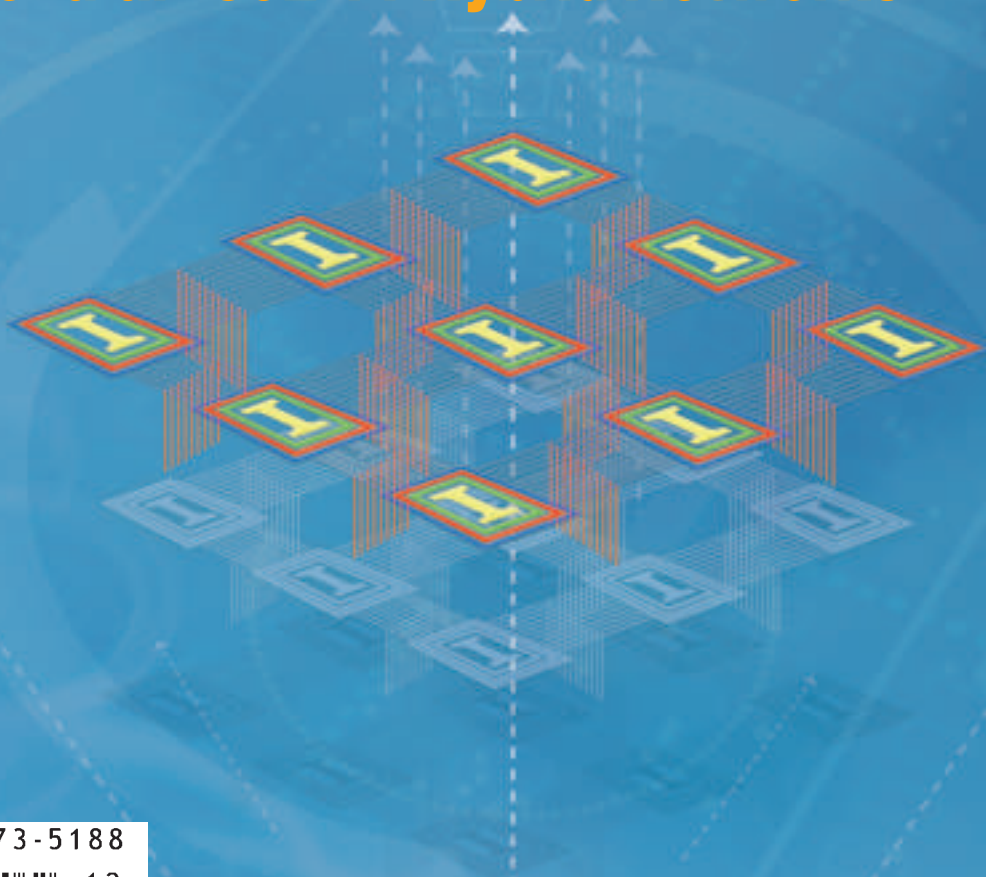


# ZTE COMMUNICATIONS

December 2012, Vol.10 No.4

## **SPECIAL TOPIC:** **Millimeter Wave Communication for Cellular and Cellular-802.11 Hybrid Networks**



ISSN 1673-5188



# New Members of ZTE Communications Editorial Board

(in alphabetical order)



**Chang Wen Chen** (F '04) has been a professor of computer science and engineering at the University at Buffalo, State University of New York, since 2008. From 2003 to 2007, he was Allen S. Henry Endowed Chair Professor in the Department of Electrical and Computer Engineering, Florida Institute of Technology. From 1996 to 2003, he was a member of the Faculty of Electrical and Computer Engineering, University of Missouri. From 1992 to 1996, he was a member of the Faculty of Electrical and Computer Engineering, University of Rochester, NY.

From 2000 to 2002, he was the head of the Interactive Media Group at the David Sarnoff Research Laboratories, Princeton, NJ.

He has received numerous awards, including the Sigma Xi Excellence in Graduate Research Mentoring Award in 2003, Alexander von Humboldt Research Award in 2009, and SUNY–Buffalo Exceptional Scholar—Sustained Achievement Award in 2012. He received his BS degree from the University of Science and Technology of China in 1983. He received his MSEE degree from the University of Southern California in 1986 and his PhD degree from the University of Illinois, Urbana–Champaign, in 1992. He is an IEEE Fellow and SPIE Fellow.



**Connie Chang–Hasnain** is the John R. Whinnery Chair Professor in the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. She received her BS degree from the University of California, Davis, in 1982. She received her MS degree and PhD degree in electrical engineering and computer sciences from the University of California, Berkeley, in 1984 and 1987. Prior to joining UC Berkeley, she was a member of the technical staff at Bellcore and an assistant professor of electrical engineering at Stanford University. She is currently the chair of the Nanoscale Science and Engineering (NSE) Graduate Group and holds a Chang Jiang Scholar Endowed Chair at Tsinghua University, PR. China.

Professor Chang–Hasnain's research interests include vertical cavity surface–emitting lasers, MEMS tunable optoelectronic devices and nanostructured materials, and nano–optoelectronic devices. She has co–authored more than 400 research papers and seven book chapters. She also holds 36 patents. She is a fellow of the IEEE, OSA and IEE and is an honorary member of A.F. Ioffe Institute.



**Jianhua Ma** received his BS and MS degrees in communication systems from the National University of Defense Technology (NUDT), China, in 1982 and 1985. He received his PhD degree in information engineering from Xidian University, China, in 1990. He has worked at Hosei University since 2000 and is currently a professor in the Digital Media Department in the Faculty of Computer and Information Sciences. Prior to joining Hosei University, he spent 15 years teaching and researching at NUDT, Xidian University, and the University of Aizu, Japan. Since 2003, his research has been focused on “smart worlds,” which are pervaded with smart/intelligent ubiquitous things (u–things). These include three essential elements—smart object, smart space/hyperspace and smart system. He envisages that future ubiquitous intelligence or pervasive intelligence will be used to solve crucial problems caused by the fast progress of semiconductors, MEMS, NEMS, sensors, RFIDs, embedded devices, ubiquitous computers, pervasive networks, and universal services. Dr. Ma is a member of IEEE and ACM. He has edited more than 15 books and proceedings and published more than 200 academic papers.



**Jiannong Cao** is chair professor and head of the Department of Computing, Hong Kong Polytechnic University. His research interests include parallel and distributed computing, computer networks, mobile and pervasive computing, fault tolerance, and middleware. He has co–authored four books, co–edited nine books, and published more than 300 papers in international journals and conference proceedings. He has directed and participated in numerous research and development projects. As a principal investigator, he has obtained more than HK\$25 million grants. Dr. Cao is a senior member of the China Computer Federation, a senior member of IEEE, and a member of ACM. He is the chair of the Technical Committee on Distributed Computing of the IEEE Computer Society. Dr. Cao has been an associate editor and a member of the editorial boards of many international journals. He has also been the chair and member of organizing and program committees for many international conferences.

Dr. Cao received his BSc degree in computer science from Nanjing University and his MSc and PhD degrees in computer science from Washington State University.



**Jinhong Yuan** received his BE and PhD degrees in electronics engineering from Beijing Institute of Technology, China, in 1991 and 1997. From 1997 to 1999 he was a research fellow at the School of Electrical Engineering, University of Sydney, Australia. In 2000, he joined the School of Electrical Engineering and Telecommunications, University of New South Wales, and is currently a professor of telecommunications in the school. He has published two books, two book chapters, and more than 200 papers in telecommunications journals and proceedings. He has also authored 40 industrial reports. He is the co–inventor of one patent on MIMO systems and two patents on low–density–parity–check (LDPC) codes. He co–authored papers that have won Best Paper Awards and Best Poster Award. His publication list is available at <http://www2.ee.unsw.edu.au/wc/JYuan.html>.

Jinhong Yuan is the chair of the IEEE NSW Joint Communications/Signal Processing/Ocean Engineering Chapter and an associate editor of IEEE Transactions on Communications. His research interests include error–control coding and information theory, communication theory, and wireless communications.



**Victor C. M. Leung** is a professor and holds the TELUS Mobility Research Chair in Advanced Telecommunications Engineering in the Department of Electrical and Computer Engineering, University of British Columbia. He received his BSc and PhD degrees from the University of British Columbia in 1977 and 1981. He has been focused on wireless networks and mobile systems for more than 30 years and has co–authored more than 600 journal and conference papers, including several papers that won best–paper awards. Dr. Leung is a fellow of IEEE, EIC and CAE. From 2009 to 2012, he was a distinguished lecturer of the IEEE Communications Society. He is serving or has served on the editorial boards of many journals. He has contributed to the organizing committees and technical program committees of numerous conferences. Dr. Leung was the winner of an APEBC Gold Medal in 1977, an NSERC Postgraduate Scholarship, an IEEE Vancouver Section Centennial Award in 2011, and a UBC Killam Research Prize in 2012.



**Yingfei Dong** received his BS degree and MS degree in computer science from Harbin Institute of Technology, China, in 1989 and 1992. He received his PhD degree in engineering from Tsinghua University in 1996 and his PhD degree in computer and information science from the University of Minnesota in 2003. He is currently an associate professor in the Department of Electrical Engineering, University of Hawaii, Manoa. His research interests include computer networks (especially network security), smart grid communication security, cloud security, reliable real–time network communication, internet services, and distributed systems. His work has been published in many referred journals and conference proceedings. He has been an organizer and program committee member for many IEEE, ACM, and IFIP conferences. He also serves on several editorial boards for journals on security and networking. His current research is supported by the National Science Foundation.



**Zhili Sun** is a chair professor of communication networking at the University of Surrey. He received his BSc degree in mathematics from Nanjing University, China, and his PhD degree in computer science from Lancaster University. After receiving his PhD, he worked as an assistant in the Computer Centre, Southeast University, and a research fellow in the Telecommunication Research Group, Queen Mary University of London.

Professor Sun has been a principal investigator and technical coordinator in many research projects funded by the EU, UK Research Councils, European Space Agency, and industry organizations. He has published more than 125 papers in international journals and conference proceedings. He authored and contributed to four books. He was the technical reviewer of the EU framework programs and UK and other national programs. He has contributed to standardization activities in ITU–T, ETSI and IETF. He was the general chair and member of various technical committees for international conferences.

His areas of interests include IP networking protocols and architecture, satellite communications and networking, Internet and teletraffic engineering, network security, mobile and wireless ad hoc networks and mobile operating systems.

# 2013 IET International Conference on Information and Communications



## Technologies

IETICT 2013, Beijing, China

27th –29th April, 2013

[www.ietict.org](http://www.ietict.org) [www.ciict.net](http://www.ciict.net)



The 2013 IET International Conference on Information and Communications Technologies (IETICT2013) will be held 27th –29th April 2013 in Beijing, in conjunction with China–Ireland International Conference on Information and Communications Technologies (CICT2013). CICT has been an international conference since its inception in 2006, and has been held alternately every year in China and Ireland, focusing on a wide range of areas in information and communication technology.

IETICT 2013 aims to provide an international platform for both academic scholars and industry leaders in fields of information and communication technologies to exchange novel ideas and latest research results. The conference includes not only technical sessions, but also invited sessions and keynote addresses.

You are invited to submit original papers to the conference. Submitted papers should not have been previously published or currently under review for any other publication. All submitted papers will be reviewed. All accepted papers will be published by The Institution of Engineering and Technology (IET), and will be included in the IEEE Xplore and IET INSPEC, and then submitted to EI Compendex.

### TOPICS OF INTEREST INCLUDE, BUT NOT LIMITED TO:

#### Circuits and Systems:

System on Chip  
Large Scale Integrated Circuit  
RF Circuit and System  
Nonlinear Circuits and Systems  
Power and Energy Circuits and Systems

#### Communication Technology:

Fiber & cable Communications  
Network Architectures  
Secure Communications  
Video and Broadcasting  
Wireless and Satellite Communications  
Ad-hoc, Sensor and Mesh Networking  
Communication Software

#### Network and Information Security:

Network Security  
Cryptology and Information Security  
Mobile and Wireless Security  
Information Hiding and Watermarking

#### Information Theory and Application:

Digital Systems  
Electronics, Computing and Control  
Quality, Reliability, Security & Safety  
Semiconductors and device  
Signal Processing

#### Information Engineering

#### Computational Intelligence:

Artificial Intelligence and Its Application  
Intelligent Data Management  
Information Retrieval  
Grid Computing  
Natural Language Processing

#### Computer Science:

Computer Design  
Data Mining & Data Engineering  
Software Engineering  
Computer Modeling and Simulation  
Computer Application

---

#### Important Dates:

Paper Submission Deadline: Feb.25, 2013  
Acceptance Notification: Mar.15, 2013  
Final Manuscript Deadline: Mar.29, 2013  
Conference Date: April.27th–29th, 2013

---

#### Contact Us:

Website: [www.ietict.org](http://www.ietict.org), [www.ieccr.net](http://www.ieccr.net)  
E-mail: [ietict2013@gmail.com](mailto:ietict2013@gmail.com)  
Tel: 86–15712895816

# C CONTENTS

[Http://www.zte.com.cn/magazine/English](http://www.zte.com.cn/magazine/English)  
Email: [magazine@zte.com.cn](mailto:magazine@zte.com.cn)



## ZTE Communications Editorial Board

### Chairman:

Houlin Zhao  
International Telecommunication Union (ITU)

### Vice Chairmen:

Lirong Shi  
ZTE Corporation (China)

Cheng-Zhong Xu  
Wayne State University (USA)

### Members (in Alphabetical Order):

Chang wen Chen  
The State University of New York (USA)

Cheng-Zhong Xu  
Wayne State University (USA)

Connie Chang-Hasnain  
University of California, Berkeley (USA)

Houlin Zhao  
International Telecommunication Union (ITU)

Huifang Sun  
Mitsubishi Electric Research Laboratories (USA)

Jianhua Ma  
Hosei University (Japan)

Jiannong Cao  
Hong Kong Polytechnic University (Hong Kong)

Jinhong Yuan  
University of New South Wales (Australia)

Ke-Li Wu  
The Chinese University of Hong Kong (Hong Kong)

Lirong Shi  
ZTE Corporation (China)

Shiduan Cheng  
Beijing University of Posts and Telecommunications (China)

Victor C. M. Leung  
The University of British Columbia (Canada)

Wen Gao  
Peking University (China)

Wenjun (Kevin) Zeng  
University of Missouri (USA)

Yingfei Dong  
University of Hawaii (USA)

Zhenge (George) Sun  
ZTE Corporation (China)

Zhengkun Mi  
Nanjing University of Posts and Telecommunications (China)

Zhili Sun  
University of Surrey (UK)

## Special Topic

### Millimeter Wave Communication for Cellular and Cellular-802.11 Hybrid Networks

#### 01 ..... Guest Editorial

by Philip Pietraski and I-tai Lu

#### 03 ..... Millimeter Wave and Terahertz Communications:

Feasibility and Challenges

by Phil Pietraski, David Britz, Arnab Roy, Ravi Pragada, and Gregg Charlton

#### 13 ..... WiGig and IEEE 802.11ad for Multi-Gigabyte-Per-Second WPAN and WLAN

by Sai Shankar N, Debashis Dash, Hassan El Madi, and Guru Gopalakrishnan

#### 23 ..... Modeling Human Blockers in Millimeter Wave Radio Links

by Jonathan S. Lu, Daniel Steinbach, Patrick Cabrol, and Philip Pietraski

#### 29 ..... 60 GHz SIW Steerable Antenna Array in LTCC

by Bahram Sanadgol, Sybille Holzwarth, Peter Uhlig, Alberto Milano, and Rafi Popovich

#### 33 ..... Light-of-Sight MIMO for Next-Generation Microwave Transmission Systems

by Xianwei Gong, Zhifeng Yuan, Jun Xu, and Liujun Hu

## Research Papers

39 ..... Terabit Superchannel Transmission: A Nyquist–WDM Signals Approach

*by Hung–Chang Chien, Jianjun Yu, Zhensheng Jia, and Ze Dong*

45 ..... Parallel Web Mining System Based on Cloud Platform

*by hengmei Luo, Qing He, Lixia Liu, Xiang Ao, Ning Li, and Fuzhen Zhuang*

54 ..... Hierarchical Template Matching for Robust Visual Tracking with Severe Occlusions

*by Lizuo Jin, Tirui Wu, Feng Liu, and Gang Zeng*

60 ..... Design and Implementation of ZTE Object Storage System with Severe Occlusions

*by Huabin Ruan, Xiaomeng Huang, and Yang Zhou*

## Roundup

2 ZTE Launches Innovative Energy–Saving Solution for LTE Networks

38 ZTE Communications Guidelines for Authors

59 ZTE Launches the First PC–Based CPT for LTE Networks

64 Ad Index

I Table of Contents for Volume 10, Numbers 1–4, 2012

## ZTE COMMUNICATIONS

Vol. 10 No.4 (Issue 36)

Quarterly

First Issue Published in 2003

### Supervised by:

Anhui Science and Technology Department

### Sponsored by:

ZTE Corporation and Anhui Science and Technology Information Research Institute

### Staff Members:

Editor–in–chief: Xie Daxiong

Associate Editor–in–chief: Zhao Jinming  
Executive Associate

Editor–in–chief: Huang Xinming

Editor in Charge: Zhu Li

Editors: Paul Sleswick, Xu Ye, Yang Qinyi, Lu Dan

Producer: Yu Gang

Circulation Executive: Wang Pingping

Assistant: Wang Kun

### Editorial Correspondence:

Add: 12/F Kaixuan Building,  
329 Jinzhai Road,  
HeFei 230061, P. R. China

Tel: +86–551–65533356

Fax: +86–551–65850139

Email: magazine@zte.com.cn

### Published and Circulated (Home and Abroad) by:

Editorial Office of  
ZTE COMMUNICATIONS

### Printed by:

Hefei Zhongjian Color Printing Company

### Publication Date:

December 25, 2012

### Publication Licenses:

ISSN 1673–5188

CN 34–1294/TN

### Advertising License:

皖合工商广字 0058 号

### Annual Subscription Rate:

USD\$50

Responsibility for content rests on authors of signed articles and not on the editorial board of ZTE COMMUNICATIONS or its sponsors. All rights reserved.



# Millimeter Wave Communication for Cellular and Cellular-802.11 Hybrid Networks

*Philip Pietraski*



*I-tai Lu*



The demand for wireless data has been driving network capacity to double about every two years for the past 50 years, if not 100 years, and this has come to be known as Cooper's Law. In recent years, this trend has accelerated as a greater proportion of the population adopts wireless devices with ever greater capabilities, including tablets that support HD video and other advanced capabilities. Many cellular operators have tried to adapt this trend by throttling data rates, backing away from all-you-can-eat data plans, and offloading to WiFi. Over the next decade, further increases in demand are expected, and this issue of *ZTE Communications* examines millimeter wave communications as one technology that may answer the call.

Historically, the ever-growing demand for data capacity has been met by adding more spectrum and improving spectral efficiency, but spectrum reuse employing smaller cells has been by far the most popular means of adding network capacity. Deploying cells with ever greater density is a simple way of adding capacity to a network. Increasing the number of cells in the network increases the network capacity without increasing the capacity per cell. However, this approach becomes cost-prohibitive in part because it is expensive to roll out all these cells and provide them with a quality backhaul connection, for example, fiber. A less-expensive means of adding network capacity is needed in the long term.

As cells become smaller and link distances have become shorter, an alternative to adding capacity and reducing deployment costs is to use much higher carrier frequencies. Shorter link distances, which come with smaller cells, combined with recent advances in millimeter wave transceivers and antennas opens the door for the use of millimeter wave spectrum in cellular systems. An obvious benefit to this is the availability of a huge amount of spectrum.

The 60 GHz unlicensed band alone offers 5–9 GHz of bandwidth (the exact amount depends on country), and there are many other millimeter wave and terahertz bands that have potential. Another great benefit of millimeter wave carriers is that high-gain, highly directional, electrically steerable antennas can be very small and greatly reduce interference. The wide bandwidths and narrow steerable beams enable low-cost deployment based on a wireless backhaul.

WirelessHD devices with 60 GHz phased array antennas are already on the market, and WiGig/802.11ad devices are on their way. ABI research predicts that by 2016 one third of all WiFi products will be tri-band (2.4/5/60 GHz). Although WiGig is intended to be an indoor, short-link technology (~10m), it may be an important standard used as a starting point for larger networks to use millimeter wave communications. Mass production of devices such as these will continue to drive costs down for millimeter wave radios and antennas that should extend to longer links.

The 60 GHz unlicensed band is of particular interest because of the growing ecosystem being built around consumer electronics that support WirelessHD and WiGig. However, the fact that the band is unlicensed means that it is riskier for cellular service providers to adopt. Molecular oxygen absorption at 60 GHz creates further confusion as some argue that these losses limit link distance. Others argue that reduced interference is worth it. Below 60 GHz, the LMDS bands are of interest and are underutilized; however, they offer less total spectrum than the 60 GHz unlicensed band. Recent technological advances may soon enable communications well above 100 GHz and into the terahertz region above 300 GHz, where allocations have not yet been made by regulators, and even greater bandwidths could become available. Some agreement on a band will be needed in order to make good progress.

Of course, there are also great challenges with millimeter

wave systems. Although link distances in a line-of-sight environment might be easily closed with millimeter wave technology, the environment poses particular problems. Millimeter waves do not generally penetrate through buildings or diffract around them. Furthermore, humans are great blockers of millimeter waves and tend to move around more than buildings. The problem of cost-effective routing around buildings and people will be one of the larger problems.

In this special issue, we examine the role that millimeter wave communication could play in cellular and cellular hybrid networks in access and backhaul. The first paper provides an introduction to the potential use of millimeter waves in a large network context and provides a preliminary simulation study. The second paper provides an overview of the 802.11ad/WiGig MAC and PHY. The third paper provides an experimental study of human blocking of millimeter wave propagation. The fourth paper describes the design and measurements of a 60 GHz LTCC phased array antenna with integrated waveguide distribution network that could be suitable for backhaul applications. The fifth paper considers the use of MIMO techniques for millimeter wave in line-of-sight conditions.

We are grateful to the authors who made contributions to this special issue and to the reviewers who spent their valuable time to provide valuable and constructive feedback. We hope that you find this special issue interesting and useful.

We are grateful to the authors who made contributions to this special issue and to the reviewers who spent their

valuable time to provide valuable and constructive feedback. We hope that you find this special issue interesting and useful.

### Biographies

**Phil Pietraski** (philip.pietraski@interdigital.com) received his BSEET from DeVry University in 1987. He received his BSEE, MSEE, Grad.Cert. in wireless communications, and PhD EE from Polytechnic University (now NYU-Poly), Brooklyn, in 1994, 1995, 1996, and 2000.

He joined InterDigital Communications in 2001 and is currently a principal engineer leading research activity in wireless communications, most recently in millimeter wave communications and future cellular architectures. He holds more than 50 patents in wireless communications and has authored multiple conference and journal papers. He is vice chair of the MoGig (Mobile Gigabit) working group at IWPC and a trustee for DeVry NJ campuses.

Prior to his transition to wireless communications in 2000, he was a research engineer at Brookhaven National Laboratory, National Synchrotron Light Source, responsible for beam-line instrumentation and X-ray detector R&D. He has also conducted research at the Polytechnic University for the Office of Naval Research (ONR) in underwater source localization.

**I-Tai Lu** received his PhD degree in electronic engineering from Polytechnic University of New York. He is currently professor and director of the online program of the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University. He has worked in wave propagation and inverse problems with applications in underwater and structure acoustics, non-destructive testing, microwave engineering, sonar and radar. His current research interests include wireless communications, in which he has made contributions to the developments of Wireless LAN (IEEE802.11n) and 3G cellular communications (WCDMA). He is currently involved in the development and standardization of the 4G (3GPP LTE-A) and future generations of wireless communications systems. He has published more than 200 journal and proceeding papers and holds 6 patents. He has given more than 50 invited lectures and spoken at more than 200 conferences, workshops, and seminars.

## Roundup

### ZTE Launches Innovative Energy-Saving Solution for LTE Networks

19 November, 2012, Shenzhen—ZTE Corporation, a publicly listed global provider of telecommunications equipment, network solutions, and mobile devices, announced the launch of its Energy Saving Solution for operator LTE networks. According to test results, a single site employing this solution can save up to 40 percent power.

The global ICT industry accounts for as much as 2.5 percent of the world's greenhouse gas emissions according to research firm Gartner. In a typical wireless telecommunications network, about 90 percent of energy is consumed by base stations. ZTE's Energy Saving Solution offers an array of innovative technologies to reduce power consumption in base stations. It includes the industry's first integrated system for automatic/dynamic PA bias voltage control and intelligent OFDM signaling shutdown.

Test results show that ZTE's Energy Saving Solution save as much as 32 percent energy on a single site. Annual reduction of as much as 5200 Kwh can be achieved on a typical 1500 W base station. For a network comprising 1000 base stations, the annual reduction would amount to 5.2 million Kwh, reducing 4500 tons of carbon dioxide emission. When combined with the deployment of renewable energy sources such as solar energy, wind energy or bio-energy, energy savings of more than 50 per cent of the entire network can be achieved.

"ZTE will continue to develop our technology to provide customers with energy-saving products and solutions," said ZTE vice president, Wang Shouchen. "We are committed to minimizing environmental impact with green technology."

ZTE will increase resources for LTE development and green technologies. The company has won 38 LTE commercial contracts and is working with more than 100 operators in Europe, the Americas, Asia Pacific and the Middle East on trial LTE networks.

# Millimeter Wave and Terahertz Communications: Feasibility and Challenges

*Phil Pietraski<sup>1</sup>, David Britz<sup>2</sup>, Arnab Roy<sup>3</sup>, Ravi Pragada<sup>3</sup>, and Gregg Charlton<sup>3</sup>*

(1. InterDigital Communications LLC., Melville, NY 11747, USA;

2. AT&T Labs, Shannon laboratories, Florham Park, NJ 07932, USA;

3. InterDigital Communications LLC., King of Prussia, PA 19406, USA)

## Abstract

In this paper, the challenges with and motivations for developing millimeter wave and terahertz communications are described. A high-level candidate architecture is presented, and use cases highlighting the potential applicability of high-frequency links are discussed. Mobility challenges at these higher frequencies are also discussed. Difficulties that arise as a result of high carrier frequencies and higher path loss can be overcome by practical, higher-gain antennas that have the added benefit of reducing intercell interference. Simulation methodology and results are given. The results show that millimeter wave coverage is possible in large, outdoor spaces, and only a reasonable number of base stations are needed. Network throughput can exceed 25 Gbit/s, and cell-edge user throughput can reach approximately 100 Mbit/s.

## Keywords

millimeter wave; terahertz; small cells; propagation; mesh backhaul

## 1 Introduction

Today's remarkably successful wireless infrastructure, services, and customer business models are largely derived from the realities, technology, and choices of spectrum made more than thirty years ago. Ironically, the cellular industry's great technological revolution and market success may now be its undoing in the not-too-distant future. An emerging dichotomy between current infrastructure and bandwidth and surging customer services now begs the question: Will our existing wireless network spectrum and infrastructure be indefinitely scalable to support near-future consumer demands, needs, and services?

Moreover, are the current technologies and enhanced infrastructure, spectrum additions, and planned capacity enhancements enough to satisfy current and near-future network capacity requirements? In an analysis by Rysavy Research in 2010 [1], the average demand per user was graphed against the average available network capacity per user. This analysis shows that average demand will irreversibly exceed capacity sometime in mid 2013, and the two curves will rapidly diverge thereafter. What is the long-term implication of this for customer service?

Beyond the traditional cellular and Wi-Fi spectrum, which ends at around 6 GHz, there is another 294 GHz worth of

FCC-defined spectrum. Within this 294 GHz span, there is an aggregation of roughly 200 GHz worth of fixed and mobile spectrum that today is largely unused. With suitable application of current and emerging technologies, a significant proportion of this enormous radio spectrum may be harvested within a decade to support multi-gigabit wireless mobile communications. Beyond this enormous FCC-defined spectrum is an additional unregulated, unlicensed radio spectrum known as the terahertz band. This spectrum band has channel sizes capable of supporting wireless data speeds of 10 Gbit/s to 100 Gbit/s and could offer an additional 300 GHz of spectrum for small-cell wireless mobile communications.

Growth in consumer data is outstripping the pace at which infrastructure can be deployed or invented to support it. Historically, demand for bandwidth in short-range wireless communications has doubled every 18 months. At this rate, demand will be 800 MB per customer per day in less than 10 years. If, as expected, we see significant growth in wireless machine-to-machine (M2M) communications and services, wireless communications as we know it will be irreversibly changed. As customers begin to demand ubiquitous gigabit and multi-gigabit services, fundamentally new and disruptive technologies and architectures will be required.

AT&T reports that its data traffic associated with smart phones (especially iPhones) has increased 8000% in the last



four years, and global sales of smart phones and other cell phones is starting to flatten, particularly in the US and Europe. Almost every human being on the planet now owns (or has access to) a cell phone, smart phone, or dongle. The amount of data used by each person is skyrocketing, and it is predicted that each user will require 14 Gbit/s by 2016. By 2020, daily use of personal mobile broadband and dongles will exceed 800 MB. By 2020, we will require approximately 130 exabits ( $10^{18}$ ) of data per year.

The FCC is now describing this period of unparalleled growth in smart phone data and wireless network use as a “looming spectrum crisis.” In response, service providers have increased the demand for wireless capacity by deploying more spectrum-efficient technologies and taking more spectrum-efficient approaches. They are also lobbying to expand spectrum coverage from 300 to 3000 MHz to support anticipated near-future capacity needs.

Service providers are now deploying 3G and 4G cellular technologies to enhance wireless network capacity. These new higher-capacity cellular systems use microcells, LTE, femtocells, and sectored and smart antenna technologies that are designed to more efficiently use existing cellular spectrum. The new technologies significantly (but still incrementally) improve network capacity and are rapidly approaching the Shannon Limit. There is little room left to squeeze any more capacity out of available cellular frequencies.

These infrastructure fixes do little to improve network capacity for the anticipated surges in wireless data demand. Adding marginally more efficient infrastructure to the same cellular spectrum bands will do little to satisfy the anticipated long-term demand for wireless data capacity. In this context, we suggest an alternative approach that takes advantage of millimeter wave frequencies.

## **2 Use Cases and Applications**

In essence, a small-cell/nanocell is simply a superhigh-capacity wireless transport pipe. Here, we describe some use cases.

Large public spaces tend to have large gatherings of people working, engaging in conversation, and involved in common activities or events.

With easy cellular communications, there may be many people talking on cell phones and sharing data. Such environments may be plazas, parks, shopping areas, theaters, or stadiums. Although in the near term the data and voice traffic of an individual user might be relatively low-volume, a small cell covering only part of a public space might struggle to handle the aggregated traffic of a conglomeration of users. Such a scenario might be a concert or sporting event where a large number of people are simultaneously using wireless terminals to download video and pictures.

By using concatenated, self-organizing small cells and adaptive frequency tiling in the public space, a large amount of user data can be supported, even in peak periods.

Concatenated small cells can forward and route high-peak traffic between the small-cell nodes and different network points-of-presence (PoPs) to avoid capacity limitations at a small-cell-to-network connection. Flexible deployment of millimeter wave backhaul allows small cells to be deployed so that system performance and customer connectivity can be optimized. The small-cell configuration does not have to be restricted by its proximity to a network optical fiber.

Stadiums have become increasingly spectrum-challenged in recent years, and operators have been forced to use a multitude of technologies to provide wireless services at major events. Highly directional millimeter wave frequency bands would provide super-high capacity links within a stadium to route in-stadium traffic between small-cell and Wi-Fi modules. They would also provide fiber-free connectivity to and from the stadium. These high frequencies would not interfere with existing radio and cellular communication within the stadium. They would dramatically increase data transport capacity within and around the stadium but not impinge on existing services and capacity-stretched cellular and Wi-Fi bands.

Perhaps the most valuable applications for millimeter wave and terahertz technologies will be in dense urban and business locations, especially where tall buildings create deep street canyons. In such locations, cellular service is often very limited and/or unreliable.

Traditional approaches to cellular transmission and relatively wide spacing between cell tower placements constrain system performance in these scenarios. The problem is further compounded by large numbers of mobile cell phone users moving around at street level and by greatly varying demand loads.

Densely backhauled small cells, and eventually nanocells, are being developed to address this street-canyon condition. A series of 3G, 4G, LTE, and Wi-Fi-capable low-powered small cells are deployed on a street (preferably mounted on existing powered street poles). The small cells are connected to each other for traffic and network management, and their concatenated chain is backhauled to and from the network PoP by optical fiber, high-capacity millimeter wave and terahertz radio, or a cost-effective combination of these.

Directed millimeter, terahertz, and optical methods for small-cell backhaul are well-suited to capitalize on high frequencies and street-level light-pole-to-light-pole geometries.

Growth in the number of cell phones worldwide is beginning to flatten out, but daily subscriber wireless traffic (mobile broadband) will increase 500% to 294 MB per day by 2020 [2], [3]. Wireless data associated with dongles will increase by 1000% to 503 MB by 2020 [3]. In addition, the emergence of resident or intelligent cloud-based agents will add a new human-to-machine data capacity requirement and, beyond this, pure machine-to-machine wireless network data transfer. Large cloud data farms will increasingly engage with the real world via sensor and ad hoc networks, and transfer vast amounts of data over high-speed optical and wireless networks.

Existing cellular and Wi-Fi infrastructure along with spectrum allocations will not have the capacity to support long-term mass data transport requirements. New network topologies will be required, perhaps to separate voice and small-data services from big-data services with the latter being routed in a new high-capacity wireless layer—the nanocell. This high-capacity wireless layer builds on the previously discussed small-cell infrastructure, making use similar fiber and self-aligning highly directed millimeter wave beams to backhaul between the network and nanocells. Rather than using limited-capacity cellular and Wi-Fi spectrum to connect the mobile customer to the network, the nanocell steers a highly directional high-frequency beam to the nearby customer for the short timeframe in which the data is transferred. Studies show that such high-capacity wireless links could provide 50–100 Gbit/s backhaul and throughput to the customer [4].

### 3 Challenges With Millimeter Wave and Terahertz Propagation

#### 3.1 Free Space Path Loss

Small cells allow a network planner to take advantage of shorter distances between the cells. These shorter distances provide greater opportunity for line of sight; there is more available power (street furniture); and there is greater opportunity to use high-data-carrying millimeter wave, terahertz, and optical frequencies (FSOC). However, nature increasingly limits the transmission distances of these higher frequencies. Ironically, natural limitations on transmission may make these higher frequencies ideal for inter-small-cell backhaul and relaying because they minimize spectrum overflow between small cells and allow greater spectrum reuse. Higher frequencies are increasingly scattered and or blocked by molecules in the atmosphere, and this loss must be accounted for in an intercell and end-to-end link budget.

Propagation losses at millimeter wave and terahertz frequencies are considered too large for practical communication in large networks that involve links of more than a few tens of meters. The dependency of the free-space path loss can be clearly seen in the well-known equation

$$PL = \left( \frac{4\pi df}{c} \right)^2 \quad (1)$$

Equation (1) shows that just by moving from a 2 GHz carrier to a 60 GHz carrier (a popular unlicensed millimeter wave band), the path loss increases by nearly 30 dB. This sounds disastrous for millimeter waves and terahertz, but first the assumptions in and derivation of the equation should be considered.

If we assume the transmit antenna radiates isotropically, then the power density (measured in W/m<sup>2</sup>) of the plane wave at the receive antenna at distance  $d$  from the transmit antenna is

$$\rho = \frac{P_T}{4\pi d^2} \quad (2)$$

There is no additional path loss at higher frequencies in terms of the power density versus distance. If the transmit antenna is permitted to have gain, for example, using directional antennas, the received power collected by the receive antenna is

$$P_R = \frac{P_T G_T A}{4\pi d^2} \quad (3)$$

where  $G_T$  is the gain of the transmit antenna, and  $A$  is the effective aperture of the receive antenna. So far, there is no dependency on the carrier frequency; the dependence is embedded in  $A$ , which is proportional to the square of the wavelength. In other words, by going to much higher frequencies, our antennas also become smaller, and  $A$  also becomes smaller. The effective aperture can be written as

$$A = \frac{\lambda^2 G_R}{4\pi} \quad (4)$$

where  $G_R$  is the gain of the receive antenna, and  $\lambda$  is the wavelength. The received power can then be written as

$$P_R = \frac{P_T G_T G_R \lambda^2}{(4\pi d)^2} \quad (5)$$

For simple antenna structures,  $G_T$  and  $G_R$  are generally not very large and do not depend on frequency; therefore, the received power decreases rapidly as the carrier frequency increases. However, a more practical constraint on antennas might also include the physical size of the antenna. We consider our antennas to be phased arrays of small antenna elements. The physical size of an array is fixed as the carrier frequency changes. If we consider the arrays to be uniform and rectangular, the number of elements that can be placed in a given area is  $1/\lambda^2$ . In this alternative view, phased arrays of fixed size are used to create narrow beams at the transmitter and receiver, and there is an additional gain proportional to the number of antenna elements in each array. The received power can now be written as

$$P_{R, \text{fixed\_area}} = C \frac{P_T G_T G_R}{(4\pi d\lambda)^2} \quad (6)$$

where  $G_T$  and  $G_R$  are the gains of each antenna element, and  $C$  is a constant depending on the chosen fixed area. The key thing to notice is that  $\lambda^2$  is now in the denominator. In (6), increased carrier frequency is an advantage not a disadvantage. This may be an optimistic view, but there is no free space path loss for higher frequencies if advanced antenna architectures, such as phased arrays, are used.

#### 3.2 Atmospheric Absorption

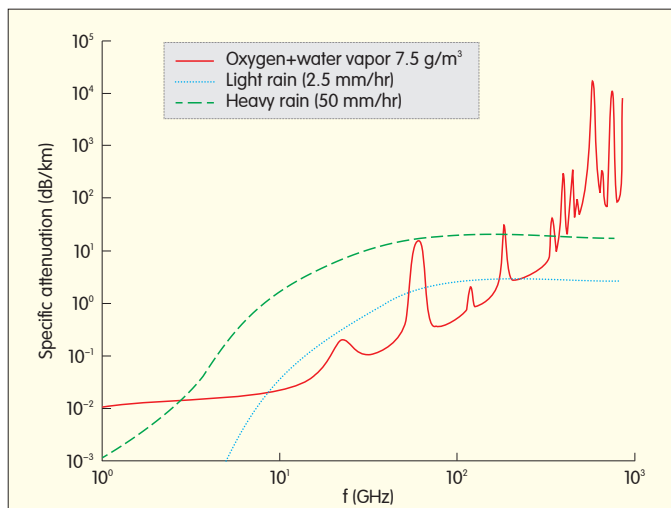
Atmospheric absorption is also often cited as a reason that millimeter wave and terahertz communications are impractical. Atmospheric absorption in much of the millimeter wave and terahertz spectrum is much higher than that of conventional cellular spectrum. This should, however, be seen in context. A traditional macro cell may have to cover link distances of several kilometers. At these distances, gaseous water and molecular oxygen absorption (terrestrial) can be

prohibitive. The unlicensed 60 GHz band is intentionally positioned at an oxygen absorption line, and this results in approximately 15 dB/km attenuation. This is a big loss at macro cell distances, but for small cells, the loss is quite manageable in terms of link budget. For example, if a circular cell needed to have a coverage distance of 150 m, the worst-case gaseous absorption loss at 60 GHz is only 2.25 dB. The gaseous absorption drops off at higher frequencies and remains below 15 dB/km until the water line is encountered at about 160 GHz. Although atmospheric absorption reduces received power, it has a benefit in the form of higher signal-to-interference ratio (SIR). Desirable signals generally emanate from base stations physically closer to the intended receiver; undesirable signals (interference) generally emanate from base stations that are further away. Thus, SIR is improved.

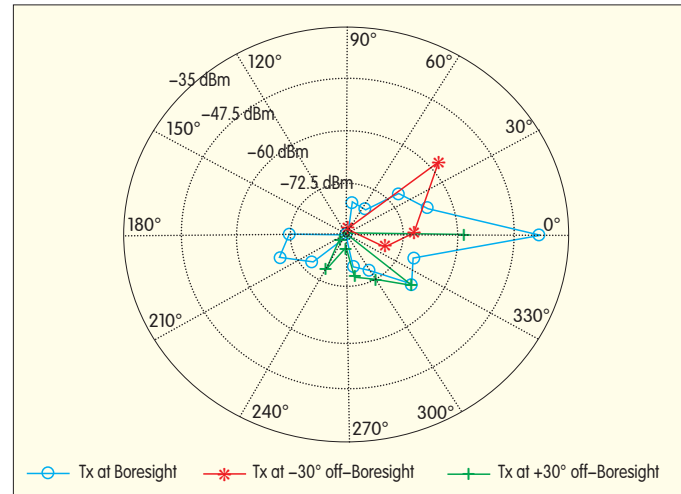
Rainfall is also often cited as a reason that millimeter wave and terahertz communications are impractical. However, the same argument regarding small cells and gaseous absorption can be made. The comparatively short link distances mean that loss from rainfall is tolerable most of the time. A 25 mm/h rainfall (heavy rainfall) could cause about 10 dB/km additional loss. At 50 mm/h, the loss is about 20 dB/km. In New York, rainfall greater than 50 mm/h occurs only 0.01% of the time and is short-lived [5]. With only 3 dB extra margin, a system in New York with 150 m links has 99.99% protection against rainfall. Fig. 1 shows the atmospheric absorption at sea level for different rates of rainfall and for oxygen and water vapor a  $7.5 \text{ g/m}^3$ .

### 3.3 Diffraction and Penetration Loss

Diffraction and penetration loss are arguably the main problems with millimeter wave and terahertz communication. Diffraction properties are almost optical, and many common building materials are almost opaque to millimeter wave and terahertz signals. This makes the channel highly specular, and such a channel is sometimes referred to as a “billiard” channel because of the occasional need to make a bank shot



▲ Figure 1. Atmospheric absorption at sea level.



▲ Figure 2. Specular channels.

to get around an impenetrable obstruction (Fig. 2).

For very high performance, links should be LoS or require a small number of reflections. Fortunately, small cell architectures and electrically steerable narrowbeam antennas can help mitigate many problems. Small cells are more likely to be serving mobiles that are reachable using either LoS transmission or a single reflection. Electrically steerable narrowbeam antennas are responsive and can quickly switch to unobstructed paths.

### 3.4 Waveforms and Multiple Access Schemes

Wide bandwidth; high-gain, narrowbeam antennas; and cost-effective electrically steered antennas influence the choice of waveform and multiple access schemes. We make the following observations:

- An advantage of the millimeter wave spectrum, with its large available bandwidths, is that there is less need for high spectral efficiency. Simple modulations may be appropriate.
- Wider bandwidths and higher data rates are also beneficial in that lower transmit power can be used to achieve the same capacity as a smaller bandwidth channel.
- Narrowbeam transmission implies that fewer mobiles are illuminated by a particular transmission. Although interference is reduced by the spatial containment of signals, narrowbeam transmission limits the possibility to serve more than one mobile with a single beam. The potential advantage of frequency domain scheduling is therefore limited (assuming RF beamforming is used), and this reduces the need for multicarrier waveforms. The lower peak-to-average power ratio (PAPR) of single-carrier waveforms is also more attractive for low-cost electronics.
- Narrowbeam transmission also implies mobiles may be spatially separated and serviced simultaneously using spatial division multiple access (SDMA) techniques. However, SDMA would require multiple RF and baseband processing chains, which initially may be cost prohibitive. A reasonable system might therefore have wide bandwidth, a simple low-PAPR waveform with low-order modulation, and

TDMA access schemes.

## 4 Regulation

The International Telecommunication Union (ITU) is the main telecommunications regulatory body in the world and coordinates the shared global use of radio spectrum. There are also several organizations with national jurisdiction to regulate spectrum. In the US, the Federal Communications Commission (FCC) regulates spectrum according to Title 47 (Telecommunications) of the Code of Federal Regulations (CFR). In Europe, spectrum is regulated by the European Conference of Postal and Telecommunications Administrations (CEPT) according to reports prepared by the Electronics Communications Committee (ECC) and European Radio Communications Committee (ERC) and according to standards formulated by the European Telecommunications Standards Institute (ETSI). In China, the Bureau of Radio Regulation under the Ministry of Industry and Information Technology regulates spectrum. In this section, regulatory issues concerning frequencies in the millimeter wave band are outlined.

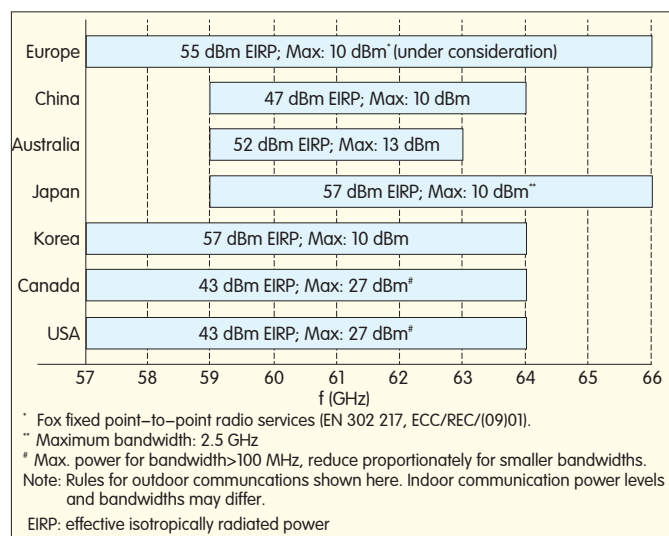
### 4.1 Unlicensed Spectrum

Unlicensed spectrum has predefined rules for both hardware and radio deployment so that interference is mitigated by the technical rules defined for the band rather than being restricted for use by entities through a spectrum licensing approach. The industrial, scientific, and medical (ISM) frequency bands are unlicensed and allow both communicating and non-communicating devices to operate. Co-channel interference from a non-communicating device can therefore affect performance of communication networks operating in such bands. However, the 60 GHz band is largely free of non-communicating devices (except on a small set of frequencies from 61 to 61.5 GHz, which is an ISM band), and because of this, the variety of potential interferers is greatly reduced.

### 4.2 Unlicensed Millimeter Wave Spectrum

Unlicensed operations on the 57–64 GHz band in the US (referred to as the 60 GHz band) are regulated by Title 47, Part 15 of the CFR (47 CFR 15). The regulations of 47 CFR 15 and the significant absorption of the 60 GHz band by atmospheric oxygen makes this band better suited for short-range point-to-point and point-to-multipoint applications. Fig. 3 shows the main regulatory provisions of the 60 GHz band in various countries. In response to a representation made by the Wireless Communications Association (WCA) in 2007, the FCC introduced a proposal seeking to change 60 GHz transmission rules [6], [7].

Table 1 summarizes the transmit power and antenna requirements for the 60 GHz band in the US and Europe. In a millimeter wave system that supplements a cellular network, an electronically steerable phase-array antenna is required to satisfy the minimum antenna gain specified in the regulations and to support user mobility. To conform to different regional



▲ Figure 3. Regulatory overview of the 60 GHz band in various countries.

specifications, antennas of different sizes and containing tens of elements may be required. Also, because the beams from these antennas have limited steerability, multiple arrays may be needed, especially at the base station, to provide omnidirectional coverage.

### 4.3 Other Bands of Interest

#### 4.3.1 The 70, 80 and 90 GHz (E/W) Bands

Table 2 summarizes the transmit power and antenna requirements for the 70, 80, and 90 GHz bands in the US and Europe. Existing E-band point-to-point links are essentially static because they use parabolic dish antennas to satisfy the

▼ Table 1. Regulatory requirements on power and antenna for 60 GHz band

Region	EIRP (dBm)	Max. TX Power (dBm)	Min. Antenna Gain (dBi)
US	43	27	16*
Europe	55	10	30*
* derived quantity		EIRP: effective isotropic radiated power	

▼ Table 2. Regulatory power and antenna requirements for 70, 80 and 90 GHz bands

Region	EIRP (dBm)	Max. TX power (dBm)	Min. Antenna Gain (dBi)	Max. Beamwidth (degrees)
70/80 GHz				
US	85 <sup>§</sup>	35	43	1.2
Europe	85*	35	38	–
90 GHz				
US	85	35*	50	0.6
Europe#	–	–	–	–
* No rules defined yet for 90 GHz band in Europe. * (50–G), when antenna gain (G) < 50 dBi. <sup>§</sup> For antenna gain ≥ 50 dBi # For antenna gain ≥ 55 dBi. EIRP [dBm] = 85–2 EIRP: effective isotropic radiated power				



## Special Topic

### Millimeter Wave and Terahertz Communications: Feasibility and Challenges

Phil Pietraski, David Britz, Arnab Roy, Ravi Pragada, and Gregg Charlton

regulatory requirements on antenna gain and beamwidth. Such antennas make network planning expensive and affect scalability. To make a network truly scalable, an electronically steerable phased-array antenna is required. This allows a transceiver using a single antenna array to dynamically steer the beam in the desired direction of another nanocell node. However, in order to meet regulatory requirements on antennas, a fairly large antenna array is required, and this drives up technological challenges and costs.

#### 4.3.2 Low Millimeter Wave Bands

There are also millimeter wave bands of interest below 60 GHz, but these tend to have smaller bandwidths (Table 3).

#### 4.3.3 High Millimeter Wave Bands

To achieve even higher capacity and data rates,

▼ Table 3. Millimeter wave frequency bands of interest

Band	Frequency Range (GHz)	Bandwidth (GHz)	Current Use
23 GHz	21.2–23.6	2.4	Fixed point-to-point wireless service (entire band not available in all regions)
LMDS	27.50–28.35	1.3	Wireless cable TV (point-to-multipoint), competitive local exchange carriers (CLEC) for business customers
	29.10–29.25		
	31.075–31.225		
39 GHz	38.6–40.00	1.4	Fixed point-to-point links for backhaul
46 GHz	45.5–46.9	1.4	Vehicle radar and cordless phones in small portions of the band, otherwise unallocated

LMDS: local multipoint distribution service

opportunities exist in the higher millimeter wave band. There are atmospheric absorption windows at 140 GHz and 240 GHz that avoid the absorption peaks at 119 GHz, 183 GHz, and 325 GHz. In these bands, there is sufficient spectrum to allocate 40 GHz of bandwidth from 125 to 165 GHz and 100 GHz of bandwidth from 200 to 300 GHz. To use this band, fragmented allocations must be consolidated and service rules must be framed while avoiding frequencies meant for passive services. There is a ten-year timeframe for new service rules once a motion has been introduced at an ITU WRC.

#### 4.4 Radio Frequency Radiation Exposure Limits

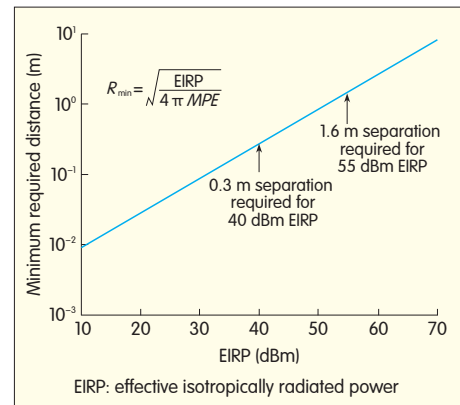
The possible health implications of human exposure to radio frequency electromagnetic (RFEM) fields have led to output power of radio transmitters being regulated. Millimeter wave frequencies are absorbed by moisture in the human body and cannot penetrate the outer layers of skin. Exposure to millimeter wave radiation must not exceed the maximum permissible exposure (MPE). The current MPE for 1.5–100 GHz is 1 mW/cm<sup>2</sup>, measured at a minimum distance of 5 cm from the radiating surface. Fig. 3 shows the minimum required separation between a human body and radiating surface to comply with MPE at 60 GHz. The minimum

separation for the maximum permissible EIRP for US and Europe are marked in Fig. 4. Antenna design and placement are crucial to MPE compliance in user devices.

## 5 High-Level System Architecture

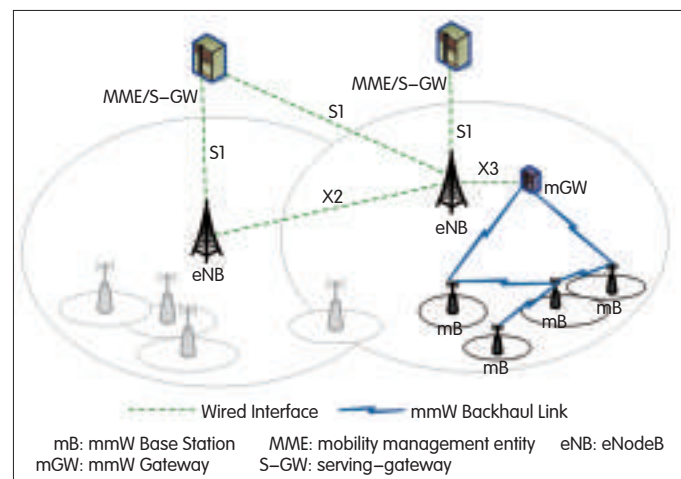
If millimeter wave and terahertz frequencies are introduced

Figure 4. ▶ Minimum distance to be maintained between a human body and radiating surface for MPE compliance.



into a cellular system, new architectural elements need to have superhigh capacity and permit connectivity with highly directional beams. An existing cellular system, such as an LTE system, provides adequate coverage, security, and reliability and is a good framework upon which a millimeter wave or terahertz layer can be added. Essentially, this is a new heterogeneous nanocell layer for 5G radio networks. All control signaling—including system information, idle-mode mobility, paging, random-access channel (RACH) access, radio resource control (RRC) and nonaccess stratum (NAS) signaling (signaling radio bearers), and multicast traffic—is provided via the existing cellular layer. The millimeter wave/terahertz layer provides high throughput for offloading traffic from the existing cellular bands below 6 GHz.

The proposed layered architecture with cellular overlay and millimeter wave/terahertz underlay as shown in Fig. 5. It provides a clear evolutionary path by building on existing



▲ Figure 5. Tiered architecture.



cellular/4G architectures using carrier aggregation, which was introduced in 3GPP Release 10. The millimeter wave architecture introduces two new nodes: a millimeter wave base station (mB) nanocell and a millimeter wave gateway (mGW). The mB primarily has millimeter wave/terahertz access links to mobiles and millimeter wave/terahertz backhaul (BH) links to other mBs and the cellular base station (eNB). However, wired connections are also possible. The mBs are expected to perform millimeter wave and terahertz physical-layer functions and possibly certain MAC-layer functions.

Apart from super-channel data processing, where the millimeter wave/terahertz link has high throughput, an mB is also expected to perform scheduling-related functions for millimeter and terahertz frequencies assigned to the mB by the eNB. To relieve the eNB of data processing and routing of user data that is carried on the millimeter wave layer, another logical node called millimeter wave gateway (mGW) node is introduced. The mGW is responsible for routing and processing user data carried over the millimeter wave overlay. This processing is done on the higher-layer access stratum (AS). An important motivation for introducing the mGW node is to support a large number of mBs in a scalable manner. This relieves the load on the eNB and provides an additional data pipe from the mBs to the evolved packet core (EPC). New network (logical) components, such as selective frequency-routing devices, are likely necessary. These frequency-routing devices are designed to separate out in real time the high data traffic from the lower-bandwidth cellular traffic. The high data traffic is routed to the superchannel. This avoids frequent hand-off overhead for control-plane bearers and low-throughput traffic.

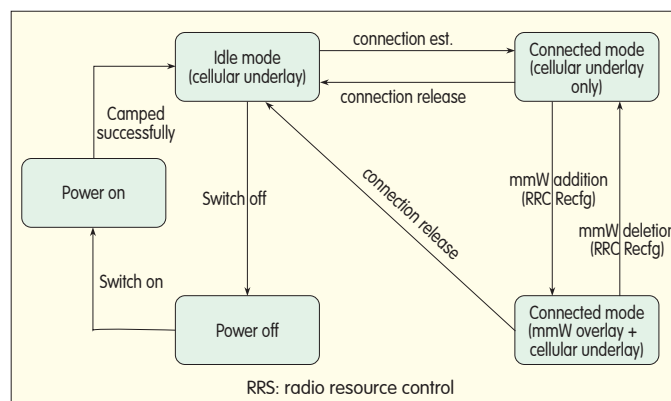
Mesh backhaul is valuable for the architecture because it increases deployment options and flexibility. Wired backhaul coverage can be scarce or cost-prohibitive. The backhaul links between mBs form a multihop mesh network, and long backhaul links are not required. This reduces capex and increases backhaul reliability by providing multiple routes. The high directionality of the millimeter wave/terahertz beam also allows the same spectrum to be used for both access and self-backhaul in the millimeter wave/terahertz layer.

For millimeter wave and terahertz technology to be widely accepted, it is imperative that mB capex and opex are low. Self-configuration, self-optimization, and self-healing are crucial to cheap mB deployment. Outdoor mB units must be small, lightweight, and “belt-able” for easy installation. They can be mounted on existing street lampposts and do not require air-conditioning or indoor housing. Because they have low energy needs, power-over-Ethernet (PoE) feeding or even local harvesting, for example, solar, may be used. Typically, mBs would be mounted on lampposts six meters or higher above the ground. At this height, there are fewer obstacles to interfere with propagation. For newly deployed mBs, an eNB can provide initial system configuration and may assist in the configuration of backhaul links with neighboring mBs. A new mB may also use information from neighboring mBs in a docitive manner in order to determine and configure

the initial set of system parameters for its operation.

### 5.1 UE Perspective

A millimeter-wave-enabled UE first needs to connect to the cellular layer before it can connect to the millimeter wave/terahertz underlay layer. Fig. 6 shows how UEs obtain millimeter wave/terahertz connectivity. The eNB coordinates



▲ Figure 6. Millimeter wave layer connectivity at the UE.

with the corresponding mB that the UE connects to. At power on, and with successful camping on cellular layer, the UE switches to idle mode and connects to the eNB. The eNB, after considering the mBs involved, will determine a suitable mB for the UE to connect to and will provide the specific millimeter wave configuration information to the UE via RRC procedures. Once the UE is finished with millimeter wave/terahertz services, it can switch to idle mode (if it is not using any cellular underlay services) or it can switch to connected mode (where only cellular underlay services are used and the millimeter wave layer is deleted).

## 6 Mobility Challenges

The directional nature of the millimeter wave link introduces mobility challenges that are unique to the millimeter wave/terahertz layer. To effectively use highly directional beams and lock onto and track the mobile user, highly directional antennas with beam steering and tracking capabilities are crucial. Traditional hand-off mechanisms need to be reassessed because of the near LoS and high directional requirements of millimeter wave links. Even when the user changes direction or orientation, hand-off procedures may be triggered. New mechanisms to reduce association and/or synchronization time with the target mB(s) need to be developed.

To alleviate hand-off issues, several new mechanisms need to be developed, and this involves significant changes to architecture. A possible solution is to separate control-plane and user-plane functionality between the eNB and mBs. With this approach, control-plane and higher-layer data-plane protocols, including radio link control (RLC) and packet data convergence protocol (PDCP), run at the eNB. Such an architecture can be easily built to coexist with the

architecture proposed in section 5. This approach minimizes data loss caused by mobility. The RLC layer is still terminated at the eNB because window-based mechanisms, such as ARQ, and buffering mechanisms are typically implemented in the RLC layer. One benefit of this approach is that security and ciphering/integrity algorithms are executed at the eNB, and this eases the burden on mBs by eliminating the need for ciphering and/or trust-zone features.

As the size of the nanocells shrink, another alternative is to have nanocells work in cooperative clusters. They can be designed to hand-off and forward customer data traffic between themselves and the next sequential group of nanocells (clusters) surrounding the customer as they move sequentially through the nanocells. This cooperative “moving” of concatenated nanocell clusters at the network edge differs from the existing method of routing and hand-off decisions at the local Central Office or deeper within the network. Because of their small coverage, these individual nanocells do not have enough time to route customer traffic through traditional circuits for hand-off decisions. They require such decisions to be executed locally within the network edge and nanocell clusters. This is the intelligent edge model.

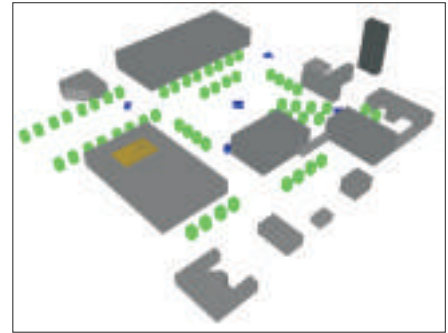
## 7 Simulations

System simulations were conducted to determine the feasibility of outdoor millimeter wave/terahertz coverage. The system simulator comprises two main modules. One is a commercially available cell-planning tool used by a number of European operators. WINPROP, from AWE Communications, makes detailed propagation predictions based on ray tracing and takes transmit antenna patterns into account. This allows accurate, static multipath channel models to be created between each transmitter and receiver pair. WINPROP determines the delay, signal strength, AoA, and AoD. If they occur, it also determines the types of interactions (for example, LoS, reflection, and diffraction) for each path arriving at a user-defined point. The other main module introduces time-varying elements, such as fading and scheduling, into the simulations and is written in MATLAB. This module creates channel models based on WINPROP data. It estimates the throughput of each UE in the system by modeling the impact of factors such as transmit power, interference, receive antenna patterns, and schedulers. The statistics are then collated to give an estimate of overall network performance.

Transmit antenna patterns are generated for the mB phased array antennas and are used in the ray-tracing simulation. We assume a  $7 \times 7$ , 49 element uniform rectangular array (URA) and classical beamforming. Each URA is limited to cover  $90^\circ$  ( $\pm 45^\circ$ ) in azimuth; that is, four arrays are used to achieve  $360^\circ$  coverage. Such an antenna provides a gain of 25.9 dBi with  $20^\circ$  beamwidth at broadside. The result is an overall transmit EIRP of 40 dBm perpendicular to the array.

The deployment was located in the central area of a large, public university (Fig. 7). Five mBs were deployed around the quadrangle—one in each corner and one in the middle.

Figure 7. ▶  
3D view of college campus.



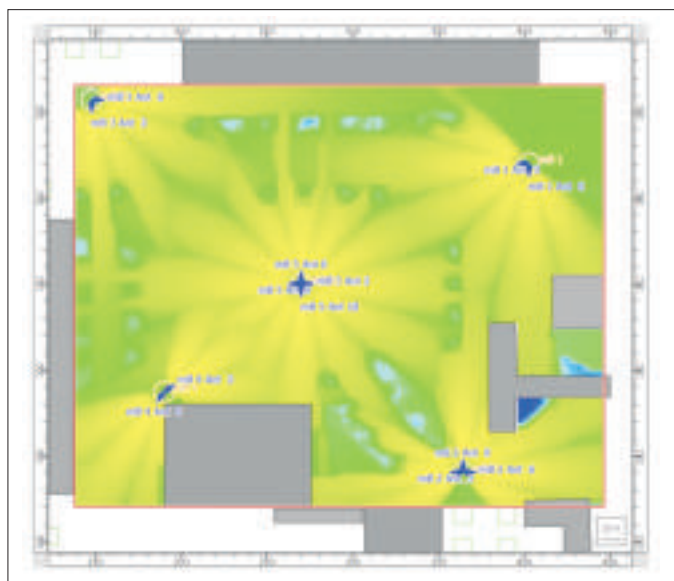
Because ray tracing is used, the RF properties of the building materials must be characterized in order to capture transmission loss, reflection loss, and diffraction loss. At the simulated carrier frequency of 60 GHz, diffraction is typically insignificant, but penetration and reflection loss must be taken into account. Simulation assumptions are given below for a variety of building materials. In the campus scenario, all buildings are assumed to have concrete walls of typical thickness, have building penetration loss of 19 dB, and have reflection loss of 14 dB [8].

The simulation assumptions are:

- 50 UEs per drop
- 10 drops
- 1000 time samples (TTIs)
- 60 GHz carrier frequency with 2 GHz bandwidth
- modulation limited to QPSK
- round robin scheduling
- each mB array chooses from one of three beam directions (all at  $0^\circ$  elevation), and each beam corresponds to  $7 \times 7$  URA
- Omni UE antenna
- 6 dB noise figure
- 5 mBs
- 40 dBm EIRP
- 4 m mB height with 1.5 m UE height
- Rician fading per path
- full buffer traffic.

The intent is to determine the feasibility of a relatively uncomplicated millimeter wave system, so complex schedulers, interference coordination, and directional receive beams are not used. Accordingly, simple QPSK modulation with round robin scheduling is assumed. The number of UEs per drop is 50 and corresponds to a 10% activity factor for a UE density of 8000 UE/km<sup>2</sup>. The 40 dBm EIRP level is consistent with FCC regulations.

Fig. 8 shows the RF coverage in the campus deployment. The yellow regions correspond to received power levels greater than  $-60$  dBm, and the green regions correspond to received power levels between  $-70$  dBm and  $-60$  dBm. There are twelve separate beams because there are four arrays in the center mB and there are three possible beam directions from each array. Only one of these three beams in an array is active in any TTI. These levels are considered sufficient to support the lowest MCS class imposed by standards such as 802.11ad. The received power is adequate across most of the



▲ Figure 8. Coverage plot for central campus area.

central quadrangle indicating that, when properly deployed, millimeter wave technology can serve outdoor users with very high data rates. Building edges and vegetation are obstacles to millimeter wave service.

Fig. 9 shows the CDF for user throughput, and Fig. 10 shows the total throughput. Approximately 95% of the UEs receive millimeter wave links, and the mean throughput is 500 Mbit/s, which is very large. In practice, the overlay LTE network is used to service the 5% of UEs that cannot receive millimeter wave links. Because the LTE network is so lightly loaded, these UEs should still receive good service. Moreover, the throughput at the millimeter wave cell edge (10th percentile) is approximately 100 Mbit/s, and the median total throughput for the campus network is 32 Gbit/s. These figures provide some numerical context to the coverage plot shown in Fig. 8. The discrepancy between the megabyte-per-second cell-edge user throughput and the gigabyte-per-second network throughput is due to the fact that there are 50 active users with a median user throughput of 500 Mbit/s. This equates to a network throughput on the order of 25 Gbit/s.

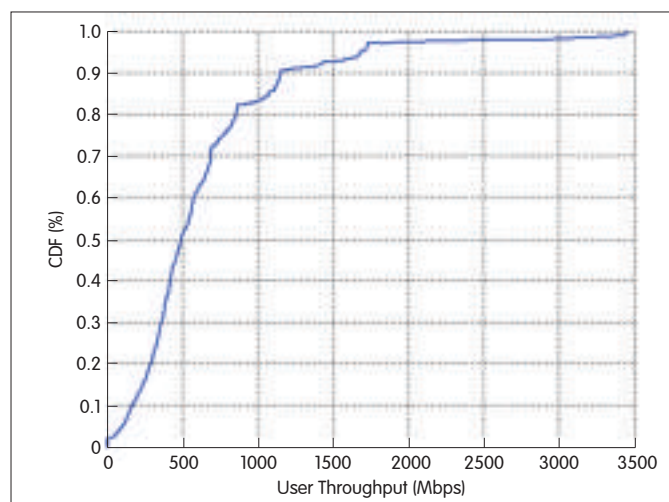
The campus could be covered by a 3.5 GHz picocell network; however, such a system would be constrained to 100 MHz bandwidth, and the necessary spectral efficiency to match the proposed millimeter wave network would have to be about 20 times greater. A combination of greater spectral efficiency and many more 3.5 GHz pico base stations could be used to achieve the same end, but such super-dense deployments should be avoided.

When human beings block the propagation paths between the transmitter and receiver, there is a significant increase in path loss and a corresponding loss in throughput. Human blockage comes in two forms that have statistically different mechanisms: self-blockage by the user (highly related to UE location) and blockage by other humans in the area (not

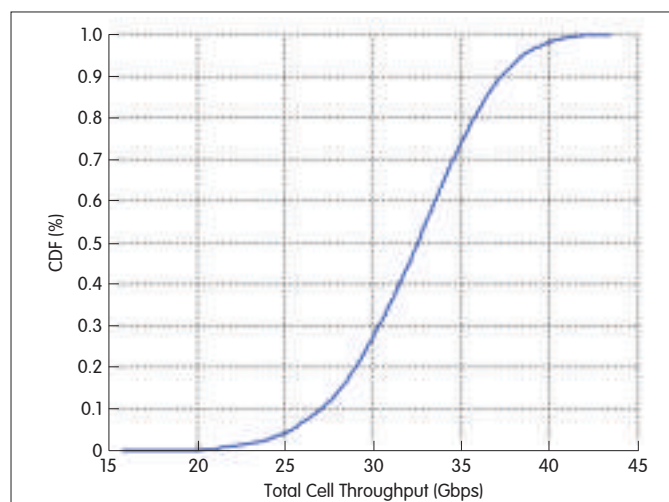
related to UE). A detailed description of this phenomenon and its effect on network and individual user performance is given in [9]. The authors of [9] used a statistical human blockage model to study the effect population density on throughput. The factors affecting blocking probability were population density, Tx–Rx distance, and Tx–Rx height difference. Fig. 11 shows the effect of human blocking for different population densities [9]. It was shown that using a pair of directional antennas in the receiver improves performance over omnidirectional receive antennas. A pair of directional antennas has a good chance of being able to point a high-gain beam at at least one reflected path, thus recovering much of the performance lost by blocking of the LoS path. Of course, some UEs are still forced out of coverage.

## 8 Conclusion

The challenges of and motivations for developing millimeter wave and terahertz communications have been presented. A



▲ Figure 9. User throughput.

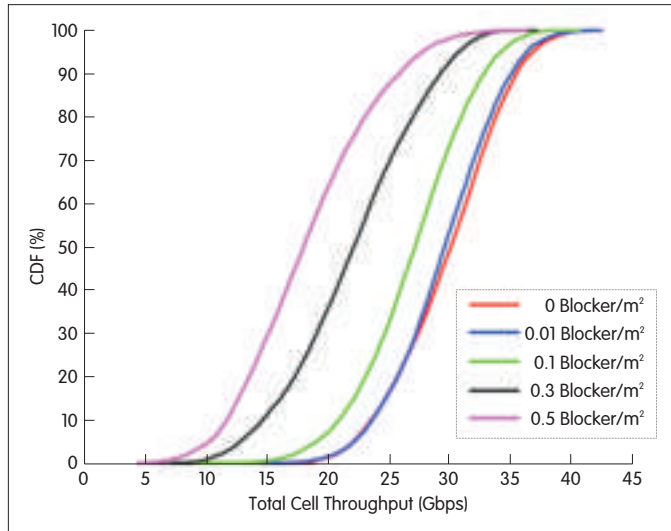


▲ Figure 10. Total cell throughput.



## Millimeter Wave and Terahertz Communications: Feasibility and Challenges

Phil Pietraski, David Britz, Arnab Roy, Ravi Pragada, and Gregg Charlton



▲ Figure 11. Network throughput as a function of blocker density.

high-level candidate architecture was also described. Use cases have been enumerated, and the potential applicability of high-frequency links in these cases has been explained. Although the path loss caused by millimeter wave and terahertz signals is significant, additional antenna gain at these wavelengths can overcome the problem, and there is even an additional benefit of reduced intercell interference.

Last, a feasibility study on 60 GHz access links on a college campus was presented. The results show that millimeter wave coverage is feasible in a large, public area, and only a limited number of base stations are required. Additionally, the 95% of users who were covered by millimeter wave links enjoy very high throughput. However, the effect of human blockage can be quite severe, reducing network throughput by about 10% when the density of humans was approximately one person per 10 m<sup>2</sup>. Arguably, this is a reasonable density to assume on campus. Using a pair of directional receive antennas restores much of the loss caused by human blocking.

### References

- [1] Rysavy Research. (2010, February). Mobile Broadband Capacity Constraints And the Need for Optimization [Online]. Available: [http://www.rysavy.com/Articles/2010\\_02\\_Rysavy\\_Mobile\\_Broadband\\_Capacity\\_Constraints.pdf](http://www.rysavy.com/Articles/2010_02_Rysavy_Mobile_Broadband_Capacity_Constraints.pdf)
- [2] Paul Rasmussen. (2009, May). Fierce Wireless article [Online]. Available: <http://www.fiercewireless.com/europe/story/orange-reports-500-increase-dongle-subs-data-growth-booms/2009-05-15#ixzz25WgMwQ7U>
- [3] UMTS Forum, "Mobile traffic forecasts: 2010–2020 report," UMTS Forum Report 44, London, 2011.
- [4] Sebastian Priebe, David M. Britz, Martin Jacob, Stephen Sarkozy, and Thomas Kurner, "Interference Investigations of Active Communications and Passive Earth Exploration Services in the THz Frequency Range," *IEEE Trans. on THz Sci. and Tech.*, vol. 2, no. 5, pp. 525–537, September 2012.
- [5] M. Marcus, "Millimeter Wave Propagation: Spectrum Management Implications," *IEEE Microwave Magazine*, vol. 6, no. 2, pp. 54–62, June 2005.
- [6] Revision of the Commission's Rules Regarding Operations in the 57–64 GHz [Online]. Available: [http://hraunfoss.fcc.gov/edocs\\_public/attachmatch/FCC-07-104A1.pdf](http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-07-104A1.pdf)
- [7] Comments of IEEE 802.18 [Online]. Available: [http://www.ieee802.org/18/Meeting\\_documents/2007\\_Sept/18-07-0082-01-0000\\_d1\\_ET\\_07-113\\_NPRM.pdf](http://www.ieee802.org/18/Meeting_documents/2007_Sept/18-07-0082-01-0000_d1_ET_07-113_NPRM.pdf)
- [8] B. Langen, G. Lober, and W. Herzig, "Reflection and Transmission Behavior of Building Materials at 60 GHz," in *Proc. IEEE 5th International Conference on Personal, Indoor, and Mobile Radio Communications*, The Hague, Netherlands, 1994, pp. 505–509.
- [9] M. Abouelseoud and G. Charlton, "The Effect of Human Blockage on the

Performance of Millimeter-wave Access Link for Outdoor Coverage," submitted to IEEE VTC2013–Spring Conference, 2013.

- [10] W. Jing, R. Prasad, and I. Niemegeers, "Analyzing 60 GHz radio links for indoor communications," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 4, pp. 1832–1840, Nov 2009.
- [11] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, UK: Cambridge University Press, 2005.

Manuscript received: August 14, 2012

### Biographies

**Phil Pietraski** (philip.pietraski@interdigital.com) received his BSEET from DeVry University in 1987. He received his BSEE, MSEE, Grad.Cert. in wireless communications, and PhD EE from Polytechnic University, Brooklyn (now NYU–Poly) in 1994, 1995, 1996, and 2000. He joined InterDigital Communications in 2001 and is currently a principal engineer leading research activity in wireless communications, most recently in millimeter wave communications and future cellular architectures. He holds more than 50 patents in wireless communications and has authored multiple conference and journal papers. He is vice chair of the MoGig (Mobile Gigabit) working group at IWPC and a trustee for DeVry NJ campuses. Prior to his transition to wireless communications in 2000, he was a research engineer at Brookhaven National Laboratory, National Synchrotron Light Source, responsible for beam-line instrumentation and X-ray detector R&D. He has also conducted research at the Polytechnic University for the Office of Naval Research (ONR) in underwater source localization.

**David Britz** (dbritz@research.att.com) is AT&T's subject matter expert on free-space optical communications (FSOC) and the Terahertz Communications initiative. He is a principal member of the technical staff at AT&T Labs Research (Shannon Labs) and has been with AT&T contiguously for 28 years. His earlier work encompassed forward-looking technologies and advanced design and development of public communications products, ISDN telephones, and advanced speakerphones. From the mid 1990s, he worked on in-building and terrestrial optical wireless. Since 2007, he has been the lead researcher on terahertz technologies and network applications for multi-gigabit nanocell mobile networks. He was a founding member and chairman of the FSO Alliance, founding member of IEC–TC76 Working Group 5 part 12, and delegate to the IEC on laser safety. He is also a founding member and current chairman of the MoGig Working Group and the founding member and vice chair of IEEE 802.15 Terahertz Interest Group. He is currently engaged with the ITU/WRC and US CORF delegation on impending terahertz spectrum usage and allocations. He has been granted nineteen patents over his career at AT&T. He graduated from Rhode Island School of Design receiving his Masters degree in industrial design in 1980.

**Arnab Roy** (Arnab.roy@interdigital.com) received his BE degree in electronics engineering from Mumbai University, India, in 2001. HE received his MS and PhD degrees in electrical engineering from Penn State University in 2004 and 2011. His general interests include signal processing and communication systems engineering. He is currently working on millimeter wave global spectrum harmonization and associated system development at InterDigital Communications.

**Ravi Pragada** (Ravi.Pragada@interdigital.com) received his BEEng. degree from Andhra University, India. He received his MS degree in communication systems engineering from SUNY, Buffalo, in 1999. He joined InterDigital in 2001 where he began working on several 3GPP FDD/TDD development projects, both for handset and infrastructure products. He was also deeply involved with device-to-device communications. Currently, he is a principal engineer at InterDigital and is focused on millimeter wave communications and beyond-4G architectures. Prior to working at InterDigital, he was part of a Motorola team that developed RNC and NodeB infrastructure for 3GPP UMTS systems.

**Gregg Charlton** (Gregg.charlton@interdigital.com) received his B.S.E.E. and M.S.E. degrees from Carnegie Mellon University and the University of Michigan in 1982 and 1986. He has worked in the areas of fiber optic, cellular, and satellite communications over the course of his career at companies such as TRW, Lockheed Martin, and AT&T Bell Laboratories. He joined InterDigital in 2000 and is currently a member of technical staff there. He has worked on establishing system requirements and characterizing GSM, UMTS, and LTE system performance. His current research at InterDigital is on millimeter wave communications systems analysis and simulation.

# WiGig and IEEE 802.11ad for Multi-Gigabyte-Per-Second WPAN and WLAN

Sai Shankar N, Debashis Dash, Hassan El Madi, and Guru Gopalakrishnan

(Tensorcom Inc., 5900 Pasteur Court, Carlsbad, CA 92008, USA)

## Abstract

The Wireless Gigabit Alliance (WiGig) and IEEE 802.11ad are developing a multigigabit wireless personal and local area network (WPAN/ WLAN) specification in the 60 GHz millimeter wave band. Chipset manufacturers, original equipment manufacturers (OEMs), and telecom companies are also assisting in this development. 60 GHz millimeter wave transmission will scale the speed of WLANs and WPANs to 6.75 Gbit/s over distances less than 10 meters. This technology is the first of its kind and will eliminate the need for cable around personal computers, docking stations, and other consumer electronic devices. High-definition multimedia interface (HDMI), display port, USB 3.0, and peripheral component interconnect express (PCIe) 3.0 cables will all be eliminated. Fast downloads and uploads, wireless sync, and multi-gigabit-per-second WLANs will be possible over shorter distances. 60 GHz millimeter wave supports fast session transfer (FST) protocol, which makes it backward compatible with 5 GHz or 2.4 GHz WLAN so that end users experience the same range as in today's WLANs. IEEE 802.11ad specifies the physical (PHY) sublayer and medium access control (MAC) sublayer of the protocol stack. The MAC protocol is based on time-division multiple access (TDMA), and the PHY layer uses single carrier (SC) and orthogonal frequency division multiplexing (OFDM) to simultaneously enable low-power, high-performance applications.

## Keywords

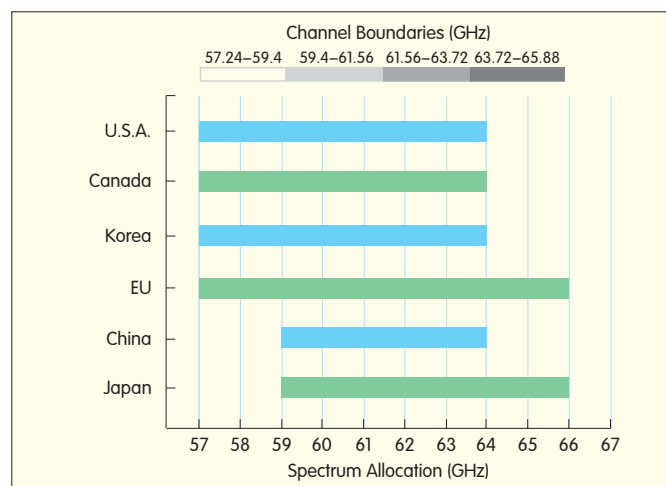
60 GHz communications; IEEE standards; WiGig; 802.11ad; contention based access protocol; scheduled protocol; Beamforming; power save

## 1 Introduction

There is a huge amount of unlicensed spectrum available worldwide in the 60 GHz band. Academia and industry have turned to the 60 GHz spectrum because of the universal availability of unlicensed spectrum, the ever-growing number of user applications creating heavy data traffic, and the need to reduce data transfer times. Considerable efforts have been made to use this spectrum and spur the development of silicon, similar to what happened with the 2.4 GHz ISM band 15 years ago. 60 GHz millimeter wave technologies offers a way to provide end users with guaranteed quality of service (QoS) for different applications. Fig. 1 shows the allocation for 60 GHz in different countries [1]–[3].

60 GHz millimeter wave technologies create significant problems in designing the radio frequency (RF) front-end, processing gigabit-per-second data, and migrating to 40 nm and 28 nm low-power technologies in designing the silicon, considerable progress have been made in making it practical

and feasible [4]–[6]. 60 GHz millimeter wave systems are needed to cater for newer applications, such as streaming video in the home or office, that have flourished as a result of



▲ Figure 1. Spectrum allocation for WiGig.



last-mile access provided by internet service providers (ISPs). Such systems will also eliminate the need for cables around docking stations, and this will reduce clutter and allow easier connection between devices. There are multiple industry organizations involved in 60 GHz standardization, the notable ones being Wireless HD [7], IEEE 802.15.3c [8], WiGig [9], and IEEE 802.11ad [10]. The last two of these organizations involve a large number of silicon, OEM, and telecom companies that are motivated to have a single worldwide 60 GHz standard. WiGig began standardization in 2008 and has recently released the WiGig 1.0 standard. IEEE 802.11ad also began standardization in 2008 and has recently released IEEE 802.11ad Draft 9.0 standard. These standards are similar, and in this paper, we will refer to 802.11ad as the representative of both standards, pointing out when there is a feature that is unique to the WiGig standard. Similar standardization efforts have been made by ECMA-387 and CMMW Study Group [2], [3]. 60 GHz millimeter wave is the next wireless networking technology and will appear in the market around 2014 [11]. It is poised to repeat the successes of Bluetooth and Wi-Fi [12]. This explosive growth of the wireless industry in such a short time can also be attributed to the opening of unlicensed bands in 60 GHz by the Federal Communications Commission (FCC).

802.11ad aims to develop the protocol adaptation layers (PALs) to support a plethora of applications that will arise from the elimination of cables and from fast wireless sync and transfer. The PALs being considered by WiGig include wireless serial extension (WSE), which eliminates USB 3.0 cables; wireless bus extension (WBE), which eliminates PCIe 3.0 cables; wireless display extension (WDE), which eliminates high-definition multimedia interface (HDMI) and display port cables; and wireless secure digital (WSD), which makes secure digital input/output card (SDIO) disks wireless. The first important 60 GHz millimeter wave application to enter the market as wireless docking based on PCIe 3.0—with one second-generation lane (also called x2)—or USB 3.0. All devices with 802.11ad MAC/PHY/Radio use the corresponding PALs between the application and MAC layers to seamlessly transfer information between devices as if the devices were connected by wires. Another 60 GHz application is wireless HDMI based on WDE, which allows transfer of uncompressed bits from devices such as set top boxes and blue ray disc players to television screens and from laptops, desktops, or ultrabooks to monitors via a display port cable replacement. The WDE also supports H264 compressed rates for handling variations in the wireless channel and to ensure seamless content delivery to the end users. Performance of the PHY and MAC protocols is analyzed in [13] and [14].

In this paper, we describe the novel features of the MAC and PHY sublayers of the protocol stack defined in 802.11ad. In section 2, we describe the TDMA protocol and the need for directionality in 60 GHz. In section 3, we outline the 802.11ad PHY layer, and in section 4, we outline the MAC layer. In section 5, we outline the beamforming protocol, and in section 6 we outline the power-saving protocol. In section 7, we

describe the fast session transfer, and in section 8, we show achievable rates using different MAC- and PHY-layer packet transmission options. Section 9 concludes the paper.

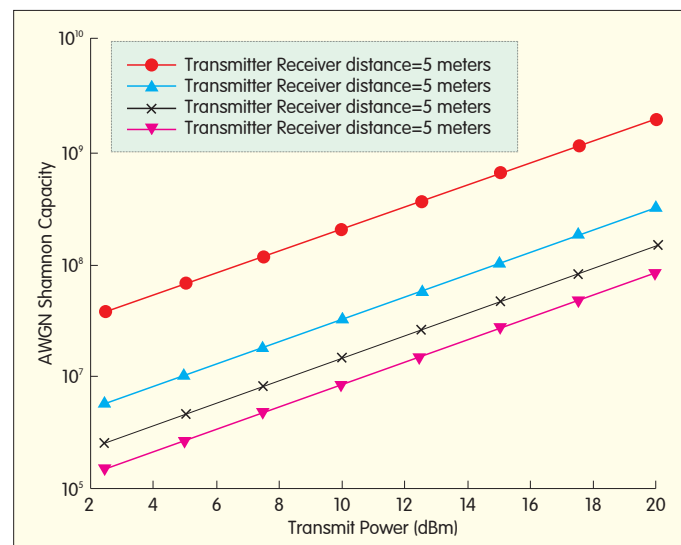
## 2 TDMA Protocol and the Need for Directionality

Interest in the 30–300 MHz millimeter wave spectrum has increased significantly because of low-cost, high-performance CMOS technology and because of low-loss, low-cost organic packaging. A millimeterwave radio can be empowered for the same cost as a radio operating in the 5 GHz band or lower. This advantage, combined with wide available bandwidth, makes the millimeterwave spectrum more attractive than ever before for supporting new systems and applications. A millimeter wave signal can propagate over a few kilometers at lower frequencies, penetrating through different construction materials and deriving advantages from reflection and refraction; however, they are highly directional and can be sustained only over short distances. The reason for this directionality is explained by the Friis free space equation:

$$P_R = P_T G_T G_R \lambda^2 / 4\pi R^2 \quad (1)$$

where  $P_R$ ,  $P_T$ ,  $G_T$ ,  $G_R$ ,  $\lambda$  and  $R$  is the receive power, transmit power, transmit antenna gain, receive antenna gain, wavelength, and distance between the transmitter and receiver, respectively. There is a 22 dB loss when we move from 5 GHz to 60 GHz. This loss is due to lower wavelength and can be offset by using directional antennas with higher gains. If 2 GHz bandwidth was used in 60 GHz and  $P_T = 10$  dBm, the noise figure  $Nf = 10$  dB, and the shadow fading margin  $\sigma = 6$  dB, 1 Gbit/s throughput could not be achieved (Fig. 2) [1]. Therefore, the gains of directional antennas must be exploited to achieve higher rates.

Directional communication requires complex discovery and beamforming protocols to establish links between different



▲ Figure 2. Shannon capacity versus transmit power at 60 Hz.

devices. Scheduled protocols such as TDMA are needed at the MAC layer to guarantee QoS at multi-gigabit-per-second rates. Randomized access protocols such as CSMA come with a random overhead that can depend on the number of users contending for the channel. Although CSMA/CA is still used to handle bursty traffic, allocation of contention-based access periods (CBAPs) is based on TDMA.

### 3 Physical Layer

802.11ad defines four different PHY layers: Control PHY, SC PHY, OFDM PHY and low-power SC PHY (LPSC PHY). Control PHY is MCS 0. SC starts at MCS 1 and ends at MCS 12; OFDM PHY starts at MCS 13 and ends at MCS 24; and LPSC starts at MCS 25 and ends at MCS 31. MCS 0 to MCS 4 are mandatory PHY MCSs. Here, we briefly describe the different PHYs and their packet structures. The system clock rate is 2640 MHz, and this rate is used for OFDM also. Control, SC and LPSC PHYs have a clock rate of  $2/3 \times 2640 = 1760$  MHz.

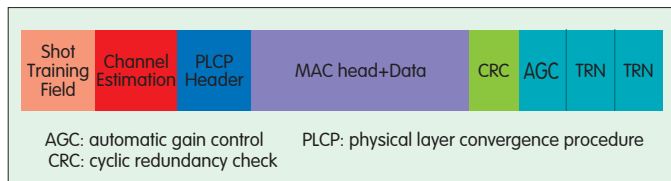
#### 3.1 General Packet Structure

As is common with all 802.11 packet formats, the packet consists of a short training sequence, a channel estimation sequence, the physical layer convergence procedure (PLCP) header, MAC packet, and cyclic redundancy check (CRC). Although there are different PHYs, they all have this unique structure, which ensures implementers do not need to change packet formats when using different PHYs. The only difference is that each PHY is a different size and uses a different Golay code.

The short training field (STF) and channel estimation field (CEF) help signal acquisition, automatic gain control training, predicting the characteristics of the channel for the decoder, frequency offset estimation and synchronization. Both STF and CEF sequences use Golay codes. The PLCP header indicates the size of the packet as well as the modulation structure (MCS) of the packet. The MAC packet comprises MAC header and data and contains information about the destination. CRC ensures that the packet is not corrupted while being transmitted through the air. 802.11ad has a TRN field comprising Golay codes. This field is used in beam tracking and refinement and is described in section 5. Fig. 3 shows a typical packet in 802.11ad.

#### 3.2 Control PHY

The control PHY, MCS 0, is the minimum rate that all devices use to communicate with before establishing a



▲ Figure 3. IEEE 802.11ad packet structure.

high-rate beamformed link. To help discovery and detection, the control PHY has an STF comprising 50 Golay sequences, each of which is 128 samples long. The CEF that follows the STF has nine Golay sequences. The STF comprises  $48 \times Gb128(n)$  (each 128 samples long) followed by a single repetition of  $-Gb128(n)$ . The CEF comprises  $Gu512(n)$  and  $Gv512(n)$  followed by  $Gv128(n)$ . These sequences are represented as function of  $Ga128(n)$  and  $Gb128(n)$  in section 21.11 of the IEEE 802.11ad draft specification [10].  $Gu512(n)$  and  $Gv512(n)$  are given by:

$$Gu512(n) = [-Gb128 - Ga128Gb128Ga128] \quad (2)$$

$$Gv512(n) = [-Gb128Ga128 - Gb128 - Ga128] \quad (3)$$

The control PHY uses BPSK with a code rate of 1/2 and is spread using a 32 code to create a PHY rate of 27.5 Mbit/s. The Control PHY is used for transmitting and receiving frames such as beacons, information request and response, probe request and response, sector sweep, sector sweep feedback, and other management and control frames. It provides reliability and exploits gain of the transmit antenna. Additionally, the first frame transmitted during the beam refinement protocol (BRP) phase is also a control PHY frame.

#### 3.3 Single-Carrier PHY, OFDM PHY and LPSC PHY

MCSs 1 to 4 are mandatory and ensure that all devices, irrespective of their PHY, are interoperable. All the MCSs, with the exception of LPSC PHY, use LDPC code, and the LPSC uses Reed Solomon (RS) codes. The following two subsections describe the MAC header and data packet encoding process for SC PHY. Other PHYs use a similar encoding process. All packets are modulated using BPSK, QPSK, 16-QAM and 64-QAM (Table 1).

#### 3.4 Header Encoding

The header is encoded using a single SC block of  $N_{CBPB}$  symbols with  $N_{GI}$  guard symbols. The bits are scrambled and encoded in the following steps:

- 1) The input header bits  $(b_1, b_2, \dots, b_{LH})$   $LH = 64$  are scrambled, starting from the eighth bit, in order to create  $d_{1s} = (q_1, q_2, \dots, q_{LH})$ .
- 2) The LDPC code word  $c = (q_1, q_2, \dots, q_{LH}, 01, 02, \dots, 0504 - LH, p_1, p_2, \dots, p_{168})$  is created by concatenating  $504 - LH$  zeros to the  $LH$  bits of  $d_{1s}$  and then generating the parity bits  $p_1, p_2, \dots, p_{168}$  so that  $Hc^T = 0$ , where  $H$  is the parity-check matrix for the 3/4 LDPC code specification in 802.11ad.
- 3) Bits  $LH + 1$  through 504 and bits 665 through 672 of the code word  $c$  are removed to create the sequence  $cs_1 = (q_1, q_2, \dots, q_{LH}, p_1, p_2, \dots, p_{168})$ .
- 4) Bits  $LH + 1$  through 504 and bits 657 through 664 of the code word  $c$  are removed to create the sequence  $cs_2 = (q_1, q_2, \dots, q_{LH}, p_1, p_2, \dots, p_{152}, p_{161}, p_{162}, \dots, p_{168})$  and then to create  $XOR$  with a PN sequence. The PN sequence is generated from the LFSR used for data scrambling, and the LFSR is initialized to the all-ones vector.
- 5)  $cs_1$  and  $cs_2$  are concatenated to form the sequence  $(cs_1, cs_2)$ . The resulting 448 bits are then mapped as  $\pi/2$ -BPSK, and the NGI guard symbols are prepended to

▼ Table 1. PHY modulation and coding scheme table

MCS Index	Modulation	NCBPS	Repetitions	Code Rate	NBPCS	NDBPS/Coding	DataRate (Mbit/s)
0	$\pi/2$ BPSK	1	32	1/2	1	168	27.5
1	$\pi/2$ BPSK	1	2	1/2	1	168	385.0
2	$\pi/2$ BPSK	1	1	1/2	1	168	770.0
3	$\pi/2$ BPSK	1	1	5/8	1	168	962.5
4	$\pi/2$ BPSK	1	1	3/4	1	168	1155.0
5	$\pi/2$ BPSK	1	1	13/16	1	168	1251.25
6	$\pi/2$ QPSK	2	1	1/2	1	168	1540.0
7	$\pi/2$ QPSK	2	1	5/8	1	168	1925.0
8	$\pi/2$ QPSK	2	1	3/4	1	168	2310.0
9	$\pi/2$ QPSK	2	1	13/16	1	168	2502.5
10	$\pi/2$ 16 QAM	4	1	1/2	1	168	3080.0
11	$\pi/2$ 16 QAM	4	1	5/8	1	168	3850.0
12	$\pi/2$ 16 QAM	4	1	3/4	1	168	4620.0
13	SQPSK	336	1	1/2	1	168	693.0
14	SQPSK	336	1	5/8	1	210	866.25
15	QPSK	672	1	1/2	2	336	1386.0
16	QPSK	672	1	5/8	2	420	1732.5
17	QPSK	672	1	3/4	2	504	2079.0
18	16-QAM	1344	1	1/2	4	672	2772.0
19	16-QAM	1344	1	5/8	4	840	3465.0
20	16-QAM	1344	1	3/4	4	1008	4158.0
21	16-QAM	1344	1	13/16	4	1092	4504.0
22	64-QAM	2016	1	5/8	6	1260	5179.0
23	64-QAM	2016	1	3/4	6	1512	6237.0
24	64-QAM	2016	1	13/16	6	1638	6756.75
25	$\pi/2$ BPSK	392	1	13/16	6	RS(224,208)+BS(16,8)	626.0
26	$\pi/2$ BPSK	392	1	13/16	6	RS(224,208)+BS(12,8)	834.0
27	$\pi/2$ BPSK	392	1	13/16	6	RS(224,208)+SPC(9,8)	1112.0
28	$\pi/2$ QPSK	392	1	13/16	6	RS(224,208)+BS(16,8)	1251.0
29	$\pi/2$ QPSK	392	1	13/16	6	RS(224,208)+BS(12,8)	1668.0
30	$\pi/2$ QPSK	392	1	13/16	6	RS(224,208)+SPC(9,8)	2224.0
31	$\pi/2$ QPSK	392	1	13/16	6	RS(224,208)+BC(8,8)	2503.0

the resulting NCBPB symbols.

### 3.5 Data Encoding

The data packet is encoded using LDPC, which includes deciding the number of shortening/repetition bits in every code word, shortening, coding each word, and repetition of bits. Data packet encoding occurs in the following steps:

- 1) The number of LDPC code words is given by  $N_{CW} = (\text{length} \times 8p) / (L_{CW} \times R)$ . This is used to calculate the number of datapad bits given by  $NDATA_{pad} = (N_{CW} \times L_{CW} \times R) / p - (\text{length} \times 8)$ , where  $L_{CW} = 672$  is the LDPC code word length; length is the length of the PSDU defined in the header field (in octets);  $p$  is the repetition factor (1 or 2); and  $R$  is the code rate. The scrambled PSDU is

concatenated with  $NDATA_P$  AD zeros, which are scrambled using the continuation of the sequence that scrambled the PSDU input bits.

- 2) The output stream of the scrambler is broken into blocks of  $L_{CWD} = L_{CW} \times R$  bits so that the  $m$ th data word is  $b_1^m, b_2^m, \dots, b_{L_{CWD}}^m$   $m < N_{CW}$ .
- 3) To each data word,  $n-k = L_{CW} - (R \times L_{CW})$  parity bits  $p_1^m, p_2^m, \dots, p_{n-k}^m$  are added to create the code word  $c^m = b_1^m, b_2^m, \dots, b_{L_{CWD}}^m, p_1^m, p_2^m, \dots, p_{n-k}^m$  so that  $H_c^{(m)} = 0$ .
- 4) The code words are concatenated one after the other to create the coded bitstream  $c_1, c_2, \dots, c_{L_{CWD} \times N_{CW}}$ . The number of symbol blocks is given by  $N_{BLK} = (N_{CW} \times L_{CW}) / N_{CBPB}$ , and the number of symbol block padding bits is given by  $N_{BLKPAD} = (N_{BLK} \times N_{CBPB}) - (N_{CW} \times L_{CW})$ , where  $N_{CBPB}$  is the number of coded bits per symbol block.
- 5) The coded bitstream is concatenated with  $N_{BLKPAD}$  zeros, which are scrambled using the continuation of the scrambler sequence that scrambled the PSDU input bits. Table 1 shows the MCS values allowed in 802.11ad.

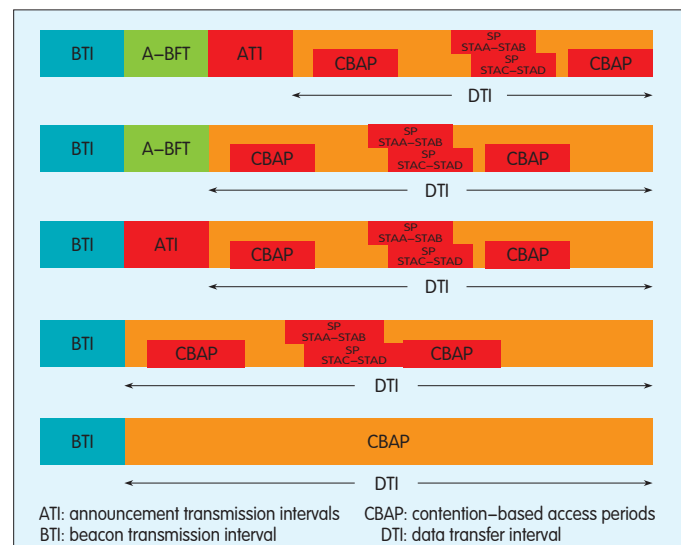
## 4 Overview of 802.11AD MAC Protocol

### 4.1 802.11ad Superframe

The 802.11ad superframe is called the beacon interval and comprises a beacon transmission interval (BTI), a data transfer interval (DTI), and optional association beamforming training (A-BFT) or announcement transmission intervals (ATI) (Figs. 4 and 10). The DTI can include one or more service periods (SPs) and CBAPs.

### 4.2 Service Period Channel Access

An SP is a scheduled access period between two stations: a transmitter and receiver. SPs are suitable for directional (high-gain) antenna use. New features introduced by 802.11ad include spatial sharing, where SPs need not be out of synch, that is, an SP between stations A and B may occur at



▲ Figure 4. Examples of 802.11ad beacon intervals.

the same time as another SP between stations C and D. IEEE 802.11ad also introduces dynamic SP allocation, truncation, and extension. An SP allocation is dynamic if it is not initially scheduled by the personal basic service set (PBSS) control point/access point (PCP/AP) but is scheduled during an existing SP or CBAP. SP truncation occurs when the transmitter station relinquishes the remaining time in its SP. SP extension occurs when the transmitter station extends the SP duration it had been allocated.

### 4.3 CBAP Channel Access

During a CBAP, all stations contend for channel access using a hybrid TDMA-CSMA/CA scheme based on 802.11 enhanced EDCA. 802.11ad provides physical carrier sensing mechanism, provided by the physical layer, and a virtual carrier-sensing mechanism, provided by the MAC layer. Physical carrier sensing uses clear channel assessment (CCA).

Virtual carrier sensing uses a timer called network allocation vector (NAV). NAV indicates, in microseconds, how long the channel is reserved by another station and counts down to 0. The virtual carrier-sensing mechanism uses request-to-send/directional multigigabit clear-to-send (RTS/DMG CTS) frames. When a station receives an RTS/DMG CTS frame, it sets its NAV to the value in the Duration field in the MAC header of the RTS/DMG CTS. Stations also use the Duration field of other frames to update their NAVs; however, the frame's destination address must be different from the receiving station's MAC address, and the value of the Duration field must be greater than the current NAV value. Stations may have a unique NAV or may have one NAV per sector. If a station has multiple NAVs and at least one has a non-zero value, the virtual carrier-sensing mechanism considers the medium busy. The medium is considered busy if either the physical or virtual carrier-sensing mechanism indicates it is busy; otherwise, it is considered idle. 802.11ad defines four different access categories (ACs) that have different priorities based on the user priority (UP) of the data being transferred. In order of increasing priority, these ACs are background (BK), best effort (BE), video (VI), and voice (VO). Only BE is mandatory in the standard, that is, only BE is implemented or all four are. When all four ACs are implemented, all ACs within a given station have to contend with each other and with other stations for channel access. Each AC contends for a channel in the following way: After the medium has been idle for a period of time, called the arbitration interframe space ( $AIFS[AC]$ ), the station contending for access randomly sets its backoff timer to a between 0 and the contention window ( $CW[AC]$ ).  $CW[AC]$  is initialized to  $CW_{min}[AC]$  and is updated after every transmission. In case of transmission failure,  $CW[AC]$  is updated using

$$CW[AC] = 2 \times CW[AC] + 1 \quad (4)$$

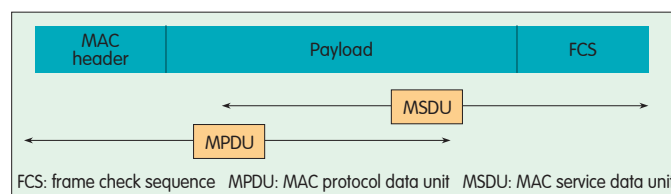
When  $CW[AC]$  reaches  $CW_{max}[AC]$ , it remains unchanged for any remaining retries. In the case of transmission success,  $CW[AC]$  is reset to  $CW_{min}[AC]$ . At every slot time boundary,

the medium is sensed. If the medium is found to be idle, the backoff timer is decremented; otherwise, it is suspended. When the backoff timer for a particular AC reaches 0, that AC obtains exclusive channel access for a period of time called the transmit opportunity ( $TXOP[AC]$ ). During the  $TXOP[AC]$ , only frames with UP mapping to that AC may be transmitted. If the backoff timers of two or more ACs reach zero at the same time, channel access is granted to the AC with the highest priority, and the other ACs treat this occurrence as if it were an external collision that happened in the wireless medium. The other ACs then enter backoff phase. For each AC, EDCA parameters such as  $CW_{min}[AC]$ ,  $CW_{max}[AC]$ ,  $AIFS[AC]$  and  $TXOP[AC]$  are calculated by the PCP/AP and included in the DMG beacon, probe response, or (re)association response frames transmitted by the PCP/AP. Higher-priority ACs are granted lower values for  $CW_{min}[AC]$ ,  $CW_{max}[AC]$ ,  $AIFS[AC]$  so that they can gain channel access while lower-priority ACs are still in backoff phase.

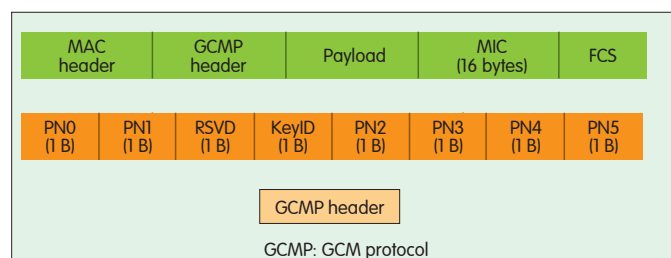
### 4.4 Packet Aggregation

802.11ad is capable of a multi-gigabyte-per-second data rate because of features such as packet aggregation and block acknowledgement in addition to directional antennas. The basic MAC data unit is called MAC protocol data unit (MPDU) and comprises a MAC header and a MAC service data unit (MSDU) or MAC payload. A PHY header and an MPDU comprise a PHY PDU (PPDU). 802.11ad uses the Galois/counter mode (GCM) protocol for data encryption. This protocol was designed for encryption at multi-gigabytes-per-second data rates.

An encrypted MPDU includes a GCM protocol header and a MIC field. Fig. 5 shows the MPDU structure. Fig. 6 shows the MPDU structure with encryption turned on. Packet aggregation involves combining several packets into a single packet. When several MSDUs or MPDUs are combined, the resulting packet is called aggregated MSDU (A-MSDU) (Fig. 7) or aggregated MPDU (A-MPDU) (Fig. 8). 802.11ad uses a new type of packet aggregation called aggregated

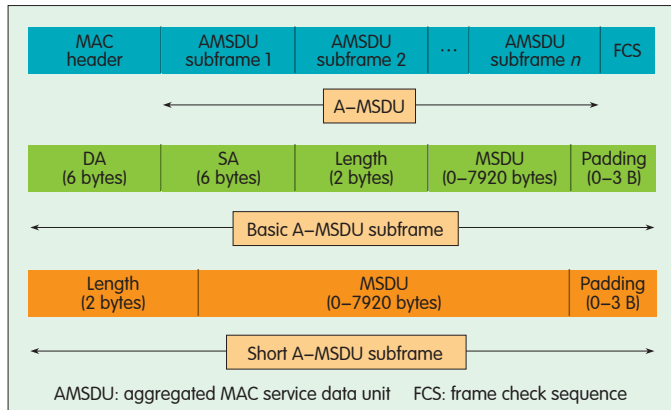


▲ Figure 5. 802.11ad MPDU structure.

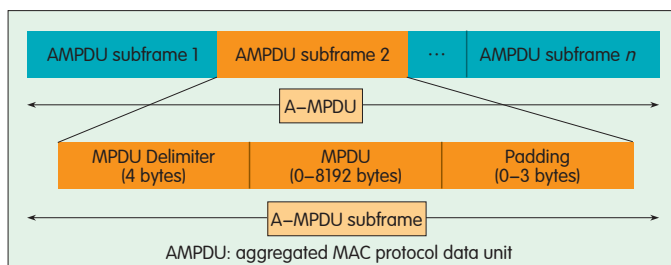


▲ Figure 6. 802.11ad MPDU structure with encryption turned on.





▲ Figure 7. 802.11ad A-MSDU structure.



▲ Figure 8. 802.11ad A-MPDU structure.

PPDU (A-PPDU). In an A-PPDU packet, several PPDU's are transmitted back-to-back without interframe spacing (IFS) and preamble in between. A-PPDU reduces overhead associated with IFS and MAC/PHY header processing.

### 4.5 Acknowledgement Policies

802.11ad defines a frame acknowledgement (ACK) policy called block acknowledgement. When block ACK is enabled, the transmitting station transmits a block of frames one frame at a time immediately after each other without waiting for the receiver to acknowledge the previous frame. After the entire block of frames has been transmitted, the receiving station sends a control frame called block ACK that includes a bitmap. The bitmap, in which each bit corresponds to a frame, indicates which frames were received successfully and which ones were not. The receiver knows to send a block ACK frame when it receives a block ACK request frame from the transmitter. This ACK policy allows the transmitter to use shorter IFS between frames, and it eliminates the IFS between each frame and its individual ACK frame, as in a typical stop-and-wait protocol. Fig. 9 shows normal ACK and block ACK policies.

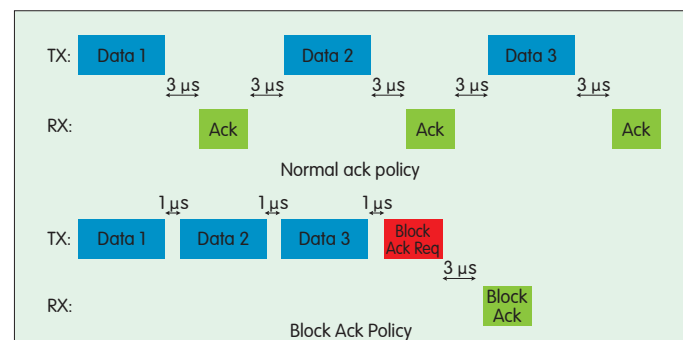
## 5 Beamforming Protocol

Because of the highly directional nature of 60 GHz communications, the transmitter and receiver antennas need to be aligned in the right direction to obtain maximum gain. 802.11ad supports up to four transmitter antennas, four receiver antennas, and 128 sectors. Beamforming is

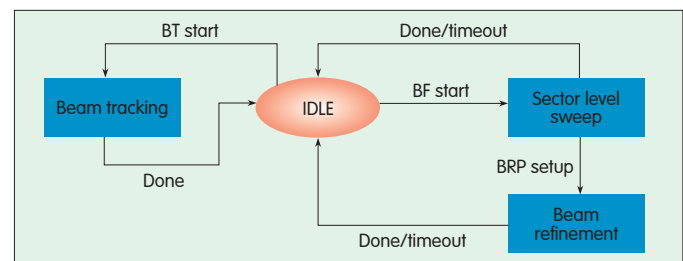
mandatory in 802.11ad, and both transmitter-side and receiver-side beamforming are supported.

Beamforming can be done at the transmitter side, receiver side, or at both sides [15]. Transmitter-side beamforming usually requires feedback from the receiver, especially when the transmitter-to-receiver and receiver-to-transmitter channels are not reciprocal. The need for feedback can be reduced or eliminated by using space time codes; however, this can cause considerable overhead in the setting-up beamforming [16]. 802.11ad uses a selection-based protocol in which the transmitter sends training from certain sectors that are pre-defined according to distinct antenna patterns created by changing the antenna weights [17]. The receiver antenna maintains an omnidirectional pattern and measures the strength of the received signal from the different sectors. It responds with information about the best sector and measured quality. With this feedback, the transmitter chooses the best sector to use while transmitting to the receiver. Similarly, in receiver-side training, the receiver repeats the training from the transmitter, which is sent using an omnidirectional antenna pattern, and measuring the strength of the received signal through pre-defined receive sectors.

The station that starts the beamforming training is called the initiator, and the recipient is called the responder. Beamforming in 802.11ad involves sector-level sweep (SLS), BRP, and beam tracking (BT). Fig. 10 shows the sequence of this beamforming. Each of the steps in the sequence is allowed in a particular part of the beacon interval (Fig. 11). SLS enables reliable communication at the lowest supported rate (called MCS0 in 802.11ad). Usually, transmitter-side training is done during the SLS. The BRP enables receiver training and iteratively trains the transmitter and receiver sides

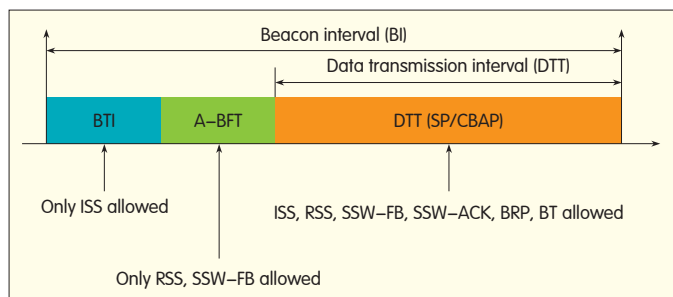


▲ Figure 9. Normal ACK policy and block ACK policy.



▲ Figure 10. The sequence in which the different phases of beamforming occur in 802.11ad/WiGig standard.





▲ Figure 11. Different transmission periods and BF phases allowed in each part of the beacon interval.

to improve on the values found during the SLS. Both SLS and BRP phases use their own special packets for beamforming training. By contrast, BT can be done during data transmission. It is used to track the beamforming state and improve it during data transmission. BT is implemented by adding training (TRN) fields to the back of a data packet.

The SLS involves initiator sector sweep (ISS), responder sector sweep (RSS), sector sweep feedback (SSW-FB), and sector sweep acknowledgement (SSW-ACK). The BRP comprises setup, multiple sector ID detection (MID), beam combining (BC), and BRP transactions. Of these, MID and BC are optional features for 802.11ad supporting stations. BT comprises BT request and BT response. The parameters for exchanging beamforming packets are obtained using the capability element in the beacon packets, probe request/response packets, or information request/response packets.

### 5.1 Beaconsing and Sector-Level Sweep

At the start of every BTI, the PCP/AP MAC schedules an initiator transmit sector sweep (TXSS) to transmit beacons through all sectors. The PCP/AP can also fragment the TXSS across multiple beacon intervals if the BTI is insufficient and cannot complete the TXSS. A station without a PCP/AP uses an omnidirectional receiving antenna configuration to scan non-associated beacons or receive associated beacons from the PCP/AP and determine the best sector/antenna ID using the sector sweep field at the end of TXSS. In a beacon, CDOWN is the number of pending beacon transmissions for completion of TXSS, with 0 being completion.

If multiple transmit antennas are supported, a PCP/AP station cannot switch its transmit antennas for beacon transmission within a BTI nor can it transmit a beacon more than once using the same antenna configuration. To minimize potential interference, the PCP/AP changes the order of sectors across beacon intervals if multiple directional beacon transmissions are required or waits for a random delay at the start of beacon interval if only a single beacon is to be transmitted (Fig. 12).

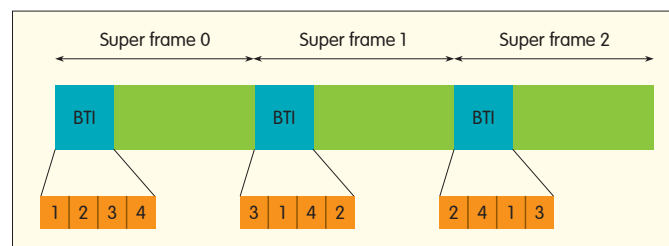
### 5.2 A-BFT Protocol

The PCP/AP announces the existence of an A-BFT period in the beacons, and this information is used for association and beamforming training for new stations, such as PBSS

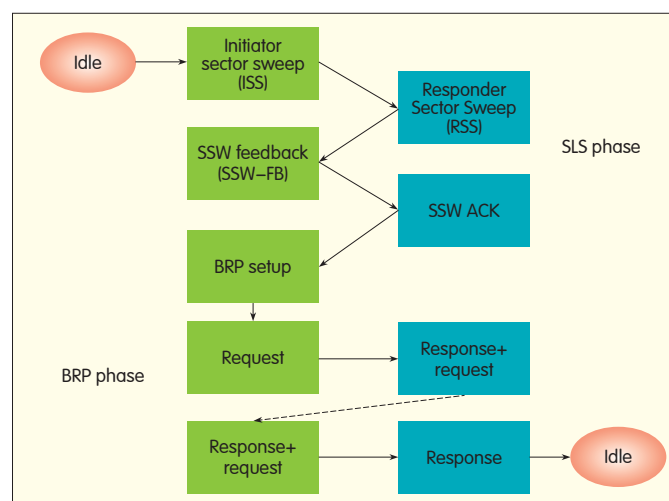
stations, that join the network. A-BFT may not be present in each beacon interval and may be periodically inserted by the AP/PCP. The A-BFT period allows stations to perform RSS and SSW-FB phases of beamforming with the PCP/AP. It is assumed that the new station has already used the beacons sent from all the transmit sectors of the PCP/AP to perform an ISS with the PCP/AP during the BTI. The A-BFT period is a slotted phase where each slot is a multiple of the time required for RSS and SSW-FB. The new stations use random backoff to select the A-BFT slot for an RSS. When beamforming is done during A-BFT, the SSW-ACK phase is skipped, and BRP is done during the data transmission interval if necessary. There may be no chance to do RSS during A-BFT because stations use random access; therefore, the AP/PCP may schedule an SP to continue beamforming with the particular station.

### 5.3 Sector-Level Sweep

The SLS is the basic type of beamforming supported by 802.11ad. It comprises ISS, RSS, SSW-FB and SSW-ACK. The link from the initiator to the responder is called the initiator link, and the link from the responder to the initiator is called the responder link. During SLS, the ISS phase is used to train the initiator link, and the RSS is used to train the responder link. The RSS contains feedback about the best sector found during ISS, and the SSW-FB contains the best sector found in



▲ Figure 12. A sample DMG beacon transmission by a PCP having one transmit antenna with four sectors.



▲ Figure 13. Parts of the sector level sweep and beam refinement phases of the beamforming in 802.11ad.

the RSS. In SLS is concluded with an SSW-ACK (Fig. 13). A station can have separate transmitter and receiver chains with their own antenna configurations. Hence, for each of the initiator and responder links, the transmitter and receiver can be trained independently to perform beamforming. The protocol used to train the transmitter during SLS is called TXSS, and the protocol used to train the receiver is called RXSS. This gives rise to four possibilities; if an ISS is used to train the transmitter side of the initiator link, the phase is called ISS TXSS. Similarly, the other three possibilities are ISS RXSS, RSS TXSS, and RSS RXSS.

During TXSS, the transmitter sends a separate SSW frame from different available transmit sectors, the number of which can be pre-negotiated between stations. The receiver maintains a quasi-omni receive configuration. If the receiver has multiple antennas, the transmitter repeats this process for each receive antenna. The receiver measures the quality of the packet received from each of the transmit sectors by cycling through all of its receive antennas in quasi-omni mode. In the end, the receiver replies with the best sector. The standard allows a vendor-specific algorithm to decide the best sector. Similarly, during RXSS, the transmitter uses an omniantenna configuration, and the receiver changes receive sectors to determine which the best receive sector.

The SLS phase can be initiated by the PCP/AP during the BTI by performing an ISS. Then, the RSS and FB phases are completed during the A-BFT announced by the PCP/AP. In this case, the BRP phase completed in an ATI or DTI. Alternatively, a station can either use a CBAP period (also announced in the beacon) or schedule an SP to perform beamforming with another station. The DTI can be used for all the phases of beamforming (Fig. 11).

The BRP phase comprises a setup phase followed by a beam-refinement phase based on request-response (Fig. 13). The request-response packets of the setup phase are exchanged until the responder (receiver) sets the capability-request field in the BRP packet at 0. This is followed by a response from the initiator (transmitter) with the capability-request field set at 0.

The beam-refinement request can be a transmitter- or receiver-refinement request. A transmitter-refinement request indicates the need for transmitter antenna training by the transmission station and vice versa. The transmitter station adds TRN-T subfields to the BRP frame. The receiver station holds data that it obtained by measuring the TRN-T fields. The receiver station responds to a receiver-refinement request (sent by the transmitter station) by appending TRN-R subfields to its response frames, that is, an ACK or block ACK frame.

The SLS and BRP phases of beamforming usually precede data transmission. They are completed right at the beginning of beamforming and are repeated periodically as needed. Beamtracking is used for beamforming training during data transmission to accommodate channel changes between two SLS/BRP beamforming training phases. In beamtracking, training fields comprising CE and STF fields are attached to the back of data packets or, for example, ACK/BA, to train the

transmitter or receiver (Fig. 3). 802.11ad allows for three types of beamtracking, and the type of beamtracking is signaled using three parameters in the PLCP header. The three parameters of interest are packet type, training length, and beamtracking request. Of these, the training length is always greater than zero. If the training length is zero, the other two fields are reserved, and the packet does not contain any beamtracking training or request. Table 2 shows the types of

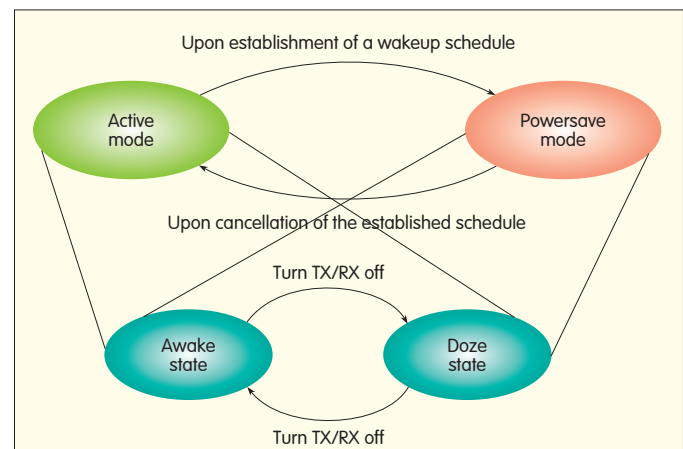
▼ Table 2. Types of beamtracking indicated by the PLCP parameters

Parameters	BT type	Explanation
BT request = 1 Packet Type = 1	Send TRN-T	The transmitter attaches TRN-T fields to the current packet and the receiver sends back a BRP frame with the feedback.
BT request = 0 Packet Type = 0	Send TRN-R	The transmitter attaches TRN-R fields to the current packet and the receiver finds its own best sector.
BT request = 1 Packet Type = 0	Request TRN-R	The transmitter is requesting receiver training. In the next packet, the receiver attaches TRN-R fields.

training indicated by the PLCP parameters.

## 6 Power-Save Protocol

Dedicated SPs in 802.11ad allow battery-powered stations to hibernate during data transmission periods that are not assigned to them. An 802.11ad station can be in one of the two power-save states: doze or awake. When awake, a station is fully powered; when in doze, the station is powered off. A station's power-save state in various sections in a beacon interval depends on whether the station is in active mode or power-save mode. In power-save mode, stations with or without PCP/AP can doze for one or more consecutive BIs, or sections of a BI, more if they were permanently in active mode. A station must check its peer's wakeup schedule before sending any individually addressed MPDUs to the peer station because it may be in doze mode. A station without PCP/AP can always use information request/response frames to request the wakeup schedule from any of its peers if required. Fig. 14 shows power management modes and state



▲ Figure 14. Power management-mode/state transitions.

transitions.

A station without PCP/AP that has not established a wakeup schedule with its peer is in active mode. To switch from active to power-save mode, the station establishes a wakeup schedule with PCP/AP. It does this by including a wakeup schedule (WS) element in its power-save configuration request. To switch from power-save mode to active mode, the station without PCP/AP sends a power-save configuration request in which the power management bit is set at 0. The station immediately switches to active mode upon receipt of the ACK frame from PCP/AP.

A PCP/AP station includes its WS in its beacon or announcement frames before switching to power-save mode. When switching back to active mode, it ceases including the WS in these frames. The PCP/AP station keeps track of the wakeup schedules of all associated stations without PCP/AP. In addition, APs also have to buffer MPDUs addressed to associated stations in doze state and forward these MPDUs at designated times.

## 7 Fast Session Transfer Protocol

Fast session transfer (FST) protocol allows different streams or sessions to transfer smoothly from one channel to another in the same band or different bands. This protocol makes 802.11ad compatible with the forthcoming 802.11ac standard and other existing standards, such as 802.11a/b/g/n. The protocol allows different radios in the same device to operate simultaneously or not simultaneously. Devices with 802.11ac and 802.11ad can have same MAC address or different MAC addresses. If the same MAC address is used for all the radios in different bands then FST is in transparent mode. If the MAC addresses differ according to channel/band, then FST is not transparent.

A simple example of FST is a video stream to be established between STA A and STA B in the 2.4 GHz band using direct-link setup (DLS). The STAs are 40 m apart. The video uses 802.11n radio at 144.4 Mbit/s and H.264 compression. After some time, the user of STA A moves very close to STA B so that the separation is less than 3 m. Both STA A and STA B understand that they have 60 GHz radio, which was discovered during in 60 GHz discovery mode.

They then transition to 60 GHz channel 2 and use an uncompressed stream by closing their link at MCS 12, which is 4.62 Gbit/s (Fig. 15).

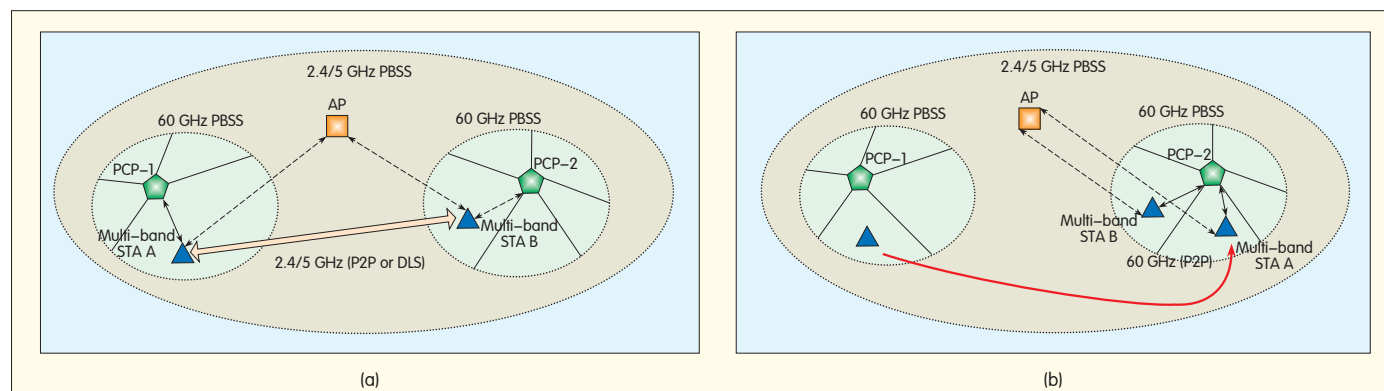
This video stream established in 60 GHz channel 2 can be moved to 60 GHz channel 1 if there is congestion in channel 2. It can also be moved to channel 4 in 5 GHz or channel 6 in 2.4 GHz when STA A starts to move away from STA B. In this example, video compression, such as H.264 and that used in the WiGig WDE specification, ensures that the session does not drop because of large range or insufficient bandwidth. The application and MAC layers interact with PHY to optimize the smooth delivery of content to the end application. They compress the video whenever the band is not 60 GHz by using IEEE 802.11ac or IEEE 802.11n and then transition to uncompressed video using 60 GHz SC or OFDM modes when the range is less than three meters. This ensures the highest QoS. The FST also ensures that a subset of streams can be transferred from one channel/band to another while the remaining streams are in the original channel/band.

## 8 Packet Throughput

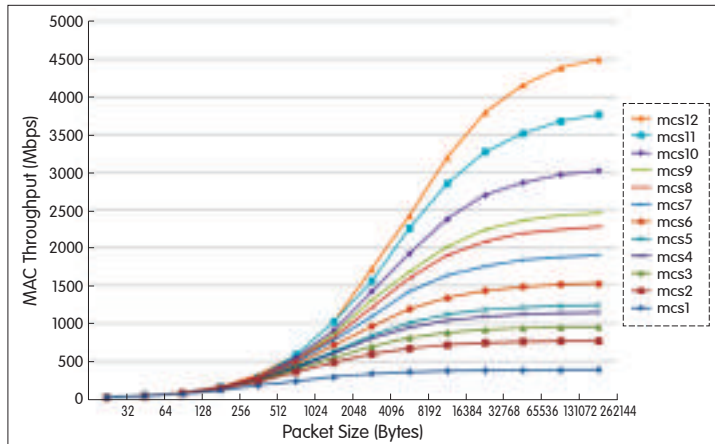
The modulation and coding schemes in section 3 along with the normal ACK, A-MSDU, and A-MPDU packet structures in section 4 allow the 60 GHz radios to change between very different achievable throughputs depending on the packet size. Fig. 16 shows the MAC layer throughput versus packet size sent at different MCS values. Similarly, Fig. 17 shows the throughput versus packet size when A-MPDU is used. In Figs. 16 and 17, BTI, ABFT and AT overheads are not taken into account. The parameters described in section 3 can also be found in section 21.3 of [9].

## 9 Conclusion

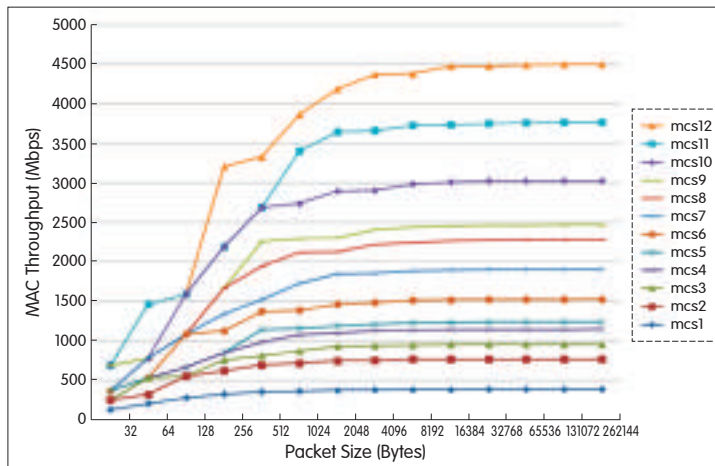
802.11ad are standardizing 60 GHz technology to facilitate multi-gigabit-per-second communications over shorter distances. This standard has many new features to improve and sustain high-speed communications with TDMA single-carrier and OFDM schemes. They allow for scheduled and contention-based access, beamforming, and power-save mechanisms that decrease power consumption



▲ Figure 15. Fast session transfer done by peer stations (STAs) A and B in an a) 2.4 GHz channel and b) a 60 GHz channel.



▲ Figure 16. Single carrier throughput as a function of the packet size for different MCS values for non-aggregated packets.



▲ Figure 17. Single carrier throughput as a function of the packet size for different MCS values for A-MPDU packets.

and increase throughput. Future evolution of 802.11ad towards full MIMO support and channel bonding can further increase its data rate. With the advent of new technologies to make these protocols practical, and with standardization by bodies such as WiGig and IEEE, truly wireless broadband will be achieved with 60 GHz, and all wires in PANs will be eliminated.

### References

- [1] C. Cordeiro and S. S. Nandagopalan, "Next generation multi-gbps wireless LANs and PANs," *Proceedings of the IEEE GLOBECOM*, 2010.
- [2] "HighRate 60 GHz PHY, MAC and HDMI PALs," *ECMA International*, December 2010. [Online]. Available: <http://www.ecma-international.org/publications/standards/Ecma-387.htm>
- [3] "China millimeter wave study group." [Online]. Available: [http://www.ieee802.org/11/Reports/cmmw\\_update.htm](http://www.ieee802.org/11/Reports/cmmw_update.htm)
- [4] R. C. Daniels and J. Robert W. Heath, "60 GHz wireless communications: emerging requirements and design recommendations," *IEEE Vehicular technology magazine*, pp. 41–50, September.
- [5] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1390–1436, August 2011.
- [6] C. J. Hansen, "WiGig: Multi-gigabit wireless communications in the 60 GHz

band," *IEEE Wireless Communications*, pp. 6–7, December 2011.

- [7] "WirelessHD Specification Overview," 9, October 2007. [Online]. Available: <http://www.wirelessHD.org>
- [8] "MAC and PHY specification for high rate wireless PANs," IEEE Std 802.15.3c–2009, pp. c1–187, Oct 2009.
- [9] "WGA-D1.0," *Wireless Gigabit Alliance draft specification*, July 2010.
- [10] "Wireless lan MAC and PHY specifications –enhancements for very high throughput in the 60 GHz band," *IEEE Std 802.11.ad/D9.0 draft specification*, pp. 1–679, July 2012.
- [11] "Marketing requirements document," WiGig, Version 1.0 2011. [Online]. Available: <http://wigig.org/>
- [12] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE 802.11 Specification*, 2012.
- [13] C. Cordeiro, D. Akhmetov, and M. Park, "IEEE 802.11ad: introduction and performance evaluation of the first multi-gbps wifi technology," in *Proceedings of the 2010 ACM international workshop on mmWave communications: from circuits to networks, ser. mmCom '10*. New York, NY, USA: ACM, 2010, pp. 3–8.
- [14] W. Zhou, S. S. Nandagopalan, and D. Qiao, "A simulation study of CSMA/CA performance in 60 GHz WPANs," *Proceedings of the IEEE GLOBECOM*, pp. 1–6, 2009.
- [15] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE journal on selected areas in communications*, vol. 16, no. 8, pp. 1451–1458, 1998.
- [16] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space–time block coding for wireless communications: performance results," *IEEE journal on selected areas in communications*, vol. 17, no. 3, pp. 451–460, 1999.
- [17] J. Wang, I. Lakkis, and et. al., "Beam codebook based beamforming protocol for multi-gbps millimeter-wave wpan systems," in *Global Telecommunications Conference*, 2009. GLOBECOM 2009. IEEE, 30 2009–dec. 4 2009, pp. 1–6.

Manuscript received: September 10, 2012

### Biographies

**Sai Shankar N** (nsai@tensorcom.com) received his PhD from the Indian Institute of Science, Bangalore, in 1998. He received a DAAD fellowship and joined the Department of Mathematics at the University of Kaiserslautern, Germany. There, he worked on queueing approaches in manufacturing. In 1999, he joined Philips Research, Eindhoven, and worked on IEEE 802.15 HFC networks and differentiated services. In 2001, he was transferred to Philips Research, New York, and worked on IEEE 802.11e, IEEE 802.11n, MBOA UWB and IEEE 802.22. EE Times nominated him as one of five finalists for his contributions to UWB MAC. In 2005, he worked at Qualcomm in San Diego on IEEE 802.11s, and RLC and MAC–hs issues in HSPA. In 2007, he worked on 802.11 AMP in Bluetooth SIG and 60 GHz at Broadcom Corporation. Currently, he is leading 60 GHz FW and MAC HW solutions at Tensorcom.

**Debashis Dash** received his B.Tech. degree from the Indian Institute of Technology, Kanpur, in 2004. He received his MS degree from Rice University, Houston, in 2007. He is a PhD candidate at the Department of Electrical and Computer Engineering, Rice University. He currently works at Tensorcom, San Diego. His research interests include information theory and graph theory and their applications in wireless systems.

**Hassan El Madi** received his BS degree in computer engineering from the University of California, San Diego, in 2007. He received his M.Eng. degree in electrical engineering with an emphasis on wireless communications from Virginia Tech, Blacksburg, in 2010. He currently works as a software staff engineer at Tensorcom, San Diego. His research interests include cognitive radios—spectrum sensing, automatic modulation classification, geolocalization—as well as design and implementation of WLAN and WPAN MAC and PHY layers.

**Guru Gopalakrishnan** received his BE degree in electronics and communication from Anna University, India, in 2006. He received his MS degree in electrical engineering (computer networks) from the University of Southern California, Los Angeles, in 2009. He earlier worked at Broadcom Corporation, San Diego in Bluetooth and Bluetooth low energy technologies and is currently at Adeptence, San Diego. His research interests include throughput and power optimizations for wireless systems.



# Modeling Human Blockers in Millimeter Wave Radio Links

*Jonathan S. Lu, Daniel Steinbach, Patrick Cabrol, and Philip Pietraski*

(InterDigital Communications, LLC, Melville, NY, 11747, USA)

## Abstract

The loss from multiple human blockers is empirically and analytically investigated at millimeter wave frequencies. Humans are modeled as absorbing screens of infinite height with two knife-edges, while physical optics is used to compute the contribution from rays diffracting around them. This model is validated with blocking gain measurements of multiple human blocking configurations on an indoor link. The blocking gains predicted from physical optics have good agreement with measurements ranging from  $-50$  dB to  $2.7$  dB, making the absorbing screen model suitable for real human blockers. Mean and standard deviation of prediction error are approximately  $-1.2$  and  $5$  dB, respectively.

## Keywords

60 GHz; diffraction; human blocking loss; human shadowing; indoor environment; millimeter wave propagation; physical optics

## 1 Introduction

In millimeter wave systems with access links to mobile users, humans are likely blockers of the radio links. This is particularly true in shopping malls, at store fronts, and in airports. Past works describe how humans can cause severe fades, that is, losses greater than  $20$  dB [1], [2]. Therefore, it is critical to include the effect of human blockers in simulations. Here, we focus on modeling and computing human blocking (shadowing) for a millimeter wave radio link.

Transmission loss caused by human blockers is very high at millimeter wave frequencies, and transmission is virtually opaque. Therefore, diffraction around human blockers and reflection and scattering by nearby objects or structures significantly affects the received power. Millimeter wave systems that include link distances greater than a few meters typically use highly directional antennas or arrays in order to overcome path loss incurred at high frequencies. Thus, the reflection and scattering by surrounding objects away from the direct line between the transmitter and receiver is greatly attenuated by the antenna patterns or array beamforming patterns. We investigate and model the diffraction around human blockers.

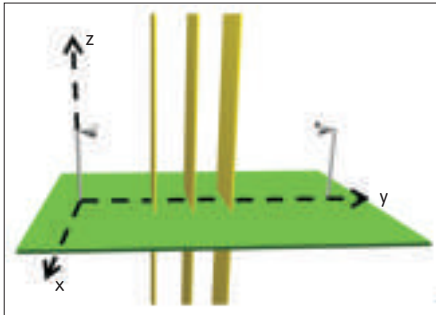
The exact electromagnetic characteristics of human bodies are not described in the literature. In previous works, researchers have attempted to model the electromagnetic properties of humans at millimeter wave frequencies. Human bodies have previously been modeled as absorbing screens

[3], [4]; water phantoms [5]; cylinders [6], [7]; and rectangular prisms [8]. These models were validated only for a single human blocker; however, to the best of our knowledge, they have not been validated for multiple human blockers. Furthermore, in previous simulations, the affect of multiple human blocking on the radio environment has been determined through geometric optics, for example, ray-tracing. However, in geometric optics, geometric theory of diffraction (GTD) and uniform theory of diffraction (UTD) do not have higher-order diffraction terms and are not valid for multiple blockers located in each other's transition regions [9]. Therefore, other approaches, such as physical optics, are recommended. Both geometric and physical optics are approximate techniques. Full-wave solutions using numerical techniques such as finite-difference time domain, finite element, and method of moments are computationally infeasible because of the small millimeter wavelength.

In some millimeter wave scenarios, the transmitting and receiving antennas are positioned low relative to the heights of the human blockers so that the predominant diffractions travel around the human blockers rather than over them. This scenario is the focus of this work. We propose using physical optics to compute the received field around the human blockers. In physical optics, points in space where there are wave fields may be considered elementary sources of radiation whose amplitudes are proportional to the amplitudes of the fields at those points. Here, we model human blockers as absorbing screens of infinite height with two knife edges similar to those in [10] (Fig. 1). This model is computationally



favorable, and if sufficiently accurate, is preferable to more complex models for rapidly simulating the radio channel. In physical optics, the received field can be expressed in the form of multiple integrals, which we numerically evaluate using Piazzzi's numerical integration method [11]–[13]. The



◀ Figure 1. Multiple blockers modeled as absorbing screens of infinite height.

predicted blocking gain from the diffracted fields is then computed using the coherent sum of fields and incoherent sum of powers. The predictions are compared with the measurements of various one-, two- and three-person blocking configurations. Here, the blocking gain is the ratio of received power with blockers to received power without blockers. Assuming there are negligible reflections from the ground and nearby objects, the blocking gain is equivalent to the ratio of received power to free-space power.

## 2 Piazzzi Physical Optics Method

We model the human blockers as absorbing screens of infinite height and with two vertical knife edges (Fig. 1). To compute diffraction gain, (the ratio of diffracted power to free-space power) from an arbitrary number of screens, we use a physical optics method developed by Piazzzi [11] to treat diffraction past multiple absorbing screens with knife-edges. The code used to evaluate multiple knife-edge diffractions is based on the same principles of physical optics used in [14]–[16]. We assume that a) the knife-edges are of infinite length and are parallel, and b) the additional diffraction gain for a point source on a plane that is perpendicular to the screens is the same as that for a line source that is parallel to the screens and intersects the plane at the source point.

With these assumptions, the physical optics description of diffraction around an absorbing screen is expressed as multiple integrations in the  $x$ - $z$  planes containing the absorbing screens. The integrations in the coordinate along a  $z$ -plane knife-edge can be approximated analytically so that we are left with integration in the  $x$ -coordinate away from the knife-edges. This is seen in the following expression for the magnetic field  $H(x_{n+1}, y_{n+1})$  in the plane containing the  $n+1$  absorbing screen [12]:

$$H(x_{n+1}, y_{n+1}) = e^{j\pi/4} \sqrt{\frac{k}{2\pi}} \int_{-\infty}^{\infty} H(x_n, y_n) \frac{e^{-jk\rho}}{\sqrt{\rho}} dx_n \quad (1)$$

where  $\rho$  is the distance from the secondary source point  $(x_n, y_n)$  on plane  $x = x_n$  to receiver point  $(x_{n+1}, y_{n+1})$  on plane  $x = x_{n+1}$ , and  $k$  is the free-space wave number. The field on

plane  $x = x_n$  containing the  $n$ th screen is given by  $H(x_n, y_n)$ . To arrive at an expression with multiple integrals, we substitute  $H(x_n, y_n)$  in (1) with  $H(x_{n-1}, y_{n-1})$ , which is the field on plane  $x = x_{n-1}$  containing the  $n-1$  screen and can similarly be written in integral form.

To predict diffraction gain from multiple screens, the integrals must be carried out numerically. The integral in (1) must be terminated with finite upper and lower limits (for the right and left sides of the screen), and the integration must be replaced by a discrete summation. An abrupt termination of the integral is equivalent to placing an absorbing screen outside the termination point and would artificially generate diffracted waves that are not part of the actual problem. In [15] and [16], quadratic approximations are used for the amplitude and phase of the integrands over intervals of less than one wavelength in order to discretize the integrals. Spurious diffraction was removed by using asymptotic approximations to analytically evaluate the integral outside the termination points of the numerical analysis. This allowed for larger intervals, but an additional cost was necessary to evaluate the complex error functions. In contrast, the Piazzzi method involves simple linear approximations of the amplitude and phase and introduces a smoothing procedure that uses a Kaiser-Bessel function to terminate the integration without introducing spurious diffraction [11]–[13].

We compare the blocking gain found by using the Piazzzi method with the time-averaged gain measurement for a given blocker configuration. To compute the complex field for each diffracted path, the previously mentioned integrals are separated. In Fig. 1, there are three absorbing screens with two edges each, and there are  $2^3 = 8$  forward diffracted paths. The gain for the diffracted path that travels along the  $x+$  sides of the screens is found by assuming that all screens are semi-infinite and have knife-edges located at  $x+$  edges. This is equivalent to setting the lower  $x$  limit of each integral to the position of the  $x+$  edges.

Human blockers in the test setup inadvertently move slightly between measurements, and these movements are reflected in the measurements. The movements are slight relative to the distances of the blockers from the transmitting and receiving antennas. Therefore, the amplitude of the field of each diffracted path can be assumed to be constant. However, the exact phase of each path cannot be known because a) these inadvertent movements cause path length differences on the order of or greater than the 5 mm wavelength at 60 GHz, and b) the exact electromagnetic interactions with the human body cannot be known. Depending on the configuration of blockers, a particular measurement may be inaccurate because of the movements. If we assume these movements are sufficient to obtain a well-mixed sample of possible phases, the time-averaged gain measurements can be approximated by assuming the phases of the different diffracted contributions are uncorrelated, uniform, random variables. Thus, the time-averaged diffraction gain can be modeled as the sum of the incoherent powers of gains from the different paths. Conversely, if we assume the phases are not random, the diffraction gain found by coherently summing

the complex magnetic fields is equivalent to not separating the previously mentioned integrals. Small, inadvertent movements may not be sufficient to produce well-mixed averages, and we could expect some hybrid model to provide more accurate predictions.

### 3 Measurements

#### 3.1 Setup

Fig. 2 shows our 60 GHz measurement setup. On the transmit side, the R&S® SMF100A microwave signal generator provides a 10 GHz sine wave to the R&S® SMZ90 frequency multiplier, which multiplies the frequency by six. The resulting 60 GHz signal then travels through a straight section waveguide to a V-band horn antenna with 24 dBi of gain and 7 degree 3 dB beamwidth. The level of the radiated signal can be adjusted via a 25 dB mechanically controlled attenuator included in the multiplier assembly.

The signal is received with an identical horn antenna that is connected to the N12-3387 low noise amplifier (LNA) with a straight waveguide section. The amplified signal is sent to the FS-Z90 harmonic mixer where it is down-converted and captured on the FSQ26 vector signal analyzer (VSA).

#### 3.2 Measurement Procedure

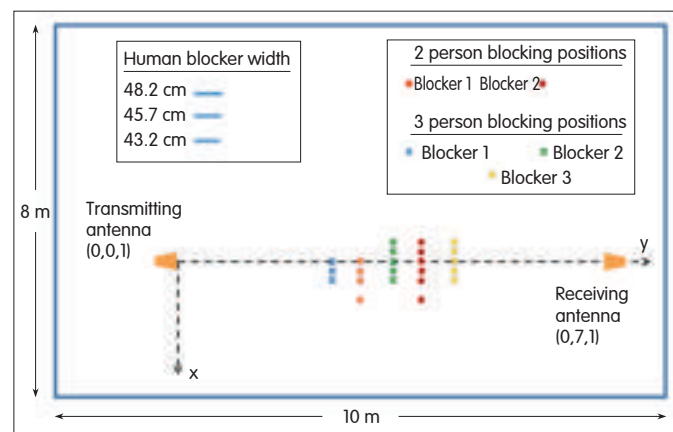
The majority of human blocking cases involved three blockers at most [3]. Therefore, we took measurements at 60 GHz for one, two and three blockers. Using the setup previously mentioned and the Cartesian coordinate system in Figs. 1 and 3, the transmitting and receiving antennas were placed 7 m apart and at a height of 1 m. The coordinates of the transmitting antenna were (0, 0, 1) and those of the receiving antenna were (0, 7, 1). The environment was an empty 8 × 10 m conference room. The reflections from the walls were heavily attenuated by the antenna patterns, longer path length, and reflection loss. The receiving antenna primarily measured the gain for propagation paths going through and around the blockers.

To measure gain in a one-person blocking scenario, a 43.2 cm wide blocker was placed halfway between the antennas. The blocker moved perpendicular to the direct line between the antennas (y-axis, Figs. 1 and 3), and measurements were taken at intervals of either 5 cm or 10 cm depending on how close the blocker was to the direct line. In the two-person scenario, 45.7 cm wide blockers were

spaced 1 m apart halfway between the antennas. The first blocker had four positions, and the second blocker had six positions. In the three-person scenario, blockers were also spaced 1 m apart halfway between the antennas. The first blocker was 48.2 cm wide and had three positions; the second and third blockers were 45.7 cm wide and had five positions. The positions of the blockers in the two- and three-person scenarios were chosen so that the line of sight (LoS) was almost always blocked, and practical blocking scenarios were considered (Fig. 3). These positions are listed in Table 1. The received powers for all possible blocking configurations in the two- and three-person scenarios were recorded. To maintain consistency, the blockers were centered over their positions and stood with their arms at their sides, facing the receiving antenna.

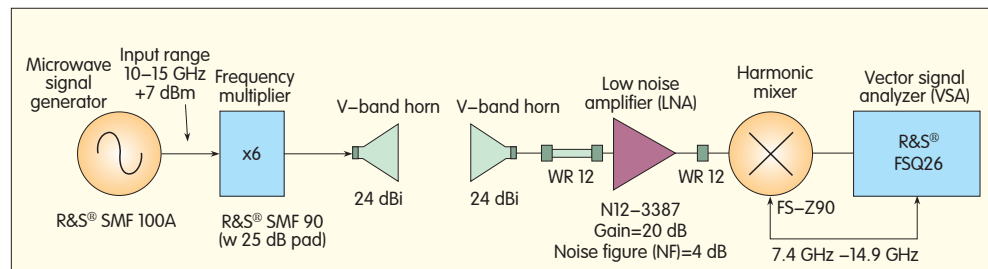
Five measurements were recorded for each blocking configuration in all scenarios. The blocking gain was then computed using the ratio of measured power with blockers to measured power without blockers. To determine the time-averaged blocking gain, the blocking gains were averaged over the five measurements.

Note that the physical optics predictions are compared with measurements where blockers have a constant separation of



▲ Figure 3. Two- and three-person blocking positions (to scale) in an 8 × 10 m room.

1 m. These comparisons are qualitatively similar to measurements where blockers have non-constant separation, because changing the blocker separation only change the diffraction angles. Our measurements consider a wide range of diffraction angles as shown in Fig. 3.



▲ Figure 2. 60 GHz Human Blocking Measurement Setup

### 4 Results and Analyses

In this section, we discuss variability in the measurements taken with various blocker configurations and compare these measurements with the diffraction gains predicted using the Piazzi method and incoherent-power-sum approach

▼ Table 1. Blocker positions using the Cartesian coordinate system of Fig. 1

Scenario	Positions (x,y) (m)					
	Blocker 1 (x)	Blocker 1 (y)	Blocker 2 (x)	Blocker 2 (y)	Blocker 3 (x)	Blocker 3 (y)
Two People	0.00	3.0	-0.31	4.0		
	0.15	3.0	-0.15	4.0		
	0.31	3.0	0.0	4.0	—	—
	0.31	3.0	0.15	4.0		
	0.61	3.0	0.31	4.0		
Three People			0.61	4.0		
	0.00	2.5	-0.31	3.5	-0.31	4.5
	0.15	2.5	-0.15	3.5	-0.15	4.5
	0.31	2.5	0.00	3.5	0.00	4.5
			0.15	3.5	0.15	4.5
			0.31	3.5	0.31	4.5

in section 2.

### 4.1 Measurement Variability

Because the blockers inadvertently moved, variability was introduced into the received signal for a given blocker configuration. The time-averaged blocking gain captured and averaged these movements. Because the ranges of blocker movements are unknown, we compare the time-averaged measurements with the blocking gains predicted using both the Piazzi method and incoherent-power-sum approach. When the blockers moved only slightly, the measurements are expected to more closely match those of the Piazzi method. When there was a greater range of movement by the blockers, the measurements are expected to more closely match the blocking gain predicted using the incoherent-power-sum approach. However, because of the limited number of measurements taken for each configuration, deviations are expected.

### 4.2 One-Person Blocking

Fig. 4 shows the measurements taken for a one-person blocking scenario. The blocking gains predicted using the Piazzi method and incoherent-power-sum approach are also shown. From Fig. 4, the sum of incoherent powers agrees fairly well with the measurements and has a maximum deviation of 4.7 dB. The measurements also agree with the predictions of the Piazzi method for different blocker positions. When the blocker's center location x coordinate is less than 0 m, the measurements fall on the Piazzi curve. When the blocker is located at (0, 3.5) in Figs. 1 and 3,  $x = 0$ . When the blocker's center location x coordinate is greater than 0 m, the measurements are slightly offset from the Piazzi curve. The reason for this may be that the blocker was not properly centered, and the blocker's effective width was actually less than 43.2 cm. These results suggest that the absorbing screens also model the phase information of a human blocker at millimeter frequencies relatively well.

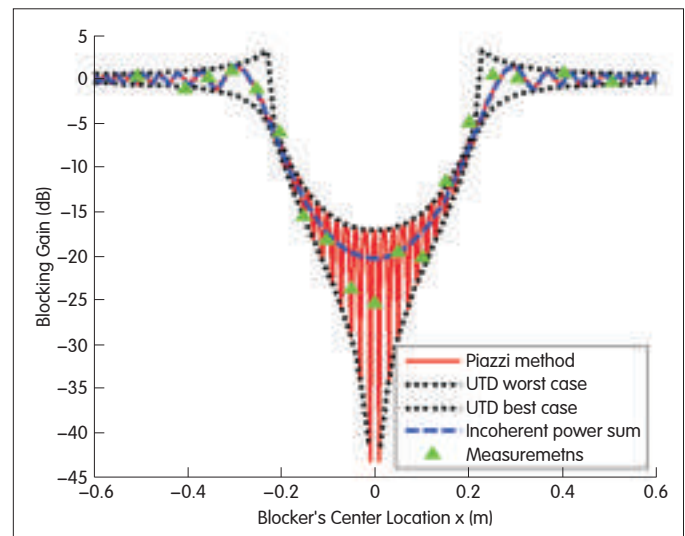
To compare with [3], we also plotted the worst-case and best-case blocking gains (computed with UTD) from a direct ray and two diffracted rays [12]. The blocking gain from the direct ray is  $-\infty$  dB when the blocker obscures the LOS and 0 dB when the blocker does not obscure the LoS. The worst-case scenario is found by assuming the secondary contributions are 180 degrees out of phase with the dominant

contribution. Conversely, the best-case scenario assumes that all contributions are in phase. The measurements mostly fall between the worst-case and best-case curves. Because the physical optics solution and UTD are nearly identical when only one screen is present, the physical optics solution lies in between the best-case and worst-case curves. The rapid variation of the Piazzi curve suggests a simplified model for human blocking. In such a model, approximate positions are used to compute the mean incoherent-power-sum gain, and the slight movements of humans are captured as a random variable with an appropriate distribution.

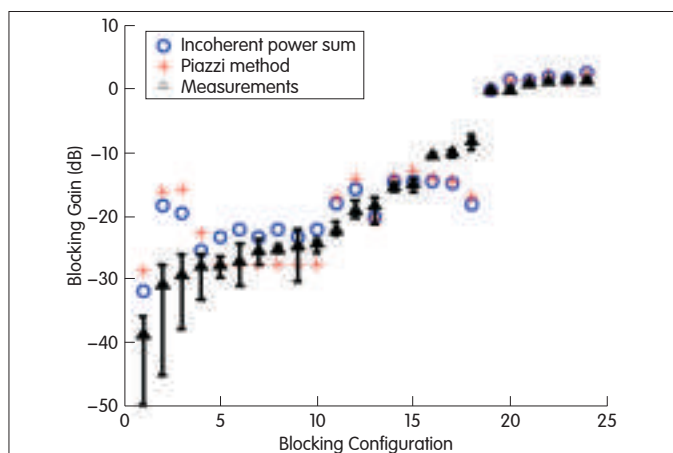
### 4.3 Blocking by Multiple People

Figs. 5 and 6 show the time-averaged blocking gain measurements for two- and three-person scenarios (in increasing order). The predicted blocking gain from Piazzi method and incoherent-power-sum approach are also plotted. Depending on the configuration in the two- and three-person blocking scenarios, there can be deep fades in the measurements when the blocking gain is less than -30 dB. The blocking gain in all scenarios ranges from 2.7 dB to -50.7 dB. This range is much larger than those in [1] and [2] and further justifies the need to include human blocking models in channel simulators.

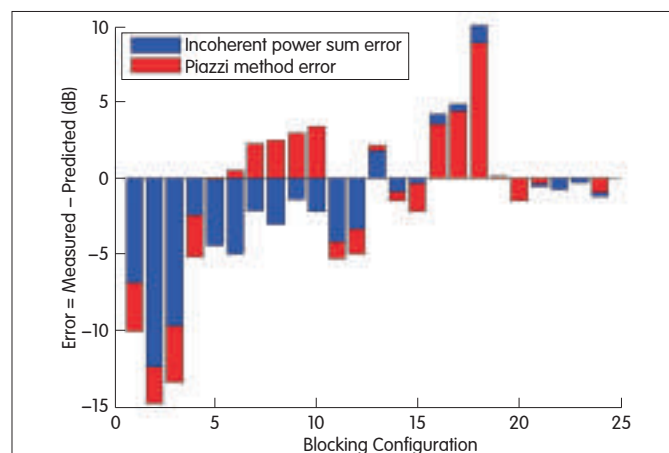
Figs. 7 and 8 show the errors in the predictions made using the Piazzi method and incoherent-power-sum approach for two- and three-person blocking scenarios. A prediction error is the predicted blocking gain minus the time-averaged blocking gain. Table 2 shows the mean and standard deviations of the prediction error using the Piazzi method and incoherent-power-sum approach. The Piazzi method has a smaller mean error but larger standard deviation compared with the incoherent-power-sum approach. This is to be expected because of uncertainty in the exact location of the blockers and time-averaging done in the measurements. We



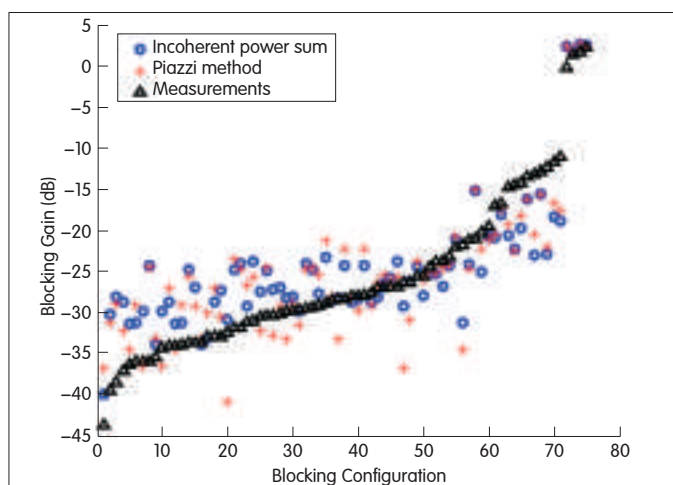
▲ Figure 4. Comparison of one-person blocking gain measurements with blocking gain predicted using the Piazzi method, incoherent-power-sum approach, and UTD.



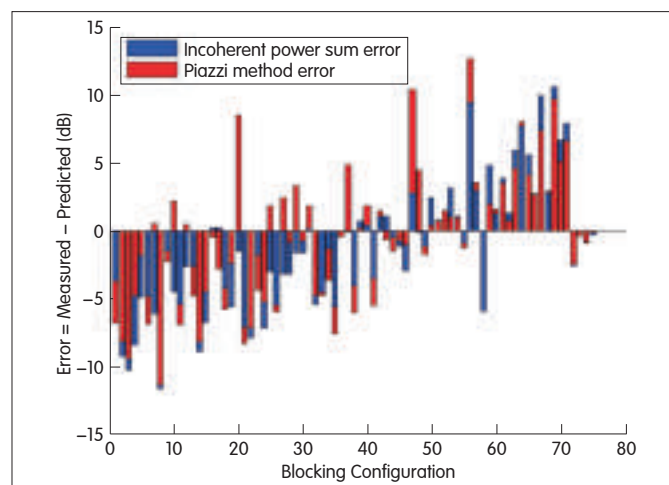
▲ Figure 5. Comparison of two-person blocking gain measurements with blocking gain predicted by Piazza method and incoherent-power-sum approach for blocker positions listed in Table 1.



▲ Figure 7. Error of blocking gain predicted using the incoherent-power-sum approach and Piazza method for two-person scenario with blocker positions listed in Table 1.



▲ Figure 6. Comparison of three-person blocking gain measurements with blocking gain predicted by Piazza method and incoherent-power-sum approach for blocker positions listed in Table 1.



▲ Figure 8. Error of blocking gain predicted using the incoherent-power-sum approach and Piazza method for three-person scenario with blocker positions listed in Table 1.

also determined the percentage of configurations with prediction error between  $\pm 5$  dB. The percentage of configurations exceeding 5 dB error for the Piazza method was 75% in a two-person scenario and 68% in a three-person scenario. The percentage of configurations exceeding 5 dB error for the incoherent-power-sum approach was 83% in a two-person and 68% in a three-person scenario.

The majority of configurations with high blocking loss typically have negative prediction errors. High blocking loss may be caused by large diffraction angles. This suggests that another model, for example, a cylinder model, may better predict diffraction around blockers at large diffraction angles. However, from the standard deviation, we conclude that an absorbing-screen model is sufficient to compute blocking gain in most applications.

In Fig. 5, the spans of the two-person blocking measurements are also plotted. The lower limit of each bar is the minimum blocking gain for a blocker configuration, and

▼ Table 2. Prediction error statistics over all configurations for two and three people blocking scenarios

Blocking Scenario	Piazza Method		Incoherent Power Sum	
	Mean Error (dB)	Standard Deviation of Error (dB)	Mean Error (dB)	Standard Deviation of Error (dB)
Two People	-1.3	5.5	-1.8	4.5
Three People	-0.7	5.0	-1.2	4.8

the upper limit is the maximum blocking gain. Because only five measurements were taken per configuration, the possible span is likely larger. In Fig. 5, most of the blocking configurations with a span larger than 10 dB have blocking gain less than -30 dB and have larger prediction errors. The variability and large negative gains could be caused by the constructive and destructive interference of the various diffracted paths and the uncertainty in the exact position and movements of the blockers. Thus, we should expect that in



these configurations, the Rician K-factor is small. The Rician K-factor is the power ratio of the dominant arrival (path with the greatest received power) to all other arrivals. It is a metric that is negatively correlated with the level crossing rate and positively correlated with the fading depth of the received field [17]. From the predicted powers of the two person blocking configurations, the computed K-factors are all less than 2 dB, which is small. However, the majority of the three-person blocking configurations have predicted K-factors greater than 10 dB. The large predicted K-factor in the three-person blocking scenarios may allude to some other significant multipath that we have not considered.

## 5 Conclusion

We have treated multiple human blockers as absorbing screens of infinite height to create a model for computing blocking gain in 60 GHz links. The model agrees well with one-, two- and three-person blocking-gain measurements in the range of 2.7 dB to -50 dB. In a few select cases with large incoherent-sum blocking gain of less than -30 dB, the predicted errors are greater than 5 dB. Coupled with a large predicted Rician K-factor, this alludes to the presence of un-accounted for multipath and/or the need for more accurate models of human blockers at large diffraction angles. However, our comparisons with the measurements show that our model is sufficient for determining the impact of multiple blocking humans.

For deterministic system-level simulations of the millimeter wave radio channel, the Piazza method can be used in conjunction with a ray-tracing simulator to determine the gain experienced by paths blocked by humans. However, in scenarios with human blockers, a statistical component in the model is preferred because it is often impractical to include the slight movements of humans in a simulation. In such scenarios, we propose the incoherent-power-sum approach coupled with a random variable to predict the gain on paths blocked by humans. This random variable is dependent on the exact complex fields of the various multipaths and the movement of the blockers. Further research needs to be done on the statistical nature of this random variable.

## Acknowledgements

The authors wish to express their gratitude and appreciation to Dr. Henry L. Bertoni of Polytechnic Institute of New York University for his helpful insights and comments on millimeter wave propagation.

## References

- [1] A. P. Garcia, W. Kotterman, U. Trautwein, D. Bruckner, J. Kunisch, and R. S. Thoma, "60 GHz Time-Variant Shadowing Characterization within an Airbus 340", in *Proc. 4th EU Conf. on Antenna And Propagation (EUCAP)*, Apr. 2010.
- [2] S. Collonge, G. Zaharia and G. E. Zein, "Influence of the Human Activity Wideband Characteristics of the 60 GHz Indoor Radio Channel", *IEEE Trans. on Wireless Comm.*, vol. 3, No. 6, 2389-2406, Nov. 2004.
- [3] M. Jacob, S. Priebe, A. Maltsev, et al., "A ray tracing based stochastic human blockage model for the IEEE 802.11ad 60 GHz channel model", in *Proc. 5th EU Conf. on Antenna And Propagation (EUCAP)*, Apr. 2011.
- [4] M. Jacob, S. Priebe, R. Dickhoff, T. Kleine-Ostmann, T. Schrader, T. Kürner,

- "Diffraction in mm and sub-mm Wave Indoor Propagation Channels", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 60, No. 3, pp.833-844, Mar. 2012.
- [5] C. Gustafson and F. Tufvesson, "Characterization of 60 GHz shadowing by human bodies and simple phantoms" in *Proc. 6th EU Conf. on Antenna And Propagation (EUCAP)*, Mar. 2012.
- [6] J. Wang, R.V. Prasad, and I. Niemegeers, "Analyzing 60 GHz radio links for indoor communications," *IEEE Trans. Consumer Electronics*, vol. 55, No. 4, pp. 1832-1840, Nov. 2009.
- [7] A. Khafaji, R. Saadane, J. El Abbadi and M. Belkasm, "Ray tracing technique based 60 GHz band propagation modelling and influence of people shadowing" World Academy of Science, Engineering and Technology, 2008.
- [8] Z. Genc, W. V. Thillo, A. Bourdoux, and E. Onur, "60 GHz PHY performance evaluation with 3D ray tracing under human shadowing," *IEEE Wireless Comm. Letters*, vol. 1, no. 2, pp. 117-120, Apr. 2012.
- [9] J. Bach Andersen, "UTD multiple-edge transition zone diffraction," *IEEE Trans. on Antennas and Propagation*, vol. 45, pp. 1093-1097, July 1997.
- [10] Kunisch and J. Pamp, "Ultra-wideband double vertical knife-edge model for obstruction of a ray by a person", in *Proc. IEEE ICUWB*, Sept. 2008.
- [11] L. Piazza, "Multiple Diffraction Modeling of Wireless Propagation in Urban Environments", Dissertation for the PhD degree in ECE, January 1998.
- [12] H. L. Bertoni, *Radio Propagation for Modern Wireless Applications*. Upper Saddle River, NJ: Prentice Hall, PTR, 2000, ch. 6.
- [13] L. Piazza and H. L. Bertoni, "Effect of terrain on path loss in urban environments for wireless applications," *IEEE Tran. on Antennas and Propagation*, vol. 46, no. 8, pp. 1138-1147, 1998.
- [14] L. E. Vogler, "An attenuation function for multiple knife-edge diffraction," *Radio Science*, vol. 17, no. 6 pp. 1541-1546, 1982.
- [15] J. H. Whitteker, "Ground wave and diffraction," *AGARD Meeting*, October 1994, pp. 2A-1 - 13.
- [16] J. H. Whitteker, "Numerical evaluation of one-dimensional diffraction integrals," *IEEE Trans. on Antennas and Propagation*, Vol. 45, No. 6, pp.1058-1061, 1997.
- [17] A. Abdi, K. Wills, H. A. Barger, M. S. Alouini, and M. Kaveh, "Comparison of the level crossing rate and average fade duration of Rayleigh, Rice, and Nakagami fading models with mobile channel data," in *Proc. IEEE Vehic. Technol. Conf.*, Boston, MA, pp. 1850-1857, 2000.

Manuscript received: August 14, 2012

## B iographies

**Jonathan S. Lu** (Jonathan.Lu@interdigital.com) received his BS and MS degrees in electrical engineering from Polytechnic Institute of New York University. He is currently working toward a PhD degree in electrical engineering at the same university. His research interests are in UHF propagation modeling for urban and rural environments, millimeter wave propagation modeling, and spectrum sensing for cognitive radio.

**Daniel Steinbach** (Daniel.Steinbach@interdigital.com) received his BSEE degree from Cornell University in 1988 and his MSEE degree from Syracuse University in 1990. He received an MBA degree from the Zarb School of Business, Hofstra University, in 2006. He worked on sonar and radar applications early in his career and later worked on data communications. He currently works in wireless communications for InterDigital Communications.

**Patrick Cabrol** (Patrick.Cabrol@interdigital.com) received his BS degree in electrical engineering from New York Institute of Technology. He is currently working toward his MS degree in electrical engineering at Polytechnic Institute of NYU. Patrick has more than 19 years' experience in RF Design and wireless communications. He works as a senior staff engineer at InterDigital Communications.

**Phil Pietraski** (philip.pietraski@interdigital.com) received his BSEET from DeVry University in 1987. He received his BSEE, MSEE, Grad.Cert. in wireless communications, and PhD EE from Polytechnic University, Brooklyn (now NYU-Poly) in 1994, 1995, 1996, and 2000.

He joined InterDigital Communications in 2001 and is currently a principal engineer leading research activity in wireless communications, most recently in millimeter wave communications and future cellular architectures. He holds more than 50 patents in wireless communications and has authored multiple conference and journal papers. He is vice chair of the MoGig (Mobile Gigabit) working group at IWPC and a trustee for DeVry NJ campuses.

Prior to his transition to wireless communications in 2000, he was a research engineer at Brookhaven National Laboratory, National Synchrotron Light Source, responsible for beam-line instrumentation and X-ray detector R&D. He has also conducted research at the Polytechnic University for the Office of Naval Research (ONR) in underwater source localization.

# 60 GHz SIW Steerable Antenna Array in LTCC

**Bahram Sanadgol<sup>1</sup>, Sybille Holzwarth<sup>1</sup>, Peter Uhlig<sup>1</sup>, Alberto Milano<sup>2</sup>, and Rafi Popovich<sup>2</sup>**

(1. IMST GmbH Carl-Friedrich-Gauss-Str. 2, 47475 Kamp-Lintfort, Germany;

2. Beam Networks, 1 Ehad Ha'am 76248 Rehovot, Israel)

## Abstract

In this paper, we present a 60 GHz substrate-integrated waveguide fed-steerable low-temperature cofired ceramics array. The antenna is suitable for transmitting and receiving on the 60 GHz wireless personal area network frequency band. The wireless system can be used for HDTV, high-data-rate networking up to 4.5 GBit/s, security and surveillance, and similar applications.

## Keywords

substrate integrated waveguide(SIW); phase shifted injected push-push oscillator(PSIPPO) ; low temperature co-fired ceramic(LTCC) ; monolithic microwave integrated chip(MMIC) ; wireless personal area network(WPAN)

## 1 Introduction

Applications of the 60 GHz band for high data-rate services have become more interesting, especially over the past few years. A comprehensive list of multimedia applications using the 60 GHz band can be found in [1]–[3]. There is a need for compact, highly efficient front-ends, but designing systems with such front-ends is challenging. This paper will start with a brief introduction of 60 GHz WPAN applications. The company Beam Networks (BN) is developing a high-performance, low-cost wireless transceiver system at 60 GHz for wireless personal area network (WPAN) applications.

The antenna design and architecture, which uses the finite difference time domain (FDTD) field solver called Empire, is then described in detail. We analyze active impedance to determine the performance of the scanned antenna array. The antenna for the system comprises four waveguide-fed columns that can be excited with different phases for beam-steering applications.

Far-field measurements of the realized antenna demonstrator are then presented. We show that the measured performance is within the required specifications, and the measurements also confirm that the antenna is low-loss.

## 2 Transceiver Architecture

Fig. 1 shows the transceiver architecture comprising a crystal-locked master transceiver on the right and a slave transceiver on the left. The frequency and phase of the slave

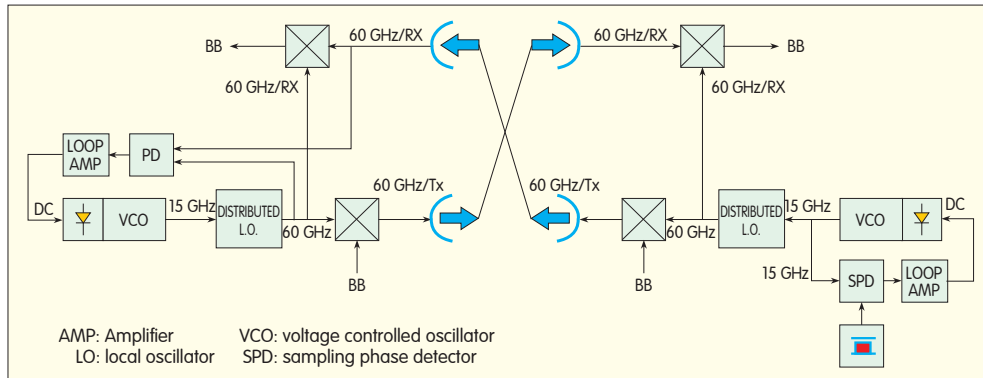
transceiver is locked to the master. Each transceiver comprises four channels that interface the antenna in order to form a steerable, focused beam. A reference signal is generated at 15 GHz, and this signal injects and locks the push-push oscillator 15–30 GHz. In this way, coherent signals of the buffer amplifiers are delivered at the output ports of the amplifier for up/down conversion. The phase of each signal in the array is adjusted by tuning the band rejection filter (BRF) of each phase-shifted injected push-push oscillator (PSIPPO) at 30–60 GHz to implement beamforming.

The transceiver can operate in full duplex or spatial-division duplexing (SDD) mode because of the high isolation that can be achieved with the presented antenna and transceiver design. In Fig. 1, two transceivers communicate over a bidirectional link. Both transmit and receive data simultaneously on carriers that use the same frequency. The transmitter and receiver can operate simultaneously as long as any reflected waves receive significant attenuation by the environment. Such SDD operation is typically possible because the channel is highly specular, the antennas are highly directional.

## 3 Antenna Design and Simulation

In this section, the design and simulation of the substrate integrated waveguide (SIW) array antenna is explained. All modelling and simulations were performed using the 3D field solver called Empire [4], which is based on finite difference time domain (FDTD).

The goals of antenna design are to achieve a bandwidth of



▲ Figure 1. Coherent down-conversion with automatic phase compensation.

10% in the 60 GHz band, a maximum scan range of  $\pm 30^\circ$ , and total antenna gain of 18–20 dBi. Bandwidth and gain (efficiency) requirements make the choice of the single element more significant. An open waveguide radiator has large bandwidth, low cross polarization, a small front-to-back radiation ratio, and high efficiency.

Loss has always been an important issue in the design of high-efficiency antennas. If designed properly, a waveguide-feeding network could be a low-loss solution compared to a microstrip transmission line. Although microstrip and stripline technologies are very appropriate from an integration perspective, the surface resistance related to these transmission lines increases with the square root of the frequency [5]. As a consequence, they are lossy at high microwave frequencies.

Integrating the antenna and feeding network was the next concern in the design process. Waveguide-like structures can be constructed using periodic metallic via posts in a substrate. Realizing an array of such antenna elements and their feeding network requires a relatively thick substrate. Low-temperature co-fired ceramic (LTCC) technology greatly benefits microwave applications and can solve this problem because a many layers, including vias and metalized surfaces in between, can be manufactured. Ceramic substrates as well as gold and silver pastes have excellent physical and electrical properties. Moreover, material and processing costs are competitive compared with substrate systems such as HTCC and printed circuit boards for high microwave frequencies [6].

Fig. 2 shows the proposed waveguide radiator. This element is designed to have the same radiation properties as an open-ended waveguide, and the dominant propagating mode is TE<sub>10</sub>. The waveguide walls comprise via fences. Between each layer, metal surfaces (usually gold or silver) connect the vias so that the current along the side walls of the waveguide is not interrupted. In a typical open-ended waveguide, the electric field of the aperture is given by

$$E_x = E_0 \cos(\pi y/a) \quad (1)$$

(1) is a good approximation of the field distribution, but it does not include all the details for an SIW radiator. In other words, it is not easy to model and analyse the SIW

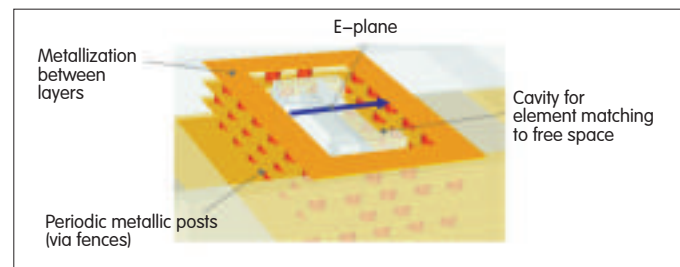
theoretically.

One of the drawbacks of waveguide radiators is the mismatch to free space. For an SIW antenna element, this mismatch is even worse because of high  $\epsilon_r$ . LTCC materials usually have high permittivity, which makes the SIW radiator suffer from high reflections at the aperture face. To improve the transition from waveguide mode to free-space mode, a substrate with lower  $\epsilon_r$  must be used or the effective dielectric constant of LTCC must

be reduced. The latter can be done by cutting out cavities in the LTCC. The special shape and dimensions of these cavities can help minimize reflection at the antenna interface (Fig. 2).

Another known problem with SIW is leakage through the gap between vias. Ideally, the vias are the waveguide walls, and there is no energy loss. However, leakage does exist and is always greater at lower frequencies. This means that denser vias do not necessarily result in better performance [7]. If the vias are not compact enough, a band-stop structure can be built in the desired frequency band. The design goal should be to minimize leakage while shifting the stop band to higher frequencies. Of course, LTCC design rules should always be kept in mind. The substrate thickness, via spacing, and via diameter should be carefully chosen. With these in mind, the single element was simulated and optimized with Empire.

The next step was to design a proper feeding network for the array. According to gain requirements, the array was set at  $4 \times 4$ , which should be able to scan in one direction. Each of the four elements in the same column shares a feeding line; therefore, we designed the feeding for one column. Like the radiating elements, the waveguides are realized in the substrate, and via fences are the waveguide walls. The feeding network should be designed in such a way that all antenna elements radiate in phase. Fig. 3 shows one array column comprising four active elements that share the same feeding network. The scan specifications require the antenna to be steerable only in one plane (the H-plane here). On the E-plane along one column, the element distance can be increased. This reduces integration complexity and improves array performance. At first, an E plane T-junction splits the power coming from a standard waveguide WR-15. Using a



▲ Figure 2. Single-element SIW radiator with cavity for improved matching.

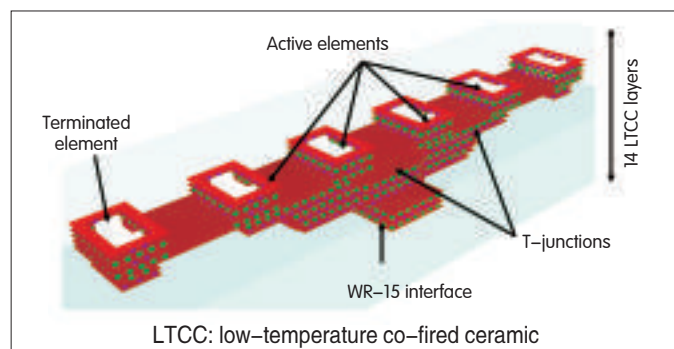
bend and another E-plane T-junction, each element can be fed with the same phase. There are two passive elements at each column that help improve side-lobe level and reduce the back radiation. These elements are terminated using resistance paste, so the reflection from the elements is minimized, and most of the coupled energy is absorbed in the resistor. A proper interface to the standard waveguide is also needed. Fig. 4 shows the electrical field of the feeding network and the active antenna elements for one column.

Fig. 5 shows the complete array configuration. To optimize the design, the array is simulated, and all four columns are active. The active impedance method guarantees the inclusion of mutual coupling in the final design [8]. This can be very important in the case of a scanning array; optimizing the array in active mode can avoid the need for post-manufacturing tuning. Using the same method, we can analyse how the array input impedance varies with the scan angle. Using this method, the array input impedance has been optimized for an angle off boresight to obtain the best bandwidth for all the scan angles (here, up to 30°). A similar array optimized with the active impedance method can be found in [9].

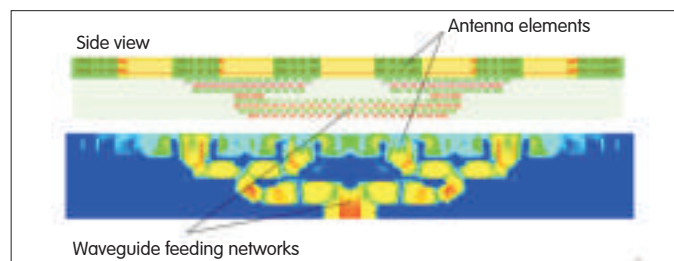
## 4 Measurement

The antenna was fabricated according to the optimized layout. Fig. 6 shows a finished prototype.

The complete LTCC tile comprises four columns, and there are only four active elements in each column. The LTCC material used here was Ferro A6-M, with  $\epsilon_r$  around 5.7. The



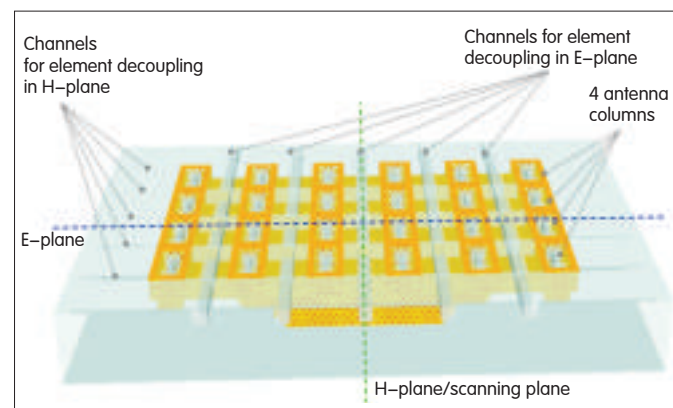
▲ Figure 3. One array column consisting of four active elements and two passive elements. The corresponding SIW feeding network is also shown.



▲ Figure 4. Electrical field of the waveguide feeding network and antenna elements.

total number of vias for one array is approximately 4300 in all 14 layers. To be able to do the reflection and far-field measurements, a special metal frame was devised. The antenna was fixed in this dedicated test frame and was then fed by the WR-15 standard waveguide from the back through an opening made for WR-15. All the prototypes manufactured in this phase contain only one active column, although the array is optimized for the case where all four columns are active simultaneously. The reason for only one active column is because the dimensions of a standard waveguide do not allow two standard waveguides beside each other. Four prototypes of the array were manufactured; two had an active outside column, and the other two had an active inside column.

Fig. 7 shows return loss against Empire simulation for one prototype. There is approximate agreement between the simulation and measurement over the entire band. The slight



▲ Figure 5. Final antenna configuration for the array scanning in H-plane.

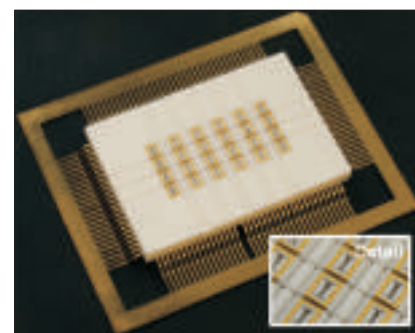


Figure 6. ▶ Manufactured tile of the four-column array (one active column only).

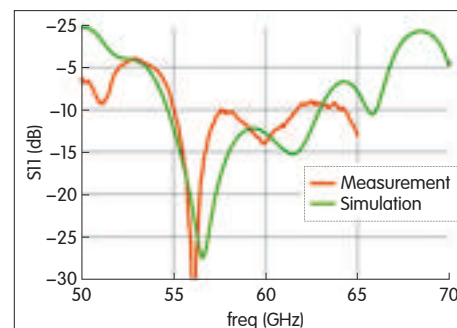


Figure 7. ▶ Reflection coefficient measurement and simulation results.



difference is usually because of LTCC design tolerances. Although the manufacturing has been done as accurately as possible, the LTCC process is naturally more sensitive to tolerances than other types of PCB manufacturing. Nevertheless, considering the number of vias and layers, LTCC seems to be a robust and reasonable solution. All the manufactured prototypes were measured and show a very good reproducibility, which is crucial for the final series production.

Farfield measurement was done in an anechoic chamber with a special setup and appropriate isolation. Each column was measured separately; that is, all the other columns were terminated passively. By applying the field superposition from each column and the required phase for the desired scan angle, the farfield diagram of the whole array was calculated. The final farfield diagrams for the array steered to  $\pm 30^\circ$  are shown in Fig. 8.

## 5 Conclusion

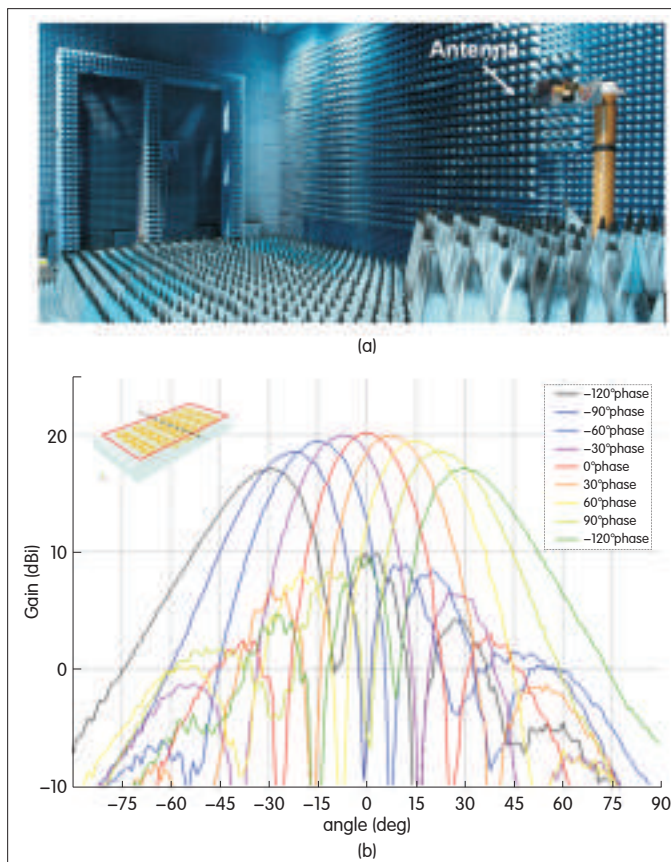
In this paper, the design, fabrication, and measurement of an array antenna for WPAN 60 GHz application was described. The antenna elements were substrate-integrated waveguide radiators and were fed by a SIW feeding network. The design was realized in LTCC, and the measurements

were taken using standard waveguide. There was very good agreement between the simulations and measurements. In the future, this array antenna will be integrated with the MMIC chip to build a commercial WPAN system.

## References

- [1] P. Smulders, "Exploiting the 60 GHz band for local wireless multimedia access: prospects and future directions", *IEEE Communications Magazine*, vol. 40, no. 1, pp. 140–147.
- [2] S. Holzwarth, R. Baggen, "Planar antenna design at 60 GHz for high data rate point-to-point connections," APS July 2005.
- [3] J. Laskar, S. Pinel, D. Dawn, S. Sarkar, P. Sen, B. Perunama, D. Yeh, and F. Barale, "60 GHz Entertainment Connectivity Solution," ICUWB September 2009.
- [4] 3IMST GmbH [online]. Available: <http://www.empire.de>
- [5] G. Ponchak, N. Dib, L. Katehi, "Design and analysis of transitions from rectangular waveguide to layered ridge dielectric waveguide", *IEEE Transactions on Microwave Theory and Technique*, July 1996
- [6] IMST GmbH [online]. Available: <http://www.ltcc.de>
- [7] Feng Xu, Ke Wu, "Guided-wave and leakage characteristics of substrate integrated waveguide", *IEEE Transactions on Microwave Theory and Technique*, January 2005.
- [8] D. M. Pozar, "A relation between the active input impedance and the active element pattern of a phased array", *IEEE Transactions on Antenna and Propagation*, September 2003.
- [9] B. Sanadgol, S. Holzwarth, J. Kassner, "30 GHz Liquid Crystal Phased Array", *Loughborough Antennas & Propagation conference*, Loughborough, UK, November 2009.

Manuscript received: September 10, 2012



▲ Figure 8. Farfield measurements for the scanned array with the proper phase for each column. a) setup, b) results.

## Biographies

**Bahram Sanadgol** received his MS degree in electrical and electronics engineering from the University of Duisburg–Essen, Germany, in 2007. During his studies, his main research focus was phased-array antennas. In November 2005, he joined the Antenna and EM Modelling Department of IMST. He currently researches phased-array design, simulation methods for planar arrays, and advanced measurement techniques for antennas.

**Sybille Holzwarth** received her Dipl.–Ing. degree in electrical engineering from the University of Ulm, Germany, in 1997. She then joined the Antenna and EM Modelling Department, IMST GmbH, Kamp–Lintfort, Germany, and currently works there as a development engineer and project manager. Her main interests are EM modelling of antennas and passive components, with special emphasis on microwave antennas for radar and communications, planar antennas and arrays, and electronically steerable antennas.

**Peter Uhlig** (uhlig@imst.de) received his degree in electrical engineering in 1984. He began his career at Wandel & Goltermann in 1984 developing microwave components for spectrum analysers. From 1988 onwards, he was responsible for developing thin-film circuits of the microwave front-end in spectrum analysers. He joined IMST GmbH in 1993 and now heads the hybrid microelectronics laboratory, which includes the LTCC prototyping line.

**Alberto Milano** (alberto@beamnetworks.com) is the CTO and co-founder of Beam Networks. He received his PhD in Telecommunications Engineering from the Polytechnic Institute of Milan. Throughout his career, he has designed and developed T/R modules, and MMIC and RFIC circuits for military and commercial use. He has worked at GE–Italy, Optomic, ELTA–IAI, and Agilent (all in Israel). Dr. Milano holds seven patents in the areas of microwave RF transmission and analog phased-array antenna design.

**Rafi Popovich** (rafi@beamnetworks.com) is an engineering manager at Beam Networks. He received his BSEE degree in electrophysics from City College, New York, in 1984. He received his MSEE degree in microwave engineering from Polytechnic University, New York, in 1986. Throughout his career, he has been involved in microwave and millimeter wave R&D and production. HE has designed, developed, and produced discrete control components, MMIC and RFIC circuits and packaging, and supercomponents and modules. He has published several papers and holds patents on microwave and millimeter wave technology.

# Line-of-Sight MIMO for Next-Generation Microwave Transmission Systems

Xianwei Gong, Zhifeng Yuan, Jun Xu, and Liujun Hu

(Wireless Technology Advance Research Team, ZTE, Shenzhen 518057, P. R. China)

## Abstract

Line-of-sight MIMO (LoS MIMO) is not applicable in scattering wireless transmission scenarios, but it may be applied in LoS microwave transmission scenarios if antenna spacing (within transmit and/or receive arrays) is suitable and there is one hop distance. LoS MIMO can improve channel capacity and performance of a transmission system. In this paper, we discuss factors affecting channel capacity and performance in LoS MIMO. We also discuss the feasibility LoS MIMO applications.

## Keywords

line of sight MIMO; microwave; channel capacity

## 1 Introduction

**M**ultiple-input multiple-output (MIMO) transmission systems with more than one transmitting (Tx) antenna and receiving (Rx) antenna can have improved channel capacity (throughput) by using the spatial dimension. By using the spatial dimension, such systems can be freely applied in wireless scenarios. However, they are mainly non-line of sight MIMO (NLoS MIMO) systems; line-of-sight MIMO (LoS MIMO) is not currently used for wireless because many wireless environments are scattering environments. However, LoS MIMO may be used for LoS microwave transmission if the antennas are suitably spaced within transmit and/or receive arrays and at one-hop distance. LoS MIMO can improve channel capacity and performance of a transmission system.

According to Shannon's law, channel capacity is given by

$$C/W = \log \left( 1 + \frac{P}{N_0} \right) \text{ bits/s/Hz} \quad (1)$$

where  $P$  is total Tx power and  $N_0$  is average noise power spectral density.

A MIMO channel can be equivalent to a vector Gaussian channel, and the capacity can be computed by decomposing the vector channel into a set of parallel, independent scalar Gaussian subchannels. The channel capacity is

$$C/W = \sum_{i=1}^n \log \left( 1 + \frac{P_i \lambda_i^2}{N_0} \right) \text{ bits/s/Hz} \quad (2)$$

where the  $i$ th subchannel power;  $P_i$  is allocated injection power according to waterfilling theorem;  $\lambda_i$  is the non-zero singular value of the channel matrix  $\mathbf{H}$ ; and  $n$  is the number of subchannels or number of singular values of  $\mathbf{H}$ .

When the MIMO channel has high SNR, the water level is deep, and allocating equal power on the non-zero eigenmodes is asymptotically optimal. The MIMO channel capacity is also asymptotically optimal [2], and is given as

$$C/W = \sum_{i=1}^n \log \left( 1 + \frac{\bar{P} \lambda_i^2}{N_0} \right) = \sum_{i=1}^n \log \left( 1 + \frac{P \lambda_i^2}{n N_0} \right) \text{ bits/s/Hz} \quad (3)$$

where  $P_i = \bar{P} = (P/n)$ .

In section 2, we discuss the principle of a  $2 \times 2$  LoS MIMO system that includes two Tx antennas and two Rx antennas. This principle is often discussed in microwave communications. In section 3, we discuss some antenna location factors that affect channel capacity in a  $2 \times 2$  LoS MIMO system. We also describe some simulations. Section 4 concludes the paper.

## 2 Principle of $2 \times 2$ LoS MIMO

In an LoS environment, all paths suffer almost equal path loss; therefore, we do not take path loss into consideration. For a  $2 \times 2$  LoS MIMO system, the normalized channel matrix  $\mathbf{H}$  can be described as

$$\mathbf{H}_{\text{LoS}} = \begin{bmatrix} \exp(jkd_{11}) & \exp(jkd_{12}) \\ \exp(jkd_{21}) & \exp(jkd_{22}) \end{bmatrix} \quad (4)$$

where  $k = 2\pi/\lambda$ ;  $\lambda$  is the carrier wavelength;  $d_{ij}$  is the length

from Tx antenna  $i$  to Rx antenna  $j$ ;  $i=1,2$ ; and  $j=1,2$ . Then the deduced eigenvalues of  $\mathbf{W}=\mathbf{H}_{\text{LoS}}\mathbf{H}_{\text{LoS}}^H$  are

$$\begin{aligned}\lambda_1^2 &= 2+[2+2\cos(kA)]^{1/2} \\ \lambda_2^2 &= 2-[2+2\cos(kA)]^{1/2}\end{aligned}\quad (5)$$

where  $A = d_{11}+d_{22}-d_{12}-d_{21} = (d_{11}-d_{21}) + (d_{22}-d_{12})$ , and  $\lambda_1, \lambda_2$  are also singular values of  $\mathbf{H}_{\text{LoS}}$  [3].

The channel capacity of a  $2 \times 2$  LoS MIMO system is obtained by putting the eigenvalues of (5) into (3). From linear algebra,  $\text{trace}\{\mathbf{W}\} = \sum_{i=1}^2 \lambda_i^2 = N \times M = 2 \times 2 = 4$  for the normalized unit matrix  $\mathbf{H}_{\text{LoS}}$ , where  $N(=2)$ ,  $M(=2)$  are the number of rows and columns of  $\mathbf{H}_{\text{LoS}}$ . Then,  $\lambda_1^2 = \lambda_2^2$  can maximize the result of (3).

When the eigenvalues are equal, the optimal channel capacity of a  $2 \times 2$  LoS MIMO system is

$$(C/W)_{\text{optimal}} = 2 \times \log\left(1 + \frac{P}{N_0}\right) \text{ bits/s/Hz} \quad (6)$$

In an  $N \times M$  MIMO system, we define the rows of  $\mathbf{H}_{\text{LoS}}$  as  $h_i$ . All rows in  $\mathbf{H}_{\text{LoS}}$  are mutually orthonormal and can be described by  $\langle h_i, h_j \rangle = 0, i \neq j$ . This condition can be used to produce the formula for the optimal antenna spacing [4]:

$$d_i d_r = \frac{\lambda D}{n \cos(\theta_i) \cos(\theta_r)} K \quad (7)$$

where

- $d_i$  is Tx adjacent antenna spacing
- $d_r$  is Rx adjacent antenna spacing
- $\lambda$  is carrier wavelength
- $D$  is the distance of one hop
- $n = \min(N, M)$
- $\theta_i$  is the angle between the Tx antenna array and Z axis
- $\theta_r$  is the angle between the Rx antenna array and Z axis
- $K$  is a positive odd number (usually 1 because that gives the smallest optimal antenna spacing).

According to (7), for a  $2 \times 2$  LoS MIMO system, the requirement on the antenna spacing is

$$d_i d_r = \frac{\lambda D}{2 \cos(\theta_i) \cos(\theta_r)} \quad (8)$$

Fig. 1 shows the Tx antennas and Rx antennas of a  $2 \times 2$  LoS MIMO system organized in an array.  $\phi_i$  is the angle between the Tx antenna and x-z plane, and  $L$  is the mast length.

Assuming that  $\theta_i = \theta_r = 0$  and  $d_i = d_r$ , then (8) becomes

$$d_i^2 = d_r^2 = \lambda D / 2 \quad (9)$$

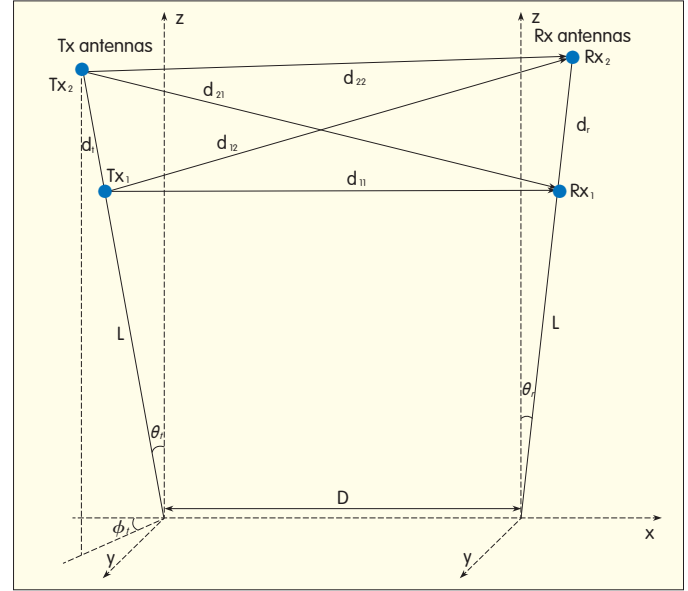
because

$$d_{11} = d_{22} = D \quad (10)$$

and  $D \gg \lambda/4$ . Then,

$$\begin{aligned}d_{12} = d_{21} &= \sqrt{d_{11}^2 + d_r^2} = \sqrt{D^2 + (\lambda D / 2)} \\ &\approx \sqrt{D^2 + (\lambda D / 2) + (\lambda / 4)^2} = D + (\lambda / 4)\end{aligned}\quad (11)$$

Substituting (11) and (12) into (5), we get



▲ Figure 1. Tx antennas and Rx antennas in  $2 \times 2$  LoS MIMO system.

$$\begin{aligned}A &= d_{11}+d_{22}-d_{12}-d_{21} = (d_{11}-d_{21}) + (d_{22}-d_{12}) \\ &= 2 \times (D - (D + \lambda/4)) = -\lambda/2 = -\pi/(2\pi/\lambda) = -\pi/k\end{aligned}\quad (12)$$

Then, the eigenvalues of  $\mathbf{W}$  are equal, and the optimal channel capacity of  $2 \times 2$  LoS MIMO can be achieved.

From (9)–(11), we obtain the principle of  $2 \times 2$  LoS MIMO system for optimal channel capacity (Fig. 2).

### 3 Factors Affecting the Channel Capacity of $2 \times 2$ LoS MIMO

Channel capacity of  $2 \times 2$  LoS MIMO system is not optimal unless antenna spacing  $d_i$  and  $d_r$ , carrier wavelength  $\lambda$ , one-hop distance  $D$ , and the angle  $\theta_i, \theta_r$  between antenna arrays and z axis (8). But in practical implementation, these factors do not satisfy (8) exactly and channel capacity is not optimal.

In this section, we discuss factors that affect LoS MIMO channel capacity of LoS MIMO in practical applications. These factors are: offset of hop distance  $D$ , offset of antenna spacing, Tx antenna and Rx antenna not on the same horizon, Tx and Rx antenna arrays are not on the same plane.

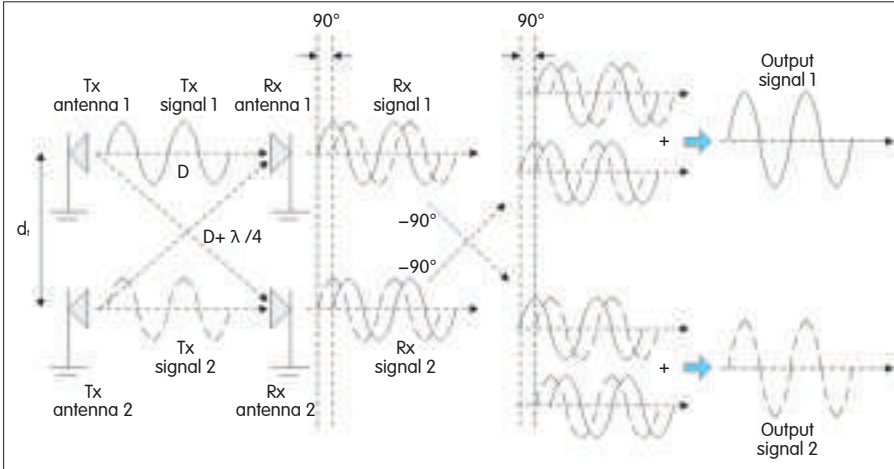
#### 3.1 Offset of Hop Distance $D$

If the offset of hop distance  $D$  ( $D$ -offset) is  $\Delta$  and  $d_i = d_r$  (Fig. 3), then

$$\begin{aligned}d_{11} &= D + \Delta & d_{21} &= \sqrt{(D+\Delta)^2 + (d_i)^2} \\ d_{22} &= D + \Delta & d_{12} &= \sqrt{(D+\Delta)^2 + (d_i)^2}\end{aligned}\quad (13)$$

Substituting (13) into (5) and (3), we get the channel capacity within  $D$ -offset. We use a simulation to show how the  $D$ -offset affects channel capacity.

Fig. 4 shows a comparison of optimal channel capacity and channel capacity within  $D$ -offset (range  $[-1000 \text{ m}, 1000 \text{ m}]$ , interval  $50 \text{ m}$ ) when carrier frequency  $f_c = 18 \text{ GHz}$ ,

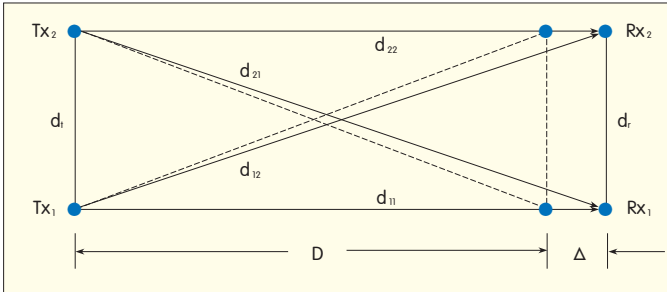


▲ Figure 2. A  $2 \times 2$  LoS MIMO system for optimal channel capacity.

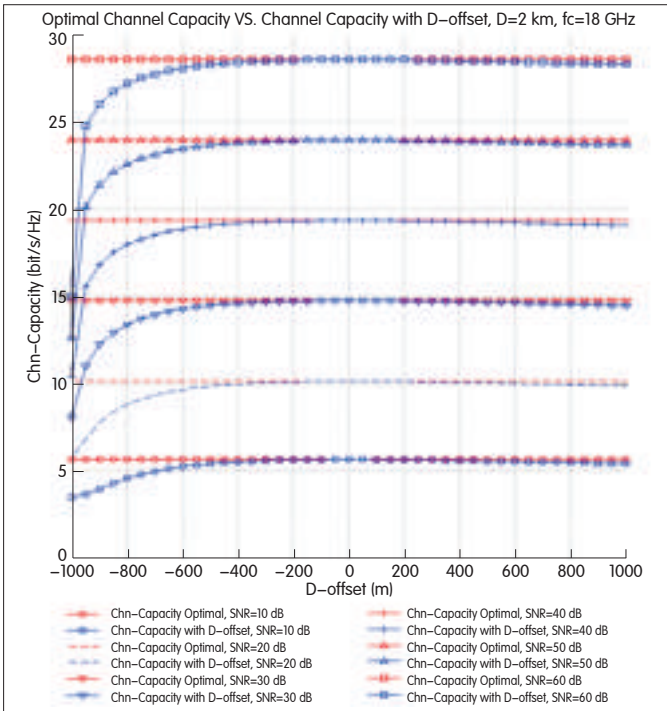
$D = 2000$  m,  $\theta_t = \theta_r = 0$ , and  $d_t = d_r$ , which satisfies (10).

The change of channel capacity corresponding to negative D-offset is larger than that corresponding to positive D-offset in  $[-1000$  m,  $1000$  m] and when  $f_c = 18$  GHz. The larger the absolute value of the negative D-offset, the greater the difference of the channel capacity within D-offset from the optimal channel capacity. As the SNR increases, this tendency becomes greater.

In practice, the D-offset factor, which influences channel capacity, should be applied during project implementation to reduce costs while maintaining performance.



▲ Figure 3. Offset of one hop distance  $D$  in  $2 \times 2$  LoS MIMO system.



▲ Figure 4. Channel capacity of optimal antenna parameters settings and antenna parameters settings within D-offset in different SNR (from 10 dB to 60 dB and interval is 10 dB).

### 3.2 Offset of Antenna Spacing

In Fig. 5, the offset of antenna spacing  $d_t$  ( $d_t$ -offset) is  $\Delta_{dt}$  and  $d_t = d_r$ . Then,

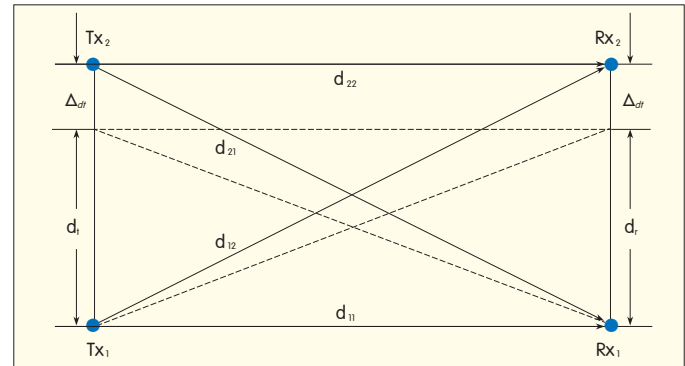
$$\begin{aligned} d_{11} &= D & d_{21} &= \sqrt{D^2 + (d_t + \Delta_{dt})^2} \\ d_{22} &= D & d_{12} &= \sqrt{D^2 + (d_t + \Delta_{dt})^2} \end{aligned} \quad (14)$$

Substituting (14) into (5) and (3), the channel capacity within  $d_t$ -offset can be obtained.

We use a simulation to show how  $d_t$ -offset influences channel capacity. Fig. 6 shows a comparison of optimal channel capacity and channel capacity within  $d_t$ -offset (range  $[-2$  m,  $2$  m], interval  $0.1$  m) when frequency  $f_c = 18$  GHz,  $D = 2000$  m,  $\theta_t = \theta_r = 0$ , and  $d_t = d_r = 4.0825$  m, which satisfies (10).

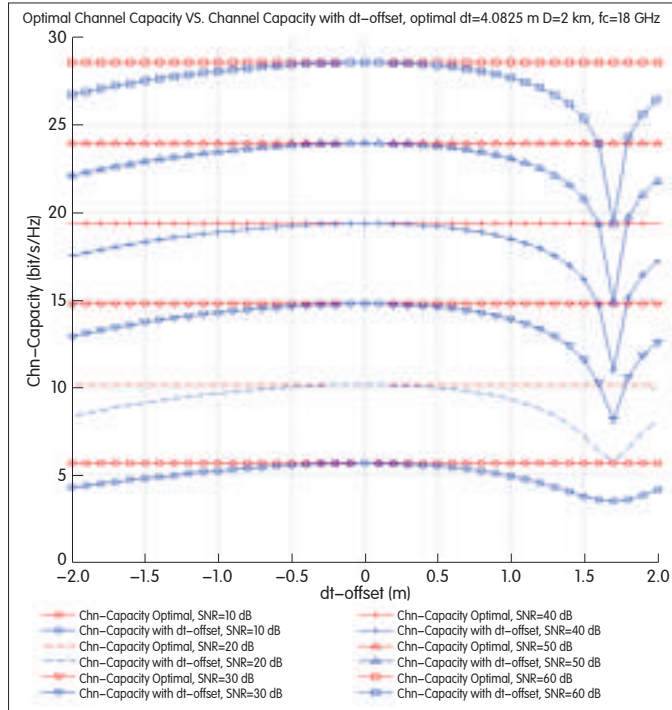
In Fig. 6, when  $f_c = 18$  GHz,  $D = 2000$  m,  $\theta_t = \theta_r = 0$ , then the optimal antenna spacing  $d_t = 4.0825$  m, which raises the question of practicality. If narrowing antenna spacing does not result in large loss of channel capacity away from the optimum channel capacity, it can be accepted.

The change of channel capacity within the negative  $d_t$ -offset is smaller than that within the positive  $d_t$ -offset, and the difference in channel capacity within the negative  $d_t$ -offset from the optimal channel capacity is acceptable when  $d_t$ -offset is in the range  $[-1.5$  m,  $0$  m]. As the SNR increases, this tendency becomes more apparent. This is



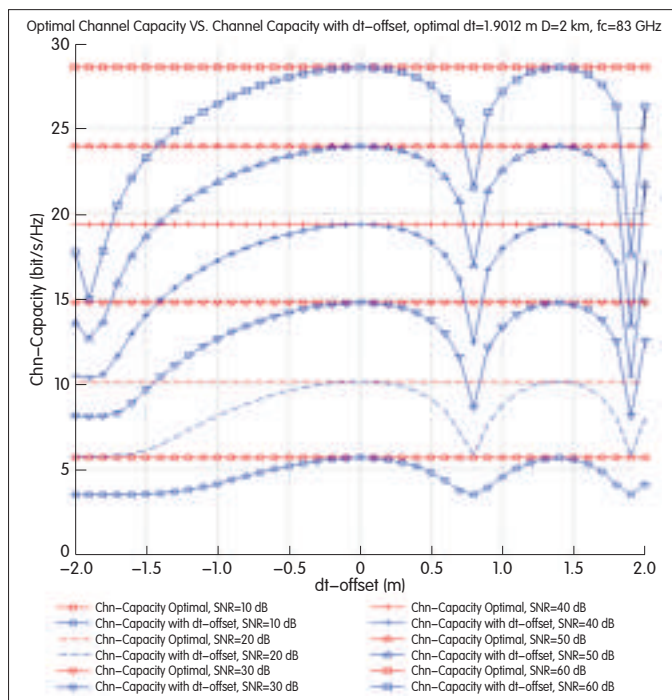
▲ Figure 5. Offset of antenna spacing  $d_t$  in  $2 \times 2$  LoS MIMO system.





▲ Figure 6. Channel capacity with optimal antenna parameters and antenna parameters within dt-offset in different SNR (10–60 dB, interval 10 dB).

conductive to antenna installation and only results in small loss from the optimal channel capacity. If the antenna spacing  $d_t$  is 2.5 m, the channel capacity may be acceptable.



▲ Figure 7. Channel capacity with optimal antenna parameters and antenna parameters within dt-offset in different SNR (10–60 dB, 10 dB).

Fig. 7 shows an enhancement on Fig. 6. As carrier frequency increases, the required antenna spacing to achieve meaningful MIMO gain decreases, and the range of accepted dt-offset is smaller. The dt-offset range  $[-1.5 \text{ m}, 0 \text{ m}]$  at 18 GHz may be acceptable, but at 83 GHz, only  $[-0.5 \text{ m}, 0 \text{ m}]$  is acceptable.

Another case of unequal  $d_t$  and  $d_r$  is also interesting; however, it will not be discussed here. Some conclusions can be drawn with similar analysis to the above.

### 3.3 Tx Antenna and Rx Antenna not on the Same Horizon

Without loss of generality, if we assume that the angle between horizontal line and line from  $\text{Tx}_1$  to  $\text{Rx}_1$  is  $\gamma$ , and  $d_t = d_r$  (Fig. 8), then

$$\begin{aligned} d_{11} &= \frac{D}{\cos \gamma} & d_{21} &= \sqrt{D^2 + (d_r + D \tan \gamma)^2} \\ d_{22} &= \frac{D}{\cos \gamma} & d_{12} &= \sqrt{D^2 + (d_r - D \tan \gamma)^2} \end{aligned} \quad (15)$$

Substituting (15) into (5) and (3), we can obtain channel capacity within angle  $\gamma$ .

We use a simulation to show how this factor influences channel capacity. Fig. 9 shows a comparison of optimal channel capacity and channel capacity within angle  $\gamma$  (range  $[-60^\circ, 60^\circ]$ , interval  $5^\circ$ ) when frequency  $f_c = 18 \text{ GHz}$ ,  $D = 2000 \text{ m}$ ,  $\theta_r = \theta_t = 0$ , and  $d_t = d_r$ , which satisfies (10).

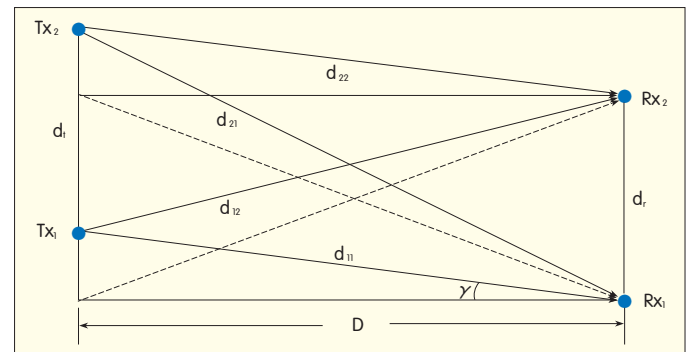
The channel capacity within angle  $\gamma [-25^\circ, 25^\circ]$  is almost optimal. This means that even if the ground is hilly, the system can work normally.

### 3.4 Tx Antenna Array and Rx Antenna Array not on the Same Plane

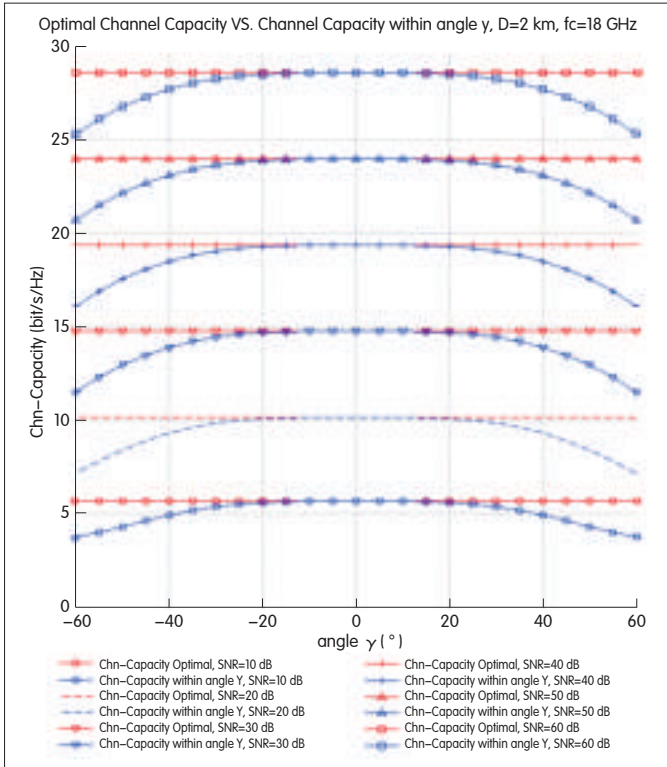
Without loss of generality, we can assume that

- the angle between the  $xz$ -plane and line from  $\text{Tx}_1$  to  $\text{Tx}_2$  is  $\phi_t$ ,
- the angle between the projection of line  $\text{Tx}_1$  to  $\text{Tx}_2$  on the  $xz$ -plane and  $z$ -axis is  $\theta_t$ ,
- the angle between the line from  $\text{Rx}_1$  to  $\text{Rx}_2$  and  $z$ -axis is  $\theta_r$ ,
- $d_t = d_r$ ,
- mast length is  $L$  (Fig. 10).

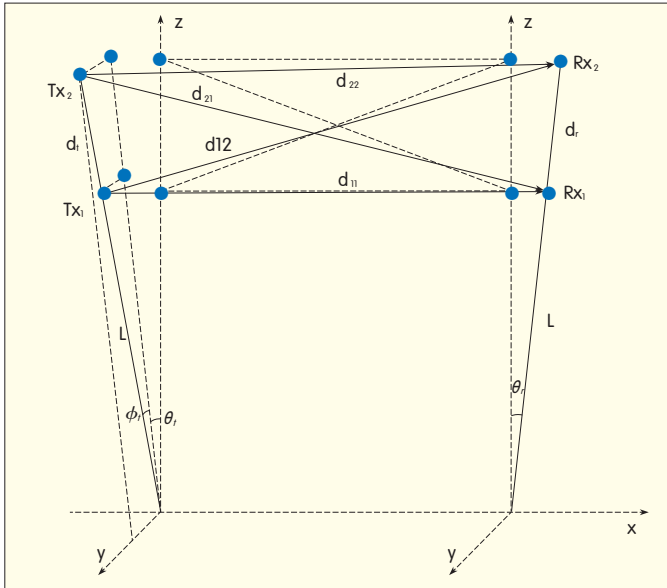
Then, the coordinate of  $\text{Tx}_1$  is  $(x_{1x1}, y_{1x1}, z_{1x1}) = (L \cos \phi_t \sin \theta_t, L \sin \phi_t, L \cos \phi_t \cos \theta_t)$   
the coordinate of  $\text{Tx}_2$  is



▲ Figure 8. Tx antenna and Rx antenna are not on the same horizon in  $2 \times 2$  LoS MIMO system.



▲ Figure 9. Channel capacity with optimal antenna parameters and antenna parameters with angle  $\gamma$ .



▲ Figure 10. Tx antenna array and Rx antenna array are not on the same plane in a 2 x 2 LoS MIMO system.

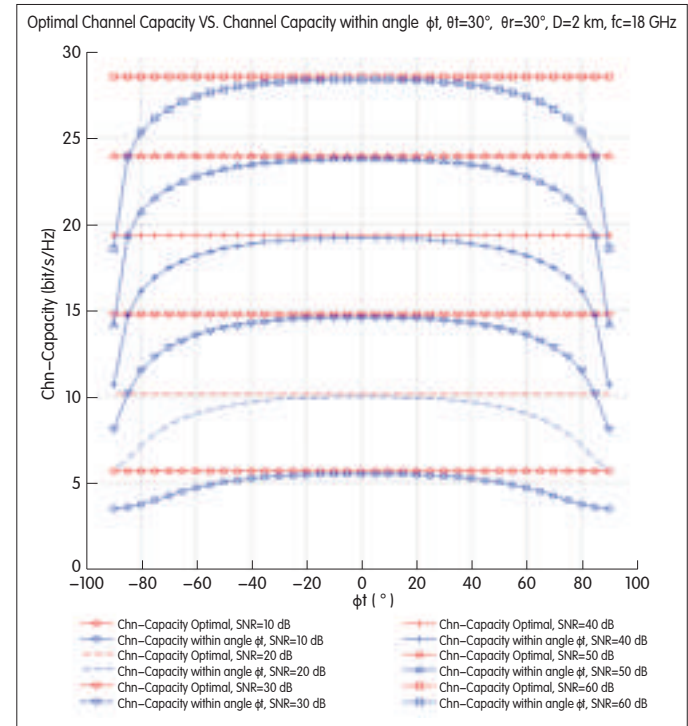
$(x_{tx2}, y_{tx2}, z_{tx2}) = ((L + d_t) \cos \phi_t \sin \theta_t, (L + d_t) \sin \phi_t, (L + d_t) \cos \phi_t \cos \theta_t)$   
 the coordinate of Rx<sub>1</sub> is  $(x_{rx1}, y_{rx1}, z_{rx1}) = (D + L \sin \theta_r, 0, L \cos \theta_r)$   
 the coordinate of Rx<sub>2</sub> is  $(x_{rx2}, y_{rx2}, z_{rx2}) = (D + (L + d_r) \sin \theta_r, 0,$

$(L + d_r) \cos \theta_r)$   
 and

$$\begin{aligned} d_{11} &= \sqrt{(x_{tx1} - x_{rx1})^2 + (y_{tx1} - y_{rx1})^2 + (z_{tx1} - z_{rx1})^2} \\ d_{12} &= \sqrt{(x_{tx1} - x_{rx2})^2 + (y_{tx1} - y_{rx2})^2 + (z_{tx1} - z_{rx2})^2} \\ d_{21} &= \sqrt{(x_{tx2} - x_{rx1})^2 + (y_{tx2} - y_{rx1})^2 + (z_{tx2} - z_{rx1})^2} \\ d_{22} &= \sqrt{(x_{tx2} - x_{rx2})^2 + (y_{tx2} - y_{rx2})^2 + (z_{tx2} - z_{rx2})^2} \end{aligned} \quad (16)$$

Substituting (16) into (5) and (3), we obtain the channel capacity within the factor of which Tx antenna array and Rx antenna array are not in the same plane.

We use a simulation to show how this factor influences channel capacity. Fig. 11 shows a comparison of optimal



▲ Figure 11. Channel capacity with optimal antenna parameters and antenna parameters with angle  $\phi_t$ .

channel capacity and channel capacity within angle  $\phi_t$  (range is  $[-90^\circ, 90^\circ]$ , interval is  $5^\circ$ ) when frequency  $f_c = 18$  GHz,  $D = 2000$  m,  $\theta_t = \theta_r = 30^\circ$ ,  $L = 15$  m, and  $d_t = d_r$ , which satisfies (10).

Channel capacity is not very sensitive to the angle between the mast of the antenna array and z-axis, and this makes it easier to install the mast of the antenna array. The channel capacity within angle  $\phi_t$   $[-25^\circ, 25^\circ]$  is almost optimal. This means that even if the ground is flat, the system can work normally, which is useful for practical antenna installation.

## 4 Conclusion

In this paper, we have discussed the principle of LoS MIMO and factors that affect LoS MIMO channel capacity. LoS MIMO performs well in specific scenarios and improves

transmission channel capacity. The simulation results show that LoS MIMO is not very sensitive to antenna parameters, and as long as these parameters do not differ too much from the optimal parameters, then there is only small loss of channel capacity, which is acceptable. This is useful for reducing costs in practical antenna installation without limiting performance.

LoS MIMO is relatively simple and easy to implement and should have good prospects.

#### References

- [1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005, p. 294.
- [2] E. Telatar, "Capacity of multi-antenna Gaussian channels," in *European Trans. Telecommunications*, vol. 10, no. 6, Dec 1999, pp. 585–595.
- [3] I. Sarris and A.R. Nix, "Maximum MIMO Capacity in Line-of-Sight Sarris," in *Proc. 5th Int. Conf. Information, Communications and Signal Processing (ICIS '05)*, Bangkok, Thailand, pp. 1236–1240.
- [4] T. Ingason and L. Haonan, *Line-of-Sight MIMO for Microwave Links Adaptive Dual Polarized and Spatially Separated Systems*, MSc Thesis in communications engineering, Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden, July 2009, pp. 27–30.

Manuscript received: September 7, 2012

#### Biographies

**Xianwei Gong** (gong.xianwei@zte.com.cn) received his MS degree in Computer Software and Theory from Harbin Engineering University in 2006. He has been as a member of the wireless technology advance research team at ZTE since 2008. His research involves wireless communication, MIMO system, error control coding, adaptive algorithm.

**Zhifeng Yuan** (yuan.zhifeng@zte.com.cn) received MS degree in signal and information processing from Nanjing University of Post and Telecommunications in 2005. He has been as a member of the wireless technology advance research team at ZTE since 2006. His research interests include wireless communication, MIMO systems, information theory, error control coding, adaptive algorithm, and high-speed VLSI design.

**Jun Xu** (xu.jun2@zte.com.cn) received his MS degree in signal and information processing from Nanjing University of Post and Telecommunications in 2003. He has been as a member of the wireless technology advance research team at ZTE since 2003. His research interests include error control coding, modulation, and MIMO systems.

**Liujun Hu** (hu.liujun@zte.com.cn) received his MS degree in electrical engineering from Harbin Engineering University in 1999. He has almost 13 years of experience in the telecom field. He has extensive experience in baseband signal processing and cellular network planning, especially with wireless system architecture design and key algorithms development.

## Roundup

### ZTE Communications Guidelines for Authors

#### Remit of Journal

ZTE Communications publishes original theoretical papers, research findings, and surveys on a broad range of communications topics, including communications and information system design, optical fiber and electro-optical engineering, microwave technology, radio wave propagation, antenna engineering, electromagnetics, signal and image processing, and power engineering. The journal is designed to be an integrated forum for university academics and industry researchers from around the world.

#### Manuscript Preparation

Manuscripts must be typed in English and submitted electronically in MS Word (or compatible) format. The word length is approximately 3000 to 6000, and no more than 6 figures or tables should be included. Authors are requested to submit mathematical material and graphics in an editable format.

#### Abstract and Keywords

Each manuscript must include an abstract of approximately 150 words written as a single paragraph. The abstract should not include mathematics or references and should not be repeated verbatim in the introduction. The abstract should be a self-contained overview of the aims, methods, experimental results, and significance of research outlined in the paper. Five carefully chosen keywords must be provided with the abstract.

#### References

Manuscripts must be referenced at a level that conforms to international academic standards. All references must be numbered sequentially in-text and listed in corresponding order at the end of the paper. References that are not cited in-text should not be included in the reference list. References must be complete and formatted according to IEEE Editorial Style [www.ieee.org/documents/stylemanual.pdf](http://www.ieee.org/documents/stylemanual.pdf). A minimum of 10 references should be provided. Footnotes should be avoided or kept to a minimum.

#### Copyright and Declaration

Authors are responsible for obtaining permission to reproduce any material for which they do not hold copyright. Permission to reproduce any part of this publication for commercial use must be obtained in advance from the editorial office of ZTE Communications. Authors agree that a) the manuscript is a product of research conducted by themselves and the stated co-authors, b) the manuscript has not been published elsewhere in its submitted form, c) the manuscript is not currently being considered for publication elsewhere. If the paper is an adaptation of a speech or presentation, acknowledgement of this is required within the paper. The number of co-authors should not exceed five.

#### Content and Structure

ZTE Communications seeks to publish original content that may build on existing literature in any field of communications. Authors should not dedicate a disproportionate amount of a paper to fundamental background, historical overviews, or chronologies that may be sufficiently dealt with by references. Authors are also requested to avoid the overuse of bullet points when structuring papers. The conclusion should include a commentary on the significance/future implications of the research as well as an overview of the material presented.

#### Peer-Review and Editing

All manuscripts will be subject to a two-stage anonymous peer review as well as copyediting, and formatting. Authors may be asked to revise parts of a manuscript prior to publication.

#### Biographical Information

All authors are requested to provide a brief biography (approx. 150 words) that includes email address, educational background, career experience, research interests, awards, and publications.

#### Acknowledgements and Funding

A manuscript based on funded research must clearly state the program name, funding body, and grant number. Individuals who contributed to the manuscript should be acknowledged in a brief statement.

# Terabit Superchannel Transmission: A Nyquist-WDM Approach

Hung-Chang Chien<sup>1</sup>, Jianjun Yu<sup>2</sup>, Zhensheng Jia<sup>2</sup>, and Ze Dong<sup>1</sup>

(1. Optics Lab, ZTE USA Inc., Morristown, NJ 07960, USA;

2. ZTE Corporation, Shenzhen 518057, P. R. China)

## Abstract

In this work, we focus on enhancing the network reach in terabit superchannel transmission by using a noise-suppressed Nyquist wavelength division multiplexing (NS-N-WDM) technique for polarization multiplexing quadrature phase-shift keying (PM-QPSK) subchannels at different symbol-rate-to-subchannel-spacing ratios up to 1.28. For the first time, we experimentally compare the transmission reach of this emerging technique with that of no-guard-interval coherent optical orthogonal frequency-division multiplexing (NGI-CO-OFDM) on the same testbed. At BER of  $2 \times 10^{-3}$  and 100 Gbit/s per channel, an NGI-CO-OFDM terabit superchannel can transmit over a maximum of 3200 km SMF-28 with EDFA-only amplification, and an NS-N-WDM terabit superchannel can transmit over a maximum of 2800 km SMF-28 with EDFA-only amplification. Assuming different coding gain,  $11 \times 112$  Gbit/s per channel with hard-decision (HD) forward-error correction (FEC) and  $11 \times 128$  Gbit/s per channel NS-N-WDM transmission with soft-decision (SD) FEC can be achieved over a maximum of 2100 km and 2170 km, respectively. These are almost equal and were achieved using digital noise filtering and one-bit maximum likelihood sequence estimation (MLSE) at the receiver DSP. Characteristics including the back-to-back (BTB) curves, the ADC bandwidth requirement, and the tolerance to unequal subchannel power of an NS-N-WDM superchannel were also evaluated.

## Keywords

optical OFDM; Nyquist WDM; MLSE

## 1 Introduction

Rapid growth in the amount of global IP traffic has been caused by the emergence of bandwidth-demanding network services. This has accelerated the commercialization of coherent 100G transport technology based on the widely recognized polarization multiplexing quadrature phase-shift keying (PM-QPSK) modulation format [1].

Technology for transmitting beyond 100G is being studied intensively and is based on increasing the symbol rate [2], increasing the number of subcarriers [3], or increasing the modulation levels [4], [5]. The predominant multicarrier multiplexing and transmission proposals for next-generation terabit optical transport are no-guard-interval coherent optical orthogonal frequency-division multiplexing (NGI-CO-OFDM) [6]–[8] and Nyquist wavelength-division multiplexing (N-WDM) [9], [10]. The former theoretically allows adjacent orthogonal wavelength channels to partly overlap without any crosstalk penalty, and the latter relies on pulse shaping and spectral filtering to optimize the trade-off between interchannel interference (ICI) and intersymbol interference (ISI). In principle, both techniques have the same sensitivity and spectral efficiency; however, in implementation, there are several physical limitations that cause suboptimal performance. NGI-CO-OFDM requires analog-to-digital converters (ADC) with large bandwidth and high sampling rate at the receiver. We previously demonstrated NGI-CO-OFDM PM-QPSK superchannel transmission over 3200 km SMF-28 with an oversampling rate of 3.2 GSa/s [11]. Ideal N-WDM transmission requires digital-to-analog converters (DAC) with a high sampling rate of 55–65 GSa/s for raised-cosine (RC) pulse shaping of PM-QPSK signals at 112 Gbit/s and beyond. Unfortunately, this is not widely available yet. Today, most transmission is considered quasi N-WDM; an optical RC pulse can only be loosely approximated using regular fourth-order super-Gaussian or spectrally engineered narrowband filtering, and ISI and ICI penalties are induced. In [12] and [13], a noise-suppressed N-WDM (NS-N-WDM) demodulation technique was proposed. The technique involves digital noise filtering followed by short-memory maximum likelihood sequence estimation (MLSE). It pinpoints the imperfection of the linear equalizer in the presence of a strongly filtered N-WDM subchannel and significantly increases tolerance towards noise and crosstalk.

In this paper, we describe a series of experiments on NS-N-WDM terabit superchannel transmission and compare the results with those obtained in a previously reported NGI-CO-OFDM experiment [11]. The system parameters in all experiments include PM-QPSK modulation format, 25 Gbaud symbol rate, 25 GHz subchannel spacing, and identical frequency-locked multicarrier light source [11]. We found that NS-N-WDM at 100 Gbit/s was capable of 2800 km SMF-28 transmission with BER below  $2 \times 10^{-3}$  and 3200 km SMF-28 transmission with BER below  $4.9 \times 10^{-3}$ . NGI-CO-OFDM was capable of 3200 km SMF-28 transmission with BER of  $2 \times 10^{-3}$ , which is slightly better

This study is supported by National High Technology Research and Development Program of China (No. 2012AA011303).



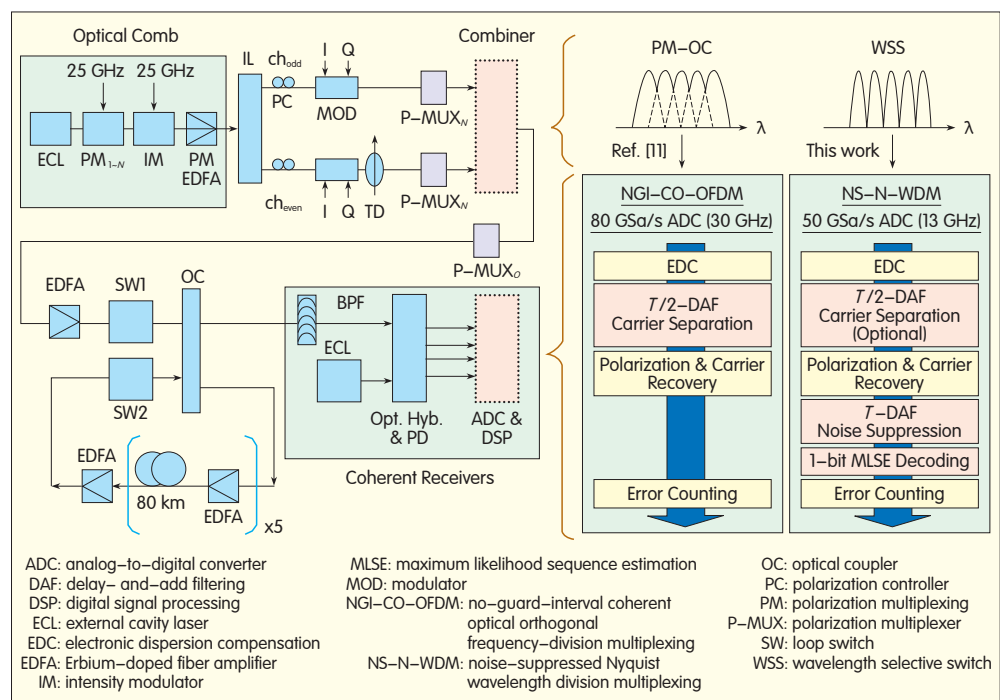
performance. For the first time, we experimentally measured and studied a 25 GHz spaced NS-N-WDM superchannel at 128 Gbit/s per channel and found that even though SD-FEC with higher coding gain was assumed, the maximum achievable transmission distance was similar to that of 112 Gbit/s per channel where HD-FEC was assumed. Transmission distance of 2100 km was achieved with 112 Gbit/s per channel and pre-FEC BER limit of  $3.8 \times 10^{-3}$ , and 2180 km was achieved with 128 Gbit/s per channel and pre-FEC BER of  $2 \times 10^{-2}$ .

Section 2 details the setups for NS-N-WDM and NGI-CO-OFDM signal generation and transmission experiments. Both setups are based on PM-QPSK at 100 Gbit/s per channel, and Fig. 1 shows the breakdown of and main differences between the digital signal processing (DSP) algorithms used in both setups. Section 3 focuses on NS-N-WDM transmission at symbol-rate-to-subchannel-spacing ratios of 1.12 and 1.28, which in principle cannot be realized in NGI-CO-OFDM transmission. We experimentally evaluated and studied symbol rate dependence, achievable distance, optimal launch power, and required ADC bandwidth for NS-N-WDM terabit superchannels.

## 2 Testbed Setups for NS-N-WDM and NGI-CO-OFDM Transmission

Fig. 1 shows the testbed setups for the NS-N-WDM transmission experiment and previous NGI-CO-OFDM transmission experiments for a clear comparison. Both setups have multicarrier light source, individual PM-QPSK modulation for even channels ( $CH_{\text{even}}$ ) and odd channels ( $CH_{\text{odd}}$ ), 400 km EDFA-only recirculating loop, and integrated optical front-end at the coherent receiver. However, NS-N-WDM requires optical spectral shaping to be done along with  $CH_{\text{even}}$  and  $CH_{\text{odd}}$  combining at the transmitter whereas NGI-CO-OFDM does not. Also, the requirements on ADC bandwidth and sampling rate for NGI-CO-OFDM transmission are much higher than that for NS-N-WDM transmission. Finally, to suppress linear crosstalk in the aggregated superchannel, the receiver DSP for NGI-CO-OFDM uses digital  $T/2$  delay-and-add filtering (DAF) for carrier separation ( $T$  is the symbol duration). In NS-N-WDM, linear crosstalk was suppressed by digital postfiltering ( $T$ -DAF) followed by MLSE with a short memory length of 1 bit. The

CW light from an external cavity laser (ECL) with a linewidth less than 100 kHz and output power of 14.5 dBm was modulated by a cascaded phase modulator (PM) and intensity modulator (IM).  $PM_1$  was driven by an RF clock at 25 GHz with a peak-to-peak voltage of 17 V, a half-wave voltage loss of 4 V, and insertion loss of 3.8 dB. When studying NGI-CO-OFDM multiterabit transmission, an additional phase modulator  $PM_2$  was previously used to generate more subcarriers, and this phase modulator was also driven by a high-level RF signal with synchronized 25 GHz clock [11]. After phase modulation, multiple coherent carriers spaced at 25 GHz were generated. Then, a rear IM driven by a synchronized RF clock at 25 GHz was connected to  $PM_2$ . This IM flattens the generated optical subcarriers to less than 2 dB. After that, the subcarriers were boosted by a polarization-maintaining Erbium-doped fiber amplifier (PM-EDFA) and then underwent  $CH_{\text{odd}}$  and  $CH_{\text{even}}$  separation by a 25/50 GHz optical interleaver (IL). The channel modulation involves an in-phase/quadrature (I/Q) modulator (MOD), a polarization controller (PC), polarization multiplexers (P-MUX), and optical combiner (OC). Each I/Q MOD was driven by two sets of 25 Gbit/s pseudorandom bit sequences (PRBSs) with word lengths of  $2^{11}-1$ , which contained two parallel Mach-Zehnder modulators, both biased at the null point and driven at full swing for zero-chirp phase modulation. For NGI-CO-OFDM,  $CH_{\text{odd}}$  and  $CH_{\text{even}}$  were combined by a polarization-maintaining optical coupler (PM-OC), and then the aggregated channel was polarization-multiplexed by P-MUX<sub>0</sub> at a time. For NS-N-WDM signal generation,  $CH_{\text{odd}}$  and  $CH_{\text{even}}$  were first individually polarization-multiplexed by P-MUX<sub>N</sub> and then



▲ Figure 1. Testbed setup and algorithms for NGI-CO-OFDM and NS-N-WDM transmission experiments.

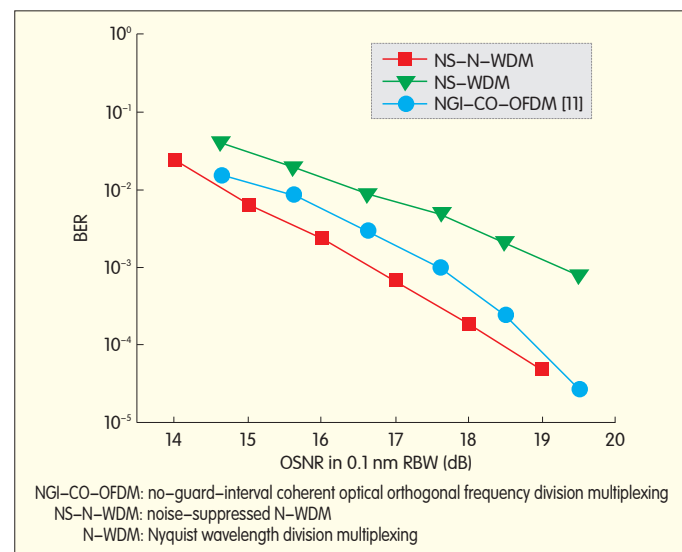
combined using a wavelength-selective switch (WSS) programmed to operate at 25 GHz interleaving mode for spectral shaping of all input PM-QPSK subchannels. P-MUX<sub>o</sub>/P-MUX<sub>N</sub> was used only for the NGI-CO-OFDM/NS-N-WDM experiment. An optical time delay (TD) was also used along with the optical path of even channels for symbol synchronization between Ch<sub>odd</sub> and Ch<sub>even</sub> tributaries. Such symbol synchronization is not necessarily required for NS-N-WDM. The loop consists of five spans of 80 km SMF-28 (with an average span loss of 16.3 dB and chromatic dispersion of 17 ps/km/nm), loop switches (SWs), optical coupler (OC), and EDFA-only amplification without optical dispersion compensation. For each span, dual-stage C-band EDFAs with mid-stage adjustable tilted filters were used to provide a flat gain. Another WSS placed in the loop was programmed to be a 9 nm wideband optical bandpass filter and was intended to block the accumulated noise peak occurring in the 1530 to 1540 nm region. At the receiver, a tunable bandpass filter (BPF) with 3 dB bandwidth of 0.4 nm was used to select the measured subchannel. The optical front-end comprises an ECL with linewidth less than 100 kHz, which was used as the fiber laser local oscillator (LO), and a polarization-diverse 90-degree optical hybrid (opt. hybrid), which was used to realize polarization and phase-diverse coherent detection of the LO and received optical signal before balanced photon detection (PD). For NGI-CO-OFDM, the ADC was operated at 80 GSa/s with a 30 GHz analog bandwidth. For NS-N-WDM, this was reduced to 50 GSa/s with a low 13 GHz bandwidth. The DSP of the NGI-CO-OFDM channel involves first extracting the clock by using "square and filter" method and then resampling the signal at twice the symbol rate based on the recovered clock. Second, a  $T/2$ -spaced time-domain finite impulse response (FIR) filter was used for electronic dispersion compensation (EDC). Third, subcarriers were separated using  $T/2$  DAF. Finally, classic constant modulus algorithm (CMA) and 21-tap,  $T/2$ -spaced adaptive FIR filters were used for polarization recovery. Carrier recovery was also performed. This included frequency offset estimation by fast Fourier transform and carrier phase recovery by fourth-power Viterbi-Viterbi algorithm. Enhanced DSP for NS-N-WDM demodulation comprised all the algorithms mentioned except the  $T/2$  DAF algorithm. T-DAF was performed after polarization and carrier recovery to suppress undesirable noise and increased linear crosstalk caused by linear equalizers in the presence of aggressive channel filtering. Such T-DAF also makes possible the use of MLSE with a short memory length of 1 bit [12], [13].

### 3 Results and Discussion

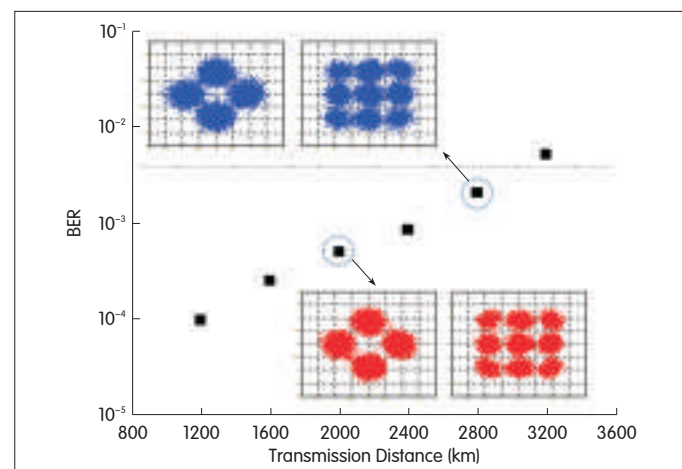
#### 3.1 NS-N-WDM versus NGI-CO-OFDM Transmission

Fig. 2 shows back-to-back (BTB) BER performance of NS-N-WDM, NGI-CO-OFDM, and N-WDM. For typical quasi N-WDM, (which has the same transmitter setup as the NS-N-WDM case but without  $T$ -DAF and MLSE at the

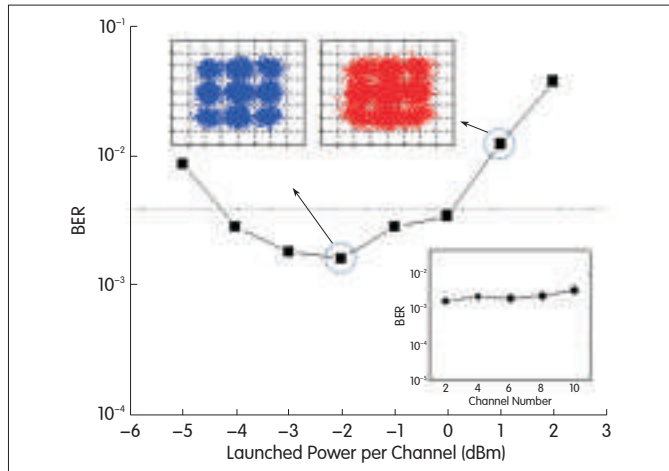
receiver DSP), we obtained a large required OSNR of 17.9 dB at BER of  $3.8 \times 10^{-3}$ . This can be significantly reduced to 15.6 dB by NS-N-WDM demodulation. In addition, NS-N-WDM (11 subchannels) had 0.8 dB less required OSNR than NGI-CO-OFDM (21 subchannels) at  $3.8 \times 10^{-3}$  BER. Fig. 3 shows BER performance as a function of transmission distance for the NS-N-WDM experiment. The received BER and delivered OSNR at 2800 km is  $2 \times 10^{-3}$  and 16.3 dB, respectively. The received BER and delivered OSNR at 3200 km is  $4.9 \times 10^{-3}$  and 14.8 dB, respectively. The insets in Fig. 3 show constellation measured at 2000 km and 2800 km and show how the received QPSK signal was shaped into a 9QAM-like signal after  $T$ -DAF. Fig. 4 shows BER versus launch power per channel over 2800 km transmission. For BER of less than  $3.8 \times 10^{-3}$ , an input dynamic range of around 4.3 dB was achieved where an



▲ Figure 2. BTB BER curves of NS-N-WDM, N-WDM, and NGI-CO-OFDM with 25 GHz-spaced 100 Gbit/s PM-QPSK subchannels.



▲ Figure 3. Transmission capability of 11 x 100 Gbit/s NS-N-WDM superchannel.



▲ Figure 4. BER performance as a function of launch power per NS-N-WDM subchannel over 2800 km SMF-28.

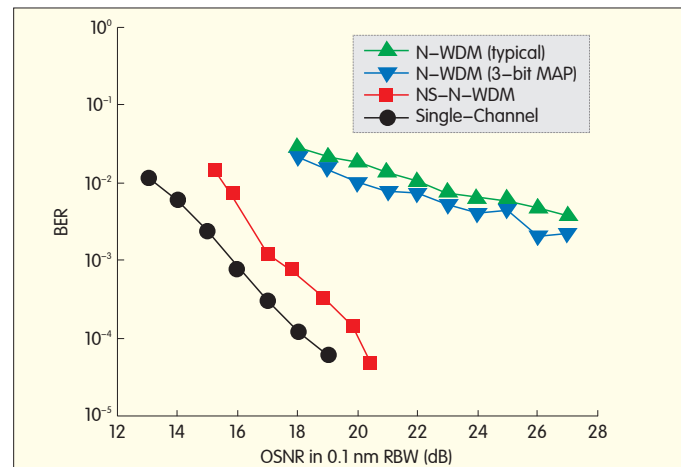
optimized operation point was found to be 2 dBm. Fig. 4 (upper left inset) shows the received 9QAM-like constellations at -2 dBm, and Fig. 4 (upper right inset) shows the received 9QAM-like constellations at 1 dBm launch power. The lower inset shows equal BER performance for five even NS-N-WDM subchannels at 2 dBm launch power.

### 3.2 25 GHz-Spaced NS-N-WDM Transmission at 112 Gbit/s per Channel

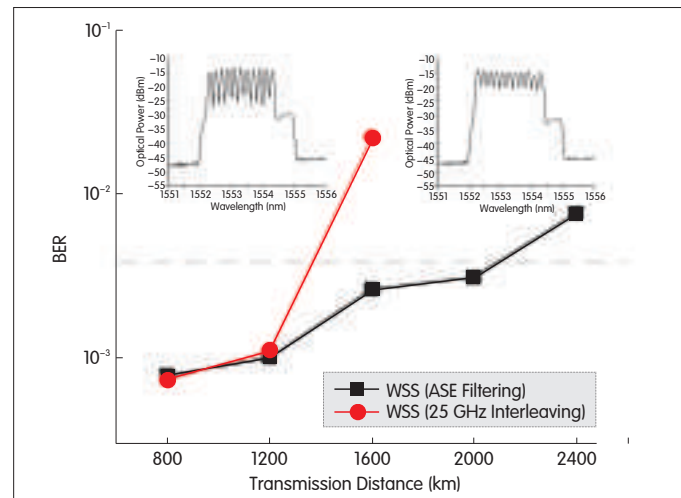
A 1T NS-N-WDM superchannel at 100 Gbit/s (25 GSa/s) per subchannel can be transmitted over 2800 km with a BER well below the pre-FEC limit. It is therefore interesting to determine the capability of the NS-N-WDM receiver in the presence of a PM-QPSK subchannel at the true OTU-4 rate of 112 Gbit/s (28 GSa/s) on a 25 GHz grid. Rather than using an optical comb as the light source, we use 11 independent ECLs. The PRBS rate was increased to 28 GSa/s without symbol alignment between  $Ch_{\text{odd}}$  and  $Ch_{\text{even}}$ , and the rest of the setup was unchanged. Fig. 5 shows BTB BER performance for N-WDM using different demodulation schemes with and without adjacent subchannels. If adjacent subchannels are turned off when  $BER = 3.8 \times 10^{-3}$ , the single PM-QPSK subchannel at 112 Gbit/s had a minimal OSNR of 14.4 dB with regular DSP. The typical N-WDM had a large 12.6 dB OSNR penalty mainly because of linear crosstalk. Although this penalty may be reduced by 3 dB by introducing three-bit maximum a-posteriori (MAP) estimation at the receiver, the required 24 dB OSNR was still too high for long-haul transmission. However, by using the NS-N-WDM receiver technique with T-DAF and 1 bit MLSE, the OSNR penalty drops significantly to 1.8 dB (16.2 dB required OSNR). Unlike with the prior NS-N-WDM subchannel at 25 GSa/s, increasing the symbol rate to 28 GSa/s only gives 0.6 dB OSNR penalty. Fig. 6 shows BER as a function of transmission distance when WSS in the loop functions as an ASE filter and a 25 GHz interleaver. After 2000 km transmission with WSS ASE filtering, BER of  $3.8 \times 10^{-3}$  can be achieved with received OSNR of 17.6 dB. However, when WSS operates in

interleaving mode as a result of cascaded filtering in the transmission path, only 1200 km is achievable. This means that the bandwidth-narrowing effect of a series of optical filters is significant, and the  $11 \times 112$  Gbit/s NS-N-WDM superchannel can pass through three 25 GHz in-line optical add-drop filters at most. Fig. 6 (left inset) shows the optical spectra at 1600 km when the WSS performs ASE filtering, and Fig. 6 (right inset) shows the optical spectra at 1600 km when the WSS performs interleaving.

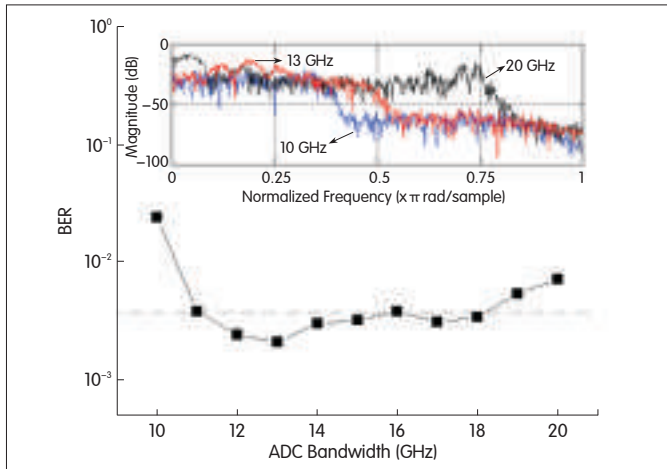
Next, we investigated the required ADC bandwidth for a single 112 Gbit/s NS-N-WDM subchannel by measuring BTB at OSNR around 17 dB. Fig. 7 shows that for BER below  $3.8 \times 10^{-3}$ , the ADC digital bandwidth ranges from 16 GHz down to 11 GHz. This has important cost implications for future hardware implementation and production. The inset shows the received RF spectra at 10 GHz, 13 GHz, and 20 GHz digital bandwidth prior to DSP. We further studied the affect of unequal power between adjacent 112 Gbit/s NS-N-WDM subchannels on BTB system performance at an



▲ Figure 5. BTB BER performance of N-WDM with different demodulation schemes at 112 Gbit/s per subchannel.



▲ Figure 6. Achievable transmission distance of  $11 \times 112$  Gbit/s NS-N-WDM superchannel with different WSS settings in the loop.

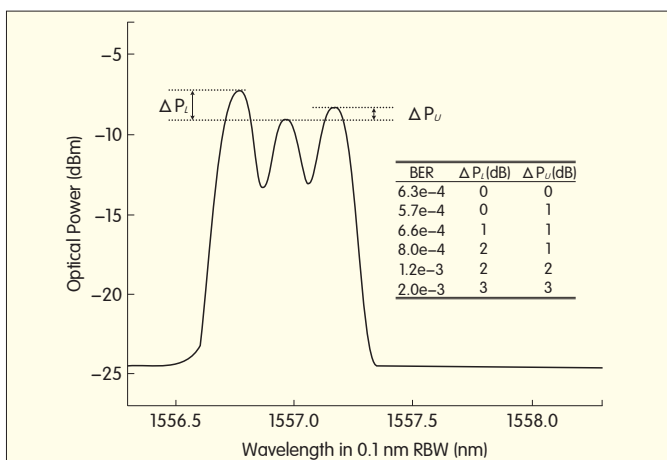


▲ Figure 7. BER performance as a function of ADC bandwidth.

OSNR of around 18 dB. For simplicity, we chose only upper, central, and lower subchannels and then varied the power increment of the upper and lower subchannels with respect to the central subchannel. Fig. 8 shows  $\Delta P_U$  and  $\Delta P_L$ , and Fig. 8 (inset) shows that the penalty is negligible when  $\Delta P_U$  and  $\Delta P_L$  are below 1 dB.

### 3.3 25 GHz-Spaced NS-N-WDM Transmission at 128 Gbit/s per Channel

Most new-generation coherent 100G PM-QPSK products come equipped with higher-coding-gain SD-FEC that allows a pre-FEC BER limit of only  $2 \times 10^{-2}$ . It is instructive to know whether such benefit applies to terabit superchannel transmission. Fig. 9 shows BTB NS-N-WDM BER curves with and without adjacent subchannels at 128 Gbit/s. Results at 112 Gbit/s (Fig. 5) are also included for comparison. At  $2 \times 10^{-2}$  BER, the required OSNR is 14.9 dB for a single channel and 15.8 dB for WDM. Then, we put the  $11 \times 128$  Gbit/s NS-N-WDM superchannel over the loop. Fig. 10 shows that below the SD pre-FEC limit, NS-N-WDM signals can potentially be delivered over 2170 km SMF-28. By

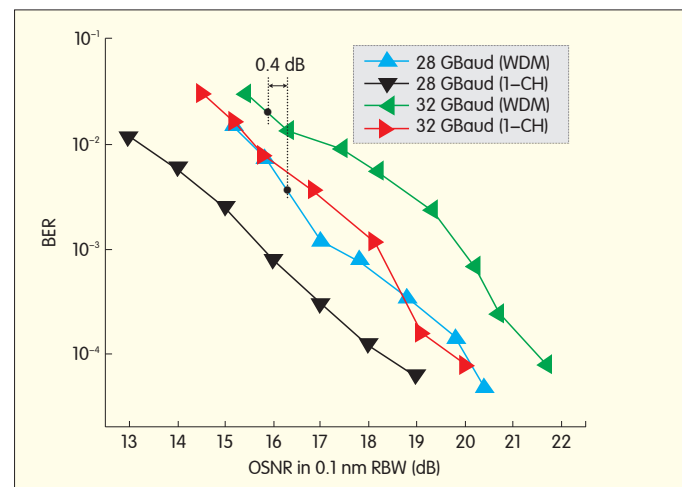


▲ Figure 8. Optical spectrum showing the unequal subchannel power and its impact on the central subchannel of interest.

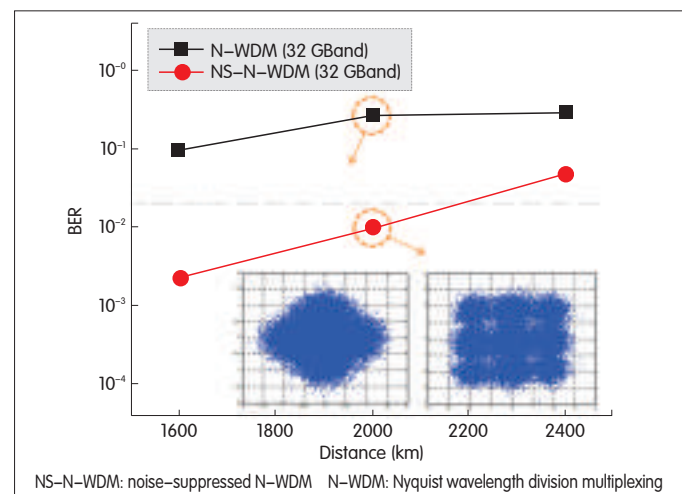
comparison, even with SD-FEC, typical N-WDM signals can barely reach 1600 km. Fig. 10 (inset) shows the received constellation diagrams after 2000 km SMF-28 with and without NS-N-WDM demodulation. In light of the results shown in Fig. 6, we found that  $11 \times 112$  Gbit/s per channel NS-N-WDM superchannel with HD and a  $11 \times 128$  Gbit/s per channel NS-N-WDM superchannel with SD-FEC are both capable of around 2100 km transmission. This may be because in the 32 Gbaud NS-N-WDM case, higher coding gain can only reduce the required OSNR by 0.4 dB. In the 28 Gbaud BTB measurement shown in Fig. 9, insufficient in-band signal bandwidth was found to be a major constraint.

## 4 Conclusion

We experimented on the transmission capability and characteristics of NS-N-WDM terabit superchannels comprising PM-QPSK subchannels operated at 100, 112,



▲ Figure 9. BTB BER of NS-N-WDM superchannel at 112 Gbit/s per channel and 128 Gbit/s per channel.



▲ Figure 10. Transmission capability of  $11 \times 128$  Gbit/s Nyquist-WDM superchannels.



128 Gbit/s per channel. All of these superchannels are on 25 GHz ITU–T grids. We compared the results with previously reported results by using no-guard-interval NGI–CO–OFDM on the same testbed. NS–N–WDM at 100 Gbit/s per channel was capable of 2800 km SMF–28 transmission with EDFA-only amplification. Although NGI–CO–OFDM exhibited a slightly better reach of 3200 km, its stringent requirement for high ADC bandwidth and sampling rate, precise symbol alignment and synchronization, and co-polarization between subchannels requires more engineering for implementation. In addition, the transmission distances for 112 Gbit/s per channel and 128 Gbit/s per channel NS–N–WDM superchannels were found to be similar, that is, 2100 km and 2170 km at pre-FEC BER limits of  $3.8 \times 10^{-3}$  and  $2 \times 10^{-2}$ , respectively. This implies that acquiring coding gain has no immediate benefit in Nyquist–WDM superchannel transmission if the in-band signal bandwidth is insufficient. In addition, larger, non-standardized grids can be used to reduce crosstalk in N–WDM transmission without DSP enhancement. However, the scope of this work is to explore the limits of terabit N–WDM transmission on the standardized 25 GHz ITU–T grid and bring about superior, constant net spectral efficiency for all 100G subchannels at different symbol rates. Although there is an increased penalty at higher symbol rate as a result of deteriorating noise and crosstalk, but this can be mitigated by using digital noise filtering and 1 bit maximum likelihood sequence estimation (MLSE) alongside typical receiver DSP. This has been demonstrated here for terabit superchannel transmission over 2000 km SMF–28.

### References

- [1] M. Birk, P. Gerard, R. Curto, L. E. Nelson, X. Zhou, and P. Magill, T. J. Schmidt, C. Malouin, B. Zhang, E. Ibragimov, S. Khatana, M. Glavanovic, R. Lo-and, R. Marcoccia, and R. Saunders, G. Nicholl, M. Nowell, and F. Forghieri, "Coherent 100 Gb/s PM–QPSK field trial," *IEEE Commun. Mag.*, vol. 48, no. 7, pp. 52–60.
- [2] C. Xie, G. Raybon, P. J. Winzer, "Transmission of mixed 224–Gb/s and 112–Gb/s PDM–QPSK at 50–GHz channel spacing over 1200–km dispersion-managed LEAF® spans and Three ROADMs," in *J. Lightwave Technol.*, vol. 30, no. 4, pp. 547–552.
- [3] J.–X. Cai, C. R. Davidson, A. Lucero, H. Zhang, D. G. Foursa, O. V. Sinkin, W. W. Patterson, A. N. Pilipetskii, G. Mohs, N. S. Bergano, "20 Tbit/s transmission over 6860 km with sub-Nyquist channel spacing," *J. Lightwave Technol.*, vol. 30, no. 4, pp. 651–657.
- [4] X. Zhou, J. Yu, M.–F. Huang, Y. Shao, T. Wang, L. Nelson, P. Magill, M. Birk, P. I. Borel, D. W. Peckham, R. Lingle, B. Zhu, "64–Tb/s, 8 b/s/Hz, PDM–36QAM transmission over 320 km using both pre- and post-transmission digital signal processing," *J. Lightwave Technol.*, vol. 29, no. 4, pp. 571–577.
- [5] A. H. Gnauck, P. J. Winzer, A. Konczykowska, F. Jorge, J.–Y. Dupuy, M. Riet, G. Charlet, B. Zhu, and D. W. Peckham, "Generation and transmission of 21.4–Gb/s PDM 64–QAM using a novel high-power DAC driving a single I/Q modulator," *J. Lightwave Technol.*, vol. 30, no. 4, pp. 532–536.
- [6] S. Chandrasekhar, X. Liu, "Experimental investigation on the performance of closely spaced multi-carrier PDM–QPSK with digital coherent detection," *Opt. Exp.*, vol. 17, no. 24, pp. 21350–21361.
- [7] B. Zhu, X. Liu, S. Chandrasekhar, D. W. Peckham, R. Lingle, Jr., "Ultra-long-haul transmission of 1.2–Tb/s multicarrier No-Guard-Interval CO–OFDM superchannel using ultra-large-area fiber," *IEEE Photon. Technol. Lett.*, vol. 22, no. 11, pp. 826–828, June 2010.
- [8] A. Sano, E. Yamada, H. Masuda, E. Yamazaki, T. Kobayashi, E. Yoshida, Y. Miyamoto, R. Kudo, K. Ishihara, and Y. Takatori, "No-Guard-Interval Coherent Optical OFDM for 100–Gb/s long-haul WDM transmission," *J. Lightwave Technol.*, vol. 27, no. 16, pp. 3705–3713.
- [9] G. Bosco, A. Carena, V. Curri, P. Poggiolini, F. Forghieri, "Performance limits of Nyquist–WDM and CO–OFDM in high-speed PM–QPSK systems," *IEEE Photon. Technol. Lett.*, vol. 22, no. 15, pp. 1129–1131.
- [10] M. Yan, Z. Tao, W. Yan, L. Li, T. Hoshida, J. C. Rasmussen, "Experimental comparison of No-Guard-Interval–OFDM and Nyquist–WDM superchannels,"

presented at the *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf. (OFC/NFOEC '12)*, Los Angeles, CA 2012, Paper OTh1B.2.

- [11] J. Yu, Z. Dong, and N. Chi, "1.96Tb/s (21×100Gb/s) OFDM Optical Signal Generation and Transmission over 3200–km Fiber," *IEEE Photon. Technol. Lett.*, vol. 23, no. 15, pp. 1061–1063.
- [12] J. Li, E. Tipsuwannakul, T. Eriksson, M. Karlsson, P. A. Andrekson, "Approaching Nyquist limit in WDM systems by low-complexity receiver-side Duobinary shaping," *J. Lightwave Technol.*, to be published.
- [13] J. Yu, Z. Dong, H.–C. Chien, Z. Jia, D. Huo, H. Yi, M. Li, Z. Ren, N. Lu, L. Xie, K. Liu, X. Zhang, Y. Xia, Y. Cai, M. Gunkel, P. Wagner, H. Mayer, A. Schippel, "Field trial Nyquist–WDM transmission of 8×216.4Gb/s PDM–CSRZ–QPSK exceeding 4b/s/Hz spectral efficiency," presented at the *Opt. Fiber Commun. Conf./Nat. Fiber Opt. Eng. Conf. (OFC/NFOEC '12)*, Los Angeles, CA 2012, Paper PDP5D.3.

Manuscript received: April 1, 2012

### Biographies

**Hung–Chang Chien** (chien.hungchang@zteusa.com) received his BS and MS degrees in electrical engineering from National Cheng Cheng University, Taiwan, in 1999 and 2001. He received his PhD degree in electro-optical engineering from National Chiao Tung University, Taiwan, in 2006. From 2007 to 2011, he was a research engineer in the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA. He is currently a senior member of technical staff at Optics Labs, ZTE USA Inc., Morristown, NJ. Dr. Chien has authored and co-authored more than 100 journal papers and conference proceedings and holds one U.S. patent with nine others pending in the fields of coherent DWDM optical transmission, microwave photonics, and passive optical networks.

**Jianjun Yu** (yu.jianjun@zteusa.com) received his PhD degree in electrical engineering from Beijing University of Posts and Telecommunications in 1999. From June 1999 to January 2001, he was an assistant research professor at the Research Center COM, Technical University of Denmark. From February 2001 to December 2002, he was a member of the technical staff at Lucent Technologies and Agere Systems, Murray Hill, NJ. He joined Georgia Institute of Technology in January 2003 as a research faculty member and director of the Optical Network Laboratory. From November 2005 to February 2010, he was a senior member of technical staff at NEC Laboratories America, Princeton, NJ. Currently, he works for ZTE Corporation as the chief scientist on high-speed optical transmission and director of optics labs in North America. He is also a chair professor at Fudan University and adjunct professor and PhD supervisor at the Georgia Institute of Technology, Beijing University of Posts and Telecommunications, and Hunan University. He has authored more than 200 papers for prestigious journals and conferences. Dr. Yu holds 11 U.S. patents with 30 others pending. He is a fellow of the Optical Society of America. He is Editor-in-chief of Recent Patents on Engineering and an associate editor for the *Journal of Lightwave Technology* and *Journal of Optical Communications and Networking*. Dr. Yu was a technical committee member at IEEE LEOS from 2005 to 2007 and a technical committee member of OFC from 2009 to 2011.

**Zhensheng Jia** (zhensheng.jia@zteusa.com) received his BE and MSE degrees in physical electronics and optoelectronics from Tsinghua University, Beijing, in 1999 and 2002. He received his PhD degree from Georgia Institute of Technology, Atlanta, in 2008. From 2002 to 2004, he worked as a research engineer on ultralong-haul optical links and backbone networks at the China Telecom Beijing Research Institute (CTBRI). From 2008 to 2011, Dr. Jia was a senior research scientist at Telecordia Technologies and worked on architecture of core optical networks and RF photonic signal processing. Currently, he is working on ultralong-haul optical transmission systems and optical transport architecture in the Optical Labs of ZTE USA.

Dr. Jia has author or co-authored more than 100 peer-reviewed journal articles and conference papers. He is also an active reviewer for many technical publications. In 2007, he won the IEEE/LEOS Graduate Students Fellowship Award, and in 2008 he won the PSC Bor–Uei Chen Memorial Scholarship Award. In 2007, he won the 2011 Telcordia CEO Award.

**Ze Dong** (dong.ze@zteusa.com) received his BS degree in electronic information science and technology from Hunan Normal University, Changsha, in 2006. He received his PhD degree in electrical engineering from Hunan University, Changsha, in 2011. From 2010 to 2011, he was a visiting scholar at Georgia Institute of Technology, Atlanta. He is currently a postdoctoral fellow in the School of Electrical and Computer Engineering, Georgia Institute of Technology. His research interests include broadband optical communication and optical coherent communications. Dr. Dong has authored and co-authored more than 35 journal papers and conference proceedings.

# Parallel Web Mining System Based on Cloud Platform

Shengmei Luo<sup>1</sup>, Qing He<sup>2</sup>, Lixia Liu<sup>1</sup>, Xiang Ao<sup>2,3</sup>, Ning Li<sup>2,3</sup>, Fuzhen Zhuang<sup>2</sup>

(1. Pre-Research department of ZTE, Nanjing, 210012, China;

2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China;

3. Graduate University of Chinese Academy of Sciences, Beijing, 100190, China)

## Abstract

Traditional machine-learning algorithms are struggling to handle the exceedingly large amount of data being generated by the internet. In real-world applications, there is an urgent need for machine-learning algorithms to be able to handle large-scale, high-dimensional text data. Cloud computing involves the delivery of computing and storage as a service to a heterogeneous community of recipients. Recently, it has aroused much interest in industry and academia. Most previous works on cloud platforms only focus on the parallel algorithms for structured data. In this paper, we focus on the parallel implementation of web-mining algorithms and develop a parallel web-mining system that includes parallel web crawler; parallel text extract, transform and load (ETL) and modeling; and parallel text mining and application subsystems. The complete system enables variable real-world web-mining applications for mass data.

## Keywords

web mining; large scale; high volume; high dimension; cloud computing

## 1 Introduction

Hadoop 1.x is a large-scale data processing platform for cloud computing. At the core of Hadoop is MapReduce, which provides users with a new parallel programming mode in the distributed environment [1]. It allows users to benefit from the advanced features of distributed computing without the need to do any programming to coordinate tasks in the distributed environment. Recently, many machine-learning algorithms have been paralleled

based on MapReduce [2]–[8]. Chu et al. developed a broadly applicable parallel programming method that can be easily applied to many different learning algorithms [2]. They have shown that algorithms fitting the statistical query model can be written in “summation form,” which allows them to be easily parallelized on multicore computers. He et al. proposed several parallel-classification algorithms, including k-nearest neighbors, naive Bayesian model, and decision tree for structured data [3]. Experimental results show the efficiency of the proposed parallel methods for handling large data sets. Zhao et al. proposed a parallel k-means clustering algorithm based on MapReduce that is scalable and can efficiently process large data sets on commercially available hardware [4]. He et al. provided a parallel incremental learning algorithm for ESVM (PIESVM) that can solve large-scale and online problems.

Previous works mainly focus on the parallel implementation of machine-learning algorithms for structured data. However, there is a large amount of unstructured data on the internet and only a few parallel text mining algorithms [9]–[10]. Elsayed et al. proposed a MapReduce algorithm for computing pair-wise document similarity in large document collections [10]. Experiments on a collection of approximately 900,000 newswire articles show that the proposed algorithm’s running time and space grows linearly with the number of articles processed. Zhang et al. introduced a novel probabilistic generative model MicroBlog-Latent Dirichlet Allocation (MB-LDA), which takes both contactor relevance and document relevance into consideration in order to improve topic mining in microblogs. They also developed distributed MB-LDA in the MapReduce framework in order to process large-scale microblogs with high scalability. However, existing learning algorithms are still very impractical for web mining simply because of the explosion in the amount of information on the internet. In this work, we concentrate on various web-mining algorithms and propose a parallel algorithm designed using MapReduce. Finally, we develop a parallel web-mining system that includes web crawler, webpage parsing, text data preprocessing, and text data mining.

In section 2, we introduce techniques related to our proposal. In section 3, we introduce the system architecture and its implementation. In section 4, we give experimental results. Section 5 concludes the paper.

## 2 Related Work

Hadoop allows programmers to easily develop and run mass-data-processing applications. The core of Hadoop is

This work is supported by the National Natural Science Foundation of China (No. 61175052, 60975039, 61203297, 60933004, 61035003), National High-tech R&D Program of China (863 Program) (No.2012AA011003).

mainly Hadoop distributed file system (HDFS), MapReduce, and HBase, which can be used as a data source for a MapReduce job.

### 2.1 Hadoop Distributed File System

HDFS is inspired by Google File System, which uses large-scale clusters to store large amounts of data [11]. Despite being similar to the existing GFS, HDFS has some differences. First, the structure of HDFS ensures that it is highly tolerant of faults compared with the low fault-tolerance of GFS. Furthermore, HDFS can be deployed on low-cost hardware. The third advantage of HDFS is that it provides high throughput for data-accessing applications whereas GFS does not. Therefore, HDFS is suitable for applications with large data sets. Fig. 1 shows the HDFS architecture. Data files are stored in blocks of the same size on clusters of DataNodes. Each block has replications to ensure fault tolerance. NameNode replicates all the file blocks and periodically receives heartbeat and block-report messages from DataNodes.

### 2.2 MapReduce

MapReduce provides a convenient programming mode and associated implementation for processing and generating large data sets using special  $\langle \text{key}, \text{value} \rangle$  pairs [1]. Specifically, users design a map function that processes  $\langle \text{key}, \text{value} \rangle$  pairs and generates a set of intermediate  $\langle \text{key}, \text{value} \rangle$  pairs, and a reduce function that merges all the intermediate values associated with the same intermediate key.

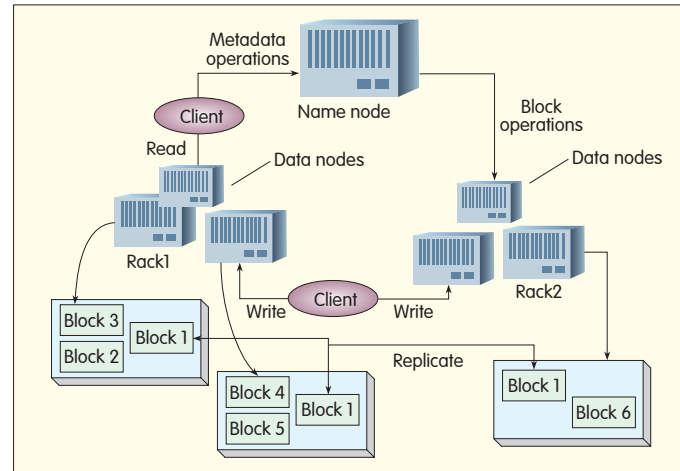
The main characteristic of MapReduce is it provides a simple but powerful interface for automatic parallelization and distribution of large-scale computation. The following is pseudocode for a simple word-count task. It illustrates the main idea of MapReduce:

```
map(Integer key, String value)
//key: offset
//value: document contents
for each word w in value: output(w, "1");
reduce(String key, Integer values):
//key: a word w
//values: a list of counts
int result = 0;
for each v in values:
result += v;
```

In the map function, we parse the input  $\langle \text{key}, \text{value} \rangle$  pairs into intermediate pairs of  $\langle \text{key}, \text{value} \rangle$ . The reduce function sums the result—the total count of a single word. The map function can be executed in parallel on non-overlapping portions of the input data, and the reduce function can be executed in parallel on each set of intermediate pairs with the same key.

### 2.3 HBase

HBase is an important Apache Hadoop-based project



▲ Figure 1. HDFS architecture.

modeled on Google's BigTable database [12]. It builds a distributed, fault-tolerant, scalable database on top of the HDFS file system and has random, real-time, read/write access to data. Each HBase table is stored as a multidimensional sparse map with rows and columns, and each cell has a timestamp. HBase has its own Java client API, and tables in HBase can be used both as an input source and output target for MapReduce jobs through TableInput/TableOutputFormat. All access to tables is by the primary key. Secondary indices are possible through additional index tables; programmers need to de-normalize and replicate.

A table is made up of regions. Each region is defined by a startKey and EndKey and may live on different nodes. A region is made up of several HDFS files and blocks, each of which is replicated by Hadoop. Columns can be added on-the-fly to tables, with only the parent column families being fixed in a schema. Each cell is tagged with column family and column name, so programs can always identify what type of data a given cell contains.

## 3 System Architecture

Our system is designed to mine web information in parallel, and we develop parallel algorithms based on MapReduce for this purpose. Fig. 2 shows the system architecture.

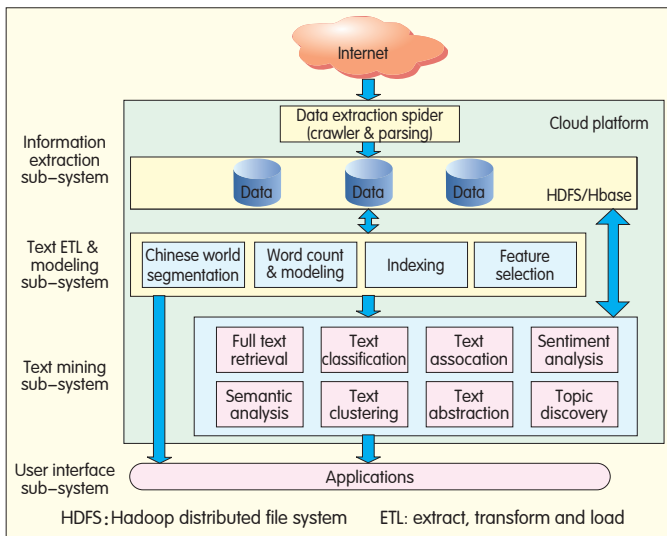
The proposed system includes four subsystems: parallel web crawler; parallel text extract, transform and load (ETL) and modeling; parallel text mining; and application subsystems.

### 3.1 Parallel Text Data Collection

The parallel text data collection subsystem crawls web pages from the internet and extract the content them. Here, we focus on extracting text information.

#### 3.1.1 Parallel Web Crawler

Intelligent Spider (HISpider) is based on Hadoop and is designed to browse the internet in a methodical, automated way. It contains site mode, keywords mode, and breakpoint



▲ Figure 2. System architecture of parallel web mining system based on cloud platform.

transmission mode. In site mode, a user-specified URL list is taken as the basis. Then, the seed pages and relative hyperlinks in these seed pages are downloaded hierarchically according to their link structure. In keywords mode, our module uses Baidu.com, Bing.com, and Sogou.com to query user-specified keywords that are stored in the keywords list file. Then, the module intelligently integrates returned pages to generate an initial URL list. Keywords mode is similar to site mode in that the seed pages and relevant hyperlinks are downloaded hierarchically according to their link structure after the initial seed URL has been constructed. The breakpoint transmission mode is used to complete a download task when the task has been unexpectedly terminated. In HISpider, update is an optional feature that is used to regularly check all downloaded data and re-extract pages that have been updated at the server side.

Using MapReduce, we assign a URL to multiple mapper classes, and web pages are downloaded in a parallel manner during the map stage. Fig. 3 shows the process of HISpider. First, HISpider acquires an initial URL list according to different extraction modes. Then, it starts a timer to time when to begin an update if update strategy has been triggered by the user. When a download is running, HISpider downloads in parallel all URLs, including webpages and documents, which are all in an incomplete URL list in the spider map. In that map, if a current iteration is less than the

extraction depth, HISpider extracts all hyperlinks at the current iteration as an incomplete URL list for the next round. After downloading all pages in the current iteration, HISpider puts metadata from downloaded pages and documents into a table in HBase, and this is used for other modules. When an update is running, the update map scans in parallel all record files that are generated by download jobs. Check whether the URL needs to be updated and re-download updated pages to overwrite original versions. The pseudo codes of the spider map and update map are shown in algorithm 1 and algorithm 2.

Algorithm 1: Spider map(*key*, *value*)

Input:

*key*: offset in bytes;

*value*: URL address

Output:

*key'*: hyperlinks in this URL address;

*value'*: null.

1: parse value to an array *valueArray*;

2: *URL* ← *valueArray*[0];

3: original last modified time *p* ← *valueArray*[1];

4: get file type  $\beta$  of URL;

5: if  $\beta$  is pdf or doc then;

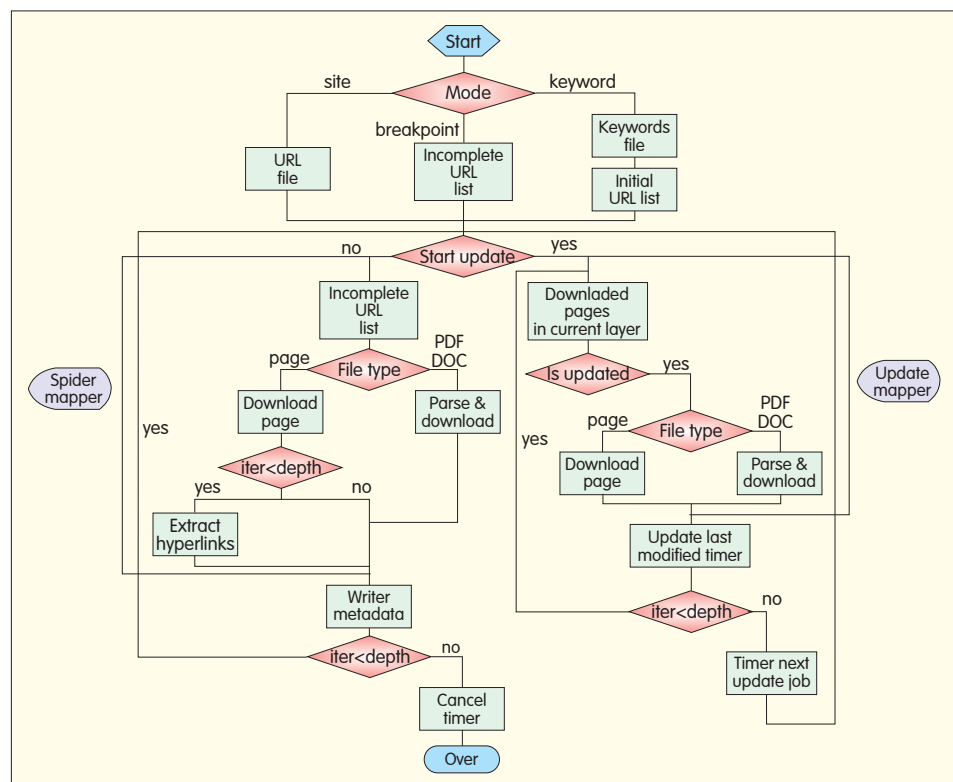
6: extract text from document and store in HDFS;

7: else;

8: get charset of webpage and transfer to UTF-8;

9: check re-directed page and go to target page;

10: get all contents of webpage;



▲ Figure 3. Parallel intelligent spider module.



```

11: output downloaded page's content into HDFS;
12: output downloaded page's URL into HDFS;
13: if iteration round  $\lambda < \text{extract depth } d$  then;
14:  $key' \leftarrow$  hyperlinks in current URL;
15: output  $key'$  as uncompleted URL into HDFS;
16: end if;
17: end if.

```

Algorithm 2: Update map( $key, value$ )

**Input:**

$key$ : offset in bytes;

$value$ : downloaded URL records with last modified time.

**Output:**

$key'$ : updated URL and its updated last modified time;  
 $value'$ : null.

1: parse value to an array  $valueArray$ ;

2:  $URL \leftarrow valueArray[0]$ ;

3: original last modified time  $\rho \leftarrow valueArray[1]$ ;

4: get last modified time  $\gamma$  of URL;

5: if  $\rho < \gamma$  then;

6: get file type  $\beta$  of URL;

7: if  $\beta$  is pdf or doc then;

8: extract text from document and store in HDFS.

### 3.1.2 Webpage Parsing

Web page parsing involves extracting text in the webpage. After parsing, web pages can be indexed for quick retrieval. It involves identifying the character set of HTML files, grasping the main structure, and extracting text according to the structure. Different websites often have different character sets, and these character sets need to be identified and converted into a uniform code for the subsequence operations. The main structure of the web page contains title, keywords, labels, pictures, and hyperlink. We extract text from the main structure.

Our parallel parsing web page algorithm and process is based on Hadoop and HBase (Fig. 4). The parsing algorithm can be easily implemented using MapReduce; therefore, we omit the detailed implementation of pseudocodes here.

Data is stored in HDFS and HBase; web page URLs for unique identification are stored in HBase, and the extracted text is store in HDFS.

### 3.2 Parallel Text ETL and Modeling Subsystem

The parallel text ETL and modeling subsystem contains four

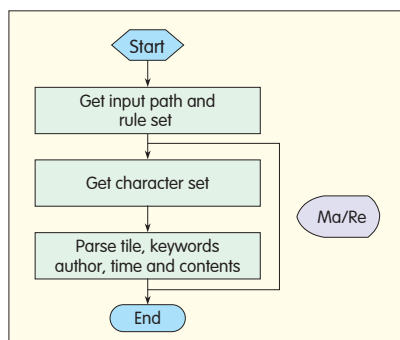


Figure 4. Parallel parsing algorithm.

modules: Chinese word segmentation, word count and modeling, indexing, and feature selection. For each module, we develop several parallel algorithms. The feature-selection algorithms include TFIDF, information gain, mutual information, and chi-square test. Here, we only detail the implementation of Chinese word segmentation and term frequency-inverse document frequency (TFIDF).

### 3.2.1 Chinese Word Segmentation

Unlike English and other European languages, there is no delimiter to mark the beginning and endings in written Chinese sentences. Therefore, word segmentation becomes the first task in processing information written in Chinese. The task is challenging because it is often difficult to define what constitutes a word in Chinese. There have been quite a few approaches, which can be roughly categorized as rules-based approaches (based on linguistics) and statistical approaches (based on a corpus and machine learning). ICTCLAS is a unified framework based on a hierarchical hidden Markov model that integrates Chinese word segmentation, part-of-speech tagging, disambiguation, and unknown-word recognition [13].

Our parallel MapReduce algorithm for Chinese word segmentation uses an open-source implementation of ICTCLAS. Fig. 5 shows our parallel Chinese word segmentation algorithm on Hadoop platform, HBase, and MapReduce. First, we upload the dictionary files from the master node to HDFS to allow the slave nodes to distributively load the dictionary before calling the word-segmentation package. Then, we configure and run a map/reduce job. This has an input HBase table, which stores the HDFS locations and IDs of the files to be processed; an output HBase table, which stores the HDFS locations; and the IDs of the files after they have been processed. A mapper class consists of three functions: setup(), map(), and cleanup(). Under the MapReduce framework, the functions of the mapper class are executed in a distributed manner. First, setup() is executed to load dictionary files from HDFS to a temporal path on each slave node. Second, map() parses a record of the input HBase table to obtain the file to be processed. It then calls the ICTCLAS package to split the sentences of the file into words. Finally, cleanup() deletes the dictionary files under the temporal path on each slave node. The pseudocode of map() of our algorithm is shown in algorithm 3.

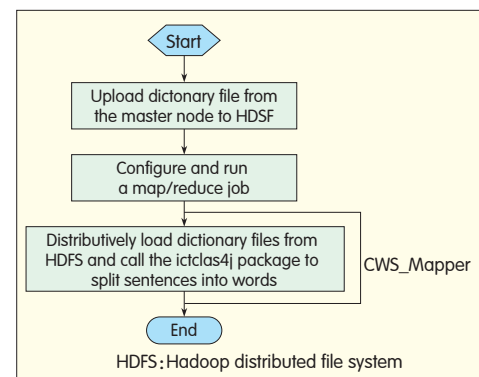


Figure 5. Flowchart of parallel Chinese word segmentation algorithm.

Algorithm 3: CWS\_map(*key*, *value*)

**Input:**

*key*: row key of the current row being processed;  
*value*: the value of the row.

**Output:**

*key'*: the same as the input key;

*put*: HDFS position and ID number of the segmentation result file.

- 1: parse a record of the input HBase table to obtain the ID number and HDFS path of the file to be processed;
- 2: read in the file to be processed;
- 3: call the `ictclas4j` package to split sentences of the file into words;
- 4: write segmentation results to some HDFS file specified by user;
- 5: take *key* as *key'*;
- 6: take the ID number and HDFS position of the segmentation result file as two columns of *put*;
- 7: write a record(*key'*, *put*) into the output HBase table.

### 3.2.2 Term Frequency–Inverse Document Frequency

TFIDF is a numerical statistic that reflects the importance of a word to a document in a collection or corpus [14]. It is often used as a weighting factor in information retrieval and text mining. The TFIDF value increases proportionally to the number of times a word appears in a document but is offset by the frequency of the word in the corpus because some words are generally more common than others. TFIDF undervalues terms that frequently appear in documents belonging to the same class and gives greater weight to terms that represent the characteristic of the documents in its class.

Term frequency (TF) is the number of times a term occurs in a document. This can be formulated as  $tf(t, d)$ , where  $t$  is the term and  $d$  is the document. Inverse document frequency (IDF) measures whether the term is common or rare across all documents. It is obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient:

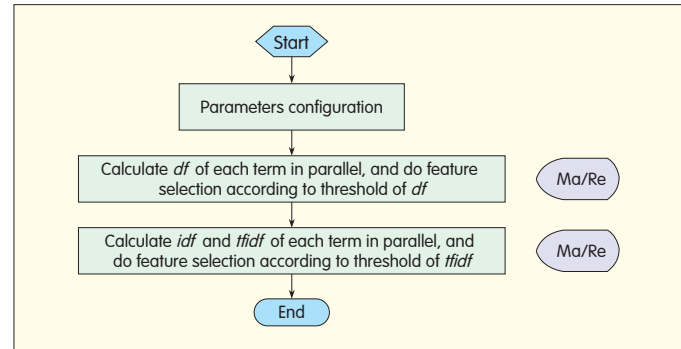
$$idf(t, D) = \log \frac{|D|}{|\{d \in D: t \in d\}|} \quad (1)$$

where  $|D|$  is the total number of documents in the corpus,  $|\{d \in D: t \in d\}|$  is the number of documents where the term  $t$  appears. If the term is not in the corpus, a division-by-zero occurs. Therefore, it is common to adjust the formula to  $1 + |\{d \in D: t \in d\}|$ . Then, the TFIDF is calculated:

$$tfidf(t, d, D) = tf(t, d) \times idf(t, D) \quad (2)$$

According to the  $tf(t, d)$  matrix, we design parallel MapReduce algorithms to calculate  $df(t, d)$  and  $idf(t, d)$  of each term, respectively. Fig. 6 shows the calculation of  $tfidf(t, d, D)$ .

There are two main MapReduce jobs. The first involves calculating  $df$  of each term and selecting features according to the  $df$  threshold; the second involves calculating  $idf$  and  $tfidf$  of each term and selecting features according to the



▲ Figure 6. Parallel TFIDF algorithm.

threshold of  $tfidf$ . Finally, a reduced document–term matrix is obtained that removes most of redundant information and keeps most of the feature information. The two MapReduce jobs in a parallel TFIDF algorithm can be implemented as follows.

The first MapReduce job is described in algorithms 4 and 5. In algorithm 4, step 1 removes the document id from the input text value, and steps 2 to 5 output all the terms with a value of 1. Steps 2 to 4 in algorithm 5 calculate the summation of each  $v$  in value, and steps 5 to 8 output all the terms with a summation value greater or equal to the DF threshold.

Algorithm 4: DFMapper

**Input:**

*key*: offset in bytes;

*value*: a text containing document id and a vector of all  $\langle \text{term}, \text{count} \rangle$  pairs of the document.

**Output:**

*key'*: a text for each term;

*value'*: an integer one.

- 1: Separate all the  $\langle \text{term}, \text{count} \rangle$  pairs from the input value;
- 2: for each  $\langle \text{term}, \text{count} \rangle$  pair;
- 3: Set *key'* as *term*, *value'* as 1;
- 4: output  $\langle \text{key}', \text{value}' \rangle$  pair;
- 5: endfor.

The second MapReduce job is described in algorithm 6. Steps 1 and 2 are preparation and involve parameter reading and preprocessing of input data. Steps 3 to 8 calculate the  $tfidf$  for each term and select features according to the TFIDF threshold. Steps 9 to 10 output the  $tfidf$  vector for each document.

Algorithm 5: DFReducer

**Input:**

*key*: a term text;

*value*: a vector of integer one with the length that the term occurs.

*key'*: the same with *key*;

*value'*: sum of integer one in *value*.

- 1: Initialize sum as zero,  $DFThreshold$  as Threshold of DF;
- 2: for each integer  $v$  in *value*;
- 3:  $sum + = v$ ;

```

4: endfor;
5: if( $sum \geq DFThreshold$ ) ;
6: Set  $key'$  as  $key$ ,  $value'$  as  $sum$ ;
7: Output  $\langle key', value' \rangle$  pair;
8: endif.

```

Algorithm 6: TFIDFMapper

**Input:**

$key$ : the offset in bytes;

$value$ : a text containing document id and a vector of all  $\langle term, count \rangle$  pairs of the document.

**Output:**

$key'$ : a text of document id;

$value'$ : a vector of all the terms with their  $tfidf$  value.

1: Read  $df$  value from temporary HDFS path, set  $N$  as the total number of documents, and  $TFIDFThreshold$  as Threshold of TFIDF;

2: split the input text  $value$ , to get document id and all the  $\langle term, count \rangle$  pairs;

3: for each  $\langle term, count \rangle$  pair;

4: calculate  $tfidf$  according to (2);

5: if( $tfidf \geq TFIDFThreshold$ ) ;

6:  $value'.append(\langle term, tfidf \rangle)$ ;

7: endif;

8: endfor;

9: set  $key'$  as document id;

10: output  $\langle key', value' \rangle$  pair.

### 3.3 Parallel Text Mining Subsystem

The parallel text mining subsystem is the core of our web mining system, and is closely related to the applications. The subsystem has eight modules: full text retrieval, text classification, text association, sentiment analysis, semantic analysis, text clustering, text abstraction, and topic discovery. In total, we have developed nineteen parallel algorithms.

Here, we only detail the implementation of two parallel algorithms.

#### 3.3.1 Co-Occurrence Analysis

Co-occurrence analysis (CoocAnalysis) is a statistical method used in text mining [15]. Generally speaking, it uses statistical theory to analyze the co-occurrence distribution characteristics of the text knowledge units, and it mines the potential association between these units. Recently, CoocAnalysis has become more important in knowledge mining and discovery. The calculation formulas of CoocAnalysis between term  $T_j$  and  $T_k$  are:

$$Weight(T_j, T_k) = \frac{\sum_{i=1}^n d_{jk}^i}{\sum_{i=1}^n d_j^i} \times Weight\ Factor(T_k) \quad (3)$$

$$Weight(T_k, T_j) = \frac{\sum_{i=1}^n d_{kj}^i}{\sum_{i=1}^n d_k^i} \times Weight\ Factor(T_j) \quad (4)$$

(3) gives the co-occurrence weight from term  $T_j$  to  $T_k$ , and (4) gives the co-occurrence weight from term  $T_k$  to  $T_j$ . The entire co-occurrence weight matrix of all terms is non-symmetric. In (3) and (4)  $d_{jk}^i$  is the co-weight of terms  $T_j$

and  $T_k$  in document  $i$  and is defined as

$$d_{jk}^i = tf_{jk}^i \times \log \frac{N}{df_{jk}} \quad (5)$$

where

- $tf_{jk}^i$  is the minor number of co-occurrences of  $T_j$  and  $T_k$  in document  $i$

- $df_{jk}$  is the number of documents where terms  $T_j$  and  $T_k$  occur together

- $N$  is the number of documents.

In the CoocAnalysis algorithm, in order to give very common terms a certain disadvantage, we multiply a weight factor to each term, which is similar to the inverse document frequency. The weight factor is given by

$$Weight(T_j) = \frac{\log \frac{N}{df_j}}{\log N} \quad (6)$$

From (6), it can be concluded that the main purpose of CoocAnalysis algorithm is to do some statistics on term frequency so that it is suitable for parallelization. We have designed parallel MapReduce algorithms for the CoocAnalysis algorithm, and the process for the CoocAnalysis algorithm is shown in Fig. 7.

There are two main MapReduce jobs. The first involves calculating document frequency and weight factor of each term. The second involves calculating the co-weight of each term pair in each document and co-occurrence weight between terms. The two MapReduce jobs of parallel CoocAnalysis algorithm can be implemented as follows.

The first MapReduce job of parallel CoocAnalysis algorithm is described in algorithms 7 and 8. In algorithm 7, step 1 obtains the document id and  $\langle term, count \rangle$  pair from the input text value, and steps 2 to 5 output all the single terms with document ids. In algorithm 8, steps 2 to 7 construct a string containing all the document ids in value and outputs them. Steps 8 to 10 calculate the weight factor of each item, which is then output.

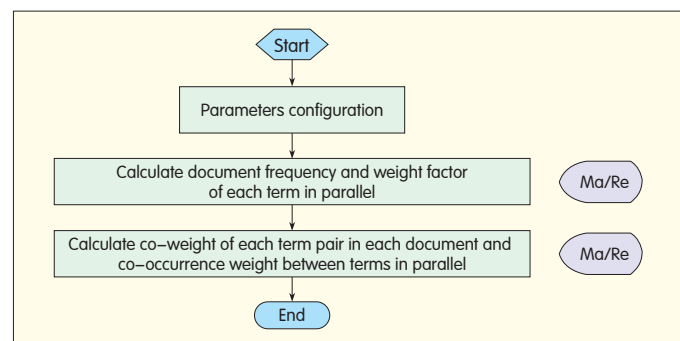
Algorithm 7: InvIndex WFMapper

**Input:**

$key$ : the offset in bytes;

$value$ : a text containing document id and a vector of all  $\langle term, count \rangle$  pairs of the document.

**Output:**



▲ Figure 7. Parallel CoocAnalysis algorithm.

*key'*: a text for each term;  
*value'*: document id for each term.  
 1: Split the input text value, to get document id and all the *<term, count>* pairs;  
 2: for each *<term, count>* pair;  
 3: set *key'* as term, *value'* as document id;  
 4: output *<key', value'>* pair;  
 5: endfor.

Algorithm 8: InvIndex WFReducer

**Input:**

*key*: a text for each term;

*value*: a vector of document ids where the term *key* occurs.

**Output:** two kinds of *<key', value'>*

For the first kind:

*key'*: the same with *key*;

*value'*: text of all document ids in *value*.

For the second kind,

*key'*: a string "WF" appended with *key*;

*value'*: weight factor of each term.

1: Initialize sum as zero, set *N* as the total number of documents;  
 2: for each document id *v* in *value*;  
 3: *value'*.append(*v*);  
 4: sum+=1;  
 5: endfor;  
 6: set *key'* as *key*  
 7: output<*key'*, *value'*> pair;  
 8: *key'*="WF"+*key*;  
 9: calculate *value'* according to (6);  
 10: output *<key', value'>* pair.

The second MapReduce job in the parallel CoocAnalysis algorithm is described in algorithms 9 and 10. In algorithm 9, steps 1 to 2 are preparation and include parameter reading and preprocessing of input data. Steps 3 to 17 calculate the *tfidf* value for each term as well as the co-occurrence TFIDF value of each term pair, which are then output. In algorithm 10, step 1 is preparation, and steps 2 to 4 calculate the co-occurrence weight between terms, which is then output.

Algorithm 9: CoocMapper

**Input:**

*key*: offset in bytes;

*value*: a text containing document id and a vector of all *<term, count>* pairs of the document.

1: Read in the output of the previous MapReduce job, initialize *TFIDF* as a list to save *tfidf* of each single term;  
 2: Split the input text *value*, to get document id and all the *<term, count>* pairs;  
 3: for each *<term, count>* pair;  
 4: Calculate *tfidf* value of each single term in *<term, count>* pair;  
 5: *TFIDF*.add(*tfidf*);  
 6: Initialize *docIDList* 1 to save all document ids where the term *term* occurs;  
 7: Put all the document ids where the term *term* occurs to

*docIDList*1, and sort it;

8: for each *<term', count'>* pair where *term'* term.

**Output:**

*key'*: a text of a term pair;

*value'*: co-occurrence TFIDF value of each term pair.

9: initialize *docIDList* 2 to save all document ids where the term *term'* occurs;

10: put all the document ids where the term *term* occurs to *docIDList* 2, and sort it;

11: according to *docIDList* 1 and *docIDList* 2, calculate the co-occurrence number of *term* and *term'*;

12: calculate co-occurrence TFIDF value of the term pair according to (5), and set *value'* as it;

13: set *key'* as *<term, term'>* pair;

14: output *<key', value'>* pair;

15: endfor;

16: endfor;

17: output *TFIDF* to a temporary HDFS path.

Algorithm 10: CoocReducer

**Input:**

*key*: a text of a term pair;

*value*: co-occurrence TFIDF value of each term pair.

**Output:**

*key'*: a text of a term pair;

*value'*: co-occurrence weight between terms.

1: Read in *TFIDF* from the temporary HDFS path;

2: Calculate *value'* according to (3)–(4);

3: Construct *key'* according to *key*;

4: Output *<key', value'>* pair.

### 3.3.2 Semantic Analysis based on PLSA

In machine learning, semantic analysis of a corpus is achieved by building structures that approximate concepts from a large set of documents. This generally does not involve prior semantic understanding of the documents. Latent semantic analysis (LSA) is a technique in natural language processing used to analyze relationships between a set of documents and terms in order to produce a set of concepts related to the documents and terms. Probabilistic latent semantic analysis (PLSA) is a popular topic-modeling technique for exploring document collections [16]. A parallel PLSA algorithm is described in [17]. We design a parallel semantic analysis algorithm (SPLSA) based on parallel PLSA.

We aim to find related words given some index words. In a corpus, PLSA algorithm can find the topics and corresponding probabilities that each word belongs to a topic. After that, we build the word-topic-probability relationship; that is, each word belongs to a topic with a probability. Algorithm 11 shows the pseudocode.

Algorithm 11: Map(*key, value*)

**Input:**

*key*: offset in bytes;

*value*: word-topics-probabilities.

**Output:**

The word-topic-probability file



```

1: Read value into array wordtopic [];
2: max = -1; // record the maximum probability with which
the word belongs to a topic;
3: topic = -1; // record the topic index to which the word
belongs;
4: for(i = 1; i < number of topics; i ++);
5: if(wordtopic[i] > max);
6: max = wordtopic[i];
7: topic = i - 1;
8: endif;
9: endfor;
10: write wordtopic[0] - topic - max; // the
word - topic - probability relationship.

```

According to this relationship, we can find all words that belong to the same topic. The words with the top- $n$  max probability in the same topic are related. The parameter given by the user is  $n$ , and the parallel analysis process is described in Fig. 8.

First, we read the index words and the word-topic file. Second, we find the topics and build the word-topic-probability relationship during the map phase. Finally, we determine related words from the relationship and output these words.

## 4 Experiments

In this work, we focus on designing parallel web-mining algorithms and therefore only test the efficiency of parallel algorithms to guarantee their correct parallel implementation.

Fig. 9 shows the simulation interface. We show how to set the parameters of the parallel CoocAnalysis algorithm, and can set the main class and jar package. It is also very convenient to set the data input path  $-i$  and output path  $-o$ , the total number of documents  $-d$ , and the number of reducers  $-r$ .

Here, we only describe the efficiency of the web crawler, web page parsing, TFIDF, CoocAnalysis according to speedup in this experiments.

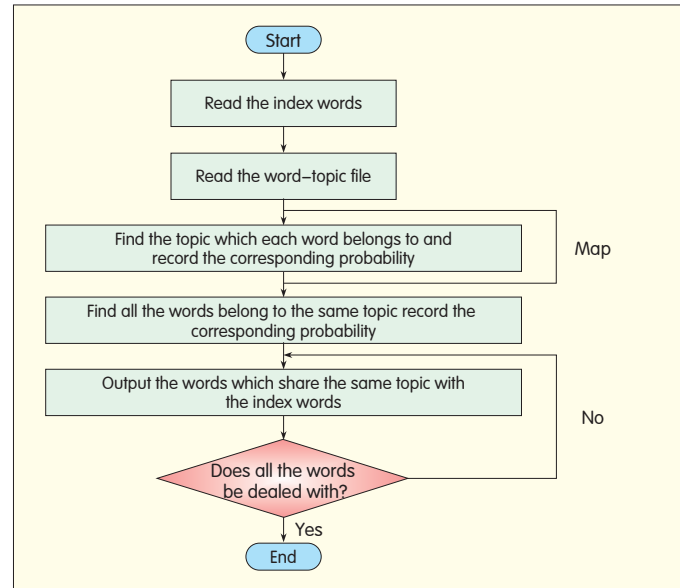
### 4.1 Experiment Preparation

The data set is from the Yahoo! News website. We use the proposed parallel web crawler to crawl the webpage. To test the web crawler, we construct a complete URL list with 20,000 URLs, and the experiments are conducted on different computing systems with different nodes. The size of the URL list is only 800 kB, and to maximize parallelization, we modify the block size rather than adopt the default value of 64 MB.

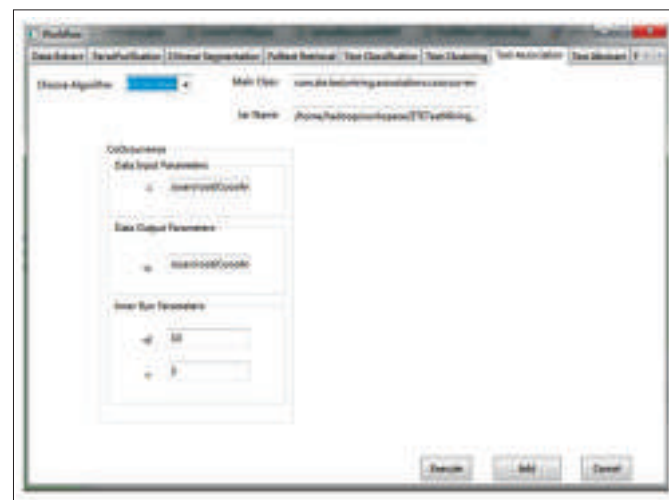
The parallel system is a cluster of four computer nodes with Linux, and each node has four 2.8 GHz cores and 4 GB memory. The MapReduce system is configured with Hadoop 0.17.0 and Java 1.6.0.22 for all experiments. We perform the experiments on computing systems with 1, 2 and 4 nodes.

We use the popular evaluation metric speedup [4] to validate our algorithms, which is defined as

$$\text{Speedup}(m) = \frac{\text{Running time on 1 node}}{\text{Running time on } m \text{ nodes}} \quad (7)$$



▲ Figure 8. Flowchart of parallel SPLSA algorithm.



▲ Figure 9. Simulation interface.

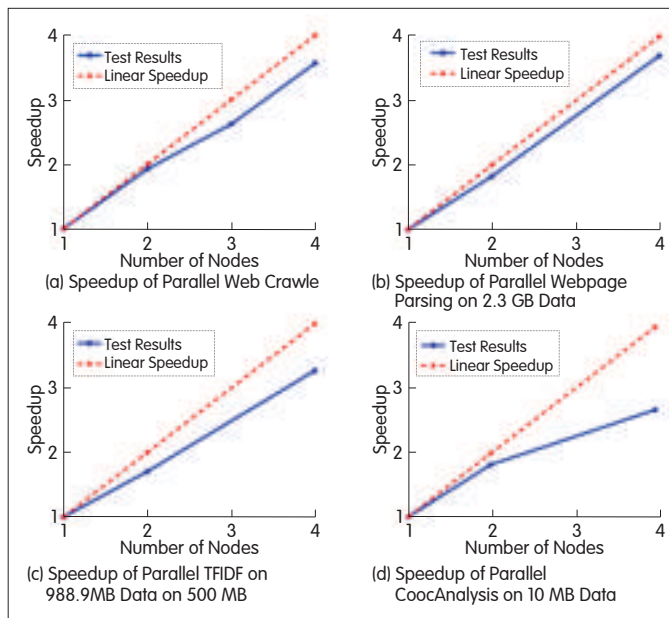
When the value of speedup is  $m$ , we obtain the linear speedup. However, in practice we cannot reach the linear speedup because of the task allocation balance, and communication cost between computer nodes.

### 4.2 Results

Fig. 10(a)–(d) shows the results. We find that all the parallel algorithms have very good speedup properties. Specifically, all the algorithms except CoocAnalysis have a speedup value higher than 3 when there are only four nodes. Furthermore, low-complexity algorithms such as webpage parsing and TFIDF have a high speedup value.

## 5 Conclusion

In this paper, we have developed a parallel web-mining



▲ Figure 10. The speedup of parallel algorithms.

system that includes more than forty machine learning algorithms for text mining. Using cloud platform and MapReduce, we analyze the mechanism of each algorithm and carefully design the  $\langle \text{key}, \text{value} \rangle$  pair for map and reduce function in order to parallelize the algorithms to the greatest possible extent. The results validate the efficiency of the proposed parallel algorithms. This work is the backbone of parallel text-mining algorithms.

#### Acknowledgements

This work is also supported by the ZTE research found of Parallel Web Mining project. We also could like to thank Wenjuan Luo, Tianfeng Shang, Changying Du, Xin Jin, Zhi Dong, YunLong Ma, Qun Wang, Shuo Han, Xinyu Wu, Xiaofeng Geng for their contributions to this paper.

#### References

- [1] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," *Communications of the ACM*, vol. 51, pp. 107–113, ACM, 2008.
- [2] C.T. Chu, S.K. Kim, Y.A. Lin, Y.Y. Yu, G. Bratski, A.Y. Ng, and K. Olukotun, "Map-reduce for machine learning on multicore," *Advances in neural information processing systems*(19), 2006.
- [3] Q. He, F. Zhuang, J. Li, and Z. Shi, "Parallel implementation of classification algorithms based on mapreduce," *Rough Set and Knowledge Technology*, pp. 655–662, 2010.
- [4] W. Zhao, H. Ma, and Q. He, "Parallel k-means clustering based on mapreduce," *Cloud Computing*, pp. 674–679, 2009.
- [5] Q. He, Q. Wang, F. Zhuang, Q. Tan, and Z. Shi, "Parallel clarans clustering based on mapreduce," *Energy Procedia*, vol. 13, pp. 3269–3279, 2011.
- [6] Q. He, Y. Ma, Q. Wang, F. Zhuang, and Z. Shi, "Parallel outlier detection using kd-tree based on mapreduce," *Cloud Computing Technology and Science (CloudCom)*, In *2011 IEEE Third International Conf. on*, pp. 75–80, IEEE, 2011.
- [7] Q. He, T. Shang, F. Zhuang, and Z. Shi, "Parallel extreme learning machine for regression based on mapreduce," *Neurocomputing*, 2012.
- [8] G. Wu, H. Li, X. Hu, Y. Bi, J. Zhang, and X. Wu, "Mrec4. 5: C4. 5 ensemble classification with mapreduce," In *ChinaGrid Annual Conf.*, 2009, ChinaGrid'09, Fourth, pp. 249–255.
- [9] C. Zhang and J. Sun, "Large scale microblog mining using distributed mb-lda," In *Pro. of the 21st international conf. Companion on World Wide Web*, PP. 1035–1042, ACM, 2012.

- [10] T. Elsayed, J. Lin, and D.W. Oard, "Pairwise document similarity in large collections with mapreduce," In *Proc. of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 265–268, Association for Computational Linguistics, 2008.
- [11] D. Borthakur, "The hadoop distributed file system: Architecture and design," Hadoop Project Website, 11:21, 2007.
- [12] F. Chang, J. Dean, S. Ghemawat, W.C. Hsieh, D.A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R.E. Gruber, "Bigtable: A distributed storage system for structured data," *ACM Transactions on Computer Systems (TOCS)*, 26(2):4, 2008.
- [13] H.P. Zhang, H.K. Yu, D.Y. Xiong, and Q. Liu, "Hhmm-based Chinese lexical analyzer ictclas," In *Proc. of the second SIGHAN workshop on Chinese language processing–Volume 17*, PP. 184–187, Association for Computational Linguistics, 2003.
- [14] Gerard Salton and Michael J. McGill, "Introduction to Modern Information Retrieval," McGraw-Hill, Inc., New York, NY, USA, 1986.
- [15] H. Chen, J. Martinez, T.D. Ng, and B.R. Schatz, "A concept space approach to addressing the vocabulary problem in scientific information retrieval: an experiment on the worm community system," 1997.
- [16] T. Hofmann, "Probabilistic latent semantic indexing," In *Proc. of the 22nd annual international ACM SIGIR conf. on Research and development in information retrieval*, PP. 50–57, ACM, 1999.
- [17] N. Li, F.Z. Zhuang, Q. He, and Z.Z. Shi, "Ppls: Parallel probabilistic latent semantic analysis based on mapreduce," In *7th International Conf. on Intelligent Information Processing(Accepted)*, 2012.

Manuscript received: July 27, 2012

#### **B** iographies

**Shengmei Luo** is chief architect at ZTE Corporation and professor at Nanjing University of Post and Telecommunications. He has been awarded prizes for scientific and technological progress, holds many patents, and has also written papers that have been published in a number of core communication journals. He is a member of the China Cloud Computing Committee. He graduated from Harbin Institute of Technology in 1996 and has been involved in telecommunication network and services development and planning for many years.

**Qing He** is a professor in the Institute of Computing Technology, Chinese Academy of Sciences. He is also a professor at the Graduate University of the Chinese Academy of Sciences. He received his BS degree in mathematics from Hebei Normal University, Shijiazhang, in 1985. He received is MS degree in mathematics from Zhengzhou University in 1987. He received his PhD degree in fuzzy mathematics and AI from Beijing Normal University in 2000. From 1987 to 1997, he has worked at Hebei University of Science and Technology. He is currently a doctoral tutor at the Institute of Computing and Technology, Chinese Academy of Sciences. His interests include data mining, machine learning, classification, and fuzzy clustering.

**Lixia Liu** is a senior engineer in the pre-research department of ZTE, and she received the M.S degree from Ocean University of China in 2008. Her research interests include natural language processing , text mining, data mining, machine learning, mathematical statistics and cloud computing.

**Xiang Ao** is a PhD candidate student in the Institute of Computing Technology, Chinese Academy of Sciences. His research interests include machine learning, data mining and cloud computing.

**Ning Li** is a PhD candidate student in the Institute of Computing Technology, Chinese Academy of Sciences. Her research interests include machine learning, data mining and cloud computing.

**Fuzhen Zhuang** is an assistant professor in the Institute of Computing Technology, Chinese Academy of Sciences. His research interests include machine learning, data mining, distributed classification and clustering, natural language processing. He has published several papers in some prestigious refereed journals and conference proceedings, such as IEEE Transactions on Knowledge and Data Engineering, Neurocomputing, ACM CIKM, SIAM SDM and IEEE ICDM.

# Hierarchical Template Matching for Robust Visual Tracking with Severe Occlusions

**Lizuo Jin<sup>1</sup>, Tirui Wu<sup>2</sup>, Feng Liu<sup>3</sup>, and Gang Zeng<sup>3</sup>**

(1. School of Automation, Southeast University, Nanjing 210096, China;

2. ZTE Corporation, Nanjing 210012, China;

3. ZTE Corporation, Chongqing 401121, China)

## Abstract

To tackle the problem of severe occlusions in visual tracking, we propose a hierarchical template-matching method based on a layered appearance model. This model integrates holistic- and part-region matching in order to locate an object in a coarse-to-fine manner. Furthermore, in order to reduce ambiguity in object localization, only the discriminative parts of an object's appearance template are chosen for similarity computing with respect to their cornerness measurements. The similarity between parts is computed in a layer-wise manner, and from this, occlusions can be evaluated. When the object is partly occluded, it can be located accurately by matching candidate regions with the appearance template. When it is completely occluded, its location can be predicted from its historical motion information using a Kalman filter. The proposed tracker is tested on several practical image sequences, and the experimental results show that it can consistently provide accurate object location for stable tracking, even for severe occlusions.

## Keywords

visual tracking; hierarchical template matching; layered appearance model; occlusion analysis

more comprehensive visual information about objects, and appearance-based object tracking methods have received greater attention in recent decades [1].

However, in real-world visual tracking applications, occlusions are inevitable and usually occur when the view of a moving object is partly or completely blocked by objects such as the static background or other foreground moving objects. When the object is partly or completely occluded, its visual appearance deviates dramatically from its appearance template. Object localization can be very imprecise, and if the appearance template badly damaged due to improper model updating, eventually object tracking is lost [2].

Much research has been down on the heavy impact of occlusions on visual tracking. Adaptive appearance-modeling algorithms deal indirectly with occlusions through statistical analysis [3]–[5]. However, the models are susceptible corruption by long-term occlusions and blind updating. In [6]–[9], the object is divided into several components or patches, and occlusions are evaluated by patch-matching and robust statistics. Using cameras is a good way to handle occlusions, but cameras cannot be applied to many visual-tracking tasks because they require a special setup and come at additional cost [10]. In [11], several algorithms are proposed to overcome occlusions in constrained conditions. In [12], occlusions among objects in the context of multiple object tracking are discussed. In [13], occlusions related to specific objects, such as human bodies, in the context of pre-defined model constraints are discussed. In [14], occlusions related to a specific scene using the depth or the motion information are discussed. A few attempts to manage occlusions and other exceptions have been made based on a spatiotemporal context [14]–[16], and they require many non-trivial observations and tracking of other objects or features outside the target objects. Machine-learning methods such as sparse learning [17], hierarchical feature learning [18], semi-supervised learning [19]–[20] and these can overcome part occlusions to various extents.

In recent years, several methods have been used to deal explicitly with object occlusions. A mixture distribution can be used to model the observed intensity of each pixel where outliers are characterized by the lost component, which has a uniform distribution [21]. Some approaches declare outlier pixels by examining whether the measurement error exceeds a predefined threshold, and they work very well only when the statistical properties of the occlusions agree with the

## 1 Introduction

Object tracking, also called visual tracking, involves automatically locating moving objects across successive frames in image sequences. Tracking objects based on their appearance is important in fields such as visual surveillance, human-computer interaction, robot navigation, and missile guidance. Appearance provides much

This work is supported by the Aeronautical Science Foundation of China under Grant 20115169016.

assumptions [22]–[24]. A general algorithm for detecting and handling occlusions is proposed in [25]. The algorithm learns a classifier by observing the likelihood of a few types of occlusion patterns based on the data gathered during the tracking of object with and without occlusions. This is an improvement on several existing tracking algorithms, but because of misclassification, error rate is not low enough.

Occlusions create four basic issues for appearance-based tracking methods [24]. The first is how to robustly determine the portion of occlusions; the second is how to accurately locate the object when the situation of occlusions is unknown; the third is how to properly update the appearance model in order to keep tracking the object while preventing damage by outliers; and the fourth is how to reliably detect the reemergence of the object and recapture it after it has been occluded completely for a period of time. To tackle these issues and handle the variations of object appearance caused by occlusions, we propose a hierarchical template-matching method for tracking. This method integrates holistic- and part-region matching in order to locate the object within a search window in a coarse-to-fine manner.

Section 2 describes the hierarchical representation of object appearance. Section 3 contains a brief review of the overall structure of the proposed tracker, and the hierarchical template-matching algorithm is detailed. The algorithm provides the accurately locates the object, even when there are severe occlusions. Section 4 contains experimental results and analysis. Section 5 concludes the paper.

## 2 Hierarchical Representation of Object Appearance

To track an object, we need to locate it first. This can be done in a predicted region far smaller than the whole field of view. This predicted region is usually based on the historical information of the object's motion and the appearance model of the object, which is adaptively updated. The success of a tracking algorithm depends greatly on the appearance model and the localization method. Some interesting ideas from object classification can be applied to object locating. In several recent schemes for object classification, the basic features used for classification are local image fragments, or patches, that depict significant object components and are chosen from training images on the learning stage. The features can be selected from a large pool of candidate image fragments or a set of regions yielded by applying interest operators. On the classification stage, the features are located in the image and are then combined using classification methods such as naïve Bayesian combination; a probabilistic model combining appearance, shape and scale; an ensemble of weak classifiers; and a SVM-based classifier. The features in these methods are non-hierarchical; that is, they are not broken down into simpler, distinct subparts, but are detected one-by-one by comparing the fragment to the image. The similarity between parts can be computed using measures such as normalized cross correlation, affine

invariant measure, and SIFT/SURF. Some visual-tracking algorithms use part-based matching techniques to locate the target object from images [6]–[9].

A number of classification schemes also use feature hierarchies rather than holistic features. Such schemes are often based on biological modeling and the structure of the primate visual system. Such a system uses a hierarchy of features of increasing complexity, from simple local features in the primary visual cortex to complex shapes and object views in higher cortical areas. In a number of biological models, the architecture of the hierarchy (size, position, shape of features and sub features) is predefined or learnt for different classification tasks. Recent advances in object classification using feature hierarchies have produced promising results on some benchmark datasets [27]–[29].

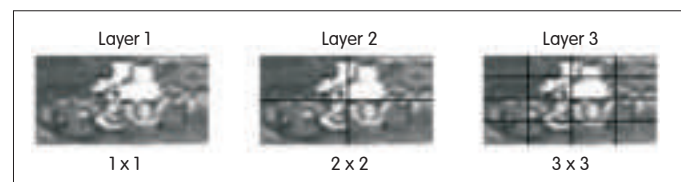
Representing object appearance using informative components is useful for handling variations in appearance during tracking. However, these components, like the objects themselves, can vary considerably in appearance. Therefore, it is natural to decompose the components into multiple informative subparts. A repeated division process results in a hierarchical representation of an object with informative parts and subparts in multiple layers. Such hierarchical decomposition can substantially improve object localization. The division process properly classifies objects by learning on a training dataset [27]–[29]. However, we apply only regular grids of different sizes to decompose the object into informative parts for balancing the computation and the discrimination capability. Finally, a layered-appearance model is built. The division process is shown in Fig. 1. The regular grids applied in the three layers are  $1 \times 1$ ,  $2 \times 2$ , and  $4 \times 4$ .

We want to use this hierarchical representation to accurately locate the object in a predicted region, not in the whole field of view. Therefore, the localization capability of each part is important and can be checked by the cornerness measure proposed in [26] that was first applied to detect corner features from images and later to choose the informative patches for object tracking. Only those parts with large cornerness measure are accepted as informative features for tracking objects. If we consider a part  $I$  in one layer, we can compute its second-moment matrix  $\mathbf{M}$  using

$$\mathbf{M} = \sum_{x,y} W_{x,y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (1)$$

where  $w_{x,y}$  is the weight,  $I_x$  is the horizontal gradient, and  $I_y$  is the vertical gradient. If the eigenvalues of  $\mathbf{M}$  are  $\lambda_1$  and  $\lambda_2$ , the

$$R = \lambda_1 \lambda_2 - \alpha (\lambda_1 + \lambda_2)^2 \quad (2)$$



▲ Figure 1. The layered-appearance model of an object formed by applying regular grids to decompose the components into parts.



corner response function  $R$  can be used to verify whether it is informative:

where  $\alpha$  is a constant with a value of 0.04 to 0.06.

### 3 Object Localization by Hierarchical Template Matching

#### 3.1 Overall Structure

The initial target object is specified by manually or automatically selecting a target region, typically covering a rectangular area. The target region is referred to as the region of interest (ROI) and tightly bounds the object to be tracked. Some background pixels might also be included in the ROI during initialization, but this does not matter too much.

We also initialize a validation mask that represents the locations of all informative parts. This mask is a binary matrix with a value of one (where the pixel belongs to the informative parts) or zero otherwise and is built using the method described in section 2. The appearance template reflects the current estimated object appearance and is initialized by sampling pixels from the initial ROI through coordinate transformation. The initial occlusion mask is just a duplication of the validation mask, which indicates that no pixel in the template is masked out except the non-informative parts.

Objects are tracked in two operation modes: normal and complete. In normal mode, the target location and ROI are predicted by first using an adaptive-velocity model when each new frame comes in. Then, the approximate target region is obtained by hierarchical template matching. We first analyze the occlusions within the ROI by matching each part in the template and updating the occlusion mask. Then, we perform masked template matching based on the result of occlusion analysis. This rectifies the target location of the object. The result that is output by template matching determines the final ROI, and within this ROI, the occlusion is analyzed again to generate the final occlusion mask. After obtaining the accurate target location and final occlusion mask, we update the masked template by temporal smoothing. When more than 80% of the target object is occluded, our proposed tracker enters complete occlusion mode, and the location of the target object is predicted by a Kalman filter. The reappearance of the target object is reliably detected by the template matching method. Once the end of a complete occlusion is declared, the occlusion mask is reinitialized, and the tracker resumes in normal mode.

#### 3.2 Evaluating the Situation of Occlusions

A layered-appearance model of parts is introduced in section 2 to represent the object, of which, the top layer is the full object, the second layer has four non-overlapping parts yield by half dividing the full object at the top layer along the x and y dimension equally, and the third layer has sixteen non-overlapping sub-parts yield by half dividing each part at the second layer in the same manner, each part has four corresponding sub-parts respectively and therefore the object can be described finely more and more.

The similarity between each corresponding part of a candidate object and target object can be measured by the normalized cross-correlation score (NCC). When the NCC is large enough, the parts of the candidate object and target object are taken to be the same. When the NCC is small enough, the parts of the candidate object and target object are taken to be different. When searching for the object at candidate locations, the NCC of the part at the top layer is computed first. If it is high enough, the object is accepted; if it is too low, the object is rejected; otherwise, NCC of the parts in the next layer is computed until an inference can be confidently made or the lowest layer is reached. After these operations, when no occlusion occurs, the object can be located by computing the NCC in the top layer. When occlusions occur, most regions with a small NCC are rejected in the top layer, and only potential regions with a mid-range NCC are analyzed.

Occlusions can be evaluated with the coverage number of parts having a high NCC. Because of the different coverage area of parts in a different layer, the coverage number of parts (CNP) is

$$CNP = \sum_{l,i} w^l \times I(NCC_i^l > \theta) \quad (3)$$

where  $w^l$  is the weight of layer  $l$ , and  $NCC_i^l$  is the NCC of object part  $i$  in layer  $l$ . The indication function is  $I(\cdot)$ , and the predefined threshold is  $\theta$ . If CNP is smaller than 90%, partial occlusions is acclaimed; if it is very low, for example smaller than 20%, complete occlusions is accepted. Accordingly, a new occlusion mask is generated where the occluded parts are set zero and others are set one.

#### 3.3 Locating the Object

For each image frame, the estimated template is mapped to the frame by coordinate transformation  $\phi(x, \alpha)$ . The type of the transformation is determined by its parameter vector  $\alpha$ . In this paper, only translation is considered; all other types of object motion are regarded as variations in object appearance. The final location of the object in frame  $n$  is determined by performing masked template matching with

$$\hat{\alpha} = \arg \min \frac{1}{|O|} \sum_{l,i} NCC_i^l(I_n(\phi(x; \alpha)), \hat{T}_i(x) \times O_i) \quad (4)$$

Where  $\hat{\alpha}$  is the estimated transformation parameter vector,  $I_n$  is frame  $n$ ,  $\hat{T}$  is the appearance template,  $l$  and  $i$  represents the part  $i$  in layer  $l$ , and  $O$  is a binary-valued occlusion mask that masks out the occluded template parts.  $|O|$  is used to calculate the number of template parts that are not occluded.

In an implementation,  $\hat{\alpha}$  is first computed by fast-searching (using hierarchical template matching) within a small region predicted by a Kalman filter. At this point, we do not have a reliable occlusion mask yet. If the object is not completely occluded, the final location is refined using (4) in a small region around  $\hat{\alpha}$  with the updated occlusion mask. Otherwise, the final location is the one predicted using the Kalman filter.

To maintain tracking even when an object's appearance varies, we need to adaptively update the appearance

template. Because blind updating can cause template drift, only the non-occluded parts should be considered. Therefore, if the object is not completely occluded, masked template updating is performed by applying temporal smoothing, given by

$$\hat{T} = (\eta \times I_n^* + (1 - \eta) \times \hat{T}) \times O + \hat{T} \times (I - O) \quad (5)$$

where  $I_n^*$  is the best-matched region in frame  $n$ ;  $\eta$  is a decaying coefficient; and  $\mathbf{I}$  is an identity matrix.

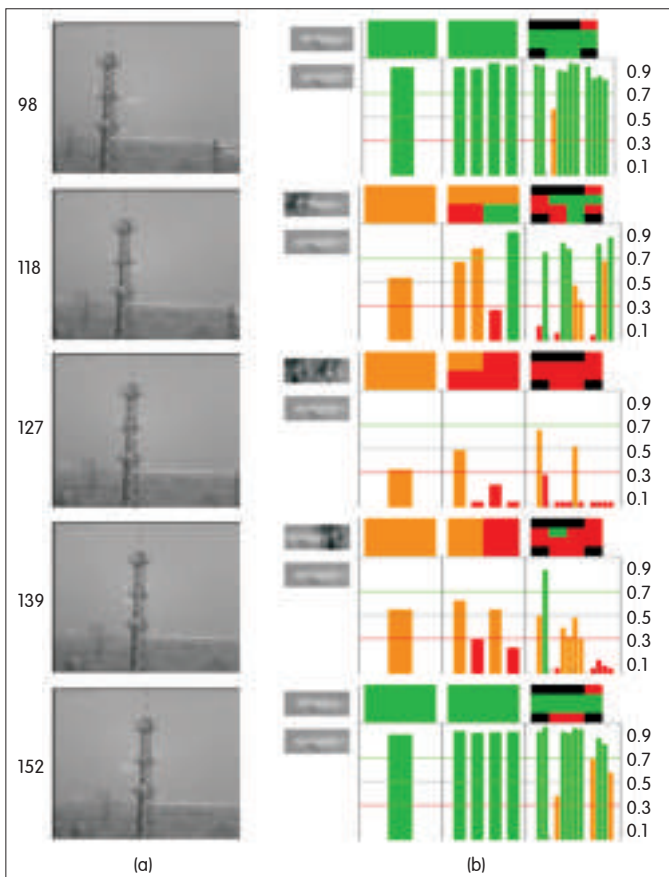
#### 4 Experimental Results

To verify performance, the proposed algorithm is tested on an infrared image sequence of a vehicle taken from the video verification of identity (VIVID) dataset, an optical image sequence of an airplane, and an infrared image sequence of a face captured by us in real-world scenarios. The vehicle sequence has heavy occlusions by tree lines; the airplane sequence has part or complete occlusions by background clutters; and the face sequence has severe occlusions by a hand or book.

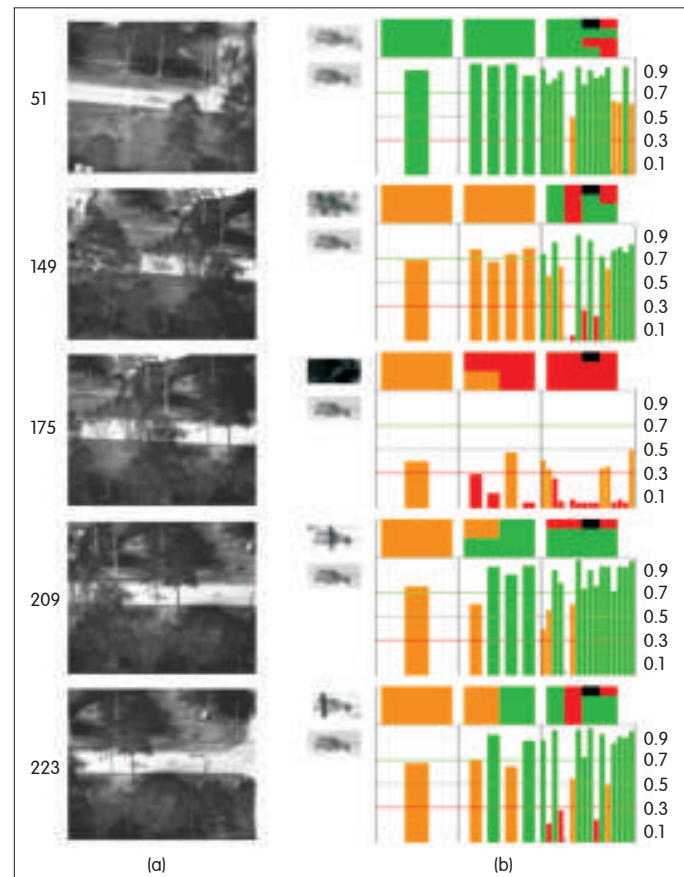
In Figs. 2 and 3, the images in row (a) are the original intensity image, and the white boxes contain the location and size of the target object. The image in row (b) shows the results obtained by applying hierarchical template matching.

The small-intensity image at the upper-left of each of the columns in row (b) is taken from the white-box region in the corresponding image of row (a). In fact, it is the best-matched image to the template or the image from the predicted target region. The small-intensity image at the bottom left of each of the columns in row (b) is the appearance template of the object. The three colored boxes at the top of each column show the results of template matching in layers one to three, from left to right respectively. The bars below these show the corresponding NCC values of template matching, and from these we can evaluate the occlusions. The confidence level of an evaluation is shown using different colors: green means high-level; amber means mid-level; red means low-level; and black means non-informative parts. Fig. 2 shows the results of tracking an airplane in an optical image sequence with severe occlusions. The frame 98, 118, 127, 139 and 152 show the tracker works very well even when the object is completely occluded. In frame 98 and 152, the NCC is very high in the first-layer matching when the object is not occluded. In frame 118 and 139, NCC is low in the first-layer matching but high in the second- or third-layer matching when the object is partly occluded. In frame 127, NCC is very low, even in the lowest third layer matching when the object is completely occluded.

Fig. 3 shows the result of tracking a vehicle in infrared



▲ Figure 2. Tracking an airplane in optical image sequence with severe occlusions.

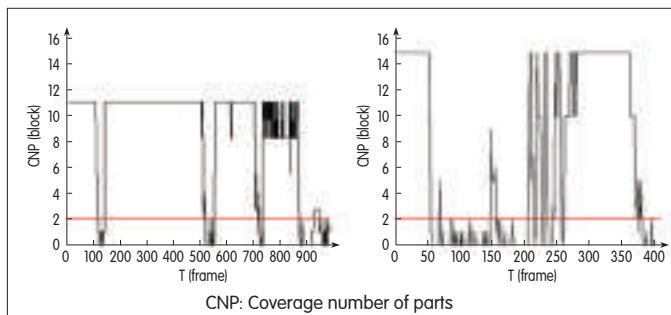


▲ Figure 3. Tracking a vehicle in infrared image sequence with severe occlusions.

image sequence with severe occlusions. The frame 51, 149, 175, 209 and 223 show that the tracker works very well, even when the object is partly or completely occluded. In frame 51, the NCC is very high in the first-layer matching when the object is not occluded. In frame 149, 209 and 139, the NCC is low in the first-layer matching but high in the second- or third-layer matching when the object is partially occluded. In frame 175, the value of NCC is very low, even in the lowest third-layer matching when the object is completely occluded.

Fig. 4 shows the coverage number of parts CNP of hierarchical template matching obtained during the tracking of object. The horizontal axis T shows the index of frames and the vertical axis CNP shows the coverage number of parts CNP. Fig. 4(a) is the result of airplane tracking, and Fig. 4(b) is the result of vehicle tracking. We can easily determine when the object is partly or completely occluded; for example, in Fig. 4(a) from frame 90 to 150, the object is occluded, and in Fig. 4(b), from frame 50 to 250, the object is heavily occluded. Although there are some mistakes, we can still take the proper operation to update the appearance template and predict the location of the object using historic information of motion when the object is occluded. This allows for stable, high-quality tracking.

We developed an active visual tracking system to follow a



▲ Figure 4. Coverage number of parts CNP of hierarchical template matching obtained during object tracking.

single moving object, and the proposed tracking algorithm was encoded in an embedded platform in order to locate the object with the core TMS320DM642 processor—a digital signal processor (DSP) chip from TI Corporation. Fig. 5 shows the result of tracking a face in infrared image sequence with severe occlusions. The sequence was captured by a thermal infrared camera mounted on a gimbal to follow single object. The offset of the location of the tracked object to the center of image yield by our tracking algorithm is fed to a controller to drive the gimbal to follow the object every 40 milliseconds. Thus, it can meet some daily-life requirement of real-time object tracking tasks. In Fig. 5, the red boxes show the location of tracked object. Fig. 5(a) and (b) shows the location of the tracked face when the size of object changes. Fig. 5(c) and (d) shows the location of the tracked face with occlusions.

## 5 Conclusion

To tackle severe occlusions, a hierarchical template



▲ Figure 5. Tracking a face in infrared image sequence with severe occlusions.

matching method for object tracking is proposed. The method integrates holistic- and part-region matching in order to locate the object in a coarse-to-fine manner. A layer-appearance model is introduced to represent the object, and the similarity between a candidate object and the target object is measured by the masked normalized cross-correlation scores of parts in each layer. Occlusions can be easily evaluated using a similarity measure. The proposed method is tested with practical image sequences, and the results show it can consistently provide accurate object location for stable tracking, even for severe occlusions.

The efficiency of the proposed tracking method comes from two aspects. First, hierarchical template matching can correctly evaluate occlusions and adapt the variations of object appearance due to various types of occlusions. Second, the historical information about motion can be applied to predict the location of object accurately, even when it is completely occluded by clutters. The success of the tracker depends on how to correctly evaluate the situation of occlusions.

## Acknowledgement

This work was supported in part by the technique cooperation project of ZTE on Intelligent Video Analysis in 2012.

## References

- [1] A. Yilmaz, O. Javed, and M. Shah, "Object Tracking: A Survey," in *ACM Computing Surveys*, vol. 38, no. 4, pp 1–45, 2006.
- [2] I. Matthews, T. Ishikawa, and S. Baker, "The template update problem," in *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 26, no. 6, pp 810–815, 2004.
- [3] D. Ross, J. Lim, R. Lin, and M. Yang, "Incremental Learning for Robust Visual Tracking," in *International Journal of Computer Vision*, vol. 77, no. 3, pp 125–141, 2008.
- [4] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based object tracking," in *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 25, no. 5, pp 564–577, 2003.
- [5] S. Avidan, "Ensemble tracking," in *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 29, no. 2, pp 261–271, 2007.
- [6] E. Maggio and A. Cavallaro, "Multi-part target representation for color tracking," in



- Proc. ICIP*, no. 1, pp 729–732, 2005.
- [7] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Proc. CVPR*, no. 1, pp 789–805, 2006.
- [8] J. Jeyakar, R. Babu, and K. Ramakrishnan, "Robust object tracking with background-weighted local kernels," in *Computer Vision and Image Understanding*, vol. 112, no. 3, pp 296–309, 2008.
- [9] B. Wu and R. Nevatia, "Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors," in *International Journal of Computer Vision*, vol. 75, no. 2, pp 247–266, 2007.
- [10] S. Dockstader and A. Tekalp, "Multiple camera tracking of interacting and occluded human motion," in *Proceedings of IEEE*, vol. 89, no. 10, pp 1441–1455, 2001.
- [11] Y. Wu, T. Yu, and G. Hua, "Tracking appearances with occlusions," in *Proc. CVPR*, no. 1, pp 789–795, 2003.
- [12] H. Lim, O. Camps, M. Sznajder, and V. Morariu, "Dynamic appearance modeling for human tracking," in *Proc. CVPR*, no. 1, pp 751–757, 2006.
- [13] H. Tao, H. Sawhney, and R. Kumar, "Object Tracking with Bayesian Estimation of Dynamic Layer Representations," in *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 24, no. 1, pp 75–89, 2002.
- [14] M. Yang, Y. Wu, and G. Hua, "Context-aware visual tracking," in *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 31, no. 7, pp 1195–1209, 2009.
- [15] L. Cerman, J. Matas, and V. Hlavac, "Sputnik tracker: Looking for a companion improves robustness of the tracker," in *Proc. Scandinavian Conf. on Image Analysis*, vol. 5575, pp 291–300, 2009.
- [16] H. Grabner, J. Matas, L. Van Gool, and P. Cattin, "Tracking the invisible: Learning where the object might be," in *Proc. CVPR*, no. 1, pp 1285–1292, 2010.
- [17] X. Mei and H. Ling, "Robust visual tracking using l1 minimization," in *Proc. ICCV*, pp 1436–1443, 2009.
- [18] N. Artner, S. Marmol, C. Beleznaï, and W. Kropatsch, "Tracking by Hierarchical Representation of Target Structure," in *Proc. SSPR & SPR 2008*, vol. 5342, pp 441–450, 2008.
- [19] T. Dinh and G. Medioni, "Co-training Framework of Generative and Discriminative Trackers with Partial Occlusion Handling," in *Proc. IEEE WACV*, no. 1, pp 642–649, 2011.
- [20] L. Jin, Z. Bian, X. Li, H. Pan, and S. Xia, "Online real AdaBoost with co-training for object tracking," in *Proc. SPIE*, vol. 7495, no. 2, pp 1–8, 2009.
- [21] A. Jepson, D. Fleet, and T. El-Maraghi, "Robust online appearance Models for Visual Tracking," in *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 25, no. 10, pp 1296–1311, 2003.
- [22] S. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," in *IEEE Trans. on Image Process*, vol. 13, no. 11, pp 1491–1506, 2004.
- [23] H. Nguyen and A. Smeulders, "Fast occluded object tracking by a robust appearance Filter," in *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 26, no. 8, pp 1099–1104, 2004.
- [24] J. Pan, B. Hu, and J. Zhang, "Robust and Accurate Object Tracking Under Various Types of Occlusions," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, no. 2, pp 223–236, 2008.
- [25] S. Kwak, W. Nam, B. Han, and J. Han, "Learning Occlusion with Likelihoods for Visual Tracking," in *Proc. ICCV*, no. 1, pp 1–8, 2011.
- [26] J. Shi and C. Tomasi, "Good Features to Track," in *Proc. CVPR*, no. 1, pp 593–600, 1994.
- [27] G. Bouchard and B. Triggs, "Hierarchical part-based visual object categorization," in *Proc. CVPR*, no. 1, pp 710–715, 2005.
- [28] S. Fidler and A. Leonardis, "Towards scalable representations of object categories: Learning a hierarchy of parts," in *Proc. CVPR*, no. 1, pp 1–8, 2007.
- [29] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," in *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 32, no. 9, pp 1627–1645, 2010.

Manuscript received: August 3, 2012

## **B** iographies

**Lizuo Jin** (jinlizuo@gmail.com) received his Ph.D. degree in Pattern Recognition and Intelligent System from Southeast University, Nanjing, China, in 2000. From 2002 to 2004, he was a post-doctoral fellow at the Institute of Industrial Technology, Tokyo University, Japan. He is now an associate professor at School of Automation, Southeast University, and also a member of Information Fusion Technical Committee of CSAA, China. His research interests include theory and methods for machine learning, pattern recognition, computer vision and embedded systems.

**Tirui Wu** (wu.tirui@zte.com.cn) received his master degree in Computer Science and Engineering from Jiangsu University, Zhenjiang, China. He is now a researcher manager with Nanjing Institute of ZTE corporation. His research interests include theory and application for computer vision, human machine interface, image fusion and GPS signal processing.

**Feng Liu** (liu.feng90@zte.com.cn) received his master degree in Computer Science and Engineering from Chongqing University of Posts and Telecommunications, Chongqing, China. He is now a researcher manager with Chongqing Institute of ZTE corporation. His research interests include theory and application for pattern recognition and artificial intelligence.

**Gang Zeng** (zeng.gang@zte.com.cn) received his master degree in Computer Science and Engineering from Chongqing University, Chongqing, China. He is now a researcher manager with Chongqing Institute of ZTE corporation. His research interests include theory and application for software engineering and computer vision.

## Roundup

### ZTE Launches the First PC-Based CPT for LTE Networks

22 October 2012, Shenzhen—ZTE Corporation announced it has launched the industry's first PC-based capacity planning tool (CPT) for LTE networks. The CPT uses an innovative concept to overcome limitations in capacity planning technology. It provides operators with a professional, systematic aid for building the highest-performance networks.

The CPT is based on a 3GPP protocol and incorporates four technology patents. With a powerful set of system-modeling and simulation features developed using data from real networks, the CPT provides operators with a competitive tool for commercial LTE network deployment. Compared with traditional solutions, ZTE's CPT is 20% more accurate and increases efficiency by more than 80%.

"ZTE's newly released CPT is based on an innovative concept and has one-stop, flexible capacity planning capability. This greatly reduces the difficulty of LTE network planning and improves our customers' operating efficiency," said Wang Shouchen, vice president of ZTE. "The tool's planning accuracy has vastly improved compared with traditional solutions. In the near future, ZTE will also launch other versions based on different network scenarios to meet customer needs."

ZTE is a world-leading wireless solution provider committed to creating resources for the development of LTE and future technologies. As of the end of the Q3 2012, ZTE has won 38 commercial LTE contracts and is working with more than 100 operators across Europe, the Americas, Asia Pacific, and Middle East on trial LTE networks.



# Design and Implementation of ZTE Object Storage System

**Huabin Ruan<sup>1</sup>, Xiaomeng Huang<sup>2</sup>, and Yang Zhou<sup>3</sup>**

(1. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China;

2. Center of Earth System Science, Tsinghua University, Beijing 100084, China;

3. Communication Services R&D Institute, ZTE Corporation, Nanjing 210012, China)

## Abstract

This paper introduces the basic concepts and features of an object storage system. It also introduces some related standards, specifications, and implementations for several existing systems. ZTE's Object Storage System (ZTE OSS) was designed by Tsinghua University and ZTE Corporation and is designed to manage large amounts of data. ZTE OSS has a scalable architecture, some open source components, and an efficient key-value database. ZTE OSS is easy to scale and highly reliable. Experiments show that ZTE OSS performs well with mass data and heavy

## Keywords

cloud storage; object storage

Compared with a file, an object is self-contained and usually contains more metadata. An object can be understood as a better encapsulation of a file. Moreover, it is intelligent because the object system itself can determine the distribution of the physical storage location of objects and the numbers of copies. The storage architecture of objects is also more elastic, so intelligent management can be implemented in the storage layer, and different QoS can be provided to various objects. Finally, objects can teach each other equally. A traditional file system is organized in a tree style, but in an object storage system, all objects are placed flat. This kind of organization is very flexible and allows the creation of different architectures, including the tree style. The container concept exists in some object storage systems, such as Amazon S3, and a container can be seen as a special file. We can also take a container as a special object; then, the object can be classified as a data object and container object.

Now that we have explained what an object is, we measure the advantages of the system. People may doubt the need to encapsulate files into objects because we already have a mature file system. An object storage system has capacity that exceeds that of a single hard disk, disk array, or even more professional storage devices. This capacity will increase markedly. S3 from Amazon emerged in early 2006, and was used to store 20 billion objects in the following two years. This number doubled each year, with more than 50 billion objects stored by 2009 and more than 100 billion objects stored by 2010. If we suppose every object is about 100 kB without considering redundancy, the total capacity required is 10 PB. More importantly, S3 is not limited in service range and is available worldwide. In order to achieve this, sufficient bandwidth guarantee, organization, and allocation are essential.

In 2007, MIT student Drew Houston set up a company to develop a product used for personalized data backup and synchronization. The product was based on Amazon S3 and is the ancestor of Dropbox [4], which attracted millions of users within a year and with only 10 staff. Because of the features of S3, Dropbox developers do not have to worry about fundamental construction and can focus on products and services. In contrast, the forerunner, Kingsoft Kuaipan [5] has to maintain its storage devices and to construct a content delivery network (CDN). Kuaipan has no advantages in terms of R&D costs.

Object storage systems relieve developers from having to construct fundamental infrastructure, but there are also benefits to enterprise customers. Siemens developed a new software distribution and upgrade platform to substitute its former system. The IT department at Siemens no longer has to

## 1 Introduction

Cloud storage has become more well-known over the past few years, and industry has taken a great interest in it. A handful of commercial cloud products have come to the fore, including S3 [1] by Amazon, Windows Azure [2] by Microsoft, and Atmos [3] by EMC. Recently, domestic Chinese companies such as China Mobile and China Telecom have developed corresponding cloud standards and prototypes. According to a report by IDC, the worldwide market for cloud storage systems was worth 1.5 billion dollars in 2009 and will climb to 7 billion dollars by 2014. In this paper, we focus on object storage, specifically, large-scale distributed object storage, which is one of the most important techniques for cloud storage systems.

First, we address the questions of what is an object, and what is the difference between objects and common files used in local file systems to store pictures and documents? We can take the object in an object storage system as the common files we use every day. An object is similar to a file in which documents, pictures, and videos are stored. However, an object also has some differences.

An object usually contains more information than a file does.

worry about maintaining three data center networks and sets of equipment, and operating costs are cut sharply. NBC and GE also set up their IT services on the storage provided by Nirvanix to reduce resource waste and cost.

## 2 Standards for Object Storage Systems

### 2.1 Amazon S3

Since 2006, Amazon has provided developers with S3, which has become the de facto standard for object storage systems. Amazon S3 is based on the idea that high-quality internet storage should be easy to obtain. Then developers no longer need to worry about security, capacity, or how to store their data. S3 frees developers from establishing and maintaining storage solutions a large-scale investment in storage is not required. Amazon S3 has simple and reliable functionality that allows cheap and secure storage of any amount of data, and data is accessible forever. With the help of Amazon S3, developers can concentrate on how to use data instead of how to store data.

An object defined by S3 contains object, bucket, and key. An object comprising object data and metadata is a basic entity stored in Amazon S3. The bucket is the container for storage objects in Amazon S3. Every object is held in a bucket. The key is the only identifier for each object in a bucket; one object in a bucket can only have one key.

For manipulating objects in S3, there are functions such as creating a bucket, writing an object, deleting an object, and listing keys. S3 is a simple storage system that provides object operation semantics to users. Users can operate S3 by putting objects into a bucket and accessing objects from a bucket. There is a simple interface for WEB service that can offer data access on the network anytime, anywhere. S3 uses highly expandable, reliable, fast, and cheap fundamental data-storage infrastructure to run its global websites. Any developer is authorized to use the same data storage infrastructure. S3 aims to expand scale to obtain benefits and pass them on to developers.

### 2.2 SINA CDMI

The storage industry standard organization SINA released a document called "Cloud Data Management Interface (CDMI)" [6]. CDMI is mainly about the client platform and data center server. It defined a standard interface for data exchange between these platforms. HTTP is used to encapsulate the representational state transfer (REST) communication command. The data center responds to client requests for web service and provides the service.

SNIA CDMI was the first cloud storage specification and contains object types such as data object, container object, domain object, queue object, and capability object. Containers can nest and contain objects.

SNIA CDMI defines capabilities, creates a container and an object in container, lists objects in a container, reads the object and deletes object.

SNIA is implementing a prototype system according to the

SNIA CDMI v1.0 specification. This prototype is based on JAVA and will soon be released to the public. China Mobile has also defined a correlated inter-enterprise specification to prepare for entry into the cloud storage field and to provide object-oriented storage.

## 3 ZTE Object Storage System

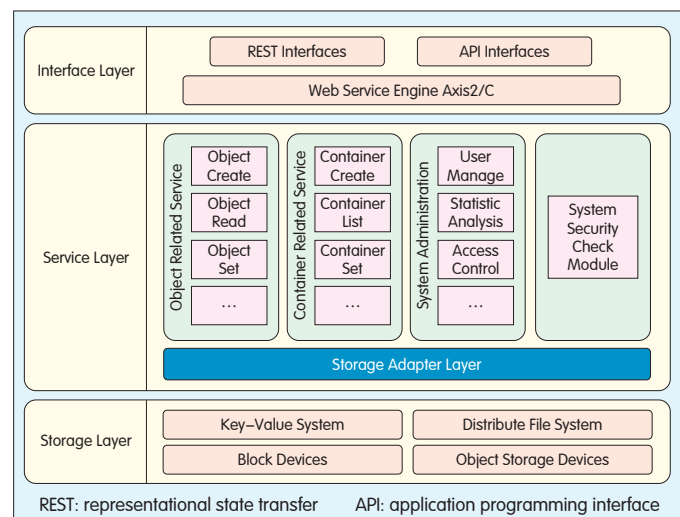
### 3.1 System Architecture

We collaborated with Tsinghua University to implement an object-oriented system [7]. Fig. 1 shows the system architecture, which contains an object interface layer, object service layer, and object storage layer.

The object interface layer provides a REST interface and API interface. The REST interface is used for accessing the object system based on HTTP protocol. The API interface is used to access a web service requested by clients. ZTE's Object Storage System (ZTE OSS) uses Apache Axis2/C, which is a web services/SOAP/WSDL engine that handles the HTTP request [8]. Apache Axis2/C provides a complete object model and a modular architecture that makes it easy to add functionality and support new web-service-related specifications and recommendations. Axis2/C allows the creation and use of a REST-based web service.

The object-service layer mainly provides object management and control, container management and control, and system management and security. To abstract storage resources, this layer provides a series of storage-adaptation interfaces called the adaptation interface layer. Using this method, the system can substitute physical storage devices without modifying the storage interface codes on the service layer, and the system can be scaled. The object-storage layer provides constant storage services for objects and containers.

Because of the interface design, the object-storage layer can adopt various kinds of physical storage, for example, local file system, network file system, distributed parallel file



▲ Figure 1. ZTE OSS architecture.

system, or storage area network.

### 3.2 Metadata Storage

Metadata storage is one of the core modules of ZTE OSS. ZTE OSS has an efficient method for managing metadata. This method requires the underlying key-value storage system to support sorting of items by key. This method is also used in the metadata storage management of PVFS2, a well-known parallel file system. High-performance key-value storage system with key sorting, such as Berkeley DB and Hadoop HBase, are relatively common. With key value and sorting, the frequent operations of an object storage system, such as creating an object, deleting an object, reading an object's data, and listing a container, a can be done efficiently.

In ZTE OSS, the object metadata is saved in a key-value table called Meta Table, and the structure of the container is saved in a key-value table called Entry Table. In Meta Table, object's ID is the key, and the object's attributes are the values. The attributes of an object include object creation time, last modified time, parent information, layout information, and access control list. In Entry Table, there is a special ID for each container, and this is used to associate a container with its subcontainers. We call this special ID the container handle ID. The items related to container handle ID in the Entry Table are

- <container ID + \$, container handle ID> (1)
- <container handle ID + @, sub objects/container counts> (2)
- <container handle ID + sub objects/containers name, sub objects/containers ID> (3)

With (1), we allocate a unique container handle ID for a specific container.

Symbol \$ in (1) is the items unique identifier. Item (2) in Entry Table saves the total number of sub-objects and sub-containers in a container so that we can easily obtain the number of items in a specific container. This is useful for the administrator. Symbol @ in (2) is used to locate this information. Item (3) is used for associating the sub-objects and sub-containers in a specific container. The key of this item is the container handle ID connected with the sub-object and container names, and the values are the IDs of sub-objects or sub-containers, which can be used to locate

▼ Table 1. Entry Table for Fig. 1

Key	Value
1+\$	5
2+\$	6
3+\$	7
5+@	Count=2
5+b	2
5+c	3
6+@	Count=1
6+d	4
7+@	Count=0

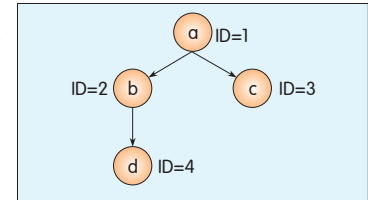
the object detail information in Meta Table.

Tables 1 and 2 are a metadata storage instances for the container structure shown in Fig. 2, where a is a container, b and c are sub-containers, and d is a data object in container b. We assume 1, 2, 3, 4 are the IDs for a, b, c, d, respectively, and 5, 6, 7 are the container handle IDs for the container a, b, c, respectively.

With the above storage

▼ Table 2. Meta Table for Fig. 1

Key	Value
1	Attributes of a
2	Attributes of b
3	Attributes of c
4	Attributes of d



▲ Figure 2. Container structure instance.

mechanism, we can perform frequent object operations efficiently. Taking Fig. 2 as example, if we want to create a new object e ( with ID 10) in container c, we only need to insert (4) in Entry Table and (5) in Meta Table and update (6) in Entry Table.

- <7+e, 10> (4)
- <10, attributes for e> (5)
- <7+@, Count=0> ==> <7 + @, Count=1> (6)

If we want to list a container, we only need to locate the items with prefix 5 and then read items in sequence until the items no longer have this prefix because our underlying key-value storage system is sorted by key.

Other operations such as delete and read objects can also be implemented efficiently with no more than three operations totally to Entry Table and Meta Table.

### 3.3 Data storage

ZTE OSS allow object data to be stored in different devices through the same internal IO interface. In this way, application developers can store their application data in a simple way using the same object interface provided by ZTE OSS with different layout information in the request message. In ZTE OSS, we use layout for the data storage mechanism. Layout is one of the object attributes that indicates where object data is to be stored. As well as layout in the request message, we also have a storage adapter layer, which is a library used to save data in different storage devices according to layout value in object operation request message. The object's public interface in the adapter layer is

```
status object_write(const void *data,
size_t data_size, const char *layout) (7)
```

where argument data is the object data to be written; data\_size is the length to be written, and layout is the value specifying the device to be written to.

ZTE OSS supports four storage devices that are currently being adapted. These devices and their layout values are shown in Table 3 and can be extended easily when new storage devices are added. We only need to add a new value for the layout and implement new private I/O interface for a new device. This is transparent to the user because the object public interface is not changed. For example, when the layout is key-value, object data will be stored together with the object attribute through the object attribute object\_data field; that is, object data is stored directly in Meta Table. Small data, such as configuration data, is suitable in this case. When layout is DFS, the object data is saved in a distributed file

system, which is suitable for storing large data such as streaming media data. Other values for the layout and corresponding storage devices are shown in Table 3.

### 3.4 Security Checking

ZTE OSS has a security mechanism that ensures data is transmitted safely. It encrypts the key content in the object

▼ Table 3. Layout for object data

Layout	Corresponding Storage Device
Key-Value	Sorted Key-Value system, such as Berkeley DB, Hadoop HBase
DFS	Distributed File System, such as PVFS2, ZTE DFS.
OSD	Object Storage Device.
BLOCK	Block Device, such Storage Area Net (SAN)

request message. The reason we only encrypt the key content and not all the content in a request message is performance. The time taken for encryption depends on the length of the message, and checking the key content is enough to determine whether the message has been modified during network transmission. Therefore, in order to reduce the affect of security checking on performance of the ZTE OSS, we only encrypt the key content in the request message. The algorithm we use for encrypting the key content is SHA1. The key content in the request message includes

- HTTP method. The method used in HTTP request message, include: PUT, GET, POST, DELETE, HEAD.
- object or container information. This is the ID or name of the object and container.
- user access key. This value can be user name or user ID in ZTE OSS.
- request time. This is the timestamp of current request generated time.

We create a container called MyContainer and assume this request is issued by user James on 03-08-2012 at 12:00:00. The string to be encrypted will be

StringToSign = "PUT" + "\n" + "MyContainer" + "\n"  
+ "James" + "\n" + "03-08-2012 12:00:00" (8)

According to the semantic of HTTP, creating a container is like a creating new resource, so we use PUT as HTTP method to create container.

### 3.5 Functions and Interfaces

The interface provides API in C language and REST in HTTP message. Object-correlated functions include object management and control. Object management involves creating an object, deleting an object, copying an object, moving an object, setting an object's metadata, reading an object, and reading the metadata of an object.

Container-correlated functions include container management and control. Container management includes creating a container, deleting a container, listing all objects of a container, reading metadata of container, setting metadata of container, and so on.

Security is ensured through validating key information in

request. The key information in request should be encrypted in MD5.

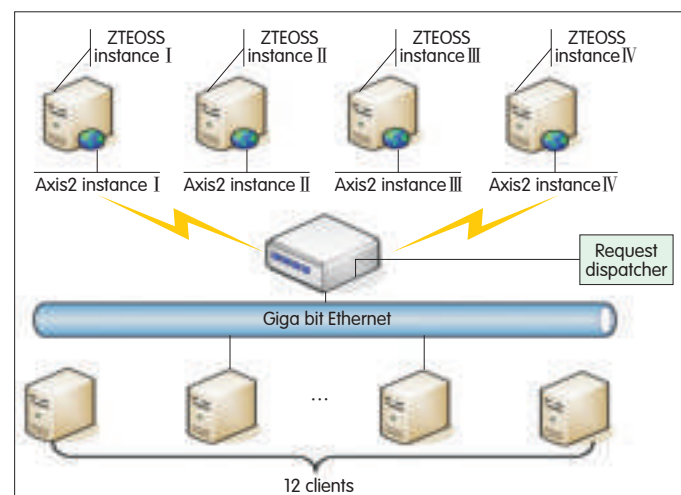
## 4 Testing and Performance

ZTE OSS parallel testing and capacity testing are based on an eight-node Linux cluster (Fig. 3). Parallel testing means testing the number of requests completed during the unit time using Request/Per Second as the unit. Capacity testing involves testing the maximum quality of the objects stored in ZTE OSS with fixed hardware resources. Every node in the Linux cluster has the same software and hardware configuration (Table 4). The object storage system was deployed on four nodes of the cluster. Then, the other four nodes were configured as the clients sent the request. The nodes from Client\_1 to Client\_12 sent the request to the nodes from OSS\_1 to OSS\_4. Every client sent the common operations to the OSS, creating an object, creating a container, deleting an object, and listing all of objects in a container.

The result of the testing showed that the degree of parallelism (DOP) between the common ZTE OSS operations—creating and deleting objects and containers, and listing all objects in a container—is the same as that using Axis2/C parsing. With the same software and hardware configuration, the DOP for Axis2/C parsing is up to 1800 requests per second, and DOP of common ZTE OSS operations (except listing the containers) is greater than 1700 request per second. The DOP of the list container depends on the number of objects in a container. When the number of objects in a container is less than or equal to 10, DOP is 1724

▼ Table 4. Software and hardware configuration of the node in the Linux cluster

Hardware	CPU Intel i5 2.8 GHz, RAM 4 GB, Hard disk 320 GB
Operating system	Redhat 5.5 Enterprise Edition 64 bits
Software	Axis2/C 1.7.0 RC1, BDB, Intel Compiler icc 12.0.2
Network	Gigabit Ethernet



▲ Figure 3. ZTE OSS testing networks and nodes.



request per second, equivalent to that of Axis2/C parsing. If the number of objects in a container is 100, the DOP is 1307 requests per second. If the number of objects in a container is 1000, the DOP is 281 requests per second. Creating a name list requires parsing the object name and merging into the name list, and it is spent on List container operation depending on the number of the objects in a container. The result of test shows that the DOP is not linear with the number of objects in a container. The result of parallel testing is shown in Table 5.

▼ Table 5. ZTE OSS parallel test result

Operation	Duration (ms)	Operation Counts	Description	Parallel (Request/Per Second)
Create Object	5980	10,000	Create 10,000 objects (size=1024 bytes)	1672
Create Container	5667	10,000	Create 10,000 containers	1764
List Container (10 objects)	580	1000	List container, 10 objects	1724
List Container (100 objects)	765	1000	List container, 100 objects	1307
List Container (1000 objects)	3550	1000	List container, 1000 objects	281
Remove Object	5679	10,000	Remove 10,000 objects	1760

Capacity testing shows that the capacity of ZTE OSS depends on the idle hardware resources (Table 6). A greater the number of idle hardware resources in the system can

▼ Table 6. Capacity test result

Operation	Duration (Hour)	Count (Times =10,000)	Description	Objects (number =10,000)
Create Object	4	2350	Create objects (size = 1024 bytes), until disk full or system hold.	2350

accommodate more objects. For example, with 1 GB free memory and 100 GB free disk in the system, 23.5 million objects can be contained in ZTE OSS. According to the test results, the performance of ZTE OSS on the parallel testing and capacity testing is good [9].

## 5 Conclusion

Popular object storage systems are imperfect in areas such

as data access, consistency, data encryption, data ownership, and data isolation. When evaluating cloud storage or object-based storage, most companies are usually concerned with safety and reliability. As a mature commercial object storage system, Amazon S3 has been serving many companies and organizations for years. However, companies using Amazon S3 have suffered losses because of breakdowns in the system. Developers need to evaluate benefits and risks and do reasonable tradeoffs.

## References

- [1] Amazon Inc. (2006). *Simple Storage Service* [Online]. Available: <http://aws.amazon.com/s3/>
- [2] Microsoft Inc. (2009). *Windows Azure* [Online]. Available: <http://www.microsoft.com/windowsazure/>
- [3] EMC Inc. (2010). *Atmos* [Online]. Available: <http://www.emc.com/storage/atmos/atmos.htm>
- [4] Dropbox. (2007). [Online]. Available: <http://www.dropbox.com/>
- [5] Kingsoft kuaipan. [Online]. Available: [www.kuaipan.cn/](http://www.kuaipan.cn/)
- [6] SNIA CDMI Working Group. (2010). *CDMI Specification* [Online]. Available: <http://cdmi.sniacloud.com/>
- [7] Wenying Zeng, Yuelong Zhao, Kairi Ou, and Wei Song, "Research on cloud storage architecture and key technologies," *Proceedings of ICIS 2009*, pp. 1044–1048.
- [8] Apache Software Foundation. (2005). *Apache Axis2/C Project* [Online]. Available: <http://axis.apache.org/axis2/c/core/>
- [9] R. Buyya, R. Ranjan, and N Rodrigo, "CalheirosInterCloud: utility-oriented federation of cloud computing environments for scaling of application services," *Lecture Notes in Computer Science*, Vol. 6081, 13–31, 2010.

Manuscript received: August 20, 2012

## Biographies

**Huabin Ruan** (ruanhuabin@gmail.com) is a PhD candidate in the Department of Computer Science and Technology, Tsinghua University, China. His research interests include acceleration of applications on multiply platforms such as CPU, GPU, FPGA. He has applied for patent on mass object storage architecture.

**Xiaomeng Huang** (hxm@tsinghua.edu.cn) received his doctoral degree from Tsinghua University. He is an associate professor at the Center for Earth System Science, Tsinghua University, China. His research interests include Earth system model, mass data processing, and distributed computing. He has published 24 papers and holds more than 10 patents.

**Yang Zhou** (zhou.yang1@zte.com.cn) received his master's degree from Huazhong University of Science and Technology, China. He is a researcher at ZTE Corporation. His research interests include object storage, data compression, distribute computing. He holds more than 20 patents.

AD Index

Back Cover:  
ZTE Corporation



# ZTE Communications

## Table of Contents, Volume 10, Numbers 1–4, 2012

Volume–Number–Page

### SPECIAL TOPICS

#### 100G and Beyond: Trends in Ultrahigh-Speed Communications (part I)

Guest Editorial.....	Gee–Kung Chang, Jianjun Yu, and Xiang Wang	10–1–01
High Spectral Efficiency 400G Transmission.....	Xiang Zhou	10–1–03
Greater than 200 Gb/s Transmission Using Direct–Detection		
Optical OFDM Superchannel .....	Wei–Ren Peng, Itsuro Morita, Hidenori Takahashi, and Takehiro Tsuritani	10–1–10
Spatial Mode–Division Multiplexing for High–Speed Optical Coherent		
Detection Systems .....	William Shieh, An Li, Abdullah Al Amin, Xi Chen, Simin Chen, and Guanjuan Gao	10–1–18
Exploiting the Faster–Than–Nyquist Concept in Wavelength–Division Multiplexing Systems		
Using Duobinary Shaping.....	Jianqiang Li, Ekawit Tipsuwannakul, Magnus Karlsson, and Peter A. Andrekson	10–1–23
Super–Receiver Design for Superchannel Coherent Optical Systems		
.....	Cheng Liu, Jie Pan, Thomas Detwiler, Andrew Stark, Yu–Ting Hsueh, Gee–Kung Chang, and Stephen E. Ralph	10–1–30
Design of a Silicon–Based High–Speed Plasmonic Modulator .....	Mu Xu, Jiayang Wu, Tao Wang, and Yikai Su	10–1–34
The Key Technology in Optical OFDM–PON .....	Xiangjun Xin	10–1–40
Compensating for Nonlinear Effects in Coherent–Detection Optical Transmission Systems.....	Fan Zhang	10–1–45
1 Tb/s Nyquist–WDM PM–RZ–QPSK Superchannel Transmission		
over 1000 km SMF–28 with MAP Equalization.....	Ze Dong, Jianjun Yu, and Hung–Chang Chien	10–1–50

#### Emerging Technologies for Multimedia Coding, Analysis and Transmission

Guest Editorial.....	Huifang Sun and Dong Wang	10–2–01
Introduction to the High–Efficiency Video Coding Standard .....	Ping Wu and Ming Li	10–2–02
Recent MPEG Standardization Activities on 3D Video Coding .....	Yichen Zhang and Lu Yu	10–2–09
AVS 3D Video Coding Technology and System .....	Siwei Ma, Shiqi Wang, and Wen Gao	10–2–13
Configurable Media Codec Framework: A Stepping Stone for Fast		
and Stable Codec Development .....	Euee S. Jang	10–2–19
Lattice Vector Quantization Applied to Speech and Audio Coding .....	Minjie Xie	10–2–25
Noise Feedback Coding Revisited: Refurbished Legacy Codecs		
and New Coding Models .....	Stéphane Ragot, Balázs Kövesi, and Alain Le Guyader	10–2–34
MMT: The Next–Generation Media Transport Standard .....	Gerard Fernando	10–2–45
Low–Complexity Error–Control Methods for Scalable Video Streaming.....	Zhijie Zhao and Jörn Ostermann	10–2–49
Key Technologies in Mobile Visual Search and MPEG		
Standardization Activities .....	Ling–Yu Duan, Jie Chen, Chunyu Wang, Rongrong Ji, Tiejun Huang, and Wen Gao	10–2–57

#### 100G and Beyond: Trends in Ultrahigh-Speed Communications (part II)

Guest Editorial.....	Gee–Kung Chang and Jianjun Yu	10–3–01
FSK Modulation Scheme for High–Speed		
Optical Transmission .....	Nan Chi, Wuliang Fang, Yufeng Shao, Junwen Zhang, and Li Tao	10–3–02
Computationally Efficient Nonlinearity Compensation for Coherent Fiber–Optic Systems .....	Kai Zhu and Guifang Li	10–3–12

# ZTE Communications

## Table of Contents, Volume 10, Numbers 1–4, 2012

### Volume–Number–Page

Flipped–Exponential Nyquist Pulse Technique to Optimize the PAPR in Optical Direct Detection OFDM System .....	Jiangnan Xiao, Zizheng Cao, Fan Li, Jin Tang, and Lin Chen	10–3–16
100Gbit/s Nyquist–WDM PDM 16–QAM Transmission over 1200 km SMF–28 with Ultrahigh Spectrum Efficiency .....	Zeng Dong	10–3–22
Field Transmission of 100G and Beyond: Multiple Baud Rates and Mixed Line Rates Using Nyquist–WDM Technology .....	Zhensheng Jia, Jianjun Yu, Hung–Chang Chien, Ze Dong, and Di Huo	10–3–28

### Millimeter Wave Communication for Cellular and Cellular–802.11 Hybrid Networks

Guest Editorial .....	Philip Pietraski and I–tai Lu	10–4–01
Millimeter Wave and Terahertz Communications: Feasibility and Challenges .....	Phil Pietraski, David Britz, Arnab Roy, Ravi Pragada, and Gregg Charlton	10–4–03
WiGig and IEEE 802.11ad for Multi–Gigabyte–Per–Second WPAN and WLAN .....	Sai Shankar N, Debashis Dash, Hassan El Madi, and Guru Gopalakrishnan	10–4–13
Modeling Human Blockers in Millimeter Wave Radio Links .....	Jonathan S. Lu, Daniel Steinbach, Patrick Cabrol, and Philip Pietraski	10–4–23
60 GHz SIW Steerable Antenna Array in LTCC .....	Bahram Sanadgol, Sybille Holzwarth, Peter Uhlig, Alberto Milano, and Rafi Popovich	10–4–29
Light–of–Sight MIMO for Next–Generation Microwave Transmission Systems .....	Xianwei Gong, Zhifeng Yuan, Jun Xu, and Liujun Hu	10–4–33

## RESEARCH PAPERS

Hardware Architecture of Polyphase Filter Banks Performing Embedded Resampling for Software–Defined Radio Front–Ends .....	Mehmood Awan, Yannick Le Moullec, Peter Koch, and Fred Harris	10–1–54
A Histogram–Based Static Error Correction Technique for Flash ADCs: Implementation .....	J Jacob Wikner, Armin Jalili, Sayed Masoud Sayedi, and Rasoul Dehghani	10–1–63
Open Augmented Reality Standards: Current Activities in Standards–Development Organizations .....	Christine Perey	10–3–39
Mobile Cloud for Personalized Any–Media Services .....	Bhumip Khasnabish	10–3–47
Multiple–Constraint–Aware RWA Algorithms based on a Comprehensive Evaluation Model: Use in Wavelength–Switched Optical Networks .....	Hui Yang, Yongli Zhao, Shanguo Huang, Daijiang Wang, Xuping Cao, and Xuefeng Lin	10–3–55
Terabit Superchannel Transmission: A Nyquist–WDM Signals Approach .....	Hung–Chang Chien, Jianjun Yu, Zhensheng Jia, and Ze Dong	10–4–39
Parallel Web Mining System Based on Cloud Platform .....	Shengmei Luo, Qing He, Lixia Liu, Xiang Ao, and Fuzhen Zhuang	10–4–45
Hierarchical Template Matching for Robust Visual Tracking with Severe Occlusions .....	Lizuo Jin, Tirui Wu, Feng Liu, and Gang Zeng	10–4–54
Design and Implementation of ZTE Object Storage System .....	Huabin Ruan, Xiaomeng Huang, and Yang Zhou	10–4–60