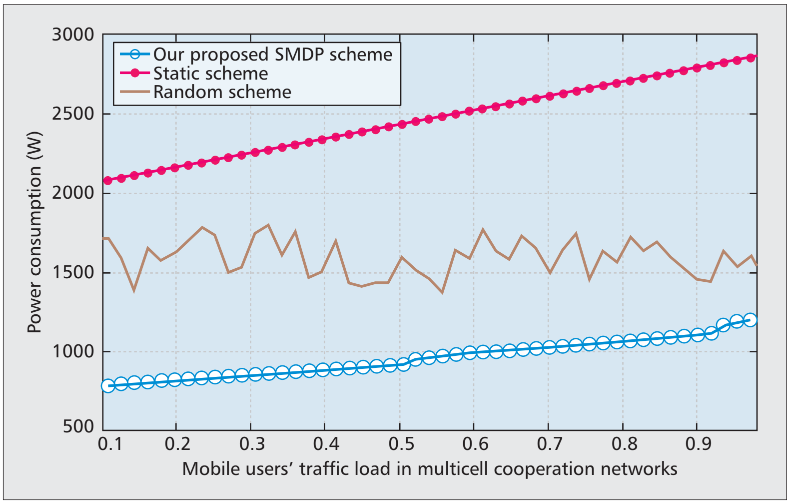


Figure 5. *Power consumption against mobile users' traffic load in multicell cooperation networks.*

Conclusions

We have proposed the semi-Markov decision process (SMDP) model-based stochastic optimization scheme for optimal trade-off between power saving and QoS provisioning over multicell cooperation networks. Our optimization objective is to minimize the power consumption at base stations while guaranteeing QoS provisioning to mobile users by using multicell cooperation power scheduling. Applying the finite-state Markov-chain techniques, we have developed a set of models to characterize mobile user density, mobility, multicell cooperation power-profile scheduling, and QoS provisioning.

[1] *We only plot two executions in Fig. 4 for lack of space. In fact, we get the numerical results for more than two executions, which all converge to the same unique optimal solution regardless of the initial states.*

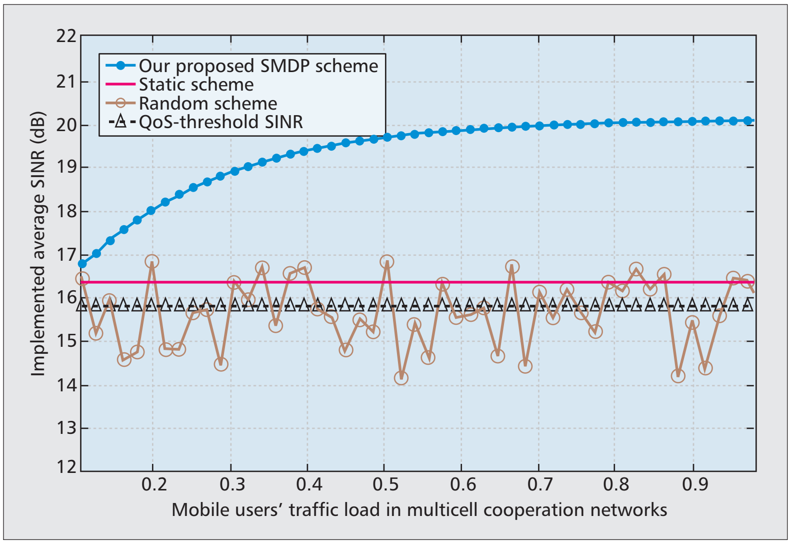


Figure 6. *The implemented average SINR of different schemes against the mobile-users' traffic load in the multicell cooperation networks.*

Based on these models, we have formulated the SMDP-based stochastic optimization problem to optimize the trade-off between the power saving at BSs and QoS provisioning to mobile users over multicell cooperation networks. We have also developed an efficient iteration algorithm to implement our proposed SMDP-based optimization scheme. Finally, we have conducted the numerical analyses to evaluate our proposed schemes, which show that our proposed SMDP-based stochastic control process converges to the unique optimal solution and significantly outperforms the other schemes in terms of minimizing the power consumption at BSs while guaranteeing the QoS provisioning to mobile users.

References

[1] 3GPP TR 32.826, "Telecommunication Management, Study on Energy Savings Management (ESM)," Rel-10, Mar. 2010, http://www.3gpp.org/ftp/Specs/html-info/32826.htm.

[2] 3GPP TS 25.214, "Universal Mobile Telecommunications System (UMTS), Physical Layer Procedures (FDD)," Rel-10, 2010.

[3] IEEE Std. 802.16-2004, "Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems," Oct. 2004.

[4] X. Zhang *et al.*, "Cross-Layer Based Modeling for Quality of Service Guarantees in Mobile Wireless Networks," *IEEE Commun. Mag.*, vol. 44, no. 1, Jan. 2006, pp. 100–06.

[5] J. Tang and X. Zhang, "Quality-of-Service Driven Power and Rate Adaptation for Multichannel Communications Over Wireless Links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, Dec. 2007, pp. 4349–60.

[6] ITU-T Focus Group on Future Networks, FG-FN OD-66, "Draft Deliverable on Overview of Energy Saving of Networks," Oct. 2010.

[7] P. Marsch and G. Fettweis, "Uplink CoMP Under a Constrained Backhaul and Imperfect Channel Knowledge," *IEEE Trans. Wireless Commun.*, vol. 10, 2011, pp. 1730–42.

[8] S. Han *et al.*, "On the Energy Efficiency of Base Station Sleeping with Multicell Cooperative Transmission," *Proc. IEEE PIMRC, 22th IEEE Int'l. Symp. Digital Object Identifier*, 2011, pp. 1536–40.

[9] B. Rengarajan, G. Rizzo, and M. A. Marsan, "Bounds on QoS-Constrained Energy Savings in Cellular Access Networks with Sleep Modes," *IEEE Proc. 23th Int'l. Teletraffic Congress*, 2011, pp. 47–54.

[10] S.-E. Elayoubi, L. Saker, and T. Chahed, "Optimal Control for Base Station Sleep Mode in Energy Efficient Radio Access Networks," *Proc. IEEE INFOCOM*, 2011, pp. 106–10.

[11] K. Dufkova *et al.*, "Energy Savings for Cellular Network with Evaluation of Impact on Data Traffic Performance," *Proc. Euro. Wireless*, 2010, pp. 916–23.

[12] M. Anderberg, *Cluster Analysis for Applications*, Academic Press, 1973.

[13] A. Jain and R. Dubes, *Algorithms for Clustering Data*, Prentice Hall, 1988.

[14] I. Humar *et al.*, "Energy Saving Modeling and Performance Analysis in Multi-Power-State Base Stations Systems," *Proc. GreenCom, IEEE/ACM Int'l. Conf. Cyber, Physical and Social Computing (CPSCom)*, 2010, pp. 474–78.

[15] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley, 1994.

Biographies

Xi Zhang [S'89, SM'98] (xizhang@ece.tamu.edu) received his Ph.D. degree in electrical engineering and computer science (electrical engineering-systems) from the University of Michigan, Ann Arbor. He is currently an associate professor and the founding director of the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station. He was a research fellow with the School of Electrical Engineering, University of Technology, Sydney, Australia, and the Department of Electrical and Computer Engineering, James Cook University, Australia. He was with the Networks and Distributed Systems Research Department, AT&T Bell Laboratories, Murray Hill, New Jersey, and AT&T Laboratories Research, Florham Park, New Jersey, in 1997. He has published more than 230 research papers on wireless networks and communications systems, network protocol design and modeling, statistical communications, random signal processing, information theory, and control theory and systems. He received the U.S. National Science Foundation CAREER Award in 2004 for his research in the areas of mobile wireless and multicast networking and systems. He is an IEEE Distinguished Lecturer. He received Best Paper Awards at IEEE GLOBECOM 2007, IEEE GLOBECOM 2009, and IEEE WCNC 2010, respectively. He also received a TEES Select Young Faculty Award for Excellence in Research Performance from the Dwight Look College of Engineering at Texas A&M University, College Station, in 2006. He is serving or has served as an Editor for *IEEE Transactions on Communications*, *IEEE Transactions on Wireless Communications*, and *IEEE Transactions on Vehicular Technology*, twice as a Guest Editor for *IEEE Journal on Selected Areas in Communications* for two special issues on "Broadband Wireless Communications for High Speed Vehicles" and "Wireless Video Transmissions," an Associate Editor for *IEEE Communications Letters*, a Guest Editor for *IEEE Communications Magazine* and *IEEE Wireless Communications*, an Editor for Wiley's *Journal on Wireless Communications and Mobile Computing*, *Journal of Computer Systems, Networking, and Communications*, and Wiley's *Journal on Security and Communications Networks*, and an Area Editor for Elsevier's *Journal on Computer Communications*, among others. He is serving or has served as TPC Chair for IEEE GLOBECOM 2011, TPC Vice-Chair IEEE INFOCOM 2010, TPC Area Chair for IEEE INFOCOM 2012, Panel/Demo/Poster Chair for ACM MobiCom 2011, General Vice-Chair for IEEE WCNC 2013, Panel/Demo/Poster Chair for ACM MobiCom 2011, and TPC/General Chair for numerous other IEEE/ACM conferences, symposia, and workshops.

Ping Wang received her B.S. degree in communication engineering from Henan University, China, in 2007. She is currently a Ph.D. student studying in the joint Ph.D. program between the Department of Electrical and Computer Engineering, Texas A&M University, and the School of Electronic Engineering, Beijing University of Posts and Telecommunications, China. She has published 11 papers in IEEE journals and conference proceedings. Her research interests are in wireless information and communication theory, underwater wireless networks, cooperative relay networks, QoS provisioning, and wireless resource management.