IEEE
Signal Processing
MAGAZINE

[VOLUME 33 NUMBER 1 JANUARY 2016]

TAKING SPS TO NEW HEIGHTS

BAYESIAN MACHINE LEARNING FOR EEG/MEG

DECISION LEARNING

BLOCK-STRUCTURED OPTIMIZATION FOR BIG DATA

COMPRESSIVE COVARIANCE SENSING

GAME THEORY FOR NETWORKS

COMBINATIONS OF ADAPTIVE FILTER

IEEE Signal Processing Society®

IEEE

SIGNAL PROCESSING

The Science Behind Our Digital Life

Join SPS or Renew Now For 2016

Membership in the IEEE Signal Processing Society (SPS) can help you lay the foundation for many years of success ahead.

CONNECT with more than 19,000 signal processing professionals through SPS conferences, workshops, and local events hosted by more than 170 SPS Chapters worldwide.

SAVE with member-only discounts on conferences, publications, SigPort document archive, and access to travel grants.

ADVANCE with world-class educational resources, awards and recognition, and society-wide volunteer opportunities in publications, conferences, membership, and more.

STUDENTS jumpstart your career with student-focused events, competitions, career opportunities, and eligibility for travel grants.

signalprocessingsociety.org

Digital Object Identifier 10.1109/MSP.2015.2502419

CONTENTS

[VOLUME 33 NUMBER 1]

FEATURES

LEARNING

14 **BAYESIAN MACHINE LEARNING**
Wei Wu, Srikantan Nagarajan, and Zhe Chen

37 **DECISION LEARNING**
Yan Chen, Chunxiao Jiang, Chih-Yu Wang, Yang Gao, and K.J. Ray Liu

BIG DATA

57 **A UNIFIED ALGORITHMIC FRAMEWORK FOR BLOCK-STRUCTURED OPTIMIZATION INVOLVING BIG DATA**
Mingyi Hong, Meisam Razaviyayn, Zhi-Quan Luo, and Jong-Shi Pang

THEORY AND METHODS

78 **COMPRESSIVE COVARIANCE SENSING**
Daniel Romero, Dyonisius Dony Ariananda, Zhi Tian, and Geert Leus

94 **GAME THEORY FOR NETWORKS**
Giacomo Bacci, Samson Lasaulce, Walid Saad, and Luca Sanguinetti

120 **COMBINATIONS OF ADAPTIVE FILTERS**
Jerónimo Arenas-García, Luis A. Azpicueta-Ruiz, Magno T.M. Silva, Vítor H. Nascimento, and Ali H. Sayed

Digital Object Identifier 10.1109/MSP.2015.2479476

COLUMNS

4 **FROM THE EDITOR**
Women in Science, Engineering, and Signal Processing
Min Wu

6 **PRESIDENT'S MESSAGE**
We Need Your Help to Take the Society to New Heights
Alex Acero

11 **SPECIAL REPORTS**
Signal Processing Research Resonates with Hearing Loss Sufferers
John Edwards

141 **REFLECTIONS**
James L. Flanagan: A Scholar and a True Gentleman
Lawrence Rabiner

144 **LECTURE NOTES**
Polyphase Channelizer Demystified
P. Murali Krishna and T.P. Sameer Babu

151 **SP TIPS&TRICKS**
Cascading Tricks for Designing Composite Filters with Sharp Transition Bands
David Shiung, Ya-Yin Yang, and Chu-Sing Yang

158 **STANDARDS IN A NUTSHELL**
Next-Generation Broadcast Television: ATSC 3.0
Richard Chernock and Jerry C. Whitaker

163 **BOOK REVIEWS**
Foundations of Signal Processing
Andres Kwasinski

Handbook of Digital Forensics of Multimedia Data and Devices
Luisa Verdoliva

166 **BOOK DIGEST**

DEPARTMENTS

7 **SOCIETY NEWS**
2016 Class of Distinguished Lecturers

IEEE Signal Processing Society Young Professionals Meet During ICIP 2015

168 **DATES AHEAD**

IEEE SIGNAL PROCESSING magazine

EDITOR-IN-CHIEF
Min Wu—University of Maryland, College Park, United States

AREA EDITORS
Feature Articles
Shuguang Robert Cui—Texas A&M University, United States

Special Issues
Wade Trappe—Rutgers University, United States

Columns and Forum
Gwenaël Doërr—Technicolor Inc., France
Kenneth Lam—Hong Kong Polytechnic University, Hong Kong SAR of China

e-Newsletter
Christian Debes—TU Darmstadt and AGT International, Germany

Social Media and Outreach
Andres Kwasinski—Rochester Institute of Technology, United States

EDITORIAL BOARD
A. Enis Cetin—Bilkent University, Turkey
Patrick Flandrin—ENS Lyon, France
Mounir Ghogho—University of Leeds, United Kingdom
Lina Karam—Arizona State University, United States
Bastiaan Kleijn—Victoria University of Wellington, New Zealand and Delft University, The Netherlands
Hamid Krim—North Carolina State University, United States
Ying-Chang Liang—Institute for Infocomm Research, Singapore
Sven Lončarić—University of Zagreb, Croatia
Brian Lovell—University of Queensland, Australia
Henrique (Rico) Malvar—Microsoft Research, United States
Stephen McLaughlin—Heriot-Watt University, Scotland
Athina Petropulu—Rutgers University, United States
Peter Ramadge—Princeton University, United States
Shigeki Sagayama—Meiji University, Japan
Eli Saber—Rochester Institute of Technology, United States
Erchin Serpedin—Texas A&M University, United States
Shihab Shamma—University of Maryland, United States
Hing Cheung So—City University of Hong Kong, Hong Kong
Isabel Trancoso—INESC-ID/Instituto Superior Técnico, Portugal
Michail K. Tsatsanis—Entropic Communications
Pramod K. Varshney—Syracuse University, United States
Z. Jane Wang—The University of British Columbia, Canada
Gregory Wornell—Massachusetts Institute of Technology, United States
Dapeng Wu—University of Florida, United States

ASSOCIATE EDITORS—COLUMNS AND FORUM
Ivan Bajic—Simon Fraser University, Canada
Rodrigo Capobianco Guido—São Paulo State University
Ching-Te Chiu—National Tsing Hua University, Taiwan
Michael Gormish—Ricoh Innovations, Inc.
Xiaodong He—Microsoft Research
Danilo Mandic—Imperial College, United Kingdom
Aleksandra Mojsilovic—IBM T.J. Watson Research Center
Douglas O'Shaughnessy—INRS, Canada
Fatih Porikli—MERL
Shantanu Rane—PARC, United States
Saeid Sanei—University of Surrey, United Kingdom
Roberto Togneri—The University of Western Australia
Alessandro Vinciarelli—IDIAP–EPFL
Azadeh Vosoughi—University of Central Florida
Stefan Winkler—UIUC/ADSC, Singapore

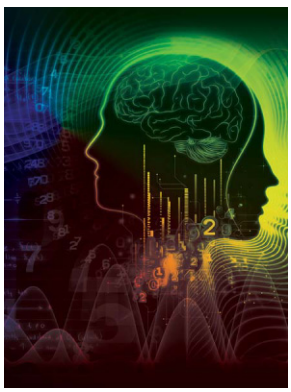
ASSOCIATE EDITORS—e-NEWSLETTER
Csaba Benedek—Hungarian Academy of Sciences, Hungary
Paolo Braca—NATO Science and Technology Organization, Italy
Quan Ding—University of California, San Francisco, United States
Pierluigi Failla—Compass Inc, New York, United States
Marco Guerriero—General Electric Research, United States
Yang Li—Harbin Institute of Technology, China
Yuhong Liu—Penn State University at Altoona, United States
Andreas Merentitis—University of Athens, Greece
Michael Muma—TU Darmstadt, Germany

IEEE SIGNAL PROCESSING SOCIETY
Rabab Ward—*President*
Ali Sayed—*President-Elect*
Carlo S. Regazzoni—*Vice President, Conferences*
Konstantinos (Kostas) N. Plataniotis—*Vice President, Membership*
Thrasyvoulos (Thrasos) N. Pappas—*Vice President, Publications*
Walter Kellerman—*Vice President, Technical Directions*

IEEE SIGNAL PROCESSING SOCIETY STAFF
Denise Hurley—Senior Manager of Conferences and Publications
Rebecca Wollman—Publications Administrator

[COVER]

©ISTOCKPHTOTO.COM/AGSANDREW



IEEE PERIODICALS MAGAZINES DEPARTMENT

Jessica Barragué
Managing Editor

Geraldine Krolin-Taylor
Senior Managing Editor

Mark David
Senior Manager, Advertising and Business Development

Felicia Spagnoli
Advertising Production Manager

Janet Dudar
Senior Art Director

Gail A. Schnitzer, Mark Morrissey
Associate Art Directors

Theresa L. Smith
Production Coordinator

Dawn M. Melley
Editorial Director

Peter M. Tuohy
Production Director

Fran Zappulla
Staff Director, Publishing Operations

IEEE prohibits discrimination, harassment, and bullying. For more information, visit http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html.

IEEE SIGNAL PROCESSING MAGAZINE (ISSN 1053-5888) (ISPREG) is published bimonthly by the Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, 17th Floor, New York, NY 10016-5997 USA (+1 212 419 7900). Responsibility for the contents rests upon the authors and not the IEEE, the Society, or its members. Annual member subscriptions included in Society fee. Nonmember subscriptions available upon request. Individual copies: IEEE Members US$20.00 (first copy only), nonmembers US$234.00 per copy. Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. Copyright Law for private use of patrons: 1) those post-1977 articles that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 USA; 2) pre-1978 articles without fee. Instructors are permitted to photocopy isolated articles for noncommercial classroom use without fee. For all other copying, reprint, or republication permission, write to IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854 USA. Copyright ©2016 by the Institute of Electrical and Electronics Engineers, Inc. All rights reserved. Periodicals postage paid at New York, NY, and at additional mailing offices. Postmaster: Send address changes to IEEE Signal Processing Magazine, IEEE, 445 Hoes Lane, Piscataway, NJ 08854 USA. Canadian GST #125634188

Printed in the U.S.A.

Digital Object Identifier 10.1109/MSP.2015.2479477

SUSTAINABLE FORESTRY INITIATIVE Certified Sourcing www.sfiprogram.org SFI-01042

SCOPE: *IEEE Signal Processing Magazine* publishes tutorial-style articles on signal processing research and applications, as well as columns and forums on issues of interest. Its coverage ranges from fundamental principles to practical implementation, reflecting the multidimensional facets of interests and concerns of the community. Its mission is to bring up-to-date, emerging and active technical developments, issues, and events to the research, educational, and professional communities. It is also the main Society communication platform addressing important issues concerning all members.

General Chair
Lina Karam
Arizona State University

Vice-General Chair
Aggelos Katsaggelos
Northwestern University

Technical Program Chairs
Fernando Pereira
Instituto Superior Técnico
Gaurav Sharma
University of Rochester

Innovation Program Chairs
Haohong Wang
TCL Research America
Jeff Bier
BDTI & Embedded Vision Alliance
Khaled El-Maleh
Qualcomm Technologies Inc.

Finance Chair
Sohail Dianat
Rochester Institute of Technology

Plenary Chairs
Michael Marcellin
University of Arizona
Sethuraman Panchanathan
Arizona State University

Special Sessions Chairs
Dinei Florencio
Microsoft Research
Chaker Larabi
Poitiers University
Zhou Wang
University of Waterloo

Challenge Sessions Chair
Dinei Florencio
Microsoft Research

Tutorials Chairs
Ghassan AlRegib
Georgia Tech
Rony Ferzli
Intel

Publicity Chair
Michel Sarkis
Qualcomm Technologies Inc.

Paper Awards Chairs
Vivek Goyal
Boston University
Ivana Tosic
Ricoh Innovations

Exhibits Chair
David Frakes
Arizona State University & Google

Publication Chairs
Patrick Le Callet
Nantes University
Baoxin Li
Arizona State University

Local Arrangement Chair
Pavan Turaga
Arizona State University

Registration Chair
Ricardo De Queiroz
Universidade de Brasilia

The 23rd IEEE International Conference on Image Processing (ICIP) will be held in the Phoenix Convention Centre, Phoenix, Arizona, USA, on September 25 - 28, 2016. ICIP is the world's largest and most comprehensive technical conference focused on image and video processing and computer vision. In addition to the Technical Program, ICIP 2016 will feature an Innovation Program focused on vision technologies and fostering innovation and networking. The conference will feature world-class speakers, tutorials, exhibits, and **a vision technology showcase**.

Topics in the ICIP 2016 Technical Program include but are not limited to:

Filtering, Transforms, Multi-Resolution Processing
Restoration, Enhancement, Super-Resolution
Computer Vision Algorithms and Technologies
Compression, Transmission, Storage, Retrieval
Computational Imaging
Color and Multispectral Processing
Multi-View and Stereoscopic Processing
Multi-Temporal and Spatio-Temporal Processing
Video Processing and Analytics
Authentication and Biometrics
Biological and Perceptual-based Processing
Visual Quality Assessment
Scanning, Display, and Printing
Document and Synthetic Visual Processing
Applications to various fields

New initiatives at ICIP 2016 include:

1) Early free access for accepted papers on IEEE Xplore; 2) Visual Innovation Award (individual or team nominations due by 31 March 2016 at conference website); 3) Support for reproducible research; 4) Support for CV uploads on the ICIP site for full-time, part-time, and consulting job opportunities; 5) Visual Technology Showcase. For more details on these and other new initiatives at ICIP 2016, visit 2016.ieeeicip.org and connect now on the ICIP 2016 social media to get automatic updates about the various deadlines, sessions and events.

Paper Submission:

Prospective authors are invited to submit full-length papers at the conference website, with up to four pages for technical content including figures and references, and with one additional optional 5th page for references only. Submission Instructions, templates for the required paper format, and information on "no show" policy are available at **2016.ieeeicip.org**.

Tutorials, Special Sessions, and Challenge Sessions Proposals:

Tutorials will be held on September 25, 2016. Tutorial proposals should be submitted at the conference website and must include title, outline, contact information, biography and selected publications for the presenter(s), and a description of the tutorial and material to be distributed to participants. For detailed submission guidelines, please refer to the tutorial proposals page. Special Sessions and Challenge Session Proposals should be submitted at conference website and must include a topical title, rationale, session outline, contact information, and a list of invited papers/participants. For detailed submission guidelines, please refer the ICIP 2016 website at **2016.ieeeicip.org**.

Important Deadlines:

Challenge Session Proposals: October 31, 2015

Special Session and Tutorial Proposals: November 16, 2015
Notification of Special Session and Tutorial Acceptance: December 18, 2015

Paper Submissions: January 25, 2016
Notification of Paper Acceptance: April 30, 2016

Visual Inovation Award Nomination: March 31, 2016

Revised Paper Upload Deadline: May 15, 2016
Authors' Registration Deadline: May 15, 2016

Digital Object Identifier 10.1109/MSP.2015.2484518

[from the EDITOR]

Min Wu
Editor-in-Chief
minwu@umd.edu

Women in Science, Engineering, and Signal Processing

It will almost be 2016 when you receive this issue of *IEEE Signal Processing Magazine*. Happy New Year!

I am writing this editorial during a season of harvesting. The scientific community of my native country of China experienced a breakthrough this past fall. Female scientist Youyou Tu, who has worked for over half a century at the intersection of pharmaceutical chemistry and Chinese traditional medicine, was awarded a Nobel Prize in Medicine. She was the very first citizen of mainland China who received a Nobel Prize in Natural Science.

Tu was recognized for her trailblazing contribution in extracting a core ingredient from Chinese herbs to treat malaria, saving millions of lives worldwide from this deadly disease. The key to the success was inspired by her studying ancient Chinese medicine literature from 340 AD, with modern chemistry interpretations and experimentations. Despite her leading role in carrying out the critical research, China's political turmoil during the cultural revolution put individual roles in research recognitions and publishing into such a taboo that Tu's contributions were marginalized for many years, and she has never been elected to the national academy. Yet her devotion to science and to saving lives as well as strong family support have propelled her throughout her career.

Speaking of women in science, another female Nobel Laureate, Marie Curie, widely known as Madame Curie, has been perhaps the most popular role model who inspired many girls to pursue science and engineering. In many ways, Marie Curie was seen almost like a goddess with incredible capabilities and personal strength: her trailblazing career as one of the first women to pursue scientific research when very few science colleges were open to women students and even fewer universities appointed females to their science faculty; her balanced life as both a leading scientist and a devoted wife and mother; and her ability to overcome the sudden tragic death of her husband, Pierre Curie, to continue making profound contributions in science while raising two daughters at the same time (her elder daughter later became a Nobel Laureate herself).

As I was reading a biography about this female pioneer, I learned that, throughout her childhood and professional development, she was surrounded by a strong support network: her father instilled in her interests in science since her childhood; she had supportive faculty mentors during her professional training; she had an incredible husband who treated her as a true partner in life and science; and the company of her family and friends gave her strength to live following the devastation of her husband's death. All played an important role to her triumph over many obstacles in her life and career.

Moving closer to home, our signal processing community has also seen a growing number of trailblazing women in recent decades. One fine representative of this is Prof. Rabab Ward, who will be taking the office of president of the IEEE Signal Processing Society (SPS) at the start of the new year. She was one of the first two female engineering Ph.D.s who graduated from the University of California at Berkeley and the first female winner of the SPS' highest honor, the Society Award. She became a role model as a female researcher and leader who is able to balance academic research, industrial impact, professional services, as well as be able to have a fulfilling family life with children, grandchildren, and a very supportive husband.

I know firsthand the challenges our female colleagues and students may face to balance career and life, as I am a mother of two young children myself. Strong support networks are crucial as we tackle these challenges. We can see from the aforementioned successful cases that such support networks do not just rely on women. Instead, the understanding, participation, and contributions from our male colleagues play an equally important role. On a related note, the SPS has formally created subcommittees for Women in Signal Processing and for Young Professionals under its Membership Board, and it has supported two Women in Signal Processing gatherings and a Young Professional reception at the ICIP and GlobalSIP conferences this past fall. As we bid farewell to Dr. Alex Acero, who is completing his term as SPS president, I express my sincere appreciation to him and the many SPS leaders and volunteers who have made these initiatives happen.

To our male colleagues—please (continue to) make supporting female colleagues and students a priority: the thoughtful encouragement and understanding you offer, the supportive and constructive critiques you make, the referral you give when a suitable opportunity comes, and the benefit of the doubt you provide, all of these contribute to building these support networks. To our female colleagues—we are not alone in career and life, and together, we can do both!

[SP]

Digital Object Identifier 10.1109/MSP.2015.2501440
Date of publication: 29 December 2015

Support the Mission with a New Perspective

At Leidos, we develop innovative solutions in radar technology, electronic warfare, radiation detection, and sensing systems for advanced threat detection and mitigation.

We are currently seeking talented signal processing, software, hardware, firmware, and system engineers and scientists to join our team.

Learn more at **leidos.com/careers/militaryjobs**

leidos

© Leidos. All rights reserved.

president's MESSAGE

Alex Acero
2014–2015 SPS President
a.acero@ieee.org

We Need Your Help to Take the Society to New Heights

Over the past two years, the IEEE Signal Processing Society (SPS) has changed significantly and grown substantially. The IEEE Signal Processing Cup started with 12 teams in 2014 and grew to 49 teams in 2015. Our Student Career Luncheon at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) is a popular venue for students and companies. Our week-long summer schools, attended mostly by students, are thriving. Our targeted campaign to increase SPS student membership has resulted in a growth of more than 1,000 undergraduate students and 900 graduate students. While average membership in IEEE Societies has dropped by 5% in the last two years, overall membership in the SPS has grown by 14%, leaving us just shy of 20,000 members as of September 2015. We have helped our Society's 168 Chapter chairs with best practices to assist them in better serving their members, and we continue to increase incentives and serve our members, awarding a total of 160 travel grants in 2015 for attendees of SPS conferences. In 2015, the Society started publishing two new journals: *IEEE Transactions on Signal and Information Processing over Networks* and *IEEE Transactions on Computational Imaging*. The total number of pages published in the Society's solely owned periodicals has grown from 20,400 in 2013 to 23,000 in 2015, all while maintaining a high impact factor.

The ChinaSIP and GlobalSIP conferences have gotten off to a flying start. In 2014, we launched SigPort, a document repository. SigView, the video tutorial online library we started in 2013, now has 87 videos. We kicked off a visibility campaign to introduce high school and college students to signal processing through two highly accessible introductory videos, which can be found on our YouTube channel. We have begun implementing our new tagline, "Signal Processing: The Science Behind Our Digital Life," paving the way for future generations of signal processing engineers. In the last two years, we have established new awards to recognize our leaders: the IEEE Fourier Award for Signal Processing, Overview Paper Award, Sustained Impact Paper Award, Signal Processing Letters Best Paper Award, Industrial Leader Award, Industrial Innovation Award, and Conference Best Paper Awards for Industry. We held our five-year Society review and five-year periodicals review with the IEEE Technical Activities Board and passed with flying colors.

None of this would have possible without the many dedicated colleagues who unselfishly volunteer their time and efforts to make our Society better. The SPS was in great shape when my term started, thanks to the incredible dedication of Past President Ray Liu, who is now the IEEE Division IX director, and José Moura, who was SPS president before Ray, became IEEE Division IX director, and is now vice president for 2016, IEEE Technical Activities Board. Rabab Ward, the Society's incoming president, is a fabulous scientist and tireless leader, and the Society is in good hands with her. Ali Sayed, SPS's president-elect, is also an outstanding researcher who is full of ideas.

SINCERE THANKS

I have been extremely fortunate to work closely with many energetic volunteers in our Executive Committee, our Board of Governors, and all of the Society's boards and committees. I am very grateful to Rich Baseil, the Society's executive director, for his quiet leadership, drafting e-mails for me, running the day-to-day operations so smoothly, and putting out fires when they arose. We're fortunate to have wonderful staff to help our volunteers run the Society. While these years have been very busy, they have passed quickly because the interactions with volunteers and staff have been very collegial, all working as a team to make the Society better and serve our members, while having fun doing it. It has been my honor and privilege to work with all of you. Please allow me to say "Thank you so much!"

I joined IEEE in 1983 as a student, started reviewing papers for ICASSP in 1991, and joined one of the Society's technical committees in 1996. I have been serving on various boards and committees ever since and just finished my term as president of SPS. I have found the experience quite rewarding, so I want to encourage all of you to consider volunteering with the IEEE and the SPS and help us take our profession to new heights. You can review conference or journal papers, be involved in your local Chapter, help organize a workshop or a summer school, or serve as president of the Society. There is much work to be done—so we need all of you!

[SP]

Digital Object Identifier 10.1109/MSP.2015.2497119
Date of publication: 29 December 2015

2016 Class of Distinguished Lecturers

The IEEE Signal Processing Society's (SPS's) Distinguished Lecturer Program provides the means for Chapters to have access to well-known educators and authors in the fields of signal processing to lecture at Chapter meetings. While many IEEE Societies have similar programs, the SPS provides financial support for the Chapters to take advantage of this service. Chapters interested in arranging lectures by the Distinguished Lecturers can obtain information from the Society's webpage (http://www.signalprocessingsociety.org/lecturers/distinguished-lecturers/) or by sending an e-mail to sp.info@ieee.org.

Candidates for the Distinguished Lecturer Program are solicited from the Society technical committees, editorial boards, Chapters, and other boards and committees by the Awards Board. The Awards Board vets the nominations and the Board of Governors approves the final selection. Distinguished Lecturers are appointed for a term of two calendar years. Distinguished Lecturers named for 2016 are as follows.

JAN P. ALLEBACH

Jan P. Allebach earned the B.S.E.E. degree from the University of Delaware (1968–1972) and the Ph.D. degree from Princeton University (1972–1976). He was a faculty member at the University of Delaware from 1976 to 1983. Since 1983, he has been at Purdue University, where he is currently the Hewlett-Packard Distinguished Professor of Electrical and Computer Engineering with courtesy appointments in computer science and psychological sciences.

Prof. Allebach is a Life Fellow of the IEEE and a fellow of the Society for Imaging Science and Technology (IS&T) (1996) and SPIE (2007). He is a recipient of the Senior Best Paper Award from the IEEE SPS (1989) and the Bowman Award from IS&T (1998). He was named Electronic Imaging Scientist of the Year by SPIE and IS&T (2004) and received honorary membership from IS&T (2007), which is its highest award. He received the Daniel E. Noble Award for Emerging Technologies and an IEEE Field Award (2013). He was elected to membership in the National Academy of Engineering (2014) and the National Academy of Inventors (2015). From Purdue University, he has received ten different awards for teaching, research, and mentorship.

Prof. Allebach was an associate editor, *IEEE Transactions on Acoustics, Speech, and Signal Processing* (1988–1990) and *IEEE Transactions on Image Processing* (1997–1999); member (1986–1995) and later chair (1990–1991), Image and Multidimensional Signal Processing Technical Committee; member, Information Forensics and Security Technical Committee (2008–2011); secretary, SPS Board of Governors (1992–1995) and later as an elected member (1996–1998); cotechnical program chair, International Conference on Acoustics, Speech and Signal Processing (ICASSP) (1993); and general cochair, Ninth Image, Video, and Multidimensional Signal Processing (IVMSP) Workshop (1996). Prof. Allebach was also a past Distinguished Lecturer, IEEE SPS (1994–1995).

Prof. Allebach has published 105 refereed journal articles and 349 conference papers. He is listed as a co-inventor on 31 issued U.S. patents and several more pending patent applications. Printing and document imaging have been central themes of his research, and he has made many seminal contributions in these areas. His lecture topics include digital printing: the transformation of a 2,000-year-old technology and what it means to you; embedding data in printed documents at the printer mechanism level; and the intertwined roles of semantics, aesthetics, and quality in color document imaging.

YORAM BRESLER

Yoram Bresler received the B.S. (cum laude) and M.S. degrees from the Technion, Israel Institute of Technology, (1974 and 1981, respectively); the Ph.D degree from Stanford University (1986), all in electrical engineering. Since 1987, he has been on the faculty at the University of Illinois, Urbana-Champaign, where he is currently a professor, Department of Electrical and Computer Engineering and the Department of Bioengineering, and at the Coordinated Science Laboratory. In 2003, Dr. Bresler cofounded InstaRecon, Inc., based in Champaign, Illinois, to commercialize breakthrough technology for tomographic reconstruction developed in his academic research. He currently serves as the company's president and chief technology officer.

Dr. Bresler was elected IEEE Fellow in 1999 "for contributions to computer-based imaging and sensor array processing," and in 2010, fellow of the American

Digital Object Identifier 10.1109/MSP.2015.2473755
Date of publication: 29 December 2015

Institute for Medical and Biomedical Engineering (AIMBE), "for pioneering contributions to fast tomographic reconstruction algorithms and fundamental contributions to sampling theory for fast dynamic imaging." He holds 11 U.S. patents and more than 20 international patents and has received the IEEE SPS Best Paper Award (1988 and 1989). He is the recipient of a National Science Foundation (NSF) Presidential Young Investigator Award (1991), the Technion (Israel Institute of Technology) Fellowship (1995), and the Xerox Senior Award for Faculty Research (1998). He was named a University of Illinois Scholar (1999); appointed as associate, the Center for Advanced Study of the University (2001–2002); and faculty fellow, the National Center for Super Computing Applications (2006).

He has served as associate editor, *IEEE Transactions on Signal Processing* (1992–1993); on editorial boards, *Machine Vision and Applications* (1987–2006), and *SIAM Journal on Imaging Science* (2007–2013); on the senior editorial board, *IEEE Journal on Selected Topics in Signal Processing* (2006–2013); and he was guest coeditor, *IEEE Transactions on Medical Imaging* (special issue on compressed sensing). He was a member, IEEE Image and Multidimensional Signal Processing Technical Committee (1993–1998) and the IEEE Bioimaging and Signal Processing Technical Committee (2005–2009), and served on the IEEE SPS Awards Board (2003–2006).

Dr. Bresler's interests are in multidimensional and statistical signal processing and their applications to inverse problems in imaging, and, in particular, compressed sensing, which he introduced with his students in the mid-1990s under the monikers of *spectrum-blind sampling* and *image compression on the fly*, as well as computed tomography (CT), magnetic resonance imaging (MRI), and learning-based signal processing.

His lecture topics include the invention of compressive sampling; breaking the speed barrier in tomography; learning sparsifying transforms for signal, image, and video processing; learning sparse representations for blind compressed sensing in MRI and CT; and blind signal processing: sparse signal reconstruction in bilinear inverse problems.

PEYMAN MILANFAR

Peyman Milanfar received his undergraduate education in electrical engineering and mathematics from the University of California (UC), Berkeley, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, in 1990 and 1993, respectively. He was a professor of electrical engineering at UC Santa Cruz (1999–2014), where he is now a visiting faculty. He was associate dean for research, School of Engineering (2010–2012). He was on leave at Google-x, from 2012 to 2014, where he helped develop the imaging pipeline for Google Glass. He currently leads the Computational Imaging team in Google Research.

Dr. Milanfar is an IEEE Fellow "for contributions to inverse problems and superresolution in imaging." He was a member, IEEE SPS Awards Board (2011–2013); editorial board member, *SIAM Journal on Imaging Science* (2010); editor, *Super-Resolution Imaging* (CRC Press, 2009); editorial board member, Image and Vision Computing (2009–2011); member, IEEE IVMSP Technical Committee (2007–2011); associate editor, *IEEE Transactions on Image Processing* (2007–2010); guest editor, *Journal of Applied Signal Processing* (special issue on superresolution imaging) (2004); outstanding reviewer of the year, *IEEE Transactions on Image Processing* (2006); paper identified by the Institute of Scientific Information as a leader of "emerging research front" in imaging (2005); graduate student Morteza Shahram winner of the Best Student Paper at ICASSP (2005); IEEE SPS Best Paper Award (2010).

Dr. Milanfar holds eight U.S. patents, several of which are commercially licensed. He has been a keynote speaker at numerous international conferences including PCS, SPIE, and the IEEE International Conference on Multimedia and Expo. Along with his students, he has won several best paper awards from the IEEE SPS. His lecture topics include computational imaging: from photons to photos; a tour of modern image filtering; and graph affinity-based image processing.

HERMANN NEY

Hermann Ney is a full professor of computer science at RWTH Aachen University, Germany. He received the diploma degree in physics from Goettingen University, Germany (1977), and the Dr.-Ing. degree in electrical engineering, Braunschweig University of Technology, Germany (1982). In 1977, he joined Philips Research Germany, and, since 1985, has headed the speech recognition group at Philips that pioneered the first prototype systems for large vocabulary continuous speech recognition and spoken dialogue systems. He was a visiting scientist, Bell Laboratories, Murray Hill, New Jersey, (1988–1989) and joined the Computer Science Department of RWTH Aachen University (1993). Since then, he has been working on a number of topics in automatic speech recognition and machine translation using statistical pattern recognition and machine learning.

Prof. Ney's professional activities include: Executive Board, Association of Computational Linguistics (2014–2016); cochair, Interspeech 2015; senior area editor, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2014–2016); IEEE SPS Speech and Language Processing Technical Committee (2012–2014); cochair plenary sessions, ICASSP 2009; cochair, IEEE-ACL Workshop on Spoken Language Technology (2006); editor, *ACM Transactions on Speech and Language Processing* (2005–2007); associate editor, *IEEE Transactions on Speech and Audio Processing* (2001–2003); IEEE SPS Speech Processing Technical Committee (1995–2000); editorial board member, *Computer, Speech and Language* (1993–2001); editorial board, *Speech Communication* (1993–2001); technical committee member, German Association for Pattern Recognition (2003–2006); executive committee member, German Section IEEE (1992–1998).

Prof. Ney is an IEEE Fellow and fellow of the International Speech Communication Association. He was the recipient of the Technical Achievement Award (2005); Distinguished Lecturer, International Speech Communication Association (2012–2013); awarded a senior DIGITEO chair at LIMIS/CNRS in Paris, France (2011–2013); and received the Award of Honour from the International Association of Machine Translation (2013).

Prof. Ney's main research interests are concerned with the application of machine-learning methods to speech recognition and machine translation. His lecture topics include: from speech recognition (ASR) to machine translation (MT)—what MT has learned from ASR; speech and language processing: achievements and challenges; the mathematical machinery of speech and language processing; discriminative training and log-linear models in ASR; generative and discriminative modeling in speech and natural language processing; the role of bayes decision rule in speech recognition and natural language processing; and error bounds on Bayes' classification error in speech and language processing.

PARIS SMARAGDIS

Paris Smaragdis is a member of faculty with the Computer Science and Electrical and Computer Engineering Departments at the University of Illinois at Urbana-Champaign, and a senior research scientist at Adobe Research. He completed his master's degree (1997), Ph.D. degree (2001), and postdoctoral studies (2002) in the Machine Listening Group, Massachusetts Institute of Technology 's (MIT's) Media Lab. He was previously a research scientist with Mitsubishi Electric Research (MERL).

Prof. Smaragdis was selected by *MIT Technology Review* in 2006 as one of the year's top young technology innovators (TR35) for his work on machine listening. In 2015, he was elevated to IEEE Fellow "for contributions in audio source separation and audio processing." He was elected as a full member, the Acoustical Society of America (2008), and he was a recipient of the C.W. Gear Outstanding Junior Faculty Award (2015), an NSF CAREER grant, and multiple teaching awards at the University of Illinois.

Prof. Smaragdis was chair, IEEE Machine Learning for Signal Processing Technical Committee (2013–2014); chair, LVA/ICA conference steering committee (2012–2015); member, IEEE Machine Learning for Signal Processing Technical Committee (2010–2015); member, Audio and Acoustic Signal Processing Technical Committee (2011–present); associate editor, *IEEE Signal Processing Letters* (2012–present); and area editor, *IEEE Transactions on Signal Processing* (2015–present). He was an organizer of the Global Conference on Signal and Information Processing (GlobalSIP) Symposium on Machine Learning for Speech Processing (2014); general cochair, Machine Learning for Signal Processing Workshop (2014); technical chair, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2011); and an organizer of multiple special sessions and tutorials at international conferences.

His research is on applications of machine-learning techniques on signal processing problems, especially as they apply to the analysis of sound mixtures. He has more than 120 publications in the areas of audio signal processing and machine learning and holds 59 patents internationally. His lecture topics include nonnegative models for source separation; machine learning for speech enhancement; and machine listening: making computers that understand sound.

JOSIANE ZERUBIA

Josiane Zerubia received the M.S. degree from the Department of Electrical Engineering at ENSIEG, Grenoble, France, in 1981, the doctor of engineering degree, Ph.D. degree, and her habilitation, in 1986, 1988, and 1994 respectively, all from the University of Nice Sophia-Antipolis, France. She has been a permanent research scientist at INRIA since 1989 and the director of research since 1995. She was head of the PASTIS Remote Sensing Laboratory (INRIA Sophia-Antipolis) (1995–1997) and of the Ariana Research Group (INRIA/CNRS/University of Nice) (1998–2011). Since 2012, she has been head of Ayin Research Group (INRIA-SAM) dedicated to models of spatiotemporal structure for high-resolution image processing with a focus on remote sensing and skincare imaging. She has been a professor, SUPAERO (ISAE) in Toulouse, France, since 1999.

Prof. Zerubia was also with the Signal and Image Processing Institute of the University of Southern California as a postdoctoral researcher. She also worked as a researcher, LASSY (University of Nice/CNRS) (1984–1988) and in the Research Laboratory of Hewlett Packard in France and in Palo-Alto, California (1982–1984).

Prof. Zerubia is an IEEE Fellow. She was a member, IEEE Image and Multidimensional Signal Processing Technical Committee (1997-2003); member, IEEE Bioimaging and Signal Processing Technical Committee (2004–2012); member, IEEE IVMSP Technical Committee (2008–2013); associate editor, *IEEE Transactions on Image Processing* (1998–2002); area editor, *IEEE Transactions on Image Processing* (2003–2006), guest coeditor, special issue of *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2003); editorial board member, IJCV (2004–2013); member-at-large, SPS Board of Governors (2002–2004); editorial board member, French Society for Photogrammetry and Remote Sensing (SFPT) (1998–present), Foundation and Trends in Signal Processing (2007–present), member-at-large, Board of Governors of the SFPT (2014–present); and associate editor, online resource Earthzine (IEEE Committee on Earth Observations and Global Earth Observing System of Systems).

Prof. Zerubia was a cochair of two workshops on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR 2001 and EMMCVPR 2003); cochair, Workshop on Image Processing and Related Mathematical Fields (IPRM 2002); technical program chair, Workshop on Photogrammetry and Remote Sensing for Urban Areas (2003); cochair, special sessions at IEEE ICASSP (2006) and IEEE ISBI (2008); publicity

chair, IEEE ICIP (2011); tutorial cochair, IEEE International Conference on Image Processing (2014); general cochair, Workshop Earthvision at IEEE Computer Vision and Pattern Recognition (2015); organizing committee member and plenary talk cochair, IEEE-EURASIP (European Signal Processing Conference 2015).

Her main research interest is in image processing using probabilistic models. She also works on parameter estimation, statistical learning, and optimization techniques. Prof. Zerubia's lecture topics include marked point processes for object detection and tracking in high-resolution images: applications to remote sensing and biology; marked point processes for object detection in high-resolution images: applications to earth observation and cartography; and high-resolution optical and synthetic-aperture radar satellite image processing for disaster management using hierarchical Markov random fields.

IEEE Signal Processing Society Young Professionals Meet During ICIP 2015

As an outreach effort to its members, the IEEE Signal Processing Society (SPS) sponsored a social reception for members of the Quebec City Chapter. The Young Professionals attending the 2015 IEEE International Conference on Image Processing (ICIP) were also invited. The IEEE Young Professionals Program (formerly known as GOLD, which stands for "Graduates of the Last Decade") is intended for IEEE Members who have received their first professional degree within the last 15 years and aims to help young professionals evaluate their career goals, polish their professional image, and create the building blocks of a lifelong and diverse professional network.

IEEE young professionals

The event was arranged by the Quebec City Chapter and hosted by Chapter Chair Dr. Paul Fortier, Laval University, and by SPS Young Professionals Subcommittee Chair Dr. Mahsa T. Pourazad, TELUS Communications, Inc. and the University of British Columbia. The predinner reception was held in the beautiful historic library of the Quebec Parliament. The iconic venue formed a wonderful atmosphere for attendees to meet, network, and exchange ideas in a friendly and relaxed manner. SPS President Dr. Rabab Ward, who was SPS president-elect at the time of this writing, welcomed attendees with a warm and informative speech. She stressed the importance of the Chapters as well as the activities of its members to nurture the liveliness of the SPS. She talked about the various new initiatives the Society is introducing to enhance the value added to members at large including 1) financial support for new initiatives by Chapters especially geared toward industry, 2) holding networking events for Chapter members in conjunction with some of the

Digital Object Identifier 10.1109/MSP.2015.2490298
Date of publication: 29 December 2015

(continued on page 150)

[special REPORTS]

John Edwards

Signal Processing Research Resonates with Hearing Loss Sufferers

Hearing enhancement is an area where signal processing technology makes a direct and positive impact on human lives. By allowing people with hearing loss to perceive voices, music, and other sounds more clearly and naturally, signal processing improves the quality of life for countless millions of people worldwide.

Hearing loss is far more than an inconvenience; it can negatively impact a person's career and social life. Significant hearing loss can also affect an individual's physical well-being. Brain shrinkage, for example, appears to be accelerated in older adults with hearing loss, according to the results of a study released in 2014 by researchers from Johns Hopkins University and the U.S. National Institute on Aging. The study added to a growing list of health issues associated with hearing loss, including an increased risk of dementia, falls, hospitalizations, and diminished overall physical and mental health.

While hearing aids are critically important instruments for the hearing impaired, the technology fails to help people in situations where background noise swamps out individual voices. "Hearing aids do a good job of amplifying sounds, but a poor job of isolating the signal of interest," observes Eric Healy, a professor of speech and hearing science at The Ohio State University and director of The Ohio State University Speech Psychoacoustics Laboratory.

Healy leads one of several academic research teams that are independently working on technologies designed to help individuals who wear hearing aids function more productively and comfortably in places where sounds generated by people, machinery, or nature interfere with signals of interest. "The difficulty of understanding speech in background noise is perhaps the number one complaint of people with hearing loss," he observes.

Digital Object Identifier 10.1109/MSP.2015.2475037
Date of publication: 29 December 2015

CUTTING THROUGH THE CHATTER

Tuning in to specific sound signals, even in places buzzing with background noise, is a task that healthy human brains and ears perform remarkably well. In a crowded room filled with voices coming from every direction, the human auditory system excels at focusing on a single voice while filtering out background chatter.

"There have been many attempts to remove background noise from speech and improve intelligibility over the past 50 years; it can be considered a 'Holy Grail'," Healy says. The inability to understand just a single voice in roomful of chatter has been dubbed the *cocktail party problem*. "It is not difficult to remove noise, but the speech often gets disturbed in that process," Healy explains.

Healy and his coresearchers believe that their approach, which uses artificial neural network technology, offers the best way to improve the recognition of spoken words by hearing-impaired people. The researchers hope that their technology will enable the development of a new generation of advanced hearing aids and cochlear implants that can better discern between speech and background chatter.

The technology is based on an algorithm developed by DeLiang "Leon" Wang, an Ohio State professor of computer science and engineering. The algorithm is designed to rapidly analyze speech and remove most of the background noise. "We began collaborating with DeLiang Wang in 2011 or 2012, but he has been working in this area for many years," Healy notes.

Wang's algorithm uses machine learning. Wang and doctoral student Yuxuan Wang are training the algorithm to separate speech from noise by exposing it to a stream of words enveloped in background clutter. A deep neural network (DNN), inspired by the deep-layered structure of the human brain, handles the processing.

Using an ideal binary mask also proved to be essential to the technology. "In this technique, the speech plus noise is divided into time-frequency units, and each unit is classified as dominated by speech or dominated by noise," Healy says. "In the simplest case, the units dominated by speech are retained and those dominated by noise are simply discarded." Using machine-learning techniques, the DNN is trained to classify units when given only the speech mixed with noise, just as the microphone on a hearing aid would receive.

The researchers (Figure 1) tested the algorithm's effectiveness at identifying voice signals against stationary noise—such as an air-conditioner's hum—and then against babbling background voices. The algorithm was most effective against background chatter, improving hearing-impaired listeners' comprehension from 25% to nearly 85% on average. Against stationary noise, the algorithm enhanced comprehension from an average of 35% to 85%. "In fact, the hearing-impaired subjects with the algorithm actually outperformed the normal-hearing subjects without the algorithm in our 2013 report," Healy says. "Thus, having the algorithm was actually more advantageous than having normal hearing."

Healy says his team encountered the same basic challenges faced by other hearing-technology researchers. The

fundamental goal, he notes, is both straightforward and elusive: eliminating extraneous information while isolating and preserving the main signal. "It is fairly easy to strip noise from a signal, but it is very difficult to have this cleaned signal result in improved intelligibility," Healy says.

Healy says that the research is still moving forward. The research team continues to refine the algorithm and conduct tests on human volunteers. "There are a few hurdles remaining before this technology can be implemented, but none seem out of reach," he states.

A METAMATERIAL METHOD

A team of Duke University researchers is taking a hardware-oriented approach to hearing-device enhancement, using a unique metamaterial sensor to help isolate speech from background noise. "We want to design a new sensing system that can mimic the functionality of so-called cocktail party listening by combining an acoustic metamaterial and computational sensing techniques, says Abel Xie, a Ph.D. student who is a member of the research team led by Steven Cummer, a professor of electrical and computer engineering at Duke.

Inspired by the frequency-dependent filtering mechanism of the human ears cochlea system, the researchers have developed a multispeaker listening system using a precisely engineered metamaterial sensor to perform dispersive frequency modulation and compressive sensing to determine the direction of a sound and extract the auditory signal from the surrounding background noise.

"This technology allows hearing aids to selectively listen to speech in a cacophony and to filter out unwanted speech and background noises from various spatial locations," Xie says. The technology can also track the real-time location of human speakers, using spatial information to further increase the signal-noise-ratio of the desired auditory signal.

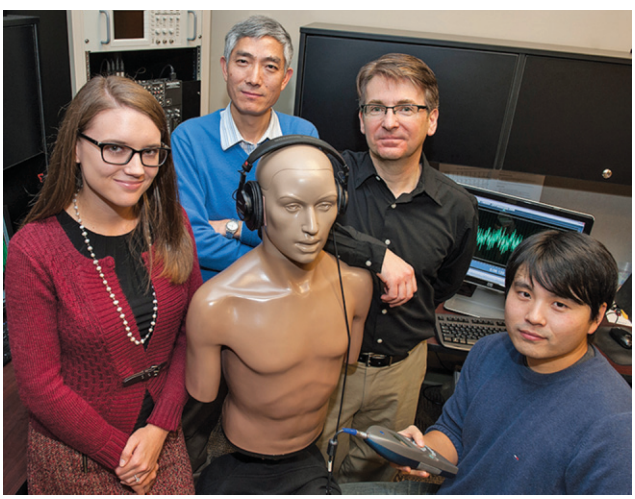
The approach incorporates a two-stage signal processing process. "In the first stage, the signal processing is analog, which means the processing is actually performed directly on the acoustic wave by the metamaterial without converting it to digital signals first," Xie explains. "The metamaterial encodes the sound signals by modulating them, so the information contained in the sound remains separable even though the sound signals from different locations are collected as a single signal." The second stage of the process separates the mixed sounds signals. "The signal processing at this stage is on the digitalized signals with computer," Xie says. "Compressive sensing algorithms are used to separate the sound."

The proof-of-concept metamaterial sensor—approximately 6 in wide—vaguely resembles a thick, plastic, pie-shaped honeycomb divided into dozens of slices (Figure 2). Although the honeycomb cavities appear to be virtually identical, their depth varies from hole to hole to create a unique pattern. When a sound wave reaches the device, it is slightly distorted by the cavities and gains a specific signature depending on what slice of the honeycomb over which it passed. After being picked up by a microphone on the other side, the sound is transmitted to a computer where the algorithms separate the jumble of noises based on the unique distortions.

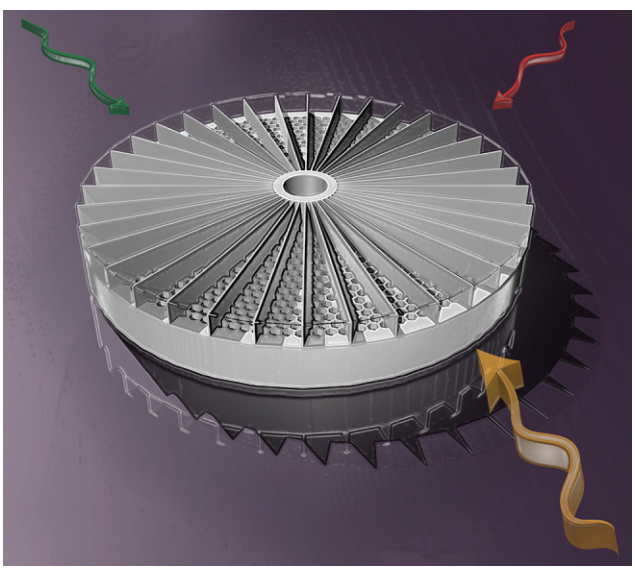
The researchers tested the system in multiple trials, simultaneously sending three identical sounds at the sensor from three different directions. The system was able to distinguish between the sounds at a 96.7% accuracy rate.

Once miniaturized, the technology could be used in various types of voice-command electronics as well as in hearing aids and cochlear implants. Since the device is formed out of plastic, and does not contain any electronic components or moving parts, it promises to be extremely efficient and reliable.

The technology also has potential consumer electronics applications. "It could possibly be combined with any medical imaging device that uses waves, such as ultrasound, to improve current sensing methods as well as to create entirely new techniques," Xie says.



[FIG1] The Ohio State University hearing researchers. From left: Sarah Yoho, doctoral student; DeLiang "Leon" Wang, professor of computer science and engineering; Eric Healy, professor of speech and hearing science; and Yuxuan Wang, doctoral student. (Photo courtesy of Jo McCulty, The Ohio State University.)



[FIG2] Developed by Duke University researchers, this prototype of a metamaterial sensor is designed to help isolate speech from background noise. (Photo courtesy of Duke University.)

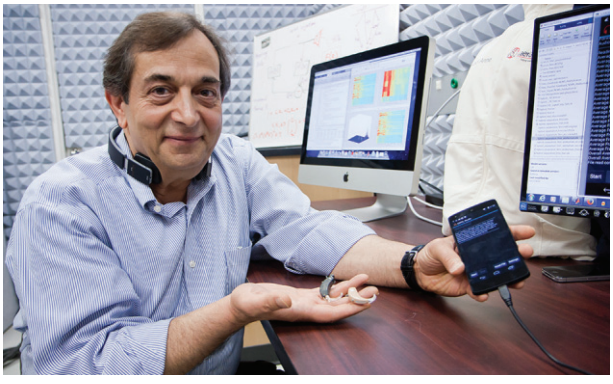
SMARTPHONE SUPPORT

A University of Texas at Dallas (UT Dallas) research project is investigating the possibility of using smartphones and similar mobile devices to augment the performance and use of hearing-assist devices such as hearing aids, cochlear implants, and personal sound amplifiers.

"Current hearing devices can easily fit inside or behind the ear, yet depend on not very powerful processors that are designed to conserve size, power, and cost," says project leader Issa Panahi, an associate professor of electrical engineering (Figure 3). The research team also includes Nasser Kehtarnavaz, a professor of electrical engineering, and Linda Thibodeau, a professor in the University of Texas School of Behavioral and Brain Sciences.

Panahi notes that current hearing aid-based enhancement algorithms are limited and imperfect. "They suppress the background noise, which is good, but at the same time they distort the intelligibility of the speech," he says. "We are trying to fix that."

Yet, developing a method to squelch background noise while preserving and enhancing the desired main signal is not easy, particularly on a platform as tiny and underpowered as the typical hearing aid. "Sophisticated algorithms are necessary, and these algorithms for noise classification and speech enhancement require more powerful processors and additional power consumption," Panahi says. "This complexity cannot be handled by existing hearing-assist devices due to their size and power constraints."



[FIG3] **Issa Panahi, an associate professor of electrical engineering, leads a UT Dallas research team looking to tap into the power of smartphones to enhance the performance of hearing devices. (Photo courtesy of UT Dallas.)**

Due to their inherent limitations, current-generation hearing devices are restricted to a relatively small set of predefined algorithms, each designed to address a particular type of noise environment. "There are settings, for example, for when you are in a car, in an area where there are lots of people talking or walking outdoors where the wind is making noise," Panahi says. "If you want to cope with other situations, then you need more advanced digital signal processing algorithms."

Offloading most of the processing work to a more powerful mobile device, linked via wireless to an ultrasmall earpiece, offers the potential to improve hearing devices performances while also enhancing user comfort,

(continued on page 162)

"Starting from basic concepts, and placing an emphasis on intuition, this book develops mathematical tools that give the readers a fresh perspective on classical results, while providing them with the tools to understand many state of the art signal representation techniques."

Antonio Ortega, University of Southern California

"Foundations of Signal Processing is a pleasure to read. This textbook strikes an excellent balance between intuition and mathematical rigor, that will appeal to both undergraduate and graduate engineering students. Students will learn not only a substantial body of knowledge and techniques, but also why things work, at a deep level, which will equip them for independent further reading and research. I look forward to using this text in my own teaching."

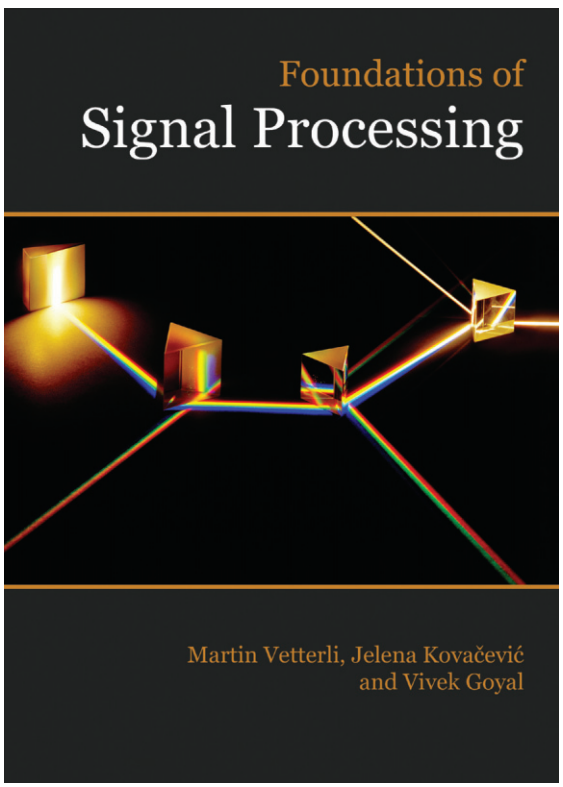
Yoram Bresler, University of Illinois at Urbana-Champaign

"A major book about a serious subject - the combination of engineering and mathematics that goes into modern signal processing. The theory is beautiful and the applications are important and widespread."

Gil Strang, Massachusetts Institute of Technology

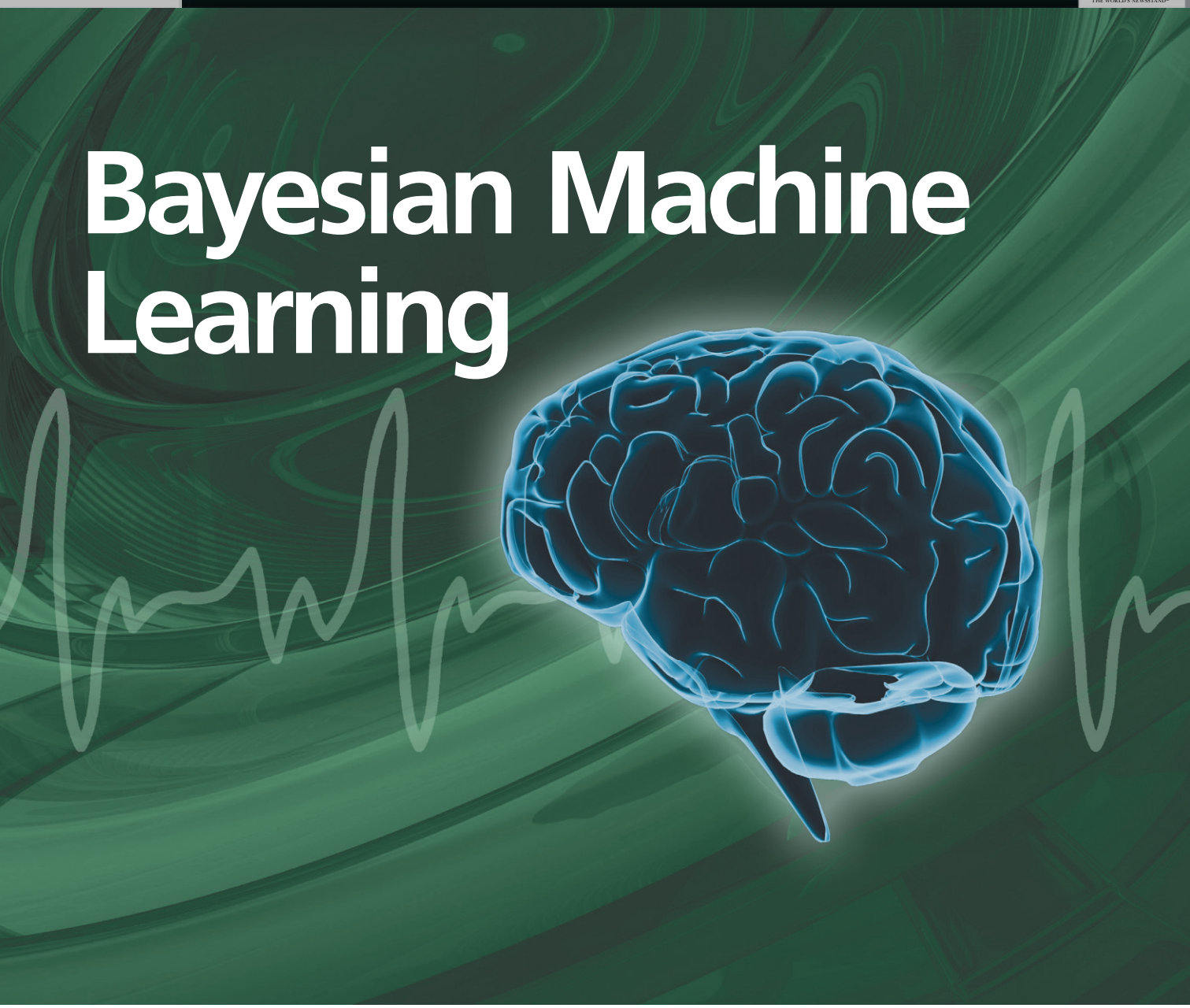
"Finally a wonderful and accessible book for teaching modern signal processing to undergraduate students."

Stéphane Mallat, Ecole Normale Supérieure



Hardcover 715p 2014 ISBN 978-1-107-03860-8 Cambridge University Press

Bayesian Machine Learning



IMAGES LICENSED BY INGRAM PUBLISHING AND GRAPHIC STOCK

[Wei Wu, Srikantan Nagarajan, and Zhe Chen]

[EEG/MEG signal processing measurements]

Electroencephalography (EEG) and magnetoencephalography (MEG) are the most common noninvasive brain-imaging techniques for monitoring electrical brain activity and inferring brain function. The central goal of EEG/MEG analysis is to extract informative brain spatio-temporal–spectral patterns or to infer functional connectivity between different brain areas, which is directly useful for neuroscience or clinical investigations. Due to its potentially complex nature [such as nonstationarity, high dimensionality, subject variability, and low signal-to-noise ratio (SNR)], EEG/MEG signal processing poses some great challenges for researchers. These challenges can be addressed in a principled manner via Bayesian machine learning (BML). BML is an emerging field that integrates Bayesian statistics, variational methods, and machine-learning

Digital Object Identifier 10.1109/MSP.2015.2481559
Date of publication: 29 December 2015

1053-5888/16©2016IEEE

techniques to solve various problems from regression, prediction, outlier detection, feature extraction, and classification. BML has recently gained increasing attention and widespread successes in signal processing and big-data analytics, such as in source reconstruction, compressed sensing, and information fusion. To review recent advances and to foster new research ideas, we provide a tutorial on several important emerging BML research topics in EEG/MEG signal processing and present representative examples in EEG/MEG applications.

INTRODUCTION

EEG and MEG are two dominant technologies for the noninvasive measurement of dynamic whole-brain electrical activity. The central goal of EEG/MEG is to extract a wealth of spatiotemporal–spectral patterns of brain activity and functional connectivity from large and complex time-series data to address important questions in neuroscience, neural engineering, and clinical studies. The analysis of EEG/MEG signals poses enormous challenges in signal processing and statistical methods because of their intrinsic high dimensionality, large or sparse sample size, low SNR, nonstationarity (across time/trials/conditions/subjects/groups), nonlinearity (of the feature space with regard to the signal channel space), and other structures afforded by a variety of signal types [e.g., event-related potentials (ERPs) and oscillations]. This article is partly motivated to address these issues in the emerging field of BML. Specifically, we focus on the probabilistic graphical models (within a generative modeling framework) and Bayesian inference algorithms that are used in EEG/MEG signal processing. We identify three emerging and distinct yet somewhat overlapping research lines as the focus of this tutorial: 1) variational Bayesian (VB) methods, 2) sparse Bayesian learning (SBL) methods, and 3) nonparametric Bayesian (NB) methods.

EEG/MEG ORIGINS AND ATTRIBUTES

Neuronal activity gives rise to extracellular electric and magnetic fields, which are detected in MEG and EEG [1]. Electric current contributions from all active cellular processes within a volume of brain tissue superimpose at a given location in the extracellular medium and generate an electric potential, V_e (a scalar measured in volts), with respect to a reference potential. The difference in V_e between two locations gives rise to an electric field (a vector whose amplitude is measured in volts per unit of distance). EEG measures the summation of the synchronous activity (i.e., population V_e in voltage) of millions of neurons that have similar spatial orientation, which can be detected on the scalp surface using conductive electrodes and sensitive bioelectric amplifiers. Synchronized neuronal currents also induce weak magnetic fields that can be detected outside the head using MEG, which makes use of highly sensitive magnetic-field sensors. Most current MEG systems use superconducting quantum interference device sensors, which have a sensitivity on the order of 5–10 fT/square root of Hz, with many more emerging magnetic-sensing technologies that have comparable sensitivity. MEG and EEG signals represent complementary information about the underlying sources that contribute to electric potentials on the scalp and magnetic fields outside the head. Importantly, both EEG and MEG measurements have an excellent temporal resolution (~1 millisecond) in sampling the rich temporal dynamics of neuronal population activity. The spatial sampling precision of EEG and MEG signals is typically around 1–2 centimeters outside the head, but the ultimate spatial resolution of the methods depends on the algorithms and applications.

MOTIVATION

At a very broad level, the two dominant signal processing applications of MEG and EEG are 1) electromagnetic brain imaging and 2) brain-state classification. Signal processing algorithms that reconstruct and visualize the neuronal activities (and functional connectivity) based on MEG/EEG sensor measurements are referred to as the *electromagnetic brain-imaging algorithms*. Electromagnetic brain imaging has diverse applications in basic and clinical neuroscience, such as neurophysiological oscillations imaging associated with normal brain function and how these processes may be altered in disease or during an intervention. Algorithms that make inferences about the behavioral/mental state of a subject are called *brain-state classification algorithms*. Brain-state classification has widespread neuroscience applications, including methods for diagnosis of abnormal brain activity, as well as the development of brain–computer interfaces (BCIs).

Classical signal processing methods for EEG/MEG analyses have mostly included digital filtering, spectral analysis, or source separation [2]. To date, EEG/MEG analyses have encountered many challenges, and the past several decades have witnessed significant progress in the two application domains mentioned previously. Despite the proliferation of BML algorithms for EEG/MEG signal processing, a comprehensive tutorial that provides an overview of important concepts and state-of-the-art development is still lacking in this fast-growing field. This article provides a review on some up-to-date BML algorithms for electromagnetic brain imaging and brain-state classification. Specifically, we focus on a few emerging theme topics, which may help researchers to gain a coherent picture of the assorted methods, develop a deep understanding of their mechanisms, appreciate the most relevant research themes, and spark further research in this area. To this end, our tutorial provides a self-contained methodological guide by disseminating representative BML methods for statistical analysis of EEG/MEG signals, with accessible yet rigorous mathematics describing their central ideas and carefully chosen examples demonstrating their efficacy in practical applications.

LATENT VARIABLE GENERATIVE MODELING FRAMEWORK FOR EEG/MEG SIGNAL PROCESSING

The main applications of this tutorial's focus are EEG/MEG-based brain imaging and brain-state classification, which can be cast under a unifying latent variable modeling framework

$$X = f(Z, \boldsymbol{\theta}) + E, \tag{1}$$

where X is the observed variables; Z is the unobserved latent variables; f is a function of Z that can be linear or nonlinear, parameterized by a known or unknown parameter $\boldsymbol{\theta}$; and E is an error term that consists of uncorrelated or correlated noise and/or interference. The analysis goal is to infer the latent

NOTATIONS

The bold font is used to denote vectors or matrices. Data and parameters are displayed by italic and regular font, respectively. Subscript index t denotes the discrete time, whereas superscript indices s, k, and l denote the subject, condition, and trial, respectively. $\mathbb{E}[\cdot]$ and $\mathbb{C}[\cdot]$ denote the expectation and covariance operators, respectively. Some common notations for probability distributions and stochastic processes used in this article are listed in Table 1.

variables and unknown parameters for data interpretation. When the data likelihood $p(X|Z,\boldsymbol{\theta})$ is a Gaussian distribution, the problem becomes regression, which includes the problem of electromagnetic imaging. When the data likelihood consists of category label c, such as $p(X|Z,\boldsymbol{\theta},c)$ (e.g., the brain-state classification problem), the inference problem becomes $p(c|X) = \int p(c,Z|X)\,dZ = \int p(c|X,Z)p(Z|X)\,dZ$, which can be decomposed into two steps: estimation of $p(Z|X)$ and estimation of $p(c|Z)$.

For electromagnetic brain imaging, f can be approximated as a linear function of Z as follows:

$$X = \mathbf{A}Z + E \text{ or } x_t = \mathbf{A}z_t + e_t, \tag{2}$$

where $X = [x_1, \cdots, x_T] \in \mathbb{R}^{N\times T}$ is the EEG/MEG sensor data matrix with N sensors and T time points. The source activity matrix, $Z = [z_1, \cdots, z_T] \in \mathbb{R}^{M\times T}$ is associated with M latent brain sources. Unless stated otherwise, throughout this tutorial, we assume that the EEG/MEG data have been preprocessed with a proper clean-up procedure to have nonbrain biological artifacts removed. In the spatiotemporal decomposition problem, $\mathbf{A} \in \mathbb{R}^{N\times M}$ is interpreted as a mixing matrix. In the inverse problem, $\mathbf{A}$ is known as the *lead-field matrix*, which can be obtained by solving the forward problem (from Maxwell's equations) based on the structural information of the subject's head, as well as electric and geometric properties of the electric sources and the volume conductor. Each unknown source Z often represents the magnitude of a neural current dipole, projecting from an rth (discretized) voxel or candidate location distributed throughout the brain. These candidate locations can be obtained by segmenting a structural magnetic resonance (MR) scan of a human subject and tesselating the brain volume with a set of vertices. Since the number of brain sources largely outnumber the sensors, i.e., $M \gg N$, reconstructing brain sources from EEG/MEG data is a highly ill-posed problem with an infinite number of solutions. Further anatomical or functional constraints should, therefore, be incorporated to restrict the solution space. Anatomically, in reasonable settings, the dipole sources are restricted to be situated on the cerebral cortex and their orientations perpendicular to the cortical surface. Functionally, spatial smoothness and sparsity are the most widely used constraints.

[TABLE 1] ABBREVIATED NOTATIONS FOR PROBABILITY DISTRIBUTIONS AND STOCHASTIC PROCESSES.

PROBABILITY DISTRIBUTION	NOTATION
$\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$	GAUSSIAN (MEAN $\boldsymbol{\mu}$, COVARIANCE $\boldsymbol{\Sigma}$)
$\mathcal{N}^+(\boldsymbol{\mu},\boldsymbol{\Sigma})$	HALF-GAUSSIAN (SAME PARAMETERIZATION AS GAUSSIAN)
$\mathcal{G}(a,b)$	GAMMA (SHAPE PARAMETER $a>0$, RATE PARAMETER $b>0$)
$\mathcal{IG}(a,b)$	INVERSE GAMMA (SHAPE PARAMETER $a>0$, SCALE PARAMETER $b>0$)
$\mathcal{W}(\nu,\mathbf{W})$	WISHART (DEGREE OF FREEDOM $\nu>0$, SCALE MATRIX $\mathbf{W}$)
$Be(p)$	BERNOULLI DISTRIBUTION ($0<p<1$)
$\mathcal{B}(a,b)$	BETA DISTRIBUTION (SHAPE PARAMETERS $a>0, b>0$)
GP($\mathbf{0},\mathbf{C}$)	GAUSSIAN PROCESS (GP) (ZERO MEAN FUNCTION, COVARIANCE FUNCTION $\mathbf{C}$)
DP(α, G_0)	DIRICHLET PROCESS (DP) (CONCENTRATION PARAMETER $\alpha>0$, BASE MEASURE G_0)

An extension of the static linear model (2) is a Markovian state-space model, also known as the *dynamic factor analysis model* [3]

$$z_t = \mathbf{F}z_{t-1} + v_t \tag{3}$$
$$x_t = \mathbf{A}z_t + e_t, \tag{4}$$

where $\mathbf{F}$ is a time-invariant state-transition matrix for the latent state z_t, $\mathbf{A} \in \mathbb{R}^{N\times M}$ can be either a factor loading matrix (for modeling low-dimensional sources, where $M \ll N$) or a lead-field matrix (for modeling high-dimensional sources, where $M \gg N$), and $v_t \sim \mathcal{N}(0,\mathbf{I})$ and $e_t \sim \mathcal{N}(0,\boldsymbol{\Sigma}_e)$ denote zero-mean Gaussian dynamic and measurement noise, respectively. Simple linear algebra will yield $x_t \sim \mathcal{N}(0, \mathbf{A}\Sigma_z\mathbf{A}^{\mathrm{T}} + \Sigma_e)$, where Σ_z denotes the marginal covariance of z_t.

For the most part, we assume that e_t is a noise-plus-interference term and, for simplicity, that e_ts are drawn independently from $\mathcal{N}(0,\boldsymbol{\Sigma}_e)$. However, temporal correlations can easily be incorporated using a simple transformation outlined in [4] or using the spatiotemporal framework introduced in [5]. Initially, we assume that $\boldsymbol{\Sigma}_e$ is known; but, in later sections, we also derive how Σ_e is estimated from data.

For brain-state classification, a labeled state or class variable, c, is additionally known given training data. The objective is to determine $p(c|X)$ of the test (unlabeled) data. Three common problems within brain-state classification are disease diagnosis, behavioral state classification, and BCIs. In disease diagnosis, the class or state variable corresponds to the disease diagnosis group, and the observed sensor data are used to make inferences about whether the given EEG/MEG observations carry a signature about the disease group. In behavioral state classification, the problem is to infer the evolving behavioral state from a subject's EEG/MEG data. The most common example is to infer the sleep stage (e.g., awake, slow-wave sleep, and rapid-eye-movement sleep). A second example is to determine the time period when abnormal epileptiform activity are present in EEG/MEG data in a patient with epilepsy. Finally, in BCIs, we learn the brain state associated with the intended state c of the user, and subsequently infer the intended brain state c from new data X.

For all of these problems, within a Bayesian generative modeling framework, all prior assumptions are embedded in the distribution $p(Z)$. If $p(Z)$ is fully or partially known under a specific experimental or clinical paradigm, then the posterior distribution $p(Z|X)$ can be computed via Bayes' rule

$$p(Z|X) = \frac{p(X|Z)p(Z)}{p(X)}. \tag{5}$$

The posterior distribution contains all possible information about the unknown Z conditioned on the observed data X. Two fundamental problems exist for computing $p(Z|X)$. First, for most prior $p(Z)$, it is impossible to analytically compute the distribution $p(X)$ by a direct integration

$$p(X) = \int p(X|Z)p(Z)\,\mathrm{d}Z. \tag{6}$$

The quantity $p(X)$, sometimes referred to as the *model evidence*, can be used to facilitate model selection. When a point estimate for Z is desired, $p(X)$ may not be needed. For example, the maximum a posteriori (MAP) estimate of $Z_{\text{MAP}} = \arg\max_z p(Z|X)$ is invariant to $p(X)$. However, Z_{MAP} may be unrepresentative of posterior mass, and it is often intractable to compute for most $p(Z)$. In addition, the prior $p(Z)$ is often assumed based on neurophysiological constraints or computational considerations, which implicitly or explicitly differentiate a wide variety of estimation methods at a very high level.

For example, we can adopt the following source prior for Z:

$$p(Z|\gamma) \propto \exp\left(-\frac{1}{2}\mathrm{trace}[Z^{\mathrm{T}}\Sigma_z^{-1}Z]\right), \Sigma_z = \sum_{i=1}^{d_\gamma}\gamma_i \mathrm{C}_i. \tag{7}$$

This is equivalent to applying independently, at each time point, a zero-mean Gaussian distribution with covariance Σ_z to each column of Z. Here, $\gamma \triangleq \mathrm{diag}(\gamma_1, \ldots, \gamma_{d_\gamma})$ is a diagonal matrix consisting of d_γ nonnegative hyperparameters that control the relative contribution of each covariance basis matrix C_i. While the hyperparameters are unknown, the set of components $\mathbf{C} \triangleq \{\mathrm{C}_i : i = 1, \ldots, d_\gamma\}$ is assumed to be fixed and known. Such a prior formulation is extremely flexible in that a rich variety of candidate covariance bases can be proposed. Moreover, this structure has been advocated in the context of neuroelectromagnetic source imaging [6], [7]. We can also assume a hyperprior on γ of the form

$$p(\gamma) = \prod_{i=1}^{d_\gamma}\frac{1}{2}\exp[-f_i(\gamma_i)], \tag{8}$$

where each $f_i(\cdot)$ is a prespecified function. Within such a hierarchical latent modeling framework, the implicit prior on Z, obtained by marginalizing the unknown γ, is known as a *Gaussian scale mixture*

$$p(Z) = \int p(Z|\gamma)p(\gamma)\,\mathrm{d}\gamma. \tag{9}$$

In the condition where the matrix $\mathbf{A}$ is unknown, one also needs to specify a prior distribution $p(\mathbf{A})$. This is used in spatiotemporal decomposition problems for both brain imaging and brain-state classification. In EEG/MEG spatiotemporal decomposition, data are decomposed into the sum of products of spatial and temporal patterns from different components, with each component representing a distinct neurophysiological process. In this case, $\mathbf{A} \in \mathbb{R}^{N\times M}$ is the spatial pattern matrix, and $Z \in \mathbb{R}^{M\times T}$ is the temporal pattern matrix. Spatiotemporal decomposition consists in finding the unknown $\mathbf{A}$ and Z from the signals, typically through some statistical constraints, e.g., statistical independence, sparsity, or nonnegativity of the components.

In the case where $f(\cdot)$ is a nonlinear function of Z in (1), the Gaussianity is lost. In addition, the mapping f may not be invertible and the inverse mapping $\phi: X \to Z$ is also nonlinear and nonunique. In general, Z can be interpreted as the nonlinear feature extracted from X, which can be further used for brain-state classification. We revisit this topic in the section "Applications."

BAYESIAN INFERENCE: A BRIEF OVERVIEW

Estimation and inference can be carried out in multiple ways, depending how the unknown quantities $\mathbf{A}$, Z, and γ are handled. This leads to a natural partitioning of a variety of inverse methods. We briefly summarize three analytical approaches and sampling approaches. For simplicity, we assume $\mathbf{A}$ is known in the following overview.

EMPIRICAL BAYESIAN APPROACHES

If γ and Σ_z are known in advance, then the conditional distribution $p(Z|X,\gamma) \propto p(X|Z)p(Z|\gamma)$ is a fully specified Gaussian distribution with mean and covariance given by

$$\begin{aligned}\mathbb{E}_{p(z_t|X,\gamma)}[z_t] &= \gamma\mathbf{A}^{\mathrm{T}}(\Sigma_e + \mathbf{A}\Sigma_z\mathbf{A}^{\mathrm{T}})^{-1}x_t\\ \mathbb{C}_{p(z_t|X,\gamma)}[z_t] &= \Sigma_z - \Sigma_z\mathbf{A}^{\mathrm{T}}(\Sigma_e + \mathbf{A}\Sigma_z\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{A}\Sigma_z,\end{aligned} \tag{10}$$

where z_t denotes the tth column of Z and is uncorrelated with time.

A common estimator for the latent sources is to use the posterior mean $\hat{Z} = \mathbb{E}_{p(Z|X,\gamma)}[Z]$. However, since γ is unknown, an approximate solution $\hat{\gamma}$ must first be found. One principled way to accomplish this is to integrate out the sources Z and then solve

$$\hat{\gamma} = \arg\max_{\gamma}\int p(X|Z)p(Z|\gamma)p(\gamma)\,\mathrm{d}Z. \tag{11}$$

This treatment is sometimes referred to as *empirical Bayes* (*EB*) because the prior $p(Z|\gamma)$ is empirically learned from the data, often using expectation-maximization (EM) algorithms. Additionally, marginalization provides a natural regularization that can shrink many elements of γ to zero, in effect pruning irrelevant covariance components from the model. Estimation under this model is sometimes called *automatic relevance determination* (*ARD*). The ARD procedure can also be leveraged to obtain a rigorous lower bound on model evidence $\log p(X)$. While knowing $p(Z|X)$ is useful for source estimation given a particular model, access to $p(X)$ [or, equivalently, $\log p(X)$] can assist model selection.

PENALIZED LIKELIHOOD APPROACHES

The second option is to integrate out the unknown γ, treat $p(Z)$ as the effective prior, and compute the MAP estimate Z_{MAP}:

$$Z_{\text{MAP}} = \arg\max_{z} \int p(Z|X)p(Z|\gamma)p(\gamma)\,\mathrm{d}\gamma$$
$$= \arg\max_{z} p(X|Z)p(Z). \quad (12)$$

Solving (12) also leads to a shrinking and pruning of superfluous covariance components. An EM algorithm can be derived for such a hierarchical model. Over the course of learning, this expectation collapses to zero for the irrelevant hyperparameters in γ.

VB APPROACHES

A third possibility involves finding approximate solutions to the posterior $p(Z|X)$ and the marginal $p(X.)$ The idea is that all unknown quantities should either be marginalized when possible or approximated with tractable distributions, and the distributions reflect underlying uncertainty and have computable posterior moments. Practically, we would like to account for the ambiguity regarding γ when estimating $p(Z|X)$, and, potentially, we would prefer a good approximation for $p(X)$ or a bound on the model evidence $\log p(X)$. Because of the intractable integration involved in obtaining either distribution, practical implementation relies on additional assumptions to yield different types of approximation strategies.

VB methods have been successfully applied to a wide variety of hierarchical Bayesian (HB) models in the machine-learning literature, which offer an alternative to EB and penalized likelihood methods [8]. Two of the most popular forms of variational approximations are the mean field approximation (VB-MF) and Laplace approximation (VB-LA). The mean field approximation assumes that the joint distribution over unknowns Z and γ is factorial: $p(Z,\gamma|X) \approx q(Z|X)q(\gamma|X)$, where $q(Z|X)$ and $q(\gamma|X)$ are chosen to minimize the Kullback–Leibler (KL) divergence between the factorized and full posterior. This is accomplished via an iterative process akin to EM, effectively using two E-steps (one for Z and one for γ). It also produces a rigorous lower bound on $\log p(X)$ similar to EB approaches. The second-order Laplace approximation assumes that the posterior on the hyperparameters (after marginalizing over Z is Gaussian, which is then iteratively matched to the true posterior; the result can then be used to approximate $p(Z|X)$ and $\log p(X)$.

Although seemingly different, there are some deep dualities between these three types of analytical approaches [9], [10]. Specifically, it can be shown that the EB approaches are equivalent to penalized likelihood methods when the prior is chosen to be nonfactorial, lead-field, and noise dependent; VB methods are equivalent to EB approaches given an appropriate hyperprior [9], [10]. Theoretical properties concerning the convergence, global and local minima, and localization bias of each of these methods have been analyzed and fast algorithms have been derived that improve upon existing methods [9], [10]. This perspective leads to explicit connections between many established algorithms and suggests natural extensions for handling unknown dipole orientations, extended source configurations, correlated sources, temporal smoothness, and computational expediency. Table 2 lists a high-level comparison of several Bayesian inference paradigms.

[TABLE 2] COMPARISON OF BAYESIAN INFERENCE APPROACHES.

BAYESIAN INFERENCE APPROACH	LATENT VARIABLE ESTIMATE	HIERARCHICAL PRIOR	SPARSITY, MODEL SELECTION	SPEED, CONVERGENCE
EB	FULL POSTERIOR	NO	YES	FAST
PENALIZED LIKELIHOOD	POINT ESTIMATE	YES	YES	FAST
VB	VARIATIONAL POSTERIOR	YES	YES	FAST
MCMC	FULL POSTERIOR	YES	YES	SLOW

SAMPLING APPROACHES

In contrast to VB approaches (deterministic version of approximate Bayesian inference), sampling approaches can be viewed as a stochastic version of approximate Bayesian inference. The basic idea is simple: instead of computing intractable integration in the posterior computation, draw random samples from the distribution and use plug-in samples for calculation. Nevertheless, drawing Monte Carlo samples from an exact target distribution P is still nontrivial.

When direct sampling from a target distribution is difficult, one can instead draw random samples from a proposal distribution Q. For instance, the importance sampling method first draws samples $\xi_i \sim Q$, and then weighs each sample with a nonnegative importance weight function $w_i(\xi_i) = P(\xi_i)/Q(\xi_i)$ [11]. Alternatively, the Markov chain Monte Carlo (MCMC) method draws samples sequentially from a Markov chain; when the Markov chain converges to the equilibrium point, the samples are viewed approximately from the target distribution. The most common MCMC sampling methods include the Metropolis–Hastings (MH) algorithm [12] and Gibbs sampling [13]. The MH algorithm is the simplest yet the most generic MCMC method to generate samples using a random walk and then to accept them with a certain acceptance probability. For example, given a random-walk proposal distribution $g(z \to z')$ (which defines a conditional probability of moving state z to z'), the MH acceptance probability $\mathcal{A}(z \to z')$ is computed as

$$\mathcal{A}(z \to z') = \min\left(1, \frac{p(z')g(z' \to z)}{p(z)g(z \to z')}\right),$$

which gives a simple MCMC implementation and satisfies the "detailed balance" condition. Gibbs sampling is another popular MCMC method that requires no parameter tuning. Given a high-dimensional joint distribution $p(z) = p(z_1, \ldots, z_M)$, the Gibbs sampler draws samples from the individual conditional distribution $p(z_i|z_{-i})$ in turn while holding others fixed (where z_{-i} denotes the $M-1$ variables in z except for z_i).

Since the random-walk model is highly inefficient in a high-dimensional space, other sampling methods attempt to exploit side information of the likelihood, such as the Hamiltonian sampling method and gradient-based Langevin MCMC [14]. The topics of interest in this article include latent variable modeling, efficient inference, model selection, sparsity (compressed sensing), group analysis, nonstationarity, and NB modeling. We will illustrate ideas using some examples and also provide

reference pointers to state-of-the-art research. Many topics are inherently interrelated (see Figure 1):

- Latent variable modeling and hierarchy are the central themes of probabilistic or Bayesian modeling.
- Hierarchy is a natural way to characterize random effects, nonstationarity, and group variability.
- Depending on the hyperprior/prior assumption or optimization criterion (e.g., posterior, marginal likelihood, and variational free energy), one can derive various exact or approximate Bayesian inference algorithms.
- Sparsity is a prior assumption imposed on the model, which can also be used for model selection.

VB METHODS

This section presents several efficient VB methods developed for EEG/MEG signal processing. Let Θ denote collectively all the unknown parameters to be inferred in (1). We focus on the VB-MF approximation, which seeks a factorable distribution q to approximate the true posterior p by minimizing the KL divergence

$$\mathrm{KL}(q\|p) = -\int q(\Theta)\log\left[\frac{p(\Theta|X)}{q(\Theta)}\right]\mathrm{d}\Theta, \tag{13}$$

where $q(\Theta) = \prod_{l=1}^{L} q(\Theta_l)$, and $\Theta_l (l = 1, \cdots, L)$ denote the disjoint groups of variables in Θ. Direct minimization of (13) is difficult, but it is easy to show that the log marginal likelihood of the data can be written as

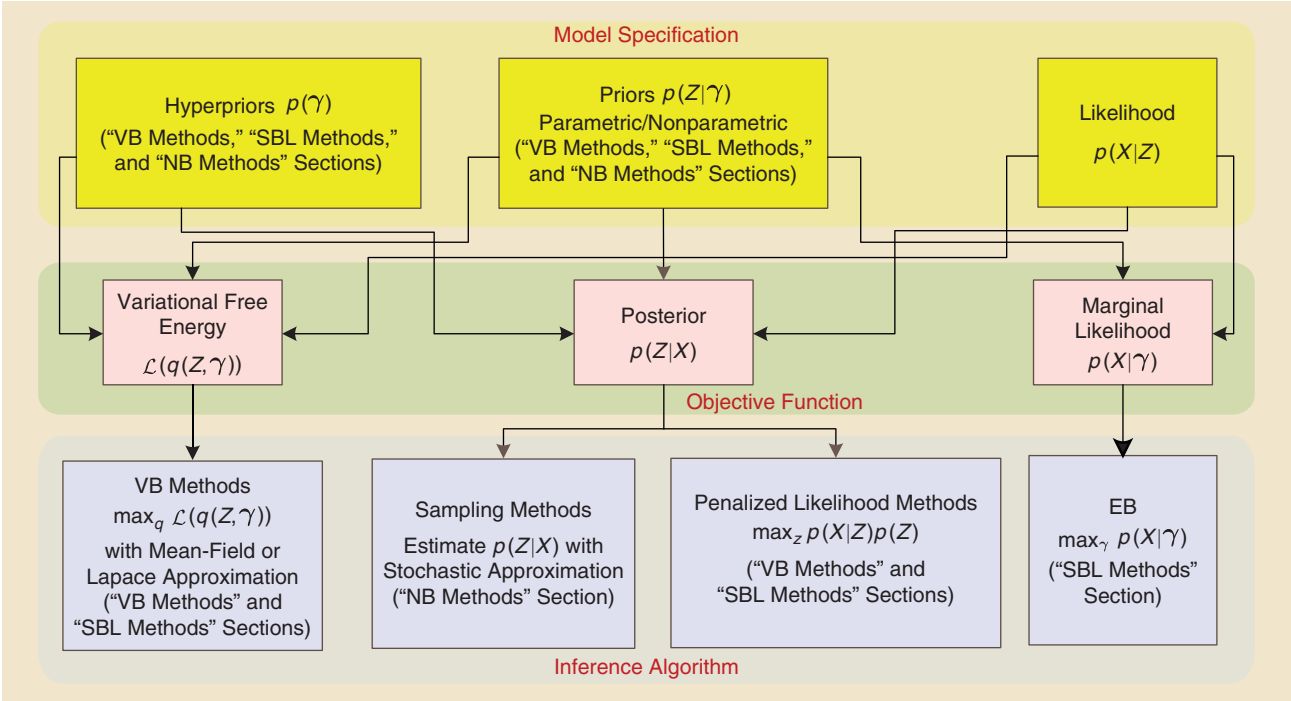
$$\log p(X) = \mathcal{L}(q) + \mathrm{KL}(q\|p), \tag{14}$$

with

$$\mathcal{L}(q) = \int q(\Theta)\log\left[\frac{p(X,\Theta)}{q(\Theta)}\right]\mathrm{d}\Theta. \tag{15}$$

Since $\mathrm{KL}(q\|p) \geq 0$, $\mathcal{L}(q)$ is a lower bound of the log marginal likelihood, which is termed the *variational free energy*. In light of (14), minimizing the KL divergence is equivalent to maximizing $\mathcal{L}(q)$, which is more computationally tractable using numerical optimization algorithms. The most widely used algorithm is coordinate ascent, which alternately updates the approximate distribution of each parameter $\{q(\Theta_l)\}$ until convergence. In brief, this coordinate ascent can be thought of as the solution for one approximate posterior marginal distribution that is expressed in terms of the others. By stepping through the different subsets of unknown parameters, we can iteratively update the approximate marginals.

In the presence of latent variable Z, the variational posterior is often assumed to have a factorial form: $q(Z,\Theta) = q(Z)q(\Theta)$. Similarly, maximizing the variational free energy with respect to two functions $q(Z)$ and $q(\Theta)$ alternatively gives rise to the so-called VB-EM (expectation-maximization) algorithm [15]. In the VB-E (expectation) step, set $[\partial\mathcal{L}/\partial q(Z)] = 0$ and update the variational posterior $q(Z)$; in the VB-M (maximization) step, set $[\partial\mathcal{L}/\partial q(\Theta)] = 0$ and update



[FIG1] **A schematic illustration of Bayesian inference and the BML techniques introduced in this tutorial. Bayesian inference is carried out via two phases: model specification and inference. In the model-specification phase, the likelihood $L(Z) = p(X|Z)$, which describes how the EEG/MEG data X are related to the unknown variables Z (including parameters and latent variables) is specified. Parametric or nonparametric priors can be imposed on Z. The hierarchy is built up by imposing hyperpriors $p(\gamma)$ on the hyperparameters γ in the priors. In the inference phase, a particular algorithm is chosen to infer Z, based on deterministic or stochastic approximations. Under specific conditions, empirical Bayesian methods, penalized likelihood methods, and VB methods are equivalent to each other. See the section "Bayesian Inference: A Brief Overview" for details.**

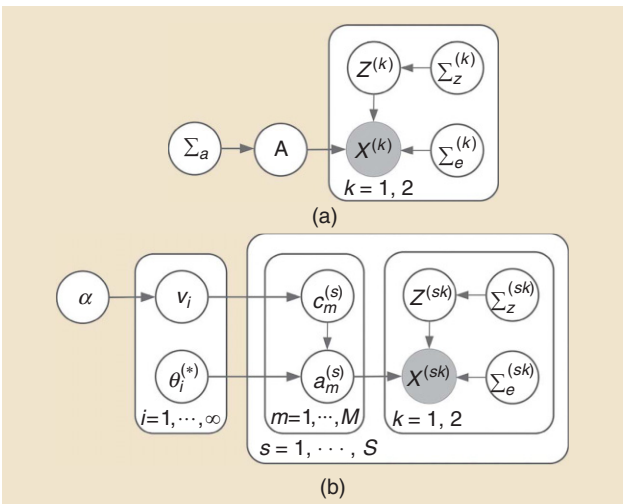
the variational posterior $q(\Theta)$; repeat these two steps until convergence.

VB FOR LEARNING COMMON EEG COMPONENTS

Learning common EEG components that distinguish different conditions is an effective feature-extraction strategy for brain-state classification. Among the various methods, the common spatial patterns (CSP) algorithm [16], which finds common EEG components with the largest variance ratios between two conditions, has attracted considerable attention as being successful in extracting sensorimotor rhythms for BCIs. Nonetheless, as a multivariate algorithm, CSP is known to suffer from overfitting, which may yield poor generalization performance [16]. Although various regularization strategies may be utilized to ameliorate CSP's overfitting, the algorithm was designed primarily for classifying instead of modeling the EEG data and, therefore, not specifically designed for exploring the underlying spatiotemporal dynamics.

A VB-EM algorithm called the *VB-CSP algorithm* was developed in [17] to address the overfitting issue of CSP. The algorithm is based on the following model that is a multicondition extension of (2):

$$\begin{aligned} X^{(k)} &= \mathrm{A}Z^{(k)} + E^{(k)} \\ \mathrm{A} &\sim \prod_{n=1}^{N} \mathcal{N}(\mathrm{a}_n \mid 0, \Sigma_a),\ Z^{(k)} \sim \prod_{t=1}^{T} \mathcal{N}(z_t^{(k)} \mid 0, \Sigma_z^{(k)}),\ E^{(k)} \\ &\sim \prod_{t=1}^{T} \mathcal{N}(e_t^{(k)} \mid 0, \Sigma_e^{(k)}) \\ \sigma_m &\sim \mathcal{IG}(\sigma_m \mid \alpha, \beta),\ \rho_m^{(k)} \sim \mathcal{IG}(\rho_m^{(k)} \mid \alpha, \beta),\ \eta_n^{(k)} \sim \mathcal{IG}(\eta_n^{(k)} \mid \alpha, \beta), \end{aligned} \tag{16}$$



[FIG2] Graphical model representations for two extended CSP models: (a) the VB-CSP model [17] and (b) the Bayesian CSP-DP model [18]. VB-CSP is parametric, while Bayesian CSP-DP is nonparametric. In VB-CSP, $X^{(k)}$ denotes the EEG data for the kth condition. A and $Z^{(k)}$ are the unknown mixing matrix and latent variables, respectively. Σ_a, $\Sigma_z^{(k)}$, and $\Sigma_e^{(k)}$ are the covariance matrices for A, $Z^{(k)}$, and $E^{(k)}$, respectively. All unknown latent variables and parameters are shown as unshaded nodes and estimated from the data using the update rules provided in Algorithm 1. In the Bayesian CSP-DP model, s represents the subject index. It is assumed that $\alpha \sim \mathcal{G}(a_0, b_0)$, $v_i \sim \mathcal{B}(1, \alpha)$, $p(c_m^{(s)}(i) = 1) = v_i \prod_{j=1}^{i-1}(1 - v_j)$, $a_m^{(s)} \mid z_m^{(s)}(i) = 1 \sim \mathcal{N}(\mu_i^*, (\Phi_i^*)^{-1})$ (see the section "DP Modeling" for details).

where $\Sigma_a \triangleq \mathrm{diag}[\sigma_{1:M}]$, $\Sigma_z^{(k)} \triangleq \mathrm{diag}[\rho_{1:M}^{(k)}]$, and $\Sigma_e^{(k)} \triangleq \mathrm{diag}[\eta_{1:N}^{(k)}]$. $\mathcal{IG}$ denotes the inverse-gamma distribution: $\mathcal{IG}(x \mid \alpha, \beta) \triangleq [\beta^\alpha / \Gamma(\alpha)] x^{-\alpha-1} \exp(-\beta/x)$. The corresponding graphical model is shown in Figure 2(a).

Equation (16) provides a principled generative framework for interpreting and improving CSP, since it can be shown that the ML estimation of A in the noiseless and square mixing setup (i.e., $E^{(k)} = 0$ and $M = N$) leads to the same cost function as CSP [19]. The overfitting issue of CSP stems from the square mixing and noiseless assumptions. The noiseless assumption implies that the EEG data are fully characterized by the estimated components and the mixing matrix. This assumption does not take into account noise or interference. The square mixing assumption is closely linked to the noiseless assumption in that, if we relax the square mixing assumption by using a smaller number of components, a model mismatch will automatically arise between the best linear fit and the EEG data.

We further assume in (16) that $\alpha \to 0, \beta \to 0$ to impose noninformative priors on σ_m, $\rho_m^{(k)}$, and $\eta_n^{(k)}$ [11]. This also allows us to leverage Bayes' rule to achieve automatic model size determination in conjunction with parameter estimation. However, exact Bayesian inference is not viable for (16) due to the product coupling of A and $Z^{(k)}$ in the likelihood, as well as the inconvenient form of the hierarchical priors.

Through the VB-MF approximation to $p(\mathrm{A}, Z^{(k)} \mid X^{(k)})$, the variational free energy $\mathcal{L}(q(\mathrm{A}, Z^{(k)}))$ can be derived as in (15), which can be further lower bounded by invoking Fenchel's duality to locally approximate the hierarchical priors on A, $Z^{(k)}$, and $E^{(k)}$, yielding

$$\begin{aligned} \tilde{\mathcal{L}}(q(\mathrm{A}, Z^{(k)})) \triangleq \min_{\substack{\Sigma_a, \Sigma_z^{(k)} \\ \Sigma_e^{(k)}}} & \\ -(\tfrac{L}{2} + \alpha) \sum_k \log |\Sigma_e^{(k)}| - (\tfrac{L}{2} + \alpha) \sum_k \log |\Sigma_z^{(k)}| & \\ -(\tfrac{N}{2} + \alpha) \log |\Sigma_a| & \\ -\frac{1}{2} \sum_k \mathbb{E}_q[\mathrm{trace}[\Sigma_e^{(k)-1}[(X^{(k)} - \mathrm{A}Z^{(k)})(X^{(k)} - \mathrm{A}Z^{(k)})^\top + 2\beta \mathrm{I}]]] & \\ -\frac{1}{2} \sum_k \mathbb{E}_q[\mathrm{trace}[\Sigma_z^{(k)-1}(Z^{(k)} Z^{(k)\mathrm{T}} + 2\beta \mathrm{I})]] & \\ -\frac{1}{2} \mathbb{E}_q[\mathrm{trace}[\Sigma_e^{(k)-1}(\mathrm{A}^\mathrm{T}\mathrm{A} + 2\beta \mathrm{I})]] & \\ -\mathbb{E}_q[\log q(\mathrm{A})] - \sum_k \mathbb{E}_q[\log q(Z^{(k)})]. & \end{aligned}$$

VB-CSP uses the coordinate descent to solve the following optimization problem:

$$\min_{q(\mathrm{A}),\, q(Z^{(k)})} \tilde{\mathcal{L}}(q(\mathrm{A}), q(Z^{(k)})). \tag{17}$$

Detailed derivations of VB-CSP are referred to in [17]. The pseudocode is provided in Algorithm 1. Because of the multiplicative structure of (16), the unknown variables can only be identified up to the permutation ambiguity unless additional constraints are imposed. Hence, the posterior distribution is intrinsically multimodal. This naturally casts doubts on the validity of VB for approximate inference, given that the resulting approximate posterior is unimodal. According to (13), there is a large positive contribution to the KL divergence from regions of $\{\mathrm{A}, Z^{(k)}\}$ space in which

Algorithm 1: The VB-CSP algorithm.

Input: $X^{(k)}$

Output: $\Sigma_a, \Sigma_z^{(k)}, \Sigma_e^{(k)}, q(Z^{(k)}), q(\mathbf{A})$

1: **Initialization**

2: **repeat**

3: $q(Z_k) = \prod_{t=1}^{T} \mathcal{N}(z_t^{(k)} \mid \mathbb{E}[z_t^{(k)}], \mathbb{C}[z_t^{(k)}])$,

where $\mathbb{E}[z_t^{(k)}] \triangleq \mathbb{C}[z_t^{(k)}]\mathbb{E}^{\mathrm{T}}[\mathbf{A}][\Sigma_e^{(k)}]^{-1}x_t^{(k)}$, $\mathbb{C}[z_t^{(k)}] \triangleq [\mathbb{E}^{\mathrm{T}}[\mathbf{A}][\Sigma_e^{(k)}]^{-1}\mathbb{E}[\mathbf{A}] + \sum_n \mathbb{C}[\mathbf{a}_n^{(k)}]/\eta_n^{(k)} + [\Sigma_z^{(k)}]^{-1}]^{-1}$

4: $q(\mathbf{A}) = \prod_{n=1}^{N} \mathcal{N}(\mathbf{a}_n \mid \mathbb{E}[\mathbf{a}_n], \mathbb{C}[\mathbf{a}_n])$,

where $\mathbb{E}[\mathbf{a}_n] \triangleq \sum_{t,k} x_{t,n}^{(k)} \mathbb{E}^{\top}[z_t^{(k)}]\mathbb{C}[\mathbf{a}_n]/\eta_n^{(k)}$, $\mathbb{C}[\mathbf{a}_n] \triangleq [\Sigma_a^{-1} + \sum_{t,k}(\mathbb{C}[z_t^{(k)}] + \mathbb{E}[z_t^{(k)}]\mathbb{E}^{\top}[z_t^{(k)}])/\eta_n^{(k)}]^{-1}$

5: $\Sigma_z^{(k)} = \frac{1}{T+2\alpha}\sum_t (\mathrm{diag}[\mathbb{C}[z_t^{(k)}] + \mathbb{E}[z_t^{(k)}]\mathbb{E}^{\top}[z_t^{(k)}]] + 2\beta\mathbf{I})$

6: $\Sigma_e^{(k)} = \frac{1}{T+2\alpha}\sum_t (\mathrm{diag}[X_t^{(k)}[X_t^{(k)}]^{\top} - 2X_t^{(k)}[Z_t^{(k)}]^{\top}\mathbb{E}^{\top}[\mathbf{A}] + \mathbb{E}[\mathbf{A}[\mathbb{C}[z_t^{(k)}] + \mathbb{E}[z_t^{(k)}]\mathbb{E}^{\top}[z_t^{(k)}]]\mathbf{A}^{\top})] + 2\beta\mathbf{I}]$

7: $\Sigma_a = \frac{1}{N+2\alpha}\sum_n (\mathrm{diag}[\mathbb{C}[\mathbf{a}_n] + \mathbb{E}^{\top}[\mathbf{a}_n]\mathbb{E}[\mathbf{a}_n]] + 2\beta\mathbf{I})$

8: **until** Convergence

$p(\mathbf{A}, Z^{(k)} \mid X^{(k)})$ is near zero unless $q(\mathbf{A}, Z^{(k)})$ is also close to zero. Thus, minimizing the KL divergence in (13) leads to distributions $q(\mathbf{A}, Z^{(k)})$ that avoid regions in which $p(\mathbf{A}, Z^{(k)} \mid X^{(k)})$ is small. This means that VB tends to find one of the modes in the posterior distribution. Hence, the unimodality of the VB posterior does not nullify the effectiveness in the inferring (16). Similar arguments apply to the other VB-based spatiotemporal decomposition algorithms discussed in this article.

Our empirical observations indicated that the algorithmic performance is only slightly affected by initialization (see sensitivity analysis in [17]). VB-CSP is particularly suited for the single-trial analysis of motor imagery EEG data, since imagined movements give rise to an attenuation of the sensorimotor rhythms in specific regions of the sensorimotor cortices, a phenomenon known as *event-related desynchronization,* which can be captured by evaluating the variance change of EEG spatial patterns across conditions.

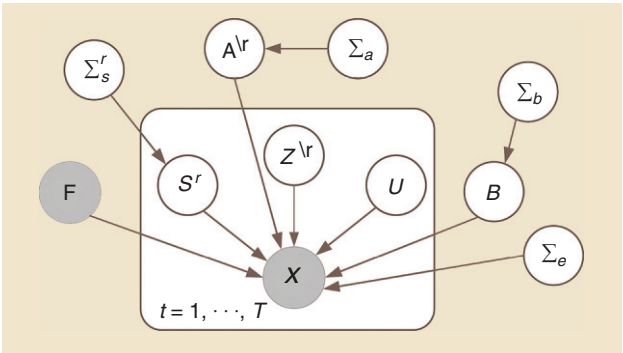
VB FOR LOCALIZING EVOKED SOURCE ACTIVITY WITH INTERFERENCE SUPPRESSION

EEG/MEG sensor measurements are often contaminated by many types of interference, such as background activity from outside the regions of interest, biological and nonbiological artifacts, and sensor noise. Source activity using knowledge of event timing for independence from noise and interference (SAKETINI) is a probabilistic graphical modeling algorithm that localizes and estimates neural source activity measured by EEG/MEG data while suppressing the interference or noise [20]. It is assumed that the recorded EEG/MEG data are separated by the zero time point (stimulus onset or experimental marker). Ongoing brain activity, biological noise, background environmental noise, and sensor noise are present in both the prestimulus (before zero) and poststimulus (after zero) periods; however, the evoked neural sources of interest are only present in the poststimulus time period. It is also assumed that the sensor data can be described as coming from four types of sources: 1) an evoked source at a particular voxel, 2) all other evoked sources not at that voxel, 3) all background noise sources with spatial covariance at the sensors (including the brain, biological, or environmental sources), and 4) sensor noise. The first step is to infer the model describing source types 3 and 4 from the prestimulus data, and the second step is to infer the full model describing the remaining source types 1 and 2 from the poststimulus data. After inference of the model, a map of the source activity is created, as well as a map of the likelihood of activity across voxels.

Let the time-ranges for prestimulus and poststimulus data be T_{pre} and T_{post}, respectively. The generative model for data $X = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T] \in \mathbb{R}^{N\times T}$ from N sensors is assumed as follows:

$$\boldsymbol{x}_t = \begin{cases} \mathbf{B}\boldsymbol{u}_t + e_t & t \in T_{\mathrm{pre}} \\ \mathbf{F}^r \boldsymbol{s}_t^r + \mathbf{A}^{\backslash r} z_t^{\backslash r} + \mathbf{B}\boldsymbol{u}_t + e_t & t \in T_{\mathrm{post}}. \end{cases} \tag{18}$$

The lead-field matrix $\mathbf{F}^r \in \mathbb{R}^{N\times q}$ represents the physical (and linear) relationship between a dipole source at voxel r for each dipole orientation q along a coordinate basis and its influence on sensors [21]. The source activity $S^r = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_T] \in \mathbb{R}^{q\times T}$ matrix of dipole time course for the voxel r. The $\mathbf{A}^{\backslash r} \in \mathbb{R}^{N\times L}$ and $Z^{\backslash r} = [z_1^{\backslash r}, \ldots, z_T^{\backslash r}] \in \mathbb{R}^{L\times T}$ represent the poststimulus mixing matrix and evoked nonlocalized factors, respectively, corresponding to source type 2 discussed previously, where the superscript $\backslash r$ indicates for all voxels not at voxel r. $\mathbf{B} \in \mathbb{R}^{N\times M}$ and $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T] \in \mathbb{R}^{M\times T}$ represent the background mixing matrix and background factors, respectively, corresponding to source type 3. Representing the sensor-level noise is e_t, which is assumed to be drawn from a Gaussian distribution with zero-mean and precision (inverse covariance) defined by the diagonal matrix Σ_e. All quantities depend on r in the poststimulus period, except for $\mathbf{B}, \boldsymbol{U},$ and Σ_e, which will be learned from the prestimulus data. Note, however, that the posterior update for $\boldsymbol{U}$ (represented by $\overline{\boldsymbol{U}}$) does depend on the voxel r. The graphical model is shown in Figure 3.



[FIG3] **A graphical model representation for SAKETINI. The observed data X are assumed to be generated from a linear mixing of several latent sources. Sources S^r correspond to brain activity that contributes to sensor data through a known mixing matrix F. Brain activity not located at the scanning location r is characterized by a set of factors $Z^{\backslash r}$ with an unknown mixing matrix $A^{\backslash r}$. Additionally, background brain activity is assumed to arise from a set of factors U, associated with an unknown mixing matrix B. The mixing matrices A and B are assumed to be drawn from a Gaussian prior with covariance Σ_a and Σ_b, respectively. Finally, sensor data is assumed to have sensor noise that is also drawn from a Gaussian prior with covariance Σ_e. All unknown latent variables and parameters are shown as unshaded nodes and estimated from the data using update rules provided in Algorithm 2.**

LEARNING BACKGROUND BRAIN ACTIVITY FROM PRESTIMULUS DATA

We use a VB factor analysis (VBFA) approach to describe the part of the sensor signals contributed by background factors, as arising from L underlying factors, via an unknown mixing matrix. In mathematical terms, let $\boldsymbol{u}_t$ denote the M-dimensional vector corresponding to the background factors. We assume Gaussian prior distributions on the background factors and sensor noise and a flat prior on the sensor noise precision. We further assume the background factors are independent and identically distributed (i.i.d.) across time. Therefore, the following is assumed for $t \in T_{\text{pre}}$:

$$p(U) = \prod_t p(\boldsymbol{u}_t); \quad p(\boldsymbol{u}_t) = \mathcal{N}(0, \mathrm{I}), \tag{19}$$

$$p(E) = \prod_t p(\boldsymbol{e}_t); \quad p(\boldsymbol{e}_t) = \mathcal{N}(0, \Sigma_e), \quad p(\Sigma_e) = \text{const.} \tag{20}$$

$$p(X \mid U, \mathrm{B}, \Sigma_e) = \prod_t p(\boldsymbol{x}_t \mid \boldsymbol{u}_t, \mathrm{B}, \Sigma_e) = \prod_t \mathcal{N}(\mathrm{B}\boldsymbol{u}_t, \Sigma_e). \tag{21}$$

We use a conjugate prior for the background mixing matrix B, as follows:

$$p(\mathrm{B}) = \prod_{n=1}^{N} \prod_{m=1}^{M} p(b_{nm}) = \prod_{n=1}^{N} \prod_{m=1}^{M} \mathcal{N}(0, \lambda_n \beta_m), \tag{22}$$

where β_m is a hyperparameter over the mth column of B, and λ_n is the precision of the nth sensor. The matrix $\Sigma_b = \text{diag}(\beta_1, ..., \beta_M)$ provides a robust mechanism for automatic model order selection (see the "Introduction" section), so that the optimal size of B is inferred from the data through Σ_b.

Exact inference on this model is also intractable, and we choose to factorize the marginal conditional posterior distribution, assuming conditional independence of the background factors U and mixing matrix B

$$p(U, \mathrm{B} \mid X) \approx q(U, \mathrm{B}) = q(U) q(\mathrm{B}). \tag{23}$$

The VB-EM algorithm is used to iteratively maximize the free-energy with respect to $q(U)$, $q(\mathrm{B})$, Σ_b, and Σ_e.

LOCALIZATION OF EVOKED SOURCES LEARNED FROM POSTSTIMULUS DATA

In the stimulus-evoked paradigm, the source strength at each voxel is learned from the poststimulus data. The background mixing matrix B and sensor noise precision Σ_e are assumed to be fixed after prestimulus estimation. We assume those quantities remain constant through the poststimulus period and are independent of source location.

The source factors have Gaussian distribution with zero mean and covariance $\Sigma_s^r \in \mathbb{R}^{q \times q}$, which relates to the strength of the dipole in each of q directions

$$p(S^r) = \prod_t p(\boldsymbol{s}_t^r); \quad p(\boldsymbol{s}_t^r) = \mathcal{N}(0, \Sigma_s^r). \tag{24}$$

The interference and background factors are assumed to be Gaussian with zero mean and identity covariance matrix (i.e., non-informative prior).

$$p(Z^{\backslash r}) = \prod_t p(\boldsymbol{z}_t^{\backslash r}); \quad p(\boldsymbol{z}_t^{\backslash r}) = \mathcal{N}(0, \mathrm{I}),$$
$$p(U) = \prod_t p(\boldsymbol{u}_t); \quad p(\boldsymbol{u}_t) = \mathcal{N}(0, \mathrm{I}). \tag{25}$$

We also use a conjugate prior for the interference mixing matrix $A^{\backslash r}$, where $\Sigma_a = \text{diag}(\alpha_1, ..., \alpha_L)$ is a hyperparameter that helps determine the size of $A^{\backslash r}$ (in a similar fashion for matrix B)

$$p(\mathrm{A}^{\backslash r}) = \prod_{n=1}^{N} \prod_{l=1}^{L} p(a_{nl}) = \prod_{n=1}^{N} \prod_{l=1}^{L} \mathcal{N}(0, \lambda_n \alpha_l). \tag{26}$$

We now specify the full model

$$p(X \mid S^r, Z^{\backslash r}, U, \mathrm{A}^{\backslash r}, \mathrm{B}, \Sigma_e) = \prod_t p(\boldsymbol{x}_t \mid \boldsymbol{s}_t^r, \boldsymbol{z}_t^{\backslash r}, \boldsymbol{u}_t, \mathrm{A}^{\backslash r}, \mathrm{B}, \Sigma_e)$$
$$= \prod_t \mathcal{N}(\mathrm{F}^r \boldsymbol{s}_t^r + \mathrm{A}^{\backslash r} \boldsymbol{z}_t^{\backslash r} + \mathrm{B}\boldsymbol{u}_t, \Sigma). \tag{27}$$

Exact inference on this model is also intractable, and we similarly use VB approximation for evoked nonlocalized factors $Z^{\backslash r}$ and mixing matrix $A^{\backslash r}$

$$p(S^r, Z^{\backslash r}, U, \mathrm{A}^{\backslash r} \mid X) \approx q(S^r, Z^{\backslash r}, U, \mathrm{A}^{\backslash r} \mid X)$$
$$= q(S^r, Z^{\backslash r}, U \mid X) q(\mathrm{A}^{\backslash r} \mid X). \tag{28}$$

All variables, parameters and hyperparameters are unknown and learned from the data using the VB-EM algorithm (Algorithm 2).

Extensions of this approach, whose performance benefits are yet to be worked out or determined, are to assume non-Gaussian prior distributions for the evoked or background factors using either mixture-of-Gaussian or Gaussian-scale mixture distributions [22], [15].

VB FOR GROUP EEG/MEG ANALYSIS

EEG/MEG recordings often arise from a multilevel structure (e.g., multiple subjects/sessions/trials), and group analysis at each level entails building statistical models that characterize the homogeneity and variability within the group. One way to impose the group structure is to use HB models [11], with the variations at

Algorithm 2: The SAKETINI algorithm.

Input: X for T_{pre} and T_{post}

Output: $q(U), q(S^r), q(Z^{\setminus r}), q(\mathrm{A}^{\setminus r}), q(\mathrm{B}), \Sigma_e, \Sigma_b, \Sigma_a$

1: **Initialization**

2: **repeat**

3: $q(\mathrm{B}\,|\,X) = \mathcal{N}(\overline{\mathrm{B}}, \psi)$, where $\overline{\mathrm{B}} = \mathrm{R}_{\mathrm{XU}}\psi, \psi = (\mathrm{R}_{\mathrm{UU}} + \Sigma_b)^{-1}, \Sigma_b^{-1} = \mathrm{diag}\left(\frac{1}{N}\overline{\mathrm{B}}^{\top}\Sigma_e\overline{\mathrm{B}} + \psi\right), \Sigma_e^{-1} = \frac{1}{T}\mathrm{diag}\left(\mathrm{R}_{\mathrm{XX}} - \overline{\mathrm{B}}\mathrm{R}_{\mathrm{XU}}^{\top}\right), \mathrm{R}_{\mathrm{XX}} = \sum_t \boldsymbol{x}_t\,\overline{\boldsymbol{x}}_t^{\top}$ and $\mathrm{R}_{\mathrm{XU}} = \sum_t \boldsymbol{x}_t\,\overline{\boldsymbol{u}}_t^{\top}$.

4: $q(U\,|\,X) = \prod_t q(\boldsymbol{u}_t\,|\,\boldsymbol{x}_t) = \prod_t \mathcal{N}(\overline{\boldsymbol{u}}_t, \gamma)$, where $\overline{\boldsymbol{u}}_t = \gamma^{-1}\overline{\mathrm{B}}^{\top}\Sigma_e\,\boldsymbol{x}_t, \gamma = \overline{\mathrm{B}}^{\top}\Sigma_e\,\overline{\mathrm{B}} + N\psi^{-1} + \mathrm{I}$

5: Let $V = (S^r\ Z^{\setminus r}\ U)$, then $q(V\,|\,X) = \prod_t q(\boldsymbol{v}_t\,|\,\boldsymbol{x}_t) = \prod_t \mathcal{N}(\overline{\boldsymbol{v}}_t, \Gamma)$, where $\overline{\boldsymbol{v}}_t = \Gamma^{-1}\overline{\mathrm{A}}_{\mathrm{aug}}^{\top}\,\Sigma_e\,\boldsymbol{x}_t, \Gamma = \overline{\mathrm{A}}_{\mathrm{aug}}^{\top}\Sigma_e\,\overline{\mathrm{A}}_{\mathrm{aug}} + N\Psi + \mathrm{I}_{\mathrm{aug}}$, where $\overline{\mathrm{A}}_{\mathrm{aug}} = (\mathrm{F}^r\overline{\mathrm{A}}^{\setminus r}\,\overline{\mathrm{B}}), \mathrm{I}_{\mathrm{aug}} = \begin{pmatrix}\Phi & 0 & 0\\ 0 & \mathrm{I} & 0\\ 0 & 0 & \mathrm{I}\end{pmatrix}, \Psi = \begin{pmatrix}0 & 0 & 0\\ 0 & \Psi_{\mathrm{AA}} & 0\\ 0 & 0 & 0\end{pmatrix}$. From marginalization, we also obtain $q(S^r\,|\,X)$ and $q(Z^{\setminus r}\,|\,X)$.

6: $q(\mathrm{A}^{\setminus r}\,|\,X) = \mathcal{N}(\overline{\mathrm{A}}, \Psi_{\mathrm{AA}})$, where $\overline{\mathrm{A}} = (\mathrm{R}_{\mathrm{XZ}} - \mathrm{F}^r\mathrm{R}_{\mathrm{SZ}} - \overline{\mathrm{B}}\mathrm{R}_{\mathrm{UZ}})\Psi_{\mathrm{AA}}, \Psi_{\mathrm{AA}} = (\mathrm{R}_{\mathrm{ZZ}} + \Sigma_a)^{-1}, \Sigma_a^{-1} = \mathrm{diag}\left(\frac{1}{N}\overline{\mathrm{A}}^{\top}\Sigma_e\overline{\mathrm{A}} + \Psi_{\mathrm{AA}}\right)$, where the covariance matrices can be defined, for instance: $\mathrm{R}_{\mathrm{SS}} = \sum_t \overline{\boldsymbol{s}}_t\,\overline{\boldsymbol{s}}_t^{\top} + N\,\Sigma_{\mathrm{SS}}$, and similarly for $\mathrm{R}_{\mathrm{SZ}}, \mathrm{R}_{\mathrm{SU}}, \mathrm{R}_{\mathrm{UU}}, \mathrm{R}_{\mathrm{ZU}}, \mathrm{R}_{\mathrm{UU}}$, where $\Sigma = \Gamma^{-1}$ is specified as: $\Sigma = \begin{pmatrix}\Sigma_{\mathrm{SS}} & \Sigma_{\mathrm{SZ}} & \Sigma_{\mathrm{SU}}\\ \Sigma_{\mathrm{SZ}}^{\top} & \Sigma_{\mathrm{ZZ}} & \Sigma_{\mathrm{ZU}}\\ \Sigma_{\mathrm{SU}}^{\top} & \Sigma_{\mathrm{ZU}}^{\top} & \Sigma_{\mathrm{UU}}\end{pmatrix}$

7: **until** Convergence

each low level described by a submodel and variations in the hyperparameters of the submodels described by an upper-level model. For instance, let $X^{(s)}(s = 1, \cdots, S)$ denote a group of S-subject EEG/MEG data sets, with s being the subject index; we may have

$$X^{(s)} \sim p(X^{(s)}\,|\,\Theta^{(s)}), \quad \Theta^{(s)} \sim p(\Theta^{(s)}\,|\,\Omega). \tag{29}$$

The upper-level prior acts as the group constraint on the parameters, $\Theta^{(s)}$, of each subject's model. The strength of HB modeling is that the hyperparameters in an upper-level model can be effectively estimated by pooling information across the data in the lower levels. As a benefit, HB models have proven to be robust, with the posterior distribution being less sensitive to the hierarchical priors. A two-level HB model was proposed to characterize the intertrial amplitude variability in the EEG signals during motor imagery tasks [19]. In this article, we use an example taken from [23] to illustrate an alternative group modeling approach to achieve multisubject electromagnetic brain imaging.

In the multisubject setup, the generative model (2) can be compactly written in the following form:

$$\begin{bmatrix}X^{(1)}\\ \vdots\\ X^{(S)}\end{bmatrix} = \begin{bmatrix}\mathrm{A}^{(1)} & & \\ & \ddots & \\ & & \mathrm{A}^{(S)}\end{bmatrix}\begin{bmatrix}Z^{(1)}\\ \vdots\\ Z^{(S)}\end{bmatrix} + \begin{bmatrix}E^{(1)}\\ \vdots\\ E^{(S)}\end{bmatrix}, \tag{30}$$

or in a compact form $\overline{X} = \overline{\mathrm{A}}\,\overline{Z} + \overline{E}$, where the overline symbol denotes the concatenated matrix form. Here, for simplicity of description, we assume that time samples are i.i.d., i.e., $E^{(s)} \sim \prod_{t=1}^{T}\mathcal{N}(e_t^{(s)}\,|\,0, \Sigma_e^{(s)})$. It is also assumed that the noise levels are proportional across subjects, i.e., $\Sigma_e^{(s)} = \sigma^{(s)}\mathrm{Q}_e$, where Q_e is a known covariance component over channels. To facilitate group analysis, a canonical mesh is warped to each individual subject's cortical sheet such that the reconstructed source activity can be assigned to the same sources over subjects. The group structure is then enforced by assuming that the source activity for each subject can be factorized into a source-specific component and a subject-specific component

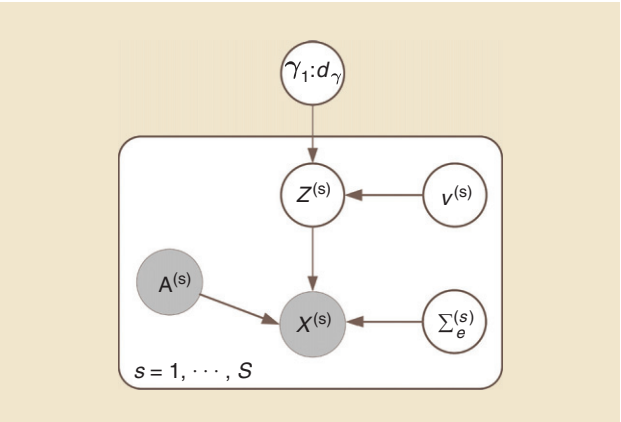
$$p(Z^{(s)}) = \prod_{t=1}^{T} p(z_t^{(s)}) = \prod_{t=1}^{T}\mathcal{N}\left(0, \upsilon^{(s)}\sum_{i=1}^{d_\gamma}\gamma_i\,\mathrm{C}_i\right), \tag{31}$$

where the source-specific scale hyperparameter γ_i represents the group constraint over the subject, and the subject-specific scale hyperparameter $\upsilon^{(s)}$ encodes the additional variability unique to each subject. This prior is a generalization of the source prior (7). In addition, log-normal hyperpriors are placed on the hyperparameters:

$$\lambda \triangleq \log([\gamma_{1:d_\gamma}, \upsilon^{(1:S)}, \sigma^{(1:S)}]^{\top}) \sim \mathcal{N}(\eta, \mathrm{R}).$$

Weakly informative hyperpriors of η and R were used in [23] to allow automatic model selection (see Figure 4 for a graphical model illustration).

For model inference, a computationally efficient two-stage source reconstruction algorithm [termed *gMSP* henceforth since multiple sparse priors akin to (7) are imposed on the source activity] was developed to estimate the hyperparameters $\boldsymbol{\lambda}$ and source activity [23]. The idea is to decompose the covariance function of the group data into a sparse set of source components and a noise component. The sparse set of source components is then used as empirical priors to estimate the subject-specific hyperparameters. More specifically, the



[FIG4] A graphical model representation of the group source imaging algorithm, gMSP, in the presence of S subjects. $X^{(s)}$ and $Z^{(s)}$ are the observed EEG data and unobserved source activity for the sth subject. $A^{(s)}$ and $\Sigma_e^{(s)}$ are the associated lead-field matrix and noise covariance matrix, respectively. $v^{(s)}$ is a subject-specific scale hyperparameter, and $\gamma_{1:d_\gamma}$ are source-specific scale hyperparameters that act as the group constraint. All unknown latent variables and parameters are shown as unshaded nodes and estimated from the data. See the section "VB for Group EEG/MEG Analysis" for details.

mean-field approximation is combined with the Laplace approximation by assuming that the posterior distribution can be factorized into Gaussian marginals (i.e., a combination of VB-MF and VB-LA; see the section "SBL Methods"):

$$q(\overline{Z}, \lambda) = q(\overline{Z})\, q(\boldsymbol{\lambda}) \triangleq \mathcal{N}(\boldsymbol{\mu}_{\tilde{z}}, \Sigma_{\tilde{z}})\, \mathcal{N}(\mu_\lambda, \Sigma_\lambda).$$

This leads to variational free energies at both the individual subject and group levels. The gMSP algorithm then proceeds as follows: in the first stage, $q(\lambda_{1:d_\gamma})$ is estimated by maximizing the group-level variational free energy; in the second stage, $q(\lambda_{(d_\gamma+1):(d_\gamma+2S+1)})$ is estimated by maximizing the subject-level variational free energy. With all the hyperparameters estimated, $q(\overline{Z})$ can be obtained as the MAP estimates of the source activity. Note that, to achieve sparsity on the source covariance components, both ARD and a greedy search approach were developed to optimize the source-specific hyperparameters. Refer to [23] and [24] for algorithmic details. Only gMSP with ARD is illustrated in this article (see the section "Group Electromagnetic Brain Imaging Using gMSP"). This hierarchical model approach allows one to place within- and between-subject constraints on the reconstructed source activity in the context of group studies.

SBL METHODS

Sparse learning, also known as *compressed sensing* in signal processing [25], is referred to as a *collection of learning methods* that seek a tradeoff between certain goodness-of-fit measure and sparsity of the solution, the latter of which allows better interpretability and enhanced generalization ability of the model. Sparse learning is particularly suited for analyzing EEG/MEG signals with high dimensionality, small sample size, and low SNR. See [26] for a recent review on the applications of sparse learning to brain signal processing.

Compared with its most common counterpart, two advantages of SBL [9], [27] are noteworthy:

- SBL allows automatic model selection. This can be achieved by both EB (see the "Introduction" section), and fully Bayesian methods. In EB, maximizing $p(X \mid \gamma)$ provides a natural regularizing mechanism that yields sparse solutions. In fully Bayesian methods, as suggested by the VB methods presented in the sections "VB for Learning Common EEG Components" and "VB for Group EEG/MEG Analysis," automatic sparse learning can be achieved by imposing noninformative priors on γ, which leads to sparsity since the hierarchical priors on the parameters are typically sparse priors. For instance, by marginalizing the variance, the normal-inverse-Gamma prior for each brain source in (16) amounts to a Student t-distribution:

$$p(z_{m,t}^{(k)}) = \int p(z_{m,t}^{(k)} \mid \rho_m^{(k)})\, p(\rho_m^{(k)})\, \mathrm{d}\rho_m^{(k)} = \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)\sqrt{2\pi}} \beta^\alpha [\beta + \frac{[z_{m,t}^{(k)}]^2}{2}]^{-(\alpha+\frac{1}{2})}, \quad (32)$$

  which is sharply peaked at zero in the noninformative case (i.e., when $\alpha \to 0, \beta \to 0$).
- SBL is more capable of finding sparse solutions than conventional methods. In typical electromagnetic brain-imaging problem setups, many columns of the lead-field matrix are highly correlated; in this case, the convex l_1-norm-based MAP (sparse) solution performs poorly since the restricted isometry property is violated. In spatiotemporal decomposition problems where $\mathbf{A}$ is unknown, the MAP estimation may suffer from too many local minima in the solution space due to the multiplicative structure of the spatiotemporal model (2). Consider the multiple covariance parameterization in (7). According to [28], if $\mathbf{C}_i = \mathbf{e}_i \mathbf{e}_i^\top$, where $\mathbf{e}_i$ is an indexing vector with one for the ith element and zero otherwise, the ARD solution can equivalently be viewed as the solution to the following MAP problem:

$$\max_{z_t} \| \boldsymbol{x}_t - \mathbf{A} z_t \|_2^2 + \lambda h^*(z_t), \quad (33)$$

  where λ denotes the regularization parameter, and $h^*(z_t) \triangleq \min_\gamma [z_t^2]^\top \gamma + \log|\Sigma_x|$ is the concave conjugate of $-\log|\Sigma_x|$. The prior distribution associated with $h^*(z_t)$ is generally nonfactorial, i.e., $p(z_t) \neq \prod_m p(z_{m,t})$, and is dependent on both $\mathbf{A}$ and λ. It has been shown that $h(z_t)$ provides a tighter approximation that promotes greater sparsity than the l_1-norm while conveniently producing many fewer local minima than using the l_0-norm [8]. In [17], we establish a similar result for VB-CSP in the spatiotemporal decomposition setup.

In this section, we review two recent works to illustrate SBL in EEG/MEG signal processing.

SBL FOR LEARNING ERPs

ERPs are stereotyped electrical activities of the brain that occur in response to specific events or stimuli. Accurate estimation of the amplitude and latency of ERPs is pivotal in delineating the successive

stages of mental processes and differentiating among various events or stimuli. In multichannel EEG recordings, since the number of ERP components can be considerably smaller than the number of the sensors, SBL is suited to automatically infer the sparse ERP components.

In [29], a Bayesian model was proposed to estimate the components specific to each experimental condition for ERPs from multicondition and multitrial EEG data. More specifically, we adapted (2) to the following linear latent variable model to accommodate the ERP estimation:

$$\boldsymbol{x}_t^{(\mathrm{kl})} = \sum_{m=1}^{M} c_m^{(k)} \mathbf{a}_m z_{m,t+\tau_m^{(k)}} + e_t^{(\mathrm{kl})}, \tag{34}$$

where $\boldsymbol{x}_t^{(\mathrm{kl})} \in \mathbb{R}^N$ and $e_t^{(\mathrm{kl})} \in \mathbb{R}^N$ denote the EEG signal and the noise-plus-interference term from the lth ($l \in \{1, \cdots, L\}$) trial of condition $k \in \{1, \cdots, K\}$, respectively; $z_{m,t}$ denotes the waveform of the stereotyped mth ERP component; $\mathbf{a}_m \in \mathbb{R}^N$ denotes the scalp map of the mth ERP component; and $c_m^{(k)}$ and $\tau_m^{(k)}$ denote the amplitude factor and the latency of the mth ERP component for condition k, respectively. We further assume the noise terms are independent and identically Gaussian distributed across time and trials $e_t^{(\mathrm{kl})} \sim \mathcal{N}(0, \Sigma_e^{(k)})$, where $\Sigma_e^{(k)}$ are the spatial covariance matrices. Note that no particular structure is assumed for $\Sigma_e^{(k)}$ (e.g., $\Sigma_e^{(k)}$ can be nondiagonal).

Two key assumptions are made in (34): 1) the waveform and spatial pattern of each ERP component are invariant across trials and conditions, and 2) intercondition ERPs differ only in their amplitudes and latencies. These assumptions are motivated by the existing experimental observations that ERPs are approximately time- and phase-locked across trials and that ERP variability is typically far less within conditions than across conditions. Furthermore, unlike fixed-effect ERP components, the noise-plus-interference activities $e_t^{(\mathrm{kl})}$ are modeled as random effects in (34), which is also in agreement with existing experimental observations. Crucially, $e_t^{(\mathrm{kl})}$ is allowed to be correlated and nonisotropic among sensors, with the covariance matrices $\Sigma_e^{(k)}$ following noninformative inverse Wishart distributions for $\nu^{-1} \to 0$:

$$[\Sigma_e^{(k)}]^{-1} \sim \mathcal{W}(\nu\mathbf{I}, M) \triangleq \frac{1}{2^{\frac{M^2}{2}} |\nu\mathbf{I}|^{\frac{M}{2}} \Gamma_M\left(\frac{M}{2}\right)} |\Sigma_e^{(k)}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathrm{trace}\left[\nu^{-1}[\Sigma_e^{(k)}]^{-1}\right]\right). \tag{35}$$

To address the inherent scaling ambiguity in (34), z_m and $\mathbf{a}_m$ are endowed with standard Gaussian priors:

$$\mathbf{a}_m \sim \mathcal{N}(0, \mathbf{I}), z_{m,t} \sim \mathcal{N}(0, 1). \tag{36}$$

To allow for automatic determination of the component number M, a noninformative hierarchical prior akin to (32) can be imposed on $c_m^{(k)}$. Here, at the first level, a half-Gaussian is assumed to account for a fixed polarity of each ERP component across conditions

$$\begin{aligned} c_m^{(k)} &\sim \mathcal{N}^+(0, \alpha_m^{-1}) \triangleq 2\sqrt{\frac{\alpha_m}{2\pi}} \exp\left(-\frac{1}{2}\alpha_m [c_m^{(k)}]^2\right), c_m^{(k)} \geq 0 \\ \alpha_m &\sim \mathcal{G}(u, v) \triangleq \frac{v^u}{\Gamma(u)} \alpha_m^{u-1} \exp(-v\alpha_m), u \to 0, v \to 0. \end{aligned} \tag{37}$$

Moreover, the latency shifts $\tau_m^{(k)}$ are integers within a preset interval $[t_1, t_2]$. The probabilistic graphical model is shown in Figure 5.

In [29], a VB-EM algorithm termed *Bayesian estimation of ERPs* (*BEEP*) was developed for inferring model (34) based on the following mean-field approximation:

Algorithm 3: The BEEP algorithm.

Input: $\{\mathrm{X}^{(\mathrm{kl})}\}$

Output: $q(\mathbf{a}_m), q(z_{m,t}), q(c_m^{(k)}), q(\alpha_m), q(\Sigma_e^{(k)}), \tau_m^{(k)}$

1: **Initialization**

2: **repeat**

3: $q(\mathbf{a}_m) = \mathcal{N}(\mathbb{E}[\mathbf{a}_m], C[\mathbf{a}_m])$, where $\mathbb{C}[\mathbf{a}_m] = [L\sum_{k,t} \mathbb{E}[[c_m^{(k)}]^2]\mathbb{E}[z_{m,t}^2]\mathbb{E}[[\Sigma_e^{(k)}]^{-1}] + \mathbf{I}]^{-1}$,
$\mathbb{E}[a_m] = R\sum_{k,t} \mathbb{C}[a_m]\mathbb{E}[c_m^{(k)}]\mathbb{E}[z_{m,t}]\mathbb{E}[[\Sigma_e^{(k)}]^{-1}]\bar{s}_{m,t}^{(k)}$, $\bar{s}_{m,t}^{(k)} \triangleq \sum_r [x_{m,t}^{(k)} - \sum_{i \neq m} \mathbf{a}_i c_i^{(k)} z_{i,t-\tau_{li}^*}]$

4: $q(z_{m,t}) = \mathcal{N}(\mathbb{E}(z_{m,t}), \mathbb{C}(z_{m,t}))$, where $\mathbb{C}[z_{m,t}] = [L\sum_k \mathbb{E}([c_m^k]^2)(\mathrm{trace}(\mathbb{C}[\mathbf{a}_m]\mathbb{E}[[\Sigma_e^{(k)}]^{-1}]) + \mathbb{E}^\top[\mathbf{a}_m]\mathbb{E}[[\Sigma_e^{(k)}]^{-1}]\mathbb{E}[\mathbf{a}_m]) + K]^{-1}$,
$\mathbb{E}[z_{m,t}] = [L\sum_k [\bar{s}_{m,t-\tau_{lj}}^{(k)}]^\top \mathbb{E}[[\Sigma_e^{(k)}]^{-1}]\mathbb{E}[\mathbf{a}_m]\mathbb{E}[c_m^{(k)}]]\mathbb{C}[z_{m,t}]$

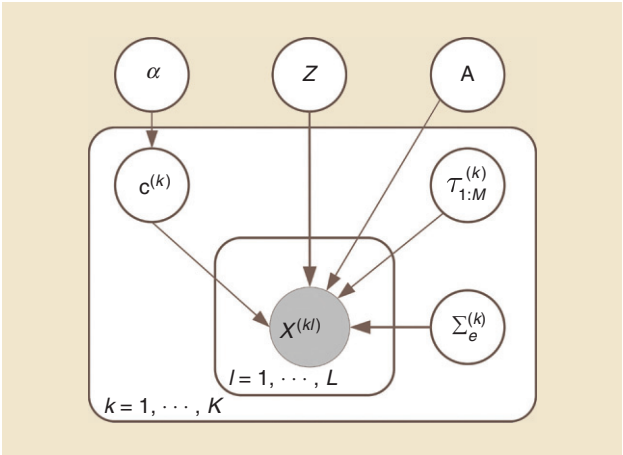
5: $q(c_m^{(k)}) = \kappa \cdot \mathcal{N}(\mu_m^{(k)}, \sigma_m^{(k)})$, $c_m^{(k)} \geq 0$, where $\sigma_m^{(k)} \triangleq \left[\alpha_m + L\sum_t \mathbb{E}([z_{m,t}]^2)(\mathrm{trace}(\mathbb{C}[\mathbf{a}_m]\mathbb{E}[[\Sigma_e^{(k)}]^{-1}]) + \mathbb{E}^\top[\mathbf{a}_m]\mathbb{E}[[\Sigma_e^{(k)}]^{-1}]\mathbb{E}[\mathbf{a}_m])\right]^{-1}$,
$\mu_m^{(k)} \triangleq [L\sum_t [\bar{s}_{m,t-\tau_{lj}}^{(k)}]^\top \mathbb{E}[[\Sigma_e^{(k)}]^{-1}]\mathbb{E}[a_m]\mathbb{E}[z_{m,t}]]\sigma_m^{(k)}$, $\kappa \triangleq 1/(1 - \Phi(-\frac{\mu_m^{(k)}}{\sqrt{\sigma_m^{(k)}}}))$.

6: $q(\alpha_m) = \mathcal{G}(\tilde{u}, \tilde{v})$, where $\tilde{u} \triangleq \frac{K}{2}$, $\tilde{v} \triangleq \frac{1}{2}\sum_k \mathbb{E}[[c_m^{(k)}]^2]$.

7: $q([\Sigma_e^{(k)}]^{-1}) = \mathcal{W}([\Gamma^{(k)}]^{-1}, LT + M)$, where
$\Gamma^{(k)} \triangleq \sum_l \sum_t (x_t^{(\mathrm{kl})}[x_t^{(\mathrm{kl})}]^\top) - 2\sum_l \sum_t \left(x_t^{(\mathrm{kl})} \sum_{j=1}^M [\mathbf{h}_{j,t}^{(k)}]^\top\right) + L\sum_t \sum_{i=1}^M \sum_{j=1, j\neq i}^M \mathbf{h}_{i,t}^{(k)}[\mathbf{h}_{j,t}^{(k)}]^\top + L\sum_t \sum_{j=1}^M \mathbb{E}[\mathbf{a}_j \mathbf{a}_j^\top]\mathbb{E}[[c_j^{(k)}]^2]\mathbb{E}[[z_{j,t}^{(k)}]^2]$,
$\mathbf{h}_{j,t}^{(k)} \triangleq \mathbb{E}[\mathbf{a}_j]\mathbb{E}[c_j^{(k)}]\mathbb{E}[z_{j,t}^{(k)}]$.

8: $\tau_m^k = \arg\max_{\bar{\tau}_m^k \in [t_1, t_2]} f(\bar{\tau}_m^k)$, where $f(\bar{\tau}_m^k) \triangleq \sum_t \mathbb{E}[c_m^{(k)}]\mathbb{E}[a_m]\mathbb{E}^\top[z_{m,t-\bar{\tau}_m^{(k)}}]\mathbb{E}[[\Gamma^{(k)}]^{-1}]\bar{s}_{m,t}^{(k)}$.

9: **until** Convergence



[FIG5] **A graphical model representation for BEEP in the presence of K conditions and L trials. $X^{(kl)}$ denotes the EEG data for the lth trial of condition k. $Z = [z_1, \cdots, z_M]^\top$ contains the time courses of the M ERP components. A denotes the mixing matrix. $\Sigma_e^{(k)}$ is the noise spatial covariance matrix for condition k. $c^{(k)} = [c_1^{(k)}, \cdots, c_M^{(k)}]$ consists of the amplitude factors of the ERP components for condition k, and $\alpha = [\alpha_1, \cdots, \alpha_M]$ are the corresponding hyperparameters. $\tau_m^{(k)}$ denotes the latency shift of the mth ERP component for condition k. All unknown latent variables and parameters are shown as unshaded nodes and estimated from the data using update rules provided in Algorithm 3.**

$$q(\Theta) = \prod_{k,m,t} q(\mathrm{a}_m)\, q(c_m^{(k)})\, q(z_{m,t})\, q(\alpha_m)\, q(\Sigma_e^{(k)}), \tag{38}$$

where $q(\mathrm{a}_m), q(c_m^{(k)}), q(z_{m,t}), q(\alpha_m), q(\Sigma_e^{(k)})$ are updated alternately, and $\tau_m^{(k)}$ is updated by maximizing the variational free energy over integer interval $[t_1, t_2]$. BEEP is summarized in Algorithm 3. Results on synthetic data and 13 subjects' EEG recordings collected in a face-inversion experiment demonstrated that BEEP was superior to several state-of-the-art ERP analysis algorithms, yielding neurophysiologically meaningful ERP components [29].

SBL FOR BRAIN IMAGING OF CORRELATED SOURCES

This section briefly describes another SBL algorithm called *Champagne* for electromagnetic brain imaging. The Champagne algorithm relies on segmenting data into pre- and poststimulus periods, learning the statistics of the background activity from the prestimulus period, and then applying the statistics of the background activity to the poststimulus data to uncover the stimulus-evoked activity. The underlying assumption is that noise and nonstimulus-locked brain activity in the prestimulus period remain unchanged in the poststimulus period, whereas the stimulus-evoked activity is linearly superimposed on the prestimulus activity. In contrast to SAKETINI (which is a scanning method and estimates the statistics of noise and interference at each voxel), in the Champagne algorithm, all voxel activity is simultaneously estimated (such as the variance parameters for each voxel). This formulation leads to implicit sparse estimation whereby many voxel variance parameters are guaranteed to be zero, if the noise and interference statistics are accurate. Therefore, the Champagne algorithm is a considerably faster method for whole-brain imaging in the presence of sparse sources.

The poststimulus sensor data X_{post} is modeled as

$$X_{\text{post}} = \sum_{r=1}^{d_z} \mathrm{A}_r\, Z_r + E, \tag{39}$$

where $X_{\text{post}} \in \mathbb{R}^{N\times T}$ represents T sample points in the poststimulus period from N sensors. $\mathrm{A}_r \in \mathbb{R}^{N\times M}$ is the lead-field matrix in M orientations for the rth voxel. Each unknown source $Z_r \in \mathbb{R}^{M\times T}$ is an M-dimensional neural current-dipole source at T time points, projecting from the ith voxel. There are d_z voxels under consideration. $E \in \mathbb{R}^{N\times T}$ is a noise plus interference term estimated from the prestimulus period using partitioned factor analysis models [30], which assumes independence over time. The first step of the Champagne algorithm is to estimate Σ_e by source localization. The second step of the Champagne algorithm is to estimate hyperparameters of Γ that govern the statistical model of the poststimulus data. We can fully define the probability distribution of the data conditioned on the sources

$$p(X_{\text{post}} \,|\, Z) \propto \exp\left(-\frac{1}{2}\left\| X_{\text{post}} - \sum_{r=1}^{d_z} \mathrm{A}_r Z_r \right\|^2_{\Sigma_e^{-1}}\right), \tag{40}$$

where $\|X\|_{\mathrm{W}}$ denotes the weighted matrix norm $\sqrt{\mathrm{trace}[X\mathrm{W}X^\top]}$.

The following source prior is assumed for Z:

$$p(Z \,|\, \Gamma) \propto \exp\left(-\frac{1}{2}\mathrm{trace}\left[\sum_{r=1}^{d_z} Z_r^\top \Gamma_r^{-1} Z_r\right]\right). \tag{41}$$

This is equivalent to applying independently, at each time point, a zero-mean Gaussian distribution with covariance $\mathbf{\Gamma}_r$ to each source Z_r. We define $\mathbf{\Gamma} = \mathrm{diag}(\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_{d_z})$ as the $d_z M \times d_z M$ block-diagonal matrix, formed by ordering each $\mathbf{\Gamma}_r$ along the diagonal of an otherwise zero-valued matrix. If the lead-field has only one orientation (scalar/orientation-constrained lead-field), $\mathbf{\Gamma}$ reduces to a diagonal matrix. Since $\mathbf{\Gamma}$ is unknown, we use the approximation $\hat{\mathbf{\Gamma}}$ by integrating out the sources Z of the joint distribution $p(Z, X_{\text{post}} \,|\, \mathbf{\Gamma}) \propto p(X_{\text{post}} \,|\, Z) p(Z \,|\, \mathbf{\Gamma})$, i.e.,

$$p(X_{\text{post}} \,|\, \mathbf{\Gamma}) = \int p(Z, X_{\text{post}} \,|\, \mathbf{\Gamma})\, \mathrm{d}Z, \tag{42}$$

and then minimizing the cost function

$$\mathcal{L}(\mathbf{\Gamma}) \triangleq -2 \log p(X_{\text{post}} \,|\, \mathbf{\Gamma}) \equiv \mathrm{trace}[\mathrm{C}_x\, \Sigma_x^{-1}] + \log|\Sigma_x|, \tag{43}$$

where $\mathrm{C}_x \triangleq (1/T) X_{\text{post}} X_{\text{post}}^\top$ is the empirical covariance and $\Sigma_x = \Sigma_e + \sum_{r=1}^{dz} \mathrm{A}_r \mathbf{\Gamma}_r \mathrm{A}_r^\top$.

Minimizing the cost function (43) with respect to $\mathbf{\Gamma}$ can be done in a variety of ways, including gradient descent or the EM algorithm, but these and other generic methods are exceedingly slow when d_z is large. Instead, we use an alternative optimization procedure that resorts to convex bounding techniques [9]. This method expands on ideas from [9], [10], [28], and [31], handles arbitrary/unknown dipole-source orientations, and converges quickly due to a tighter upper bound [10]. This optimization procedure yields a modified cost function:

$$\mathcal{L}(\{\mathbf{\Gamma}_r\}, \{Z_r\}, \{Y_r\}) = \left\| \tilde{X} - \sum_{r=1}^{d_z} \mathrm{A}_r Z_r \right\|^2_{\Sigma_e^{-1}} + \sum_{r=1}^{d_z} [\| Z_r \|^2_{\mathbf{\Gamma}_r^{-1}} + \mathrm{trace}(Y_r^\top \mathbf{\Gamma}_r)] - h^*(Y_r), \tag{44}$$

Algorithm 4: The Champagne algorithm.

Input: $X_{\text{post}}, \mathbf{A}_r, \Sigma_e$
Output: $Z_r, \mathbf{\Gamma}_r((r = 1, \ldots, d_z))$
1: **Initialization**
2: **repeat**
3: $\quad \Sigma_x = \Sigma_e + \sum_r \mathbf{A}_r \mathbf{\Gamma}_r \mathbf{A}_r^\top$
4: $\quad Z_r = \mathbf{\Gamma}_r \mathbf{A}_r^\top \Sigma_x^{-1} \tilde{X}$
5: $\quad Y_r = \nabla_{\mathbf{\Gamma}_r} \log|\Sigma_x| = \mathbf{A}_r^\top \Sigma_x^{-1} \mathbf{A}_r$
6: $\quad \mathbf{\Gamma}_r = Y_r^{-1/2} (Y_r^{1/2} Z_r Z_r^\top Y_r^{1/2})^{1/2} Y_r^{-1/2}$.
7: **until** Convergence

where $h^*(Y_r)$ is the concave conjugate of $\log|\Sigma_x|$ for auxiliary variables $Y_r = \mathbf{A}_r^\top \Sigma_x^{-1} \mathbf{A}_r$. By construction $\mathcal{L}(\mathbf{\Gamma}) = \min_{\{Z_r\}} \min_{\{Y_r\}} \mathcal{L}(\{\mathbf{\Gamma}_r\}, \{Z_r\}, \{Y_r\})$, the matrix $\tilde{X} \in \mathbb{R}^{N \times \text{rank}(X_{\text{post}})}$, $\tilde{Z}\tilde{Z}^\top = \mathbf{C}_x$. Minimizing this modified cost function yields three update rules (summarized in Algorithm 4) [31]. In summary, the Champagne algorithm estimates $\mathbf{\Gamma}$ by iterating between $\{Z_r\}$, $\{Y_r\}$, $\{\mathbf{\Gamma}_r\}$, and with each pass monotonically decreasing $\mathcal{L}(\mathbf{\Gamma})$.

NB METHODS

The historic roots of Bayesian nonparametrics date back to the late 1960s, but its applications in machine learning have not been widespread until recently [32]. In contrast to parametric models, NB models accommodate a large number of degrees of freedom (infinite dimensional parameter space) to exhibit a rich class of probabilistic structure, which make them more appealing, flexible, and powerful in data representation. The fundamental building blocks of Bayesian nonparametrics are two stochastic processes: the GP and DP. Two excellent introductory books on GP and DP are [33] and [34]. This section presents a few examples of these nonparametric modeling tools for EEG/MEG data analysis.

GP MODELING

Any finite set of random variables is a GP if it has a joint Gaussian distribution. Unlike the fixed finite-dimensional parametric model, the GP defines priors for the mean function and covariance function, where the covariance kernel determines the smoothness and (non)-stationarity between the time-series data points. To construct a GP model, given an input–output training sample set $(\boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T})$ of size T (where $\boldsymbol{x}_i \in \mathbb{R}^N$, $y_i \in \mathbb{R}$,, without loss of generality, assuming the time series has zero mean), the goal is to estimate a distribution from the data and predict the outcome of an unseen input data vector $\boldsymbol{x}_{T+1}$

$$P(y_{T+1} \,|\, \boldsymbol{x}_{T+1}, \boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T}) = \frac{1}{Z} \exp\left(-\frac{1}{2} \boldsymbol{y}_{1:T+1}^\top \mathbf{C}_{T+1}^{-1} \boldsymbol{y}_{1:T+1}\right), \quad (45)$$

where Z is a normalizing constant and $\mathbf{C}$ defines the covariance matrix for input data, characterized by an unknown hyperparameter vector θ. The covariance kernel can be either stationary or nonstationary [35], and different choices of the covariance kernel result in various types of GPs.

GP FOR EEG SEIZURE DETECTION

Important tasks of GP modeling are prediction or outlier detection [36], as well as EEG signal classification [37]. Specifically, the predictive distribution can be computed by integrating out the hyperparameters θ:

$$P(y_{T+1} \,|\, \boldsymbol{x}_{T+1}, \boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T}) = \int P(y_{T+1} \,|\, \boldsymbol{x}_{T+1}, \mathbf{C}, \theta) p(\mathbf{C}, \theta) \, \mathrm{d}\theta, \quad (46)$$

and the covariance kernel has the form $\mathbf{C}(\boldsymbol{x}_t, \boldsymbol{x}_{t'}) = \theta_0 e^{-\frac{1}{2}\sum_{i=1}^N \theta_i (x_{i,t} - x_{i,t'})^2} + \theta_\eta \delta_{tt'}$, where $\theta = [\theta_0, \theta_1, \ldots, \theta_N, \theta_\eta] \in \mathbb{R}^{N+2}$ denotes the hyperparameters, and $\delta_{tt'}$ is a Kronecker delta function (which is equal to one if and only if $t' = t$).

In most of cases, the integral of (46) is analytically intractable. Two possible solutions can be considered: a Monte Carlo method or a likelihood method that replaces the distribution of $\boldsymbol{\theta}$ with a maximum likelihood (point) estimate. For EEG outlier prediction, the variance approach and hyperparameter approach have been proposed [36]. In the variance approach, the variance of predicted output is monitored: if the variance is outside the range of the normal training samples, it will indicate the change of EEG structure or the presence of an outlier. In the hyperparameter approach, the ratio of two hyperparameters (e.g., $|\theta_0/\theta_\eta|$, which reflects the level of determinism in the EEG signal) is monitored.

MEG MODELING AND DIMENSIONALITY REDUCTION

For high-dimensional, structured EEG/EEG data, dimensionality reduction is useful for data denoising and interpretation. It is also insufficient to assume that the correlations between the elements of observations vector are static in many practical applications. Recently, GP has been used for heteroscedastic modeling of noisy high-dimensional MEG data [38]. Specifically, the N-dimensional MEG time series at the lth trial can be modeled as a non-Markovian dynamic linear factor analysis model

$$z_t^{(l)} = \psi^{(l)}(\tau_t) + v_t^{(l)} \quad (47)$$
$$\boldsymbol{x}_t^{(l)} = \mathbf{A}(\tau_t) z_t^{(l)} + e_t^{(l)}, \quad (48)$$

where $\mathbf{A}(\tau_t) \in \mathbb{R}^{N \times M}((M \ll N)$ denotes a time-evolving factor loading matrix to account for nonstationarity, and $v_t^{(l)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $e_t^{(l)} \sim \mathcal{N}(\mathbf{0}, \Sigma_0)$ denote the zero-mean Gaussian dynamic and measurement noise, respectively. The noise covariance matrix Σ_0 is assigned with an inverse gamma prior.

This model differs from (3) and (4) in that the latent process is drawn from GPs. In the state equation, the low-dimensional latent process is constructed by a hierarchy, in which the evolution of the latent factor z_t, at each trial, is governed by a collection of M dictionary functions. Each element itself is a GP [38]: $\psi^{(l)}(\tau) = [\psi_1^{(l)}(\tau), \ldots, \psi_M^{(l)}(\tau)] \in \mathbb{R}^M$, where $\psi_j^{(0)}(\cdot) \backsim \text{GP}(0, \mathbf{C}_0)$, $\psi_j^{(l)}(\cdot) \backsim \text{GP}(\psi_j^{(0)}, \mathbf{C}_1)$ are two GPs with squared exponential correlation function, with $\mathbf{C}_k(\boldsymbol{\xi}, \boldsymbol{\xi}) = d_k \exp(-\kappa \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_2^2)$ for $k = 0, 1$. Therefore, the child processes $\psi_j^{(l)}$ are centered around the parent process $\psi_j^{(0)}$, and they all share information through the parent process.

In the observation equation, the factor loading matrix $\mathbf{A} \in \mathbb{R}^{N \times M}$ is modeled by a weighted combination of R-dimensional latent covariance dictionary functions [38]

$$\mathbf{A}(\tau) = \mathbf{B}\mathbf{\Phi}(\tau),$$

where $\mathbf{B} \in \mathbb{R}^{N\times R}$ $(R \ll M \ll N)$, and $\Phi \in \mathbb{R}^{R\times M} = \{\phi_{jk}(\cdot)\} \sim \mathrm{GP}(0, \mathbf{C}_0)$ consists of a covariance dictionary. Such GP modeling of dictionary elements allows the MEG signals' long-range dependencies to be captured. Each row of $\mathbf{B}$ will be assigned with a sparse shrinkage prior (that penalizes a large coefficient). The proposed hierarchical model is able to characterize nonstationarity (via the time-varying factor loading matrix) and to share information (via coupling) between single trials.

The goal of Bayesian modeling is to infer the latent variable and parameters of the hierarchical factor analysis model. First, the componentwise observation $x_{n,t}^{(l)}$ can be written as

$$x_{n,t}^{(l)} = \sum_{m=1}^{M} z_{m,t}^{(l)} \sum_{r=1}^{R} b_{\mathrm{nr}}\phi_{\mathrm{rm}}(\tau_t) + e_{i,t}^{(l)}.$$

Marginalizing the dynamic and measurement noise $v_t^{(l)}$ and $e_t^{(l)}$ induces the following time-varying mean and covariance structure for the observed signal $\boldsymbol{x}_t^{(l)} \sim \mathcal{N}(\boldsymbol{\mu}^{(l)}(\tau_t), \Sigma(\tau_t))$, where

$$\begin{aligned}\boldsymbol{\mu}^{(l)}(\tau_t) &= \mathbf{B}\Phi(\tau_t)\psi^{(l)}(\tau_t)\\ \Sigma(\tau_t) &= \mathbf{B}\Phi(\tau_t)\Phi(\tau_t)^{\top}\mathbf{B}^{\top} + \Sigma_0.\end{aligned}$$

The time-varying covariance structure captures the heteroscedasticity of time series, and the overcomplete representation provides flexibility and computational advantages. Bayesian inference (posterior computation and predictive likelihood) of this hierarchical GP model is achieved by MCMC sampling methods of $\{\psi_j^{(l)}, v_{1:T}^{(l)}, \psi^{(0)}, \Phi, \mathbf{B}, \Sigma_0\}$ [38]. As shown in [38], such hierarchical NB modeling provides a powerful framework to characterize a signal noisy MEG recording that allows word category classification.

DP MODELING

Mixture modeling has been commonly used time-series data analysis. Unlike finite mixture models, NB models define a prior distribution over the set of all possible partitions, where the number of clusters or partitions may grow as the data samples increase. This is particularly useful for EEG/MEG applications in clustering, partition, segmentation, and classification. Examples of static or dynamic mixture models include DP mixtures, the infinite hidden Markov model, and hierarchical DP. As an illustration, we consider an example of multisubject EEG classification in the context of NB CSP, which defines the Bayesian CSP model with either a DP prior or an Indian-buffet-process prior [18], [39].

The NB CSP model extends the probabilistic CSP (see [17]) and further introduces the DP prior, which was referred to as *BCSP-DP* [18]. The BCSP-DP uses a DP mixture model to learn the number of spatial patterns among multiple subjects. The spatial patterns with the same hyperparameter are grouped in the same cluster [see the graphical model illustration in Figure 2(b)], thereby facilitating the information transfer between subjects with similar spatial patterns. Essentially, BCSP-DP is a nonparametric counterpart of VB-CSP.

For condition $k \in \{1, 2\}$ and subject $s \in \{1, \ldots, S\}$, the multisubject CSP model is rewritten as

$$X^{(\mathrm{sk})} = \mathrm{A}^{(s)}Z^{(\mathrm{sk})} + E^{(\mathrm{sk})}. \tag{49}$$

As seen in Figure 2(b), spatial patterns $\{\mathrm{a}_m^{(s)}\}$ are drawn from Gaussian distributions $\mathcal{N}(\mathrm{a}_m^{(s)} \mid \boldsymbol{\mu}_m^{(s)}, (\Phi_m^{(s)})^{-1})$, where the Gaussian parameters $(\boldsymbol{\theta}_m^{(s)} = \{\boldsymbol{\mu}_m^{(s)}, \Phi_m^{(s)}\})$ are further drawn from a random measure G in a DP, $\boldsymbol{\theta}_m^{(s)} \sim G, G \sim \mathrm{DP}(\alpha, G_0)$ for $m = 1, \ldots, M$ and $s = 1, \ldots, S$, where $\alpha > 0$ is a positive concentration parameter (which specifies how strong this discretization is: when $\alpha \to 0$, the realizations are all concentrated on a single value, when $\alpha \to \infty$, the realizations become continuous), and the base measure $G_0 = p(\boldsymbol{\mu}, \Phi)$ is set to be a Gaussian–Wishart distribution, which is the conjugate prior for Gaussian likelihood

$$p(\boldsymbol{\mu}, \Phi) = \mathcal{N}(\boldsymbol{\mu} \mid \boldsymbol{m}_0, (\beta_0\Phi)^{-1})\mathcal{W}(\Phi \mid \nu_0, \mathbf{W}_0), \tag{50}$$

where $\mathcal{W}(\Phi \mid \nu_0, \mathbf{W}_0)$ denotes a Wishart distribution with the degree of freedom ν_0 and the scale matrix $\mathbf{W}_0$. The random measure G has the following stick-breaking representation

$$G = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i^*}, \quad ;\pi_i = v_i \prod_{j=1}^{i-1}(1 - v_i), \tag{51}$$

where $v_i \sim \mathcal{B}(v_i \mid 1, \alpha)$ and $\theta_i^* \in G_0$ are independent and random variables drawn from a beta distribution and the beta measure G_0, respectively. The mixing proportions $\{\pi_i\}$ are given by successively breaking a unit-length stick into an infinite number of pieces.

Assuming hierarchical priors for $z_t^{(\mathrm{sk})}$ and $\boldsymbol{x}_t^{(\mathrm{sk})}$ [see legend of Figure 2(b) for the priors for α, v_i, $\mathbf{A}^{(s)}$]

$$\begin{aligned}z_t^{(\mathrm{sk})} &\sim \mathcal{N}(0, \Sigma_z^{(\mathrm{sk})})\\ \boldsymbol{x}_t^{(\mathrm{sk})} &\sim \mathcal{N}(A^{(s)}z_t^{(\mathrm{sk})}, \Sigma_e^{(\mathrm{sk})}),\end{aligned}$$

where Σ_z and Σ_e denote two diagonal covariance matrices with diagonal elements drawn from inverse-Gamma distributions. Finally, let Θ define the collective set of unknown variables of parameters and hyperparameters

$$\Theta = \{\{\mathrm{A}^{(s)}\}, \boldsymbol{c}_m^{(s)}, \{Z^{(\mathrm{sk})}\}, \{v_i\}, \alpha, \{\Sigma_z^{(\mathrm{sk})}\}, \{\Sigma_e^{(\mathrm{sk})}\}, \{\mu_i^*, \Phi_i^*\}\}.$$

The authors of [18] employed VB inference by assuming a factorized variational distribution $q(\Theta)$

$$\begin{aligned}q(\Theta) = q(\{\mathrm{A}^{(s)}\})q(\boldsymbol{c}_m^{(s)})q(\{Z^{(\mathrm{sk})}\})q(\{v_k\})\\ \times q(\alpha)q(\{\Sigma_z^{(\mathrm{sk})}\})q(\{\Sigma_e^{(\mathrm{sk})}\})q(\{\boldsymbol{\mu}_i^*, \Phi_i^*\}).\end{aligned}$$

The variational posterior $q(Z^{(\mathrm{sk})}) = \prod_t q(z_t^{(\mathrm{sk})})$, $q(z_t^{(\mathrm{sk})}) = \mathcal{N}(\mu_t^{(\mathrm{sk})}, \Sigma_*^{(\mathrm{sk})})$ has an analytic form, with

$$\begin{aligned}(\Sigma_*^{(\mathrm{sk})})^{-1} &= \mathbb{E}_q[(\Sigma_z^{(\mathrm{sk})})^{-1}] + \sum_{i=1}^{N}\mathbb{E}_q[(\Sigma_e^{(\mathrm{sk})})^{-1}]_i\,\mathbb{E}_q[[A^{(s)}]_{i,:}^{\top}[A^{(s)}]_{i,:}]\\ \boldsymbol{\mu}_t^{(\mathrm{sk})} &= \Sigma_*^{(\mathrm{sk})}\mathbb{E}_q[A^{(s)}]^{\top}\mathbb{E}_q[(\Sigma_e^{(\mathrm{sk})})^{-1}]x_t^{(\mathrm{sk})}.\end{aligned}$$

For inference details, see [18]. At the cost of increasing computational complexity, the BCSP-DP model achieved improved EEG classification performance compared to state-of-the-art CSP models [18].

APPLICATIONS

BML has been successfully used in a variety of EEG/MEG applications. As mentioned in the "Introduction" section, the two main applications of BML algorithms for EEG/MEG data are brain imaging and brain-state classification. Electromagnetic brain imaging has diverse applications in basic and clinical neuroscience. Brain-state classification may include development of BCIs and methods for diagnosis of abnormal brain activity. In this section, we briefly describe a few representative examples of both applications.

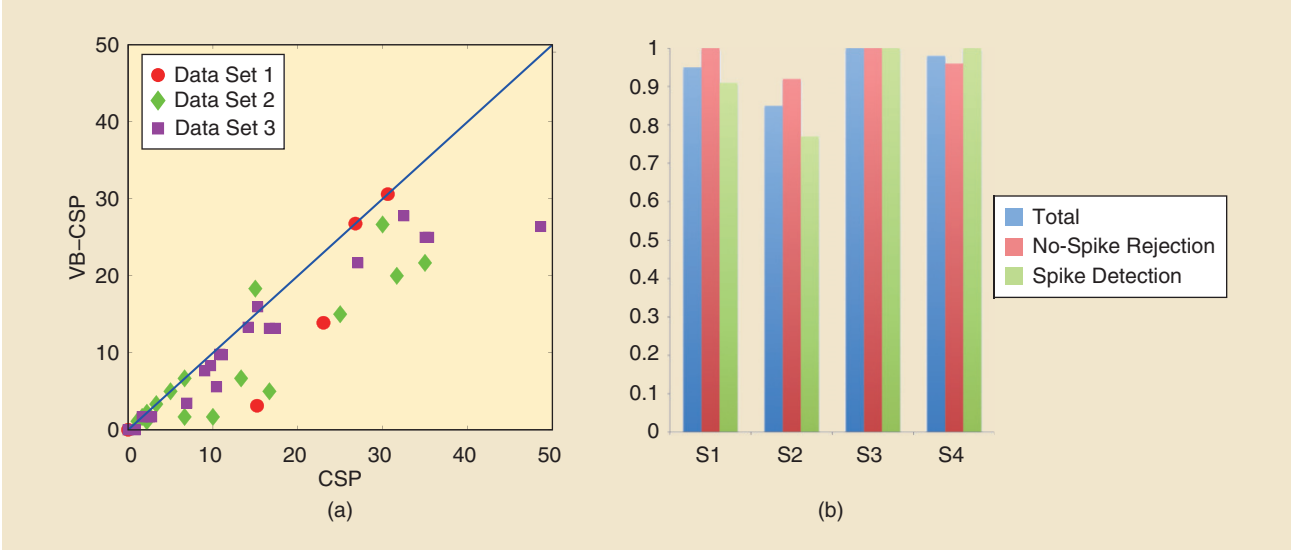
BRAIN-STATE CLASSIFICATION USING VB-CSP

The VB-CSP algorithm can be used for brain-state classification, such as the motor imagery task [17]. Imagined limb or tongue movements give rise to an attenuation of the EEG rhythms in different regions of the sensorimotor cortices, and CSP-based methods have proven to be effective in distinguishing imagined movements by extracting their distinctive EEG spatiotemporal patterns. The standard CSP methods have the overfitting problem. In our previous work, the efficacy of VB-CSP was demonstrated through applications to single-trial classifications of three motor imagery EEG data sets, two of which were taken from BCI Competition III (Data Set IIIa as data set 1, and Data Set IVa as data set 2). A total of 43 binary-class data subsets were generated from the three data sets by pairing MI tasks. The overall test error for VB-CSP is 10.36%, compared with 14.22% for CSP. Paired t-tests indicate that VB-CSP significantly outperforms CSP ($p = 1.61 \times 10^{-5}$). For data sets with a small number of training trials, which CSP has a tendency to overfit, VB-CSP alleviates the overfitting by using substantially less but automatically determined components than the available EEG channels.

The VB-CSP algorithm was also employed to identify periods of interictal epileptiform activity in four different patients with epilepsy. Generally, the challenges in analyzing EEG and/or MEG data from epilepsy patients are twofold. First, it is important to identify events in the data that represent epileptiform activity and this is often done manually by domain experts. After identification of such abnormal epileptiform activity events, then imaging algorithms are run on these events to determine origin in the brain to identify the epileptogenic zone. Algorithms like VB-CSP can be used to automatically segment long MEG or EEG traces into short pieces that can be classified as representing epileptiform activity, and such segmented data can subsequently be used with imaging algorithms. Alternative non-Bayesian approaches in the literature have included engineering solutions [40]. In this particular data set, a registered EEG/evoked potential technologist marked segments of the continuously collected data set that contained spontaneous epileptiform spikes, as well as segments that clearly contained no spikes. All of these data were split into training and test data sets. The VB-CSP was first applied to learn a generative model for spike and "nonspike" training data, and then the classification performance was evaluated in independent test data. Results from the four subjects, shown in Figure 6, demonstrated a brain-state classification accuracy of above 85% for spikes and nonspikes in all subjects.

LOCALIZATION OF INTERICTAL EPILEPTIC SPIKES WITH SAKETINI

As mentioned previously, once data segments can be classified as being epileptiform, it is important to determine the origin of such data in the brain, thereby identifying the epileptogenic zone. With regard to imaging, determining the epileptogenic source location and its time course while removing the measurement noise and artifacts is challenging. The SAKETINI algorithm is one algorithm that can be used to localize interictal spikes in patients [20]. One segment of data, identified as a spike marked at 400 milliseconds, as well as

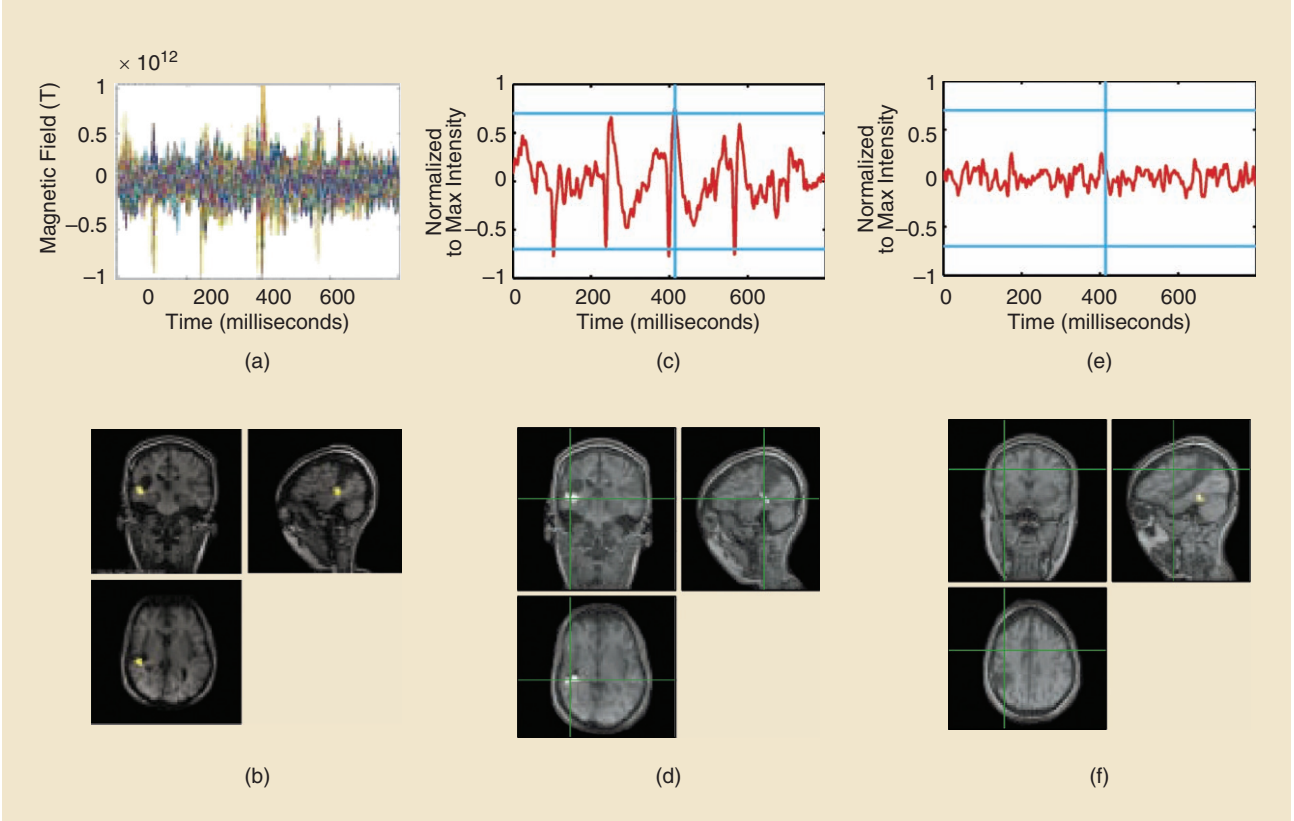


[FIG6] **The brain-state classification performance of VB-CSP. (a) Test errors (%) on three motor imagery BCI data sets (43 subsets) for VB-CSP against CSP. (b) The accuracy of VB-CSP classification of spike and nonspike segments in four epilepsy patients (S1–S4). The overall classification accuracy is well above 80% for all patients. (Figure adapted with permission from [17].)**

three additional spikes in the 800-millisecond segment, was used here as the poststimulus period, and a separate, spike-free segment of equal length was used as the prestimulus period. Figure 7 shows SAKETINI's performance estimating single spikes relative to a spike-free prestimulus period. Figure 7(a) shows the raw sensor data for the segment containing the marked spike. Figure 7(b) shows the location of the equivalent-current dipole (ECD) fit to 20 spikes from this patient. Figure 7(d) shows the SAKETINI likelihood map based on the data in Figure 7(a); the peak is in clear agreement with the standard ECD localization. Figure 7(c) shows the time course estimated for the likelihood spatial peak. The spike at 400 milliseconds is clearly visible; this cleaned waveform can be used by a clinician for analyzing the peak shape. Figure 7(e) and (f) shows a source signal's time course from a randomly selected location far from the epileptic spike source, to show the low noise level and to show the absence of cross talk onto source estimates elsewhere.

RECONSTRUCTION OF LOW SNR MEG DATA WITH SAKETINI

The SAKETINI algorithm can be used to reconstruct MEG data under a low-SNR condition [20]. In one MEG data set for localization of the primary somatosensory cortex (S1), a small diaphragm was placed on the subject's right index finger and was driven by compressed air, and the stimulus was delivered 256 times every 500 milliseconds. If we limit the available data to only a small subset of trials, the lower SNR issue became limiting for most benchmark reconstruction algorithms. We first applied SAKETINI to the average of all 256 trials to assess performance for the high-SNR case. We then applied it to the average of only the first five trials and across other sets of five-trial averages. Figure 8(a) shows typical somatosensory evoked MEG data, with the largest peak at 50 milliseconds, expected to be coming from S1 in the posterior wall of the central sulcus. Figure 8(b) shows performance of SAKETINI for the high-SNR data, which can be accurately localized to the contralateral S1. Figure 8(c) shows the sensor data averaged over the first five trials of the same data set. Therefore, SAKETINI consistently localizes S1 even in these low-SNR data, whereas the benchmark algorithms [minimum variance adaptive beamforming (MVAB) and standardized low-resolution brain electromagnetic tomography (sLORETA); results not shown—see [20] for details] perform poorly in low-SNR regimes.



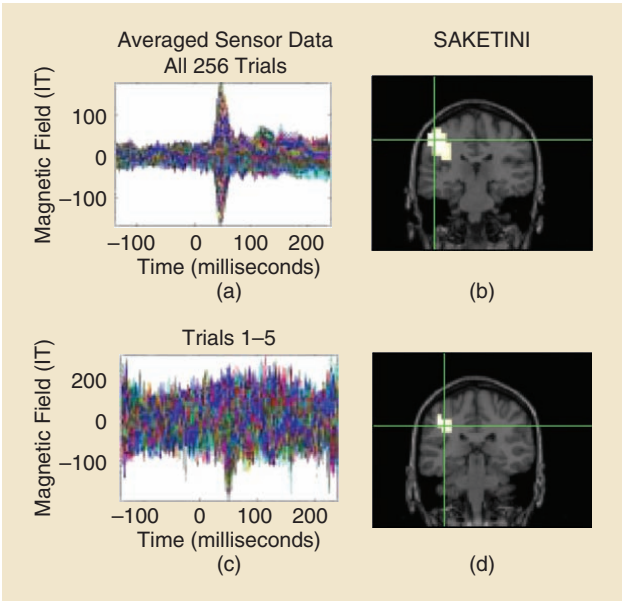
[FIG7] **Localization of interictal spike MEG data with SAKETINI. (a) MEG sensor waveforms for a 2-second segment of data from an epilepsy patient with interictal spikes. (b) Three orthogonal views of the MR imaging (MRI) for this patient, and a cluster of dipoles fit by traditional dipole fitting methods. (c) The time course of a voxel showing interictal spiking activity. (d) Three orthogonal views of the MRI and the overlay of source activity estimated by SAKETINI for this data. The source localization cluster is consistent with the more traditional methods of localization. This interictal source localizes to the margins of a prior resection cavity seen in the MRI. (e) Time-course of a voxel farther away from the interictal source. The voxel does not show clear spiking activity. (f) The location of this "control" voxel on the patient's MRI. These results demonstrate the sensitivity of SAKETINI in spike localization as well as the source time course estimation. (Figure adapted with permission from [20].)**

GROUP ELECTROMAGNETIC BRAIN IMAGING USING gMSP

Conventional source imaging algorithms are mainly designed for EEG/MEG data from a single subject or session and are not optimal for detecting group effects. The gMSP algorithm was used to reconstruct the somatosensory evoked source activity from a median nerve-stimulation EEG data set [23]. The data set consists of recordings of eleven subjects and five session recordings per subject. The reconstruction was performed on the 10–40-millisecond poststimulus data segments from 100 groups of experiments (each with $S = 11$ subjects), which were randomly sampled from the 5^{11} possible groups. The median nerve somatosensory sources have been extensively studied in humans and are known to largely reside in the hand area of the S1. The reconstruction results of gMSP were compared with those of classical minimum norm estimation (MNE) and a multiple sparse prior (MSP) modeling approach to individual subjects without the group constraint. To assess the group-level reconstructed source activity, for each method, paired t-tests were performed between the reconstructed source activities and same images flipped across the midsagittal plane to generate statistical parametric maps (SPMs). Figure 9 shows the SPMs for one randomly selected group. All SPMs exhibit maximal activation around the left sensorimotor area. However, despite MSP's regional specificity, the t-values are generally low because of the few overlaps of sparse solutions across subjects. By contrast, large t-values are dispersed over a wide range of brain regions for MNE due to its low spatial resolution. The gMSP solution attains t-values close to those of MNE and preserves sparsity consistent across subjects due to the group constraint.

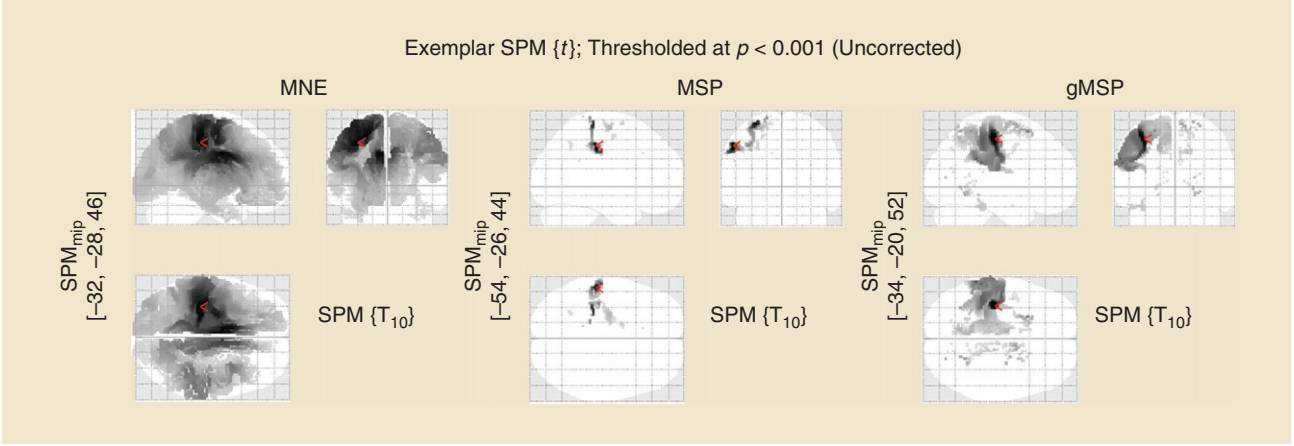
RECONSTRUCTION OF COMPLEX BRAIN SOURCE CONFIGURATIONS WITH THE CHAMPAGNE ALGORITHM

Typical brain-imaging algorithms are designed to reconstruct either isolated dipoles or extended sources with very large spatial extents. However, it is clear from functional MRI (fMRI) studies that typical source activations occur in spatially clustered sources with complex geometries of activations. Very few computational

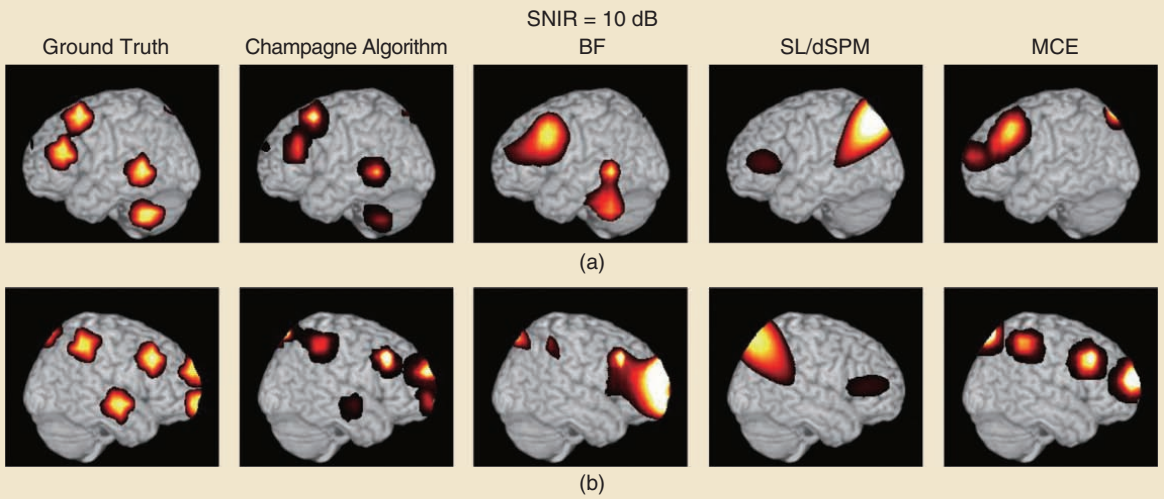


[FIG8] **Localization performance of SAKETINI on high- and low-SNR MEG data. (a) High-SNR sensor waveforms for tactile stimulation obtained by averaging 256 trials. (b) Localization of the waveforms in (a) shows a clear peak in S1. (c) Low-SNR sensor waveforms for tactile stimulation obtained by averaging only five trials. (d) Localization of low-SNR data is robust in showing S1 activity. (Figure adapted with permission from [20].)**

algorithms have attempted reconstructing such complex source configurations, especially from noisy sensor data corrupted by biological and nonbiological artifacts. Here, we apply the Champagne algorithm to localize complex distributed activity from simulated clusters of brain sources. Ground truth from one such simulation consists of ten clusters, with ten dipolar sources within each cluster. The placement of the cluster center is random, and the clusters consist of sources seeded in the nine nearest neighboring voxels. The source time courses within each cluster have an interdipole correlation coefficient of 0.8 and an intradipole correlation coefficient of 0.25. The source simulation is constructed with realistic



[FIG9] **SPMs comparing the reconstructed source activities with the midsagittal-flipped version of the same images. Results are overlaid on a glass brain. The location of the maximum t-value in each SPM is marked by a red pointer. (Figure adapted with permission from [23]).**

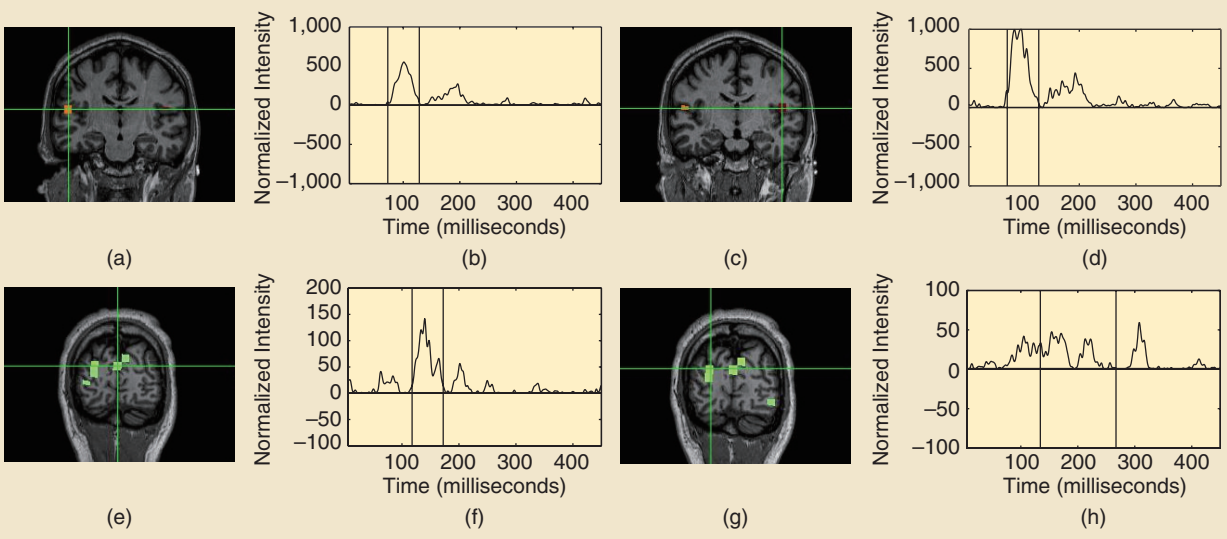


[FIG10] **The performance of the Champagne algorithm in computer simulations. The ground truth has activity in ten dipole clusters spreading throughout the brain. Reconstruction comparison using the Champagne algorithm, beamforming (BF), sLORETA (SL), dynamic statistical parameter mapping (dSPM), and MCE methods. (Figure adapted with permission from [41].)**

brain noise and 10 dB signal-to-noise-plus-interference-ratio (SNIR). Reconstruction results from this simulation with clusters are shown in Figure 10. In this example, the result derived from the Champagne algorithm is superior to other benchmark algorithms, such as MVAB, generalized minimum current estimation (MCE), and minimum norm methods (sLORETA/dSPM) [41], [31]. It can be seen that the standard benchmark algorithms are not able to reconstruct such complex brain source activity configurations. More details of benchmark algorithms and Champagne algorithm performance can be found in [41] and [31].

RECONSTRUCTION OF AUDIOVISUAL NETWORK ACTIVITY FROM MEG DATA WITH CHAMPAGNE

One example of a complex source activity configuration is the network of brain regions recruited during passive viewing of simultaneously presented audiovisual stimuli. This network consists of partially overlapping and partially segregated spatial and temporal activity in auditory and visual cortices as well as brain regions involved in integration of audiovisual information. The Champagne algorithm was used to reconstruct this audiovisual network activity from MEG data to examine the integration of auditory and visual



[FIG11] **Results of audiovisual data localization from the Champagne algorithm. The Champagne algorithm localizes a bilateral auditory response at 100 milliseconds after the simultaneous presentation of tones and a visual stimulus. (a) and (c) Localized bilateral auditory activity from the Champagne algorithm, with time courses shown in (b) and (d). The Champagne algorithm localizes an early visual response at 150 milliseconds after the simultaneous presentation of tones and a visual stimulus. The time course in (f) corresponds to the location indicated by the crosshairs in (e) the coronal sections. The Champagne algorithm localizes a later visual response later than 150 milliseconds after simultaneous presentation of tones and a visual stimulus. The time course in (h) corresponds to the location indicated by the crosshairs in (g) the coronal sections. (Figure adapted with permission from [41].)**

information [41]. In one MEG data set, a healthy control participant was presented single 35-millisecond-duration tones (1 kHz) simultaneously with a visual stimulus. The visual stimulus consisted of a white cross at the center of a black monitor screen. The data were averaged across 100 trials (after the trials were time-aligned to the stimulus). The prestimulus window was selected to be −100 to 5 milliseconds, and the poststimulus time window was selected to be 5–450 milliseconds, where zero marks the onset of the simultaneous auditory and visual stimulation. Champagne's results are presented in Figure 11. In the first and second rows, the brain activations associated with the auditory stimulus are shown. The Champagne algorithm is able to localize bilateral auditory activity in Heschel's gyrus in the window around the M100 peak. The time courses for the left and right auditory sources are shown in Figure 11(b) and (d), respectively, along with the window used around the M100 peak. The two auditory sources had the maximum power in the window around the M100 peak. The Champagne algorithm is also able to localize a visual source in medial, occipital gyrus with a peak around 150 milliseconds. The power in the window around this peak is shown in Figure 11(e), and the time course of the source marked with the crosshairs is shown in Figure 11(f).

RECONSTRUCTION OF FACE PROCESSING NETWORK ACTIVITY FROM EEG DATA WITH THE CHAMPAGNE ALGORITHM

Another example of overlapping spatial and temporal activity in a variety of brain regions occurs during the early brain response to face visual stimuli. The Champagne algorithm was used to reconstruct the face-processing network activity from EEG data. In Figure 12, we present the results from using the Champagne algorithm on a face-processing EEG data set. The 126-channel EEG data was downloaded from a public website (http://www.fil.ion.ucl.ac.uk/spm/data/mmfaces). A three-orientation component lead-field matrix was calculated in SPM8 using the coarse resolution. The EEG data paradigm involved randomized presentation of at least 86 faces, and the average across trials time-aligned to the presentation of the face was used for source reconstruction. The prestimulus window was selected to be −200 to 5 milliseconds, and the poststimulus time window was selected to be 5–250 milliseconds. Reconstructed power was plotted on a three-dimensional brain, and the time courses for the peak voxels are plotted (the arrows point from a particular voxel to its time course). In Figure 12, it is shown that the Champagne algorithm is able to localize early visual areas that have a peak around 100 milliseconds as well as activations in and around fusiform gyrus that peaks around 170 milliseconds, corresponding to the N170 seen in the sensor data. These results are consistent with those obtained in [42] using the same EEG data set but show greater spatial resolution than previously published methods.

DISCUSSION AND CONCLUSIONS

EEG and MEG signals offer a unique and noninvasive way to record brain activity at a fine temporal resolution, providing many opportunities to address important neuroscience questions. However, EEG/MEG analysis is also confronted with a multitude of challenges, such as high dimensionality, nonstationarity, and very large sample size (e.g., continuous EEG recordings) or very sparse sample size (e.g., single trial analysis, disease diagnosis). BML is an emerging research area that integrates Bayesian inference, optimization, Monte Carlo sampling, and machine-learning techniques for data analysis. The Bayesian framework enables us to capture various sources of uncertainties in the data or parameters. By taking into account (hierarchical) priors, Bayesian inference provides an optimal estimate of the model or model parameters. In addition, hierarchical modeling brings the advantages in extra flexibility and the insensitivity to priors.

Although this article focuses on brain-state classification and electromagnetic brain imaging as the two main applications, BML methods have also been successfully developed in some other EEG/MEG signal processing applications, e.g., functional brain connectivity analysis [43], multimodal brain data fusion [42], and data compression [44]. The topics introduced in this article are theoretical underpinnings of these methods and provide the key to understanding their mechanisms. Table 3 lists some useful resources (especially on software) related to the topics reviewed in this article.

CHALLENGES AND FUTURE RESEARCH DIRECTIONS

ELECTROMAGNETIC BRAIN IMAGING FOR SPATIALLY EXTENDED SOURCES

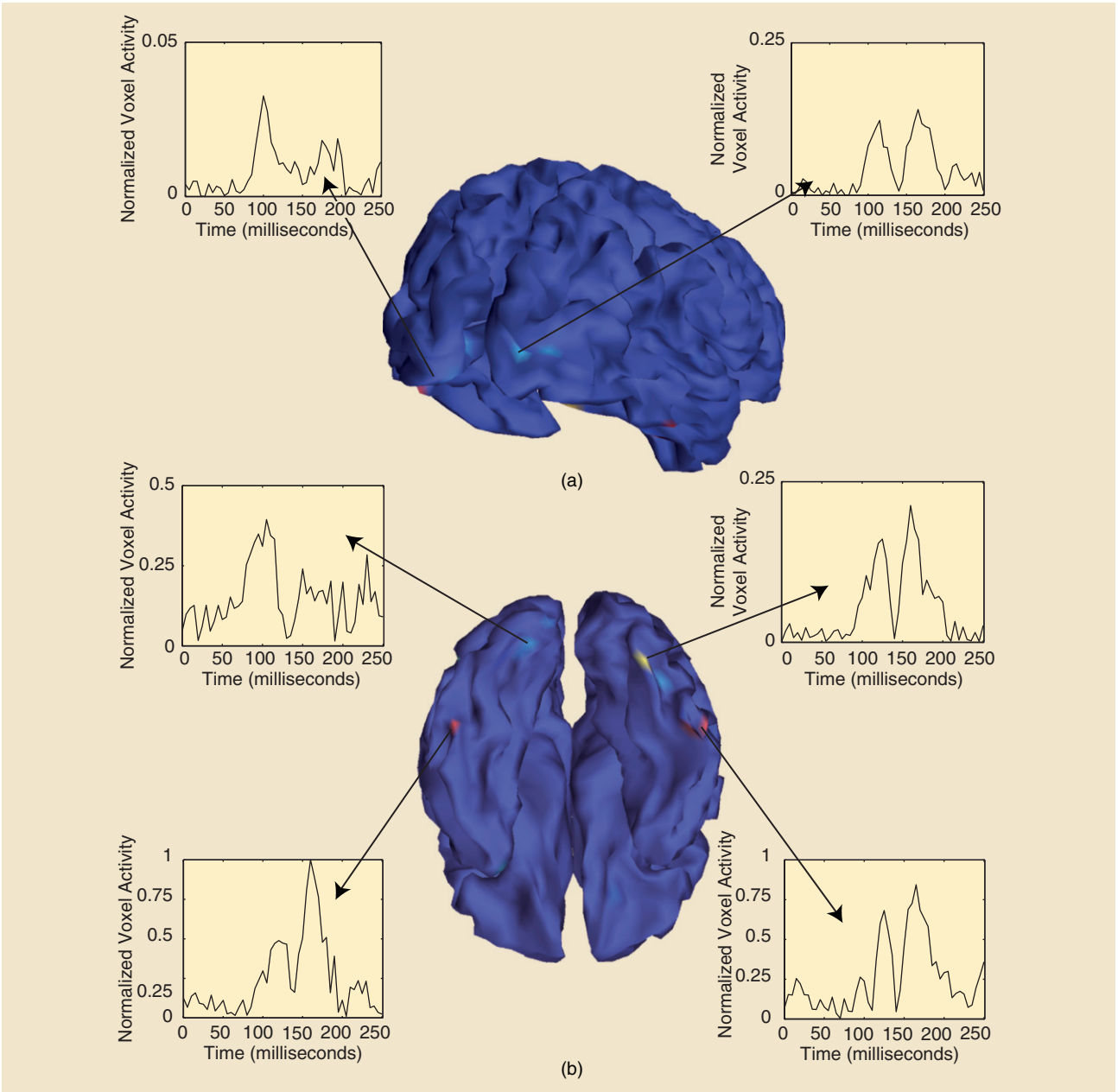
Accurate estimation of locations and spatial extents of brain sources remains a challenge for EEG/MEG source imaging. Conventional approaches yield source estimates that are either too sparse or blurry. Proper determination of the source extents requires modeling the functional interactions between sources within local patches [3]. Moreover, the majority of existing source imaging methods assume temporal independence (i.e., the source estimation is applied to each time point separately), which ignores the temporal dynamics structure clearly present in many EEG/MEG measurements. Naturally, the information afforded by this structure could be employed to further regularize the solution space and lead to performance improvement. A promising future research direction is to develop new Bayesian methods that integrate spatial and temporal modeling at the local patch level and leverage SBL techniques to automatically infer the patch size.

JOINT IMAGING OF ACTIVITY AND FUNCTIONAL CONNECTIVITY

Current approaches to functional connectivity imaging with EEG/MEG data comprise first performing imaging estimates and, subsequently, conducting functional connectivity analyses. In principle, these two steps could be integrated within a single framework. However, few efforts have been undertaken to integrate such estimates for functional connectivity due to the large number of parameters and model complexity.

LEARNING MODEL STRUCTURE

The essence of Bayesian modeling is to capture the inherent data structure. In a probabilistic graphical model, the inference task is



[FIG12] **Results for EEG face processing data from the Champagne algorithm: (a) two early visual responses in the occipital cortex with the time courses and (b) four ventral activations in (or near) the face fusiform area with time courses showing peaks around 170 milliseconds [ventral side of brain shown in (b), with the right hemisphere on the right]. (Figure adapted with permission from [41].)**

[TABLE 3] RESOURCES RELATED TO BAYESIAN INFERENCE AND RELATED TOPICS.

TOPIC	WEB RESOURCES
VB METHODS	HTTP://WWW.VARIATIONAL-BAYES.ORG
GIBBS SAMPLING (BUGS)	HTTP://WWW.MRC-BSU.CAM.AC.UK/SOFTWARE/BUGS/
SPARSE LEARNING	HTTP://WWW.MIKETIPPING.COM/DOWNLOADS.HTM
	HTTP://SPAMS-DEVEL.GFORGE.INRIA.FR
	HTTP://WWW.YELAB.NET/SOFTWARE/SLEP
NB (GP)	HTTP://WWW.GAUSSIANPROCESS.ORG
NB (DP)	HTTPS://CRAN.R-PROJECT.ORG/PACKAGE=DPPACKAGE
DEEP LEARNING	HTTP://DEEPLEARNING.NET/SOFTWARE_LINKS/

to identify the parameters or hyperparmeters specified by the model. In EEG/MEG data analysis, model specification is often empirically determined by researchers, according to certain hypotheses. This is not viable in exploratory analyses where the knowledge regarding the model structure is weak or even unavailable (e.g., in the functional connectivity analysis where the involved brain regions are unknown a priori). How to infer the graphical model structure remains an active research topic [45], [46]. For other observed input variables, it is also important to identify the causal or statistical dependency (including temporal delay) between the input and observed EEG/MEG signals.

DEVELOPMENT OF EFFICIENT INFERENCE AND OPTIMIZATION ENGINES

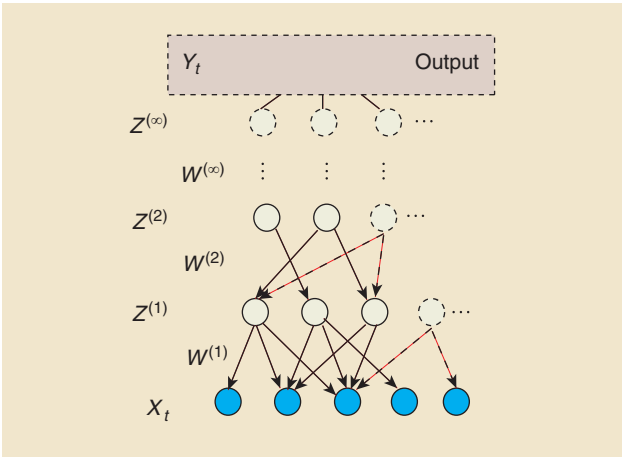
It is our belief that BML will play an important role in modern EEG/MEG signal processing applications. In comparison to the standard signal processing or likelihood-based inference, BML comes at the cost of increasing computational complexity, which is particularly problematic for EEG/MEG brain imaging in the high-resolution source space. Therefore, development of approximate, computationally efficient Bayesian inference algorithms (including parallel or distributed modules) remains an active task in the research field.

One direction is to exploit the data structure or to impose constraints (e.g., sparsity in space or time) in the model and restrict the parameter search space, in which compressed sensing idea can be further explored [47]. Another interesting direction is to employ the so-called Bayesian optimization technique, which aims to automatically tune the parameters and hyperparameters in HB models.

BAYESIAN DEEP LEARNING METHODS

Nonlinear features have been investigated and proven useful in EEG/MEG applications such as BCIs and epileptic seizure prediction. Commonly, spatiotemporal or spectral-temporal features are manually selected, which are mostly based on second-order statistics (e.g., power spectra, eigenvectors of covariance). To overcome these limitations, it is desirable to automatically extract nonlinear or high-order features that are tailored to the properties of signals of interest [48]. Deep learning is an emerging feature learning method by composing multiple nonlinear transformations of the data to produce more abstract yet potentially more useful representations [49]. It typically combines bottom-up (unsupervised, step one) inference and top-down (supervised, step two) learning procedures. Although deep learning was primarily developed for discriminative learning (which is in contrast to generative models and probabilistic inference), Bayesian analysis can be integrated with deep learning to extract optimal features.

Recently, deep learning has been demonstrated with a superior performance in an EEG-based BCI application [48]. However, to the best of our knowledge, Bayesian deep learning has not been explored in EEG/MEG applications. As deep learning often involves a multilayer network architecture (Figure 13), identifying the optimal network architecture is an important task. With the optimal or suboptimal network architecture inferred from



[FIG13] An infinite generative model for Bayesian deep learning. The shaded nodes (X_t) represent the observed random variables, the open nodes represent the latent variables ($Z^{(1)}, Z^{(2)}, \ldots, Z^{(\infty)}$), and $W^{(1)}, W^{(2)}, \ldots, W^{(\infty)}$ denote the connection weights between layers. The output layer consists of one node (Y_t) for logistic regression. (Figure adapted with permission from [50].)

Bayesian analysis, Bayesian deep learning will have great potential for EEG/MEG feature extraction.

From a Bayesian modeling viewpoint, the number of hidden layers and the number of hidden factors (i.e., latent variables) for each layer need to be determined, and both of them can be potentially infinite [49]. Each hidden factor in hidden layers may correspond to certain higher-order features extracted from the EEG time series. Hierarchical NB model provides a principled approach to this solution. In [50], the authors proposed a generative deep network architecture and imposed an infinity structure both latently and hierarchically. At the stage of unsupervised learning, Gibbs sampling can be used to infer the hidden factors and connection weights between layers. At the stage of supervised learning, supervisory signals $\boldsymbol{y}_t$ (e.g., target labels) can be fed to the output layer, followed by a discriminative learning procedure [49].

In conclusion, the BML framework provides an integrated framework for EEG/MEG data analysis. Despite many signal processing challenges, more advanced BML algorithmic development and successful EEG/MEG applications are anticipated in the forthcoming years.

ACKNOWLEDGMENTS

We would like to thank the following former and current colleagues for their contributions, cited in this article: Hagai Attias, Kensuke Sekihara, David Wipf, Julia Owen, Sarang Dalal, and Johanna Zumer. We also thank the cited publishers for permission to use previously published figures. We are grateful to Karl Friston, Vladimir Litvak, and Jose David López for helpful discussions. Wei Wu acknowledges support from the 973 Program of China (2015CB351703), the 863 Program of China (2012AA011601), the National Natural Science Foundation of China (61403114), the Guangdong Natural Science Foundation (2014A030312005 and S2013010013445), and the Steven and Alexandra Cohen Foundation. Srikantan Nagarajan received support from the U.S. National Institutes of Health (R01DC010145 and R01DC013979), the U.S.

National Science Foundation (NSF) (BCS 1262297), and the U.S. Department of Defense (W81XWH-13-1-0494). Zhe Chen was supported by a Collaborative Research in Computational Neuroscience Award (1307645) from the U.S. NSF.

AUTHORS

Wei Wu (wwumed@stanford.edu) is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou, and the Department of Psychiatry and Behavioral Sciences, Stanford University, California. He is a Member of the IEEE.

Srikantan Nagarajan (sri@ucsf.edu) is with the Department of Radiology and Biomedical Imaging, University of California at San Francisco. He is a Senior Member of the IEEE.

Zhe Chen (zhe.chen3@nyumc.org) is with the Department of Psychiatry, Neuroscience, and Physiology, School of Medicine, New York University. He is a Senior Member of the IEEE.

REFERENCES

[1] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing. Mag.*, vol. 18, no. 6, pp. 14–30, 2001.

[2] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Hoboken, NJ: Wiley, 2007.

[3] C. Lamus, M. S. Hämäläinen, S. Temereanca, E. N. Brown, and P. L. Purdon, "A spatiotemporal dynamic distributed solution to the MEG inverse problem," *Neuroimage*, vol. 63, no. 2, pp. 894–909, 2012.

[4] K. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. Trujillo-Barreto, R. Henson, G. Flandin et al., "Multiple sparse priors for the M/EEG inverse problem," *Neuroimage*, vol. 39, no. 3, pp. 1104–1120, 2008.

[5] A. Bolstad, B. V. Veen, and R. Nowak, "Space-time event sparse penalization for magneto-/electro-encephalography," *Neuroimage*, vol. 46, no. 4, pp. 1066–1081, 2009.

[6] K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: Theory," *Neuroimage*, vol. 16, no. 2, pp. 465–483, 2002.

[7] J. Mattout, C. Phillips, W. D. Penny, M. D. Rugg, and K. J. Friston, "MEG source localization under multiple constraints: an extended Bayesian framework," *Neuroimage*, vol. 30, no. 3, pp. 753–767, 2006.

[8] M. W. Seeger and D. P. Wipf, "Variational Bayesian inference techniques," *IEEE Signal Processing. Mag.*, vol. 27, no. 6, pp. 81–91, 2010.

[9] D. P. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6236–6255, 2011.

[10] D. P. Wipf and S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," *Neuroimage*, vol. 44, no. 3, pp. 947–966, 2009.

[11] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. New York: Springer-Verlag, 2007.

[12] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall/CRC, 1995.

[13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 6, no. 6, pp. 721–741, 1984.

[14] T. Marshall and G. Roberts, "An adaptive approach to Langevin MCMC," *Stat. Comput.*, vol. 22, no. 5, pp. 1041–1057, 2012.

[15] H. Attias, "Independent factor analysis," *Neural Comput.*, vol. 11, no. 4, pp. 803–851, 1999.

[16] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[17] W. Wu, Z. Chen, X. Gao, Y. Li, E. Brown, and S. Gao, "Probabilistic common spatial patterns for multichannel EEG analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 639–653, 2015.

[18] H. Kang and S. Choi, "Bayesian common spatial patterns with Dirichlet process priors for multi-subject EEG classification," in *Proc. Int. Joint Conf. Neural Networks (IJCNN)*, 2012, pp. 1–6.

[19] W. Wu, Z. Chen, S. Gao, and E. N. Brown, "A hierarchical Bayesian approach for learning sparse spatio-temporal decompositions of multichannel EEG," *Neuroimage*, vol. 56, no. 4, pp. 1929–1945, 2011.

[20] J. M. Zumer, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "A probabilistic algorithm integrating source localization and noise suppression for MEG and EEG data," *Neuroimage*, vol. 37, no. 1, pp. 102–115, 2007.

[21] J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Phys. Med. Biol.*, vol. 32, no. 1, pp. 11–22, 1987.

[22] S. S. Nagarajan, H. T. Attias, K. E. H. II, and K. Sekihara, "A graphical model for estimating stimulus-evoked brain responses from magnetoencephalography data with large background brain activity," *Neuroimage*, vol. 30, no. 2, pp. 400–416, 2006.

[23] V. Litvak and K. Friston, "Electromagnetic source reconstruction for group studies," *Neuroimage*, vol. 42, no. 4, pp. 1490–1498, 2008.

[24] J. D. López, V. Litvak, J. J. Espinosa, K. Friston, and G. R. Barnes, "Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM," *Neuroimage*, vol. 84, pp. 476–487, Jan., 2014.

[25] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[26] Y. Li, Z. Yu, N. Bi, Y. Xu, Z. Gu, and S. I. Amari, "Sparse representation for brain signal processing: A tutorial on methods and applications," *IEEE Signal Processing. Mag.*, vol. 31, no. 3, pp. 96–106, 2014.

[27] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[28] D. P. Wipf and S. Nagarajan, "A new view of automatic relevance determination," in *Proc. Advances in Neural Information Processing Systems*, 2008, vol. 20, pp. 1625–1632.

[29] W. Wu, C. Wu, S. Gao, B. Liu, Y. Li, and X. Gao, "Bayesian estimation of ERP components from multicondition and multichannel EEG," *Neuroimage*, vol. 88, pp. 319–339, Mar., 2014.

[30] S. S. Nagarajan, H. T. Attias, K. E. Hild, and K. Sekihara, "A probabilistic algorithm for robust interference suppression in bioelectromagnetic sensor data," *Stat. Med.*, vol. 26, no. 21, pp. 3886–910, 2007.

[31] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG," *Neuroimage*, vol. 49, no. 1, pp. 641–655, 2010.

[32] Z. Ghahramani, "Bayesian non-parametrics and the probabilistic approach to modelling," *Philos. Trans. R. Soc. A*, vol. 371, no. 1984, p. 20110553, 2013.

[33] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2005.

[34] N. L. Hjort, C. Holmes, P. Mülcer, and S. G. Walker, Eds., *Bayesian Nonparametrics*. Cambridge, UK: Cambridge Univ. Press, 2010.

[35] C. J. Paciorek and M. J. Schervish, "Nonstationary covariance functions for Gaussian process regression," in *Proc. Advances in Neural Information Processing Systems,* 2004, vol. 16, pp. 273–280.

[36] S. Faul, G. Gregorcic, G. Boylan, W. Marnane, G. Lightbody, and S. Connolly, "Gaussian process modeling of EEG for the detection of neonatal seizure," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2151–2162, 2007.

[37] M. Zhong, F. Lotte, M. Girolami, and A. Lecuyer, "Classifying EEG for brain computer interfaces using Gaussian process," *Pattern Recognit. Lett.*, vol. 29, no. 3, pp. 354–359, 2008.

[38] A. Fyshe, E. Fox, D. Dunson, and T. Mitchell, "Hierarchical latent dictionaries for models of brain activation," in *Proc. 15th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, 2012, pp. 409–421.

[39] H. Kang and S. Choi, "Bayesian common spatial patterns for multi-subject EEG classification," *Neural Netw.*, vol. 57, pp. 39–50, Sept. 2014.

[40] A. Ossadtchi, S. Baillet, J. C. Mosher, D. Thyerlei, W. Sutherling, and R. M. Leahy, "Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering," *Clin. Neurophysiol.*, vol. 115, no. 3, pp. 508–522, 2004.

[41] J. P. Owen, D. P. Wipf, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Performance evaluation of the CHAMPAGNE source reconstruction algorithm on simulated and real M/EEG data," *Neuroimage*, vol. 60, no. 1, pp. 305–323, 2012.

[42] R. N. Henson, G. Flandin, K. J. Friston, and J. Mattout, "A parametric empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction," *Hum. Brain Mapp.*, vol. 31, no. 10, pp. 1512–1531, 2010.

[43] S. J. Kiebel, M. I. Garrido, R. Moran, C. Chen, and K. J. Friston, "Dynamic causal modeling for EEG and MEG," *Hum. Brain Mapp.*, vol. 30, no. 6, pp. 1866–1876, 2009.

[44] Z. Zhang, T. Jung, S. Makeig, and B. D. Rao, "Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 221–224, 2013.

[45] D. Heckerman, "A tutorial on learning with Bayesian networks," in *Learning in Graphical Models*, M. I. Jordan, Ed. Cambridge, MA: MIT Press, 1999, pp. 301–354.

[46] J. Lee and T. Hastie, "Structure learning of mixed graphical models," in *Proc. 16th Int. Conf. Artificial Intelligence and Statistics*, 2013, pp. 388–396.

[47] B. Babdi, G. Obregon-Henao, C. Lamus, M. S. Hämäläinen, E. N. Brown, and P. L. Purdon, "A subspace pursuit-based iterative greedy hierarchical solution to the neuromagnetic inverse problem." *Neuroimage*, vol. 87, pp. 427–443, Feb., 2014.

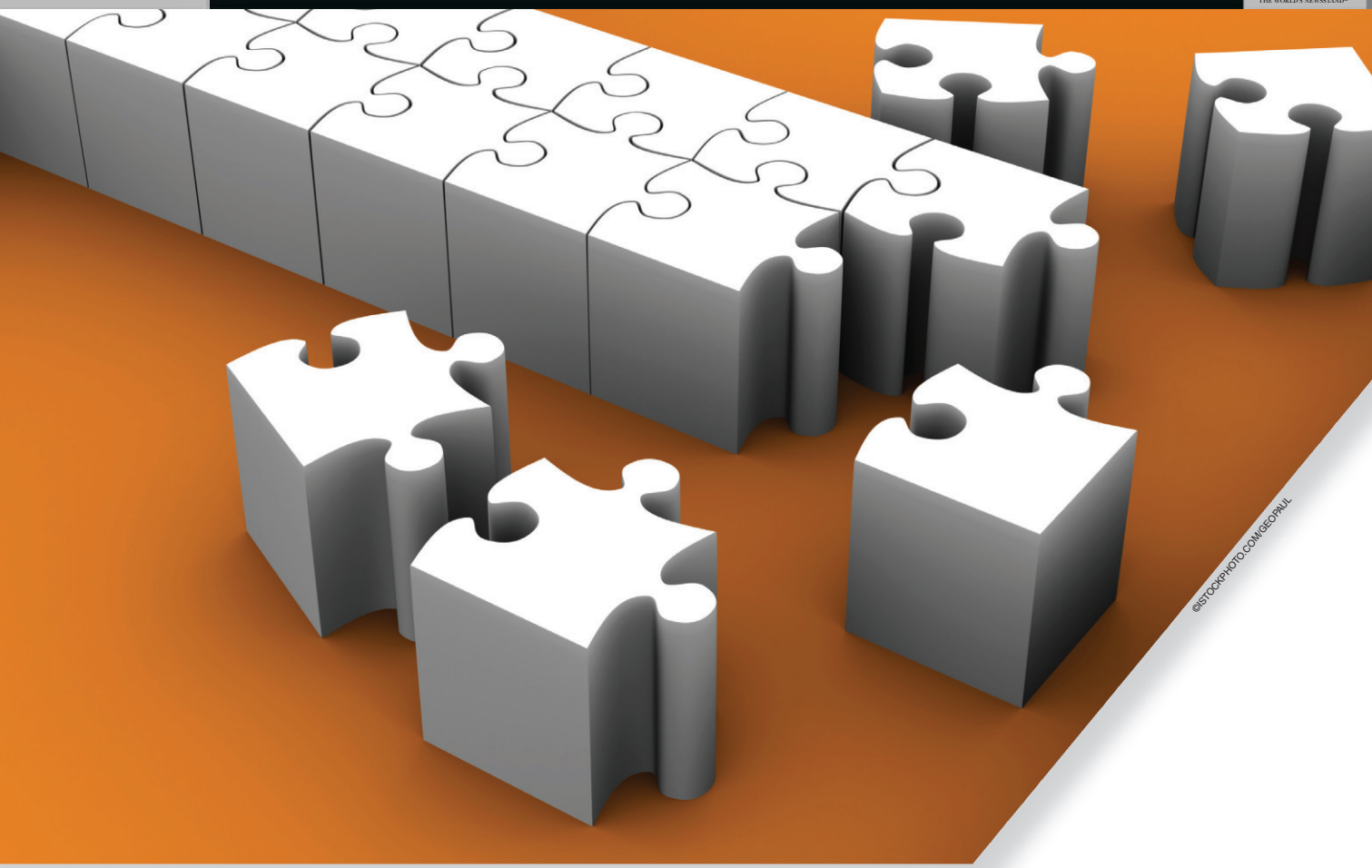
[48] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, 2011.

[49] L. Deng and D. Yu, "Deep learning: methods and applications," *Found. Trends Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[50] E. Pan and Z. Han. (2014). Non-parametric Bayesian learning with deep learning structure and its application in wireless network. [Online]. Available: https://arxiv.org/pdf/1410.4599v2.pdf

[SP]

IEEE SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®



Yan Chen, Chunxiao Jiang, Chih-Yu Wang, Yang Gao, and K.J. Ray Liu

Decision Learning

Data analytic learning with strategic decision making

With the increasing ubiquity and power of mobile devices as well as the prevalence of social systems, more activities in our daily life are being recorded, tracked, and shared, creating the notion of social media. Such abundant and still growing real-life data, known as *big data*, provide a tremendous research opportunity in many fields. To analyze, learn, and understand such user-generated data, machine learning has been an important tool, and various machine-learning algorithms have been developed. However, since the user-generated data are the outcome of users' decisions, actions, and socioeconomic interactions, which are highly dynamic, without considering users' local behaviors and interests, existing learning approaches tend to focus on optimizing a global objective function at the macroeconomic level, while totally ignoring users' local interactions at the microeconomic level. As such, there is a growing need to combine learning with strategic decision making, which are two traditionally distinct research disciplines, to be able to jointly consider both global phenomena and local effects to better understand, model, and analyze the newly arising issues in the emerging social media with user-generated data. In this article, we present an overview of the emerging notion of decision learning, i.e., learning with strategic decision making, which involves users' behaviors and interactions by combining learning with strategic decision making. We will discuss some examples from social

Digital Object Identifier 10.1109/MSP.2015.2479895
Date of publication: 29 December 2015

media with real data to show how decision learning can be used to better analyze users' optimal decision from a user's perspective, as well as design a mechanism from the system designer's perspective to achieve a desirable outcome.

INTRODUCTION

With the rapid development of communication and information technologies, the last decade has witnessed a proliferation of emerging social systems that help to promote the connectivity of people to an unprecedentedly high level. Examples of these emerging systems can be found in a wide range of domains, from online social networks like Facebook or Twitter to crowdsourcing sites like Amazon Mechanical Turk or Topcoder where people solve various tasks by assigning them among a large pool of online workers to online question-and-answer (Q&A) sites like Quora or Stack Overflow where people ask all kinds of questions, and all the way to new paradigms of power systems like smart grid. In a social system, there are two key characters: the system designer, who designs the social system, and the users, who participate in the social system and generate the data.

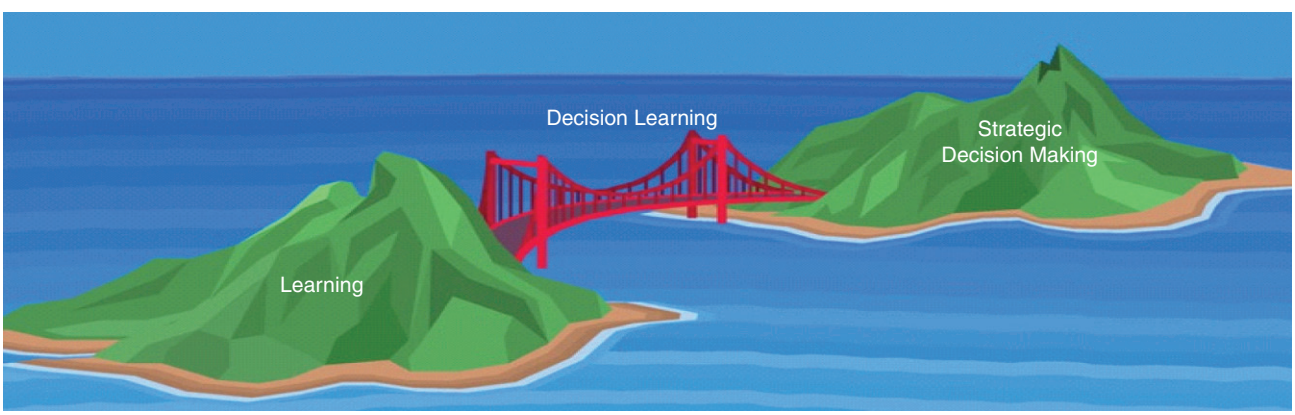
Because of the increasing ubiquity and power of mobile devices, the prevalence of social systems, and the rise of cloud data storage, our daily activities are increasingly being recorded, tracked, and shared. Big data provides a tremendous research opportunity in many fields, for example, behavior and sentiment analysis, epidemics and diseases propagation modeling, grid and network traffic management, and financial market trends tracking, just to name a few.

Machine learning has been an important tool to understand user-generated data, [1], [2]. Learning aims to use reasoning to find new, relevant information, given some background knowledge through representation, evaluation, and optimization. However, there are some limitations and constraints. For example, the generalization assumption that the training set is statistically consistent with the testing set is often not true because users behave differently at different times and under different settings. Moreover, the single-objective function cannot address all interests of users since interests vary and, thus, there will be different objective functions. Besides, users are rational and naturally selfish, so they want to optimize their own objective functions [3], [4]. In addition, the data are the outcome of users' interactions, while learning algorithms cannot naturally involve users' individual local interests. Therefore, the knowledge contained in the data is difficult to fully exploit from such a macroscopic view.

> LEARNING APPROACHES TEND TO FOCUS ON OPTIMIZING A GLOBAL OBJECTIVE FUNCTION AT THE MACROECONOMIC LEVEL BUT TOTALLY IGNORE USERS' LOCAL DECISIONS AND INTERACTIONS AT THE MICROECONOMIC LEVEL.

Existing learning approaches tend to focus on optimizing a global objective function at the macroeconomic level but totally ignore users' local decisions and interactions at the microeconomic level. Indeed, user-generated data are the outcome of users' decisions, actions, and their social–economic interactions, which are highly dynamic, and, thus, the interactions of users and their decision-making process should be taken into consideration. As such, there is a growing need to bridge learning and strategic decision making to be more effective in mining, reasoning, and extracting knowledge and information from big data.

Yet there is a missing link. Traditionally, both learning and decision making are two distinct research disciplines. Successfully bridging them allows us to jointly consider both global phenomena and local effects to better understand, model, and analyze user-generated data from social media. Besides, learning is for making optimal decisions. In essence, learning and decision making are destined to couple due to the network externality, i.e., the influence of other users' behaviors on one user's reward [5]. In this article, we describe the emerging research field of decision learning, i.e., learning with strategic decision making, which involves users' behaviors and interactions by combining learning with strategic decision making, as illustrated in Figure 1. In decision learning, there are two major elements of data-driven issues:



[FIG1] Decision learning: bridging learning and strategic decision making. Learning and decision making are traditionally two distinct research disciplines. Bridging them allows us to jointly consider both global and local effects to better understand, model, and analyze user-generated data from social media.

one is the modeling, analysis, and understanding of user behaviors and their interactions, and the other is the design of mechanisms to achieve the desired outcomes. The former considers the issues from user perspectives, while the latter is motivated by the system's point of view.

In contrast to traditional networks and systems, where users are constrained by fixed and predetermined rules, user interactions in social networks are generally self-enforcing [6], [7]. On the one hand, users in these systems have great flexibility in their actions and have the ability to observe, learn, and make intelligent decisions. On the other hand, because of their selfish nature, users will act to pursuit their own interests, which often conflict with other users' objectives and the system designer's goal. These new features call for new theoretical and practical solutions to the designs of social networks. How can system designers design their systems to resolve the conflicting interests among users? And given the various and conflicting interests among users, how is the desired system-wide performance achieved?

The aforementioned questions motivate the study of user behaviors and incentive mechanisms in social systems. Incentive mechanisms refer to schemes that aim to steer user behaviors through the allocation of various forms of rewards, such as monetary rewards, virtual points, and reputation status. Plenty of empirical evidence can be found in the social psychology literature that demonstrates user behaviors in social networks are indeed highly influenced by these rewards [8]–[13]. Although we can use the social psychology literature to learn what factors influence user behaviors and, thus, can be used as rewards, how to allocate the rewards to achieve the desired user behavior is still not well understood, which leads to ad hoc or poor designs of incentive mechanisms in many social networks in practice. How can we fundamentally understand user behavior under the presence of rewards in social networks? Moreover, based on such understanding, how should a system developer design incentive mechanisms to achieve various objectives in a systematic way?

LEARNING AIMS TO USE REASONING TO FIND NEW, RELEVANT INFORMATION, GIVEN SOME BACKGROUND KNOWLEDGE THROUGH REPRESENTATION, EVALUATION, AND OPTIMIZATION.

The focus of this article is to open a discussion in a tutorial way of an emerging field, termed *decision learning*, which jointly combines learning with decision making toward a better fundamental understanding of user behaviors embedded under the tsunami of user-generated big data. In this article, we present three game-theoretic frameworks (readers who are interested in the preliminaries of game theory are referred to [3], [4], [6], [7], and [14]) to formally model user participation and interactions under various scenarios in social networks: decision learning with evolutionary user behavior, decision learning with sequential user behavior, and decision learning with mechanism design. A high-level comparison of the three frameworks is shown in Figure 2. On the evolutionary behavior, how information diffuses over online social networks using a graphical evolutionary game will be presented [15], [16], and the focus will be on how to learn users' utility function from Twitter and MemeTracker data for understanding and modeling strategic decision making; on the sequential behavior, how customers learn and choose the "best" deals using the Chinese-restaurant-game framework will be considered [17], [18]. In addition, Groupon deals

	Sequential Behavior	Evolutionary Behavior	Mechanism Design
Perspective	User	User	System
Approach	Bottom Up	Bottom Up	Top Down
Rationality	Fully Rational	Limited	Fully Rational
User Type or Population	Static	Dynamic	Probabilistic
Solution Concept	Subgame-Perfect Nash Equilibrium	Evolutionarily Stable Strategy	Symmetric Bayesian Equilibrium
Solution Method	Backward Induction	Replicator Dynamic	Punishment Design
Performance Metrics and Desired Properties	Existence and Uniqueness of Equilibrium, Price of Anarchy	Evolutionarily Stable, Price of Anarchy	Strategy-Proof, Individual Rationality, Budget Balance
Application	Information Sharing	Deal Selection	Crowdsourcing

[FIG2] A high-level comparison of decision learning with evolutionary user behavior, decision learning with sequential user behavior, and decision learning with mechanism design.

and Yelp rating data will be used to discuss how users can learn from each other's interactions for better strategic decision making. On the mechanism design, how to design the mechanism to collect high-quality data with low cost from crowdsourcing will be illustrated [19]. Using these frameworks, we can theoretically analyze and predict user behaviors through equilibrium analysis. And, based on the analysis, one can optimize in a systematic way the design of incentive mechanisms for social networks to achieve a wide range of system objectives and analyze their performances accordingly. Finally, recent related works on the intersection of learning and strategic decision making will be surveyed and discussed.

INCENTIVE MECHANISMS REFER TO SCHEMES THAT AIM TO STEER USER BEHAVIORS THROUGH THE ALLOCATION OF VARIOUS FORMS OF REWARDS SUCH AS MONETARY REWARDS, VIRTUAL POINTS, AND REPUTATION STATUS.

USER BEHAVIOR MODELING AND ANALYSIS IN DECISION LEARNING

In this section, we will address decision learning from the user's point of view. Both the evolutionary and sequential user behaviors are commonly exhibited in social systems. How learning with strategic decision making may arise from both settings will be illustrated, first with information diffusion over online social networks using the graphical evolutionary game framework from Twitter and MemeTracker data, and then with the optimal restaurant strategy using the Chinese-restaurant-game framework from both Groupon deals and Yelp ratings, respectively.

EVOLUTIONARY USER BEHAVIOR: GRAPHICAL EVOLUTIONARY GAME FRAMEWORK

One typical user behavior in social systems is the repetitive and evolutionary decision making. A good example is that users repetitively decide whether to post information or not on online social networks. Figure 3 shows the top 50 threads in the news cycle with highest volume for the period of 1 August–31 October 2008, where each thread consists of all new articles and blog posts containing a textual variant of a particular quoted phrase. The five large peaks between late August and late September corresponding to the Democratic and Republican National Conventions illustrate the spread of comments and phrasing by candidates. Notice that the information forwarding is often not unconditional. One has to make a decision on whether or not to do so based on many factors, such as if the information is exciting or if friends are interested in it, etc. Other examples include repetitive online purchasing and review posting.

We find that, in essence, the repetitive and evolutionary decision-making process on social systems follows the evolution process in natural ecological systems [21]. It is a process that evolves from one state at a particular instance to another when information is shared and a decision is made. Thus, the evolutionary game is an ideal tool to model and analyze the social system users' repetitive and evolutionary behavior. Evolutionary game theory (EGT) is an application of the mathematical theory of games to the interaction-dependent strategy evolution in populations [21]. Arising from the realization that frequency-dependent fitness introduces a strategic aspect to evolution, EGT becomes an essential component of a mathematical and computational approach to biological contexts, such as genes, viruses, cells, and humans. Recently, EGT has also become of increased interest to economists, sociologists, anthropologists, and social scientists. Here, we show how the evolutionary game theory is deployed to study users' repetitive and evolutionary behavior in social systems.

In the setting of our consideration, the social system user topology can be treated as a graph structure, and the user with a new decision can be regarded as the mutant. By considering the decision-making process as the mutant-spreading process [to forward or not to forward when an event (mutation) takes place], the graphical evolutionary game provides us with an analytical means to find the evolutionary dynamics and equilibrium of user behavior.

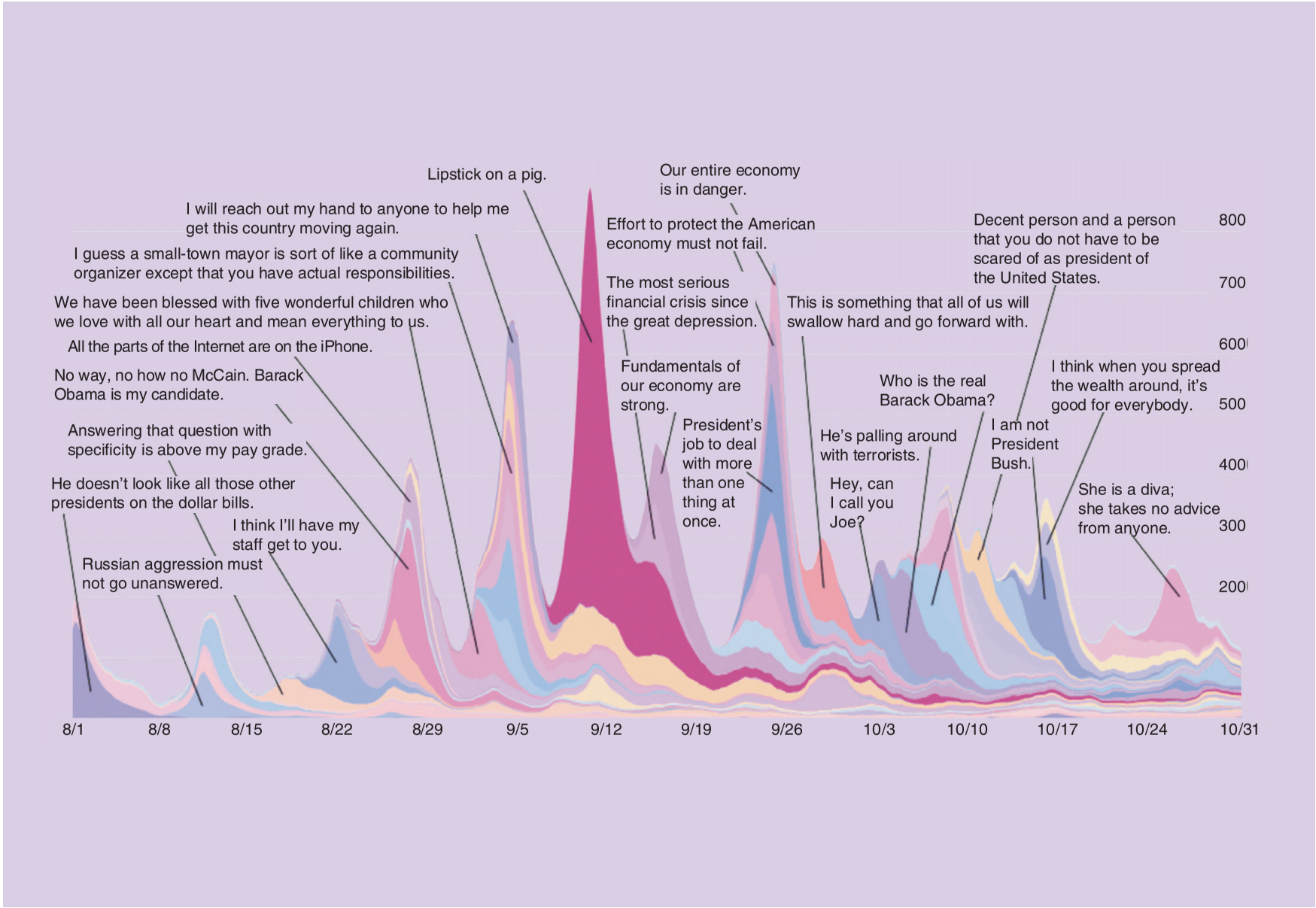
GRAPHICAL EVOLUTIONARY GAME FRAMEWORK

In EGT, the utility of a player is referred to as *fitness* [22]. Specifically, the fitness Φ is a linear combination of the baseline fitness (B) representing the player's inherent property and the player's payoff (U), which is determined by the predefined payoff matrix and the player's interactions with others as follows:

$$\Phi = (1-\alpha)B + \alpha U, \tag{1}$$

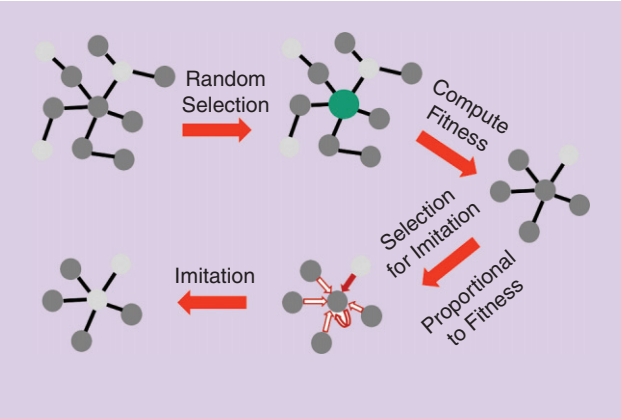
where the combining weight α is called the *selection intensity*. One can interpret that one's fitness is not only determined by one's own strength, but also from one's environment affecting with a selection intensity α. The case that $\alpha \to 0$ represents the limit of weak selection [23], while $\alpha \to 1$ denotes strong selection. The selection intensity can also be time varying, e.g., $\alpha = \beta e^{-\varepsilon t}$, which means that the contribution of game interaction decreases with time.

With the fitness function, the EGT studies and characterizes how a group of players converge to a stable equilibrium after a period of strategic interactions. Such a final equilibrium state is called the *evolutionarily stable state* (*ESS*), which is "a strategy such that, if all members of the population adopt it, then no mutant strategy could invade the population under the influence of natural selection" [21]. In other words, even if a small fraction of players may not be rational and take out-of-equilibrium strategies, ESS is still a locally stable state. How to find the ESSs is an important issue in EGT. One common approach is to find the stable points of the system state dynamic, which is known as *replicator dynamics*. The corresponding underlying physical meaning is that, if adopting a certain strategy can lead to a higher fitness than the average level, the proportion of population adopting this strategy will increase, and the increasing rate is proportional to the difference between the average fitness with this strategy and the average fitness of the whole population. Note that when the total population is sufficiently large



[FIG3] **MemeTracker phrase cluster during the 2008 U.S. presidential election: spreading of comments and phrasing by candidates [20].**

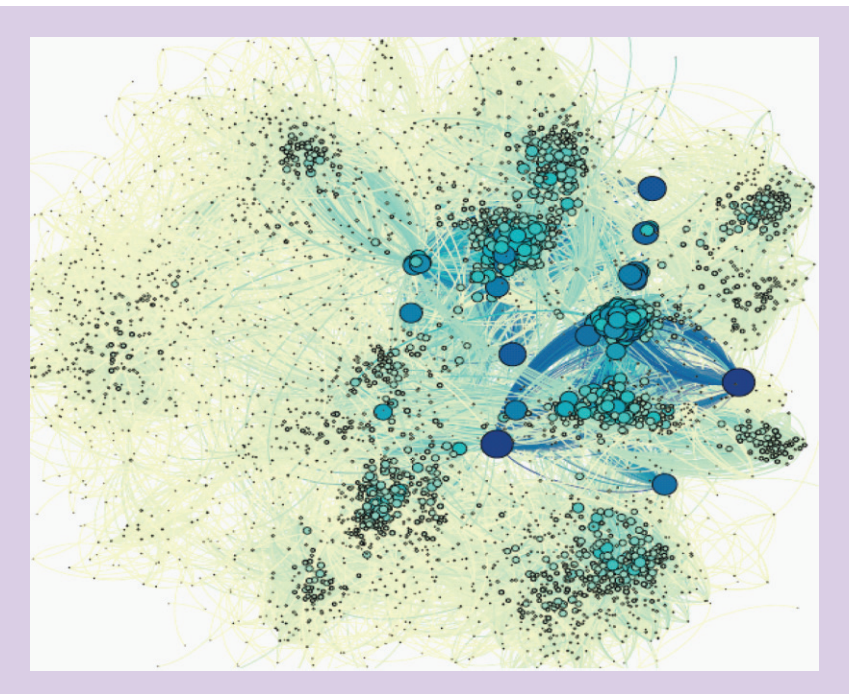
IEEE SignalProcessing
Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page



[FIG4] **The imitation strategy updating rule.**

and homogeneous, the proportion of players adopting a certain strategy is equivalent to the probability of one individual player adopting such a strategy, i.e., the strategy distribution over the whole population can be interpreted as each player's mixed strategy, and the replicator dynamics can be interpreted as each player's mixed strategy update.

Graphical EGT studies the strategies' evolution in such a structured population [24]. In the EGT, in addition to the entities of players, strategy, and fitness matrix, each game model is associated with a graph structure, where the vertices represent players and the edges determine which player to interact with. Since the players only have limited connections with others, each player's fitness is locally determined from interactions with all adjacent players.



[FIG5] **A Facebook subnetwork [16], [26].**

The commonly used strategy updating rules [25] are originated from the evolutionary biology field and used to model the mutant evolution process. Figure 4 illustrates the detailed evolution procedures of the imitation strategy update rule. In the first step, a user is randomly chosen from the population for imitation. Then, the fitness of the chosen user and all corresponding neighbors is computed. Finally, the user will, in all probability, either be imitated by one of the neighbors or remain with the current strategy, with the probability being proportional to fitness. There are also other rules, such as the birth–death strategy update rule and death–birth strategy update rule, but through theoretical analysis [15], we find that these rules are equivalent when the network degree is sufficiently large.

INFORMATION-DIFFUSION FORMULATION AND ANALYSIS

A social network is usually illustrated by a graph, e.g., a Facebook subnetwork is shown in Figure 5, where each node represents a user, and lines represent the relationships between users. When some new information is originated from one user, the information may be propagated over the network, depending on other users' actions to forward the information or not. For each user, whether he or she forwards the information is determined by several factors, including the user's own interest in the information and the neighbor's actions, in the sense that, if all the neighbors forward the information, the user may also forward the information with a relatively high probability. In such a case, the users' actions are coupled with each other through their social interactions. This is very similar to the player's strategy update in the graphical evolutionary game, where players' strategies are also influenced with each other through the graph structure. In the graphical evolutionary game, a user's strategy can influence one of the neighbors when the fitness of adopting this strategy is high. Similarly, in the information-diffusion process, when forwarding the information can bring a user more utility, the user's neighbors may also be influenced to forward the information in the near future. Therefore, the information-diffusion process can be well modeled by the graphical evolutionary game, as illustrated in Figure 6.

There are two possible actions for each user, i.e., to forward (S_f) or not forward (S_n), and the corresponding users' payoff matrix can be written as

$$\begin{pmatrix} u_{\mathrm{ff}} & u_{\mathrm{fn}} \\ u_{\mathrm{fn}} & u_{\mathrm{nn}} \end{pmatrix}, \tag{2}$$

where a symmetric payoff structure is considered, i.e., when a user with strategy S_f meets a user with strategy S_n, each of them receives the same payoff u_{fn}. Note that the payoff matrix is related to the fitness in the graphical evolutionary game according to (1). The physical meaning of

the payoff can be either the popularity of a user in a social network or the hit rate of a website. These three parameters will be learned from the data and then used for decision making. Under different application scenarios, the values of the payoff matrix may be different. For example, if the information is related to recent hot topics and forwarding of the information can attract more attentions from other users or websites, the payoff matrix should have the following characteristic: $u_{\mathrm{ff}} \geq u_{\mathrm{fn}} \geq u_{\mathrm{nn}}$. According to (1), the fitness of forwarding is larger, and, thus, the probability of forwarding will be higher. On the other hand, if the information is about useless advertisements, the payoff matrix would exhibit $u_{\mathrm{nn}} \geq u_{\mathrm{fn}} \geq u_{\mathrm{ff}}$, i.e., the fitness of not forwarding is higher, and, thus, users tend not to forward the information. Furthermore, if the information is supposed to be shared only within a circle, i.e., a small group with the same interest, the payoff matrix could exhibit $u_{\mathrm{fn}} \geq u_{\mathrm{ff}} \geq u_{\mathrm{nn}}$.

Since the player's payoff is determined by both his or her own strategy and the opponent's strategy, to characterize the global population dynamics, we need to first derive the local influence dynamics as well as the corresponding influence equilibria. We find in [15] that the local network states, i.e., the neighbors' strategy distribution given a player's strategy, evolve with a rate of order 1, while the global network state, i.e., the strategy distribution of the whole population, evolves with a rate at the order of the selection intensity α, which is much smaller than one due to the weak selection [23]. In such a case, the local network states will converge to equilibria at a much faster rate than the global network state. This is because the dynamics of local network states are only in terms of a local area, which contains only the neighbors. At such a small scale, the local dynamics can change and converge quite fast. On the other hand, if the dynamics of the global network state are associated with all users, i.e., the whole network, the dynamics would be much slower. Therefore, the global network state can be regarded as constant during the convergence of influence dynamics. By doing so, the equilibria of the local influence dynamics can be obtained, which are found to be linear functions of the global network state.

With the equilibria of the local influence dynamics, the global population dynamics can be derived through analyzing the strategy updating rules specified in the graphical evolutionary game [25]. It is found that the global population dynamics can be represented as a two-parameter, third-order polynomial function of the global network state [15]

$$\dot{p}_f(t) = \frac{\alpha(\bar{k}-1)(\overline{k^2}-2\bar{k})}{(\overline{k^2}-\bar{k})^2} p_f(t)[1-p_f(t)][ap_f(t)+b], \quad (3)$$

where $p_f(t)$ is the proportion of the population forwarding the information, $\dot{p}_f(t)$ is the corresponding dynamics, $\bar{k} = E[k]$ is the average degree of the network, $\overline{k^2} = E[k^2]$ is the second moment of the degree of the network, and a and b are two parameters determined by the payoff matrix shown in (2).

From (3), we can see that, given the characteristic of the network, i.e., the average degree $\bar{k}$ and the second moment of the degree $\overline{k^2}$, the evolution dynamics of the information diffusion can be modeled by a simple two-parameter, third-order polynomial function, where the two parameters a and b are determined by the payoff in the payoff matrix, i.e., u_{ff}, u_{fn} and u_{nn}. Therefore, by learning the payoff from the data, we are able characterize the evolution dynamics of information diffusion using the evolutionary game-theoretic framework.

By evaluating the global population dynamics at the steady state, the global population equilibria can be found [16], which is zero (no user shares the information with the neighbors), one (all users share the information with their neighbors), or only a portion of users share the information with their neighbors where the amount of such users is purely determined by the payoff matrices as follows:

$$p_f^* = \begin{cases} 0, & \text{if } u_{\mathrm{nn}} > u_{\mathrm{fn}} > u_{\mathrm{ff}}; \\ 1, & \text{if } u_{\mathrm{ff}} > u_{\mathrm{fn}} > u_{\mathrm{nn}}; \\ \dfrac{(\overline{k^2}/\bar{k}-2)(u_{\mathrm{fn}}-u_{\mathrm{nn}})+(u_{\mathrm{ff}}-u_{\mathrm{nn}})}{(\overline{k^2}/\bar{k}-2)(2u_{\mathrm{fn}}-u_{\mathrm{ff}}-u_{\mathrm{nn}})}, & \text{else.} \end{cases} \quad (4)$$

From (4), we can see that neither user forwarding the information can gain the most payoff, while both forwarding gains the least payoff, $p_f^* = 0$. This corresponds to the scenario where the released information is useless or a negative advertisement, forwarding that can only incur unnecessary cost. On the contrary, both users forwarding the information can gain the most payoff, while not forwarding gains the least payoff, $p_f^* = 1$. This corresponds to the scenario where the released information is an extremely hot topic, and forwarding it can attract more attention. For other cases, p_f^* lies between zero and one. For this third ESS, some approximations can be made as follows:

$$\begin{aligned} p_f^* &= \frac{(\overline{k^2}/\bar{k}-2)(u_{\mathrm{fn}}-u_{\mathrm{nn}})+(u_{\mathrm{ff}}-u_{\mathrm{nn}})}{(\overline{k^2}/\bar{k}-2)(2u_{\mathrm{fn}}-u_{\mathrm{ff}}-u_{\mathrm{nn}})}, \\ &\doteq \frac{1}{1+\dfrac{u_{\mathrm{fn}}-u_{\mathrm{ff}}}{u_{\mathrm{fn}}-u_{\mathrm{nn}}}}, \end{aligned} \quad (5)$$

where the last approximation is due to $\overline{k^2}/\bar{k} \geq \bar{k}$ and the assumption that the average network degree $\bar{k} \gg 2$ in real social

Graphical EGT	Social Network
Graph Structure	Social Network Topology
Players	Users in the Social Network
Strategy	S_f: Forward the Information S_n: Not Forward the Information $\begin{array}{c} \\ S_f \\ S_n \end{array} \begin{array}{c} \begin{array}{cc} S_f & S_n \end{array} \\ \begin{pmatrix} u_{ff} & u_{fn} \\ u_{fn} & u_{nn} \end{pmatrix} \end{array}$
Fitness	Utility from Forwarding or Not
ESS	Stable Information-Diffusion State

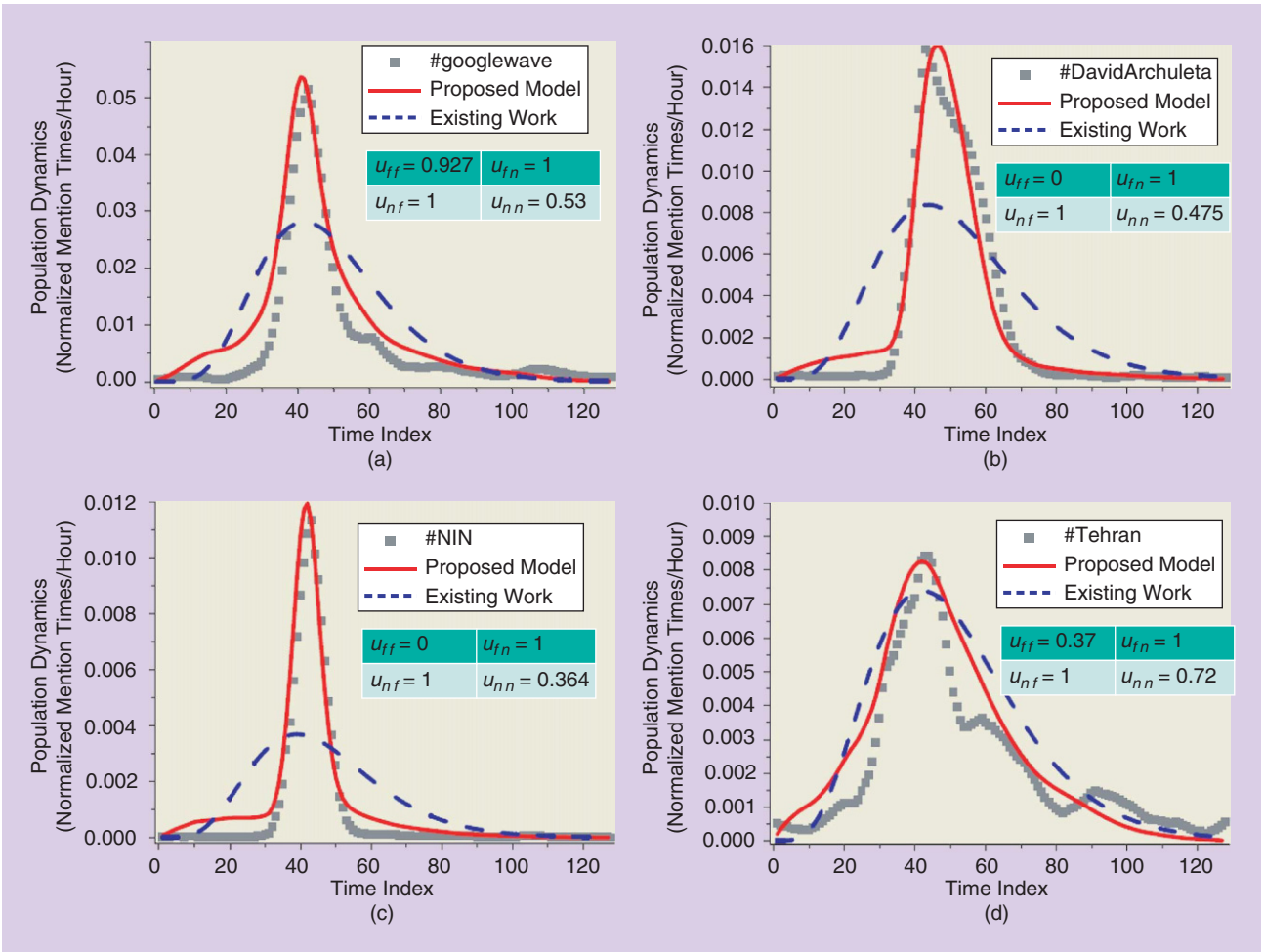
[FIG6] **Information diffusion as a graphical evolutionary game.**

networks. We can see that, when the average network degree $\bar{k}$ is sufficiently large, the information-diffusion result is independent of the network scale, i.e., there is a scale-free phenomenon for the information-diffusion equilibrium.

EXPERIMENTS WITH REAL-WORLD DATA SETS

The real-world data sets are used to validate the proposed model. We first use the Twitter hashtag data set to validate the evolutionary population dynamics [15]. Specifically, we learn the payoff matrix in (2) by fitting the real temporal dynamics with the evolution dynamics in (3) and generate the corresponding evolution dynamics based on the estimated payoff matrix. The Twitter hashtag data set contains the number of mentions per hour of 1,000 Twitter hashtags with the corresponding time series, which are the 1,000 hashtags with highest total mention times among 6 million hashtags from June to December 2009 [26]. We compare our results with one of the most related existing works using a data mining method [20]. Figure 7 shows the comparison results, where the vertical axis is the dynamics and the mention times of different hastags per hour in the Twitter data set are normalized within interval [0, 1] and denoted by solid gray square. From the figure, we can see that the game-theoretic model can fit very well the real-world information-diffusion dynamics, better than the data mining method in [20] since the users' interactions and decision-making behaviors are taken into account.

We then use the MemeTracker data set to validate the ESS [16]. The data set contains more than 172 million news articles and blog posts from 1 million online sources [20]. When a site publishes a new post, it will put hyperlinks to related posts in some other sites published earlier as its sources. Later, the site will also be cited by other newer posts as well. An example is shown in Figure 8. In such a case, the hyperlinks between articles and posts can be used to represent the spreading of information from one site to another. We extract five groups of sites, where each group includes 500 sites. Each group is regarded as a complete graph, and each site is considered a user. We divide the data set into two halves, where the first half is used to train the payoff matrix and the second half is used for testing. Figure 9 shows the results using the proposed model and the results from the real-world data set, from which we can see they match well with each other. We also depict the variances



[FIG7] **Experimental results of the evolutionary population dynamics: (a) #googlewave, (b) #DavidArchuleta, (c) #NIN, and (d) #Tehran.**

```
2  8  we're not commenting on that story i'm afraid   2131865
   3  3  we're not commenting on that    489007
      2008-08-18 14:23:05  1  M  http://business.theage.com.au/business/bb-chief-set-to-walk-plank-20080818-3xp7.html
      2008-11-26 01:27:13  1  B  http://sfweekly.com/2008-11-26/news/buy-line
      2008-11-27 18:55:30  1  B  http://aconstantineblacklist.blogspot.com/2008/11/re-researcher-matt-janovic.html
   5  2  we're not commenting on that story      2131864
      2008-12-08 14:50:18  3  B  http://videogaming247.com/2008/12/08/home-in-10-days-were-not-commenting-on-that-story-says-scee
      2008-12-08 19:35:31  2  B  http://jplaystation.com/2008/12/08/home-in-10-days-were-not-commenting-on-that-story-says-scee
```

[FIG8] An example of a MemeTracker phrase-cluster data set [20].

of the estimated results in Figure 9, which shows that the simulated results are always in the variance interval of the corresponding estimated results. Figure 9 also reveals the cohesiveness of different groups. We can see that the sites in Group 5 behave cohesively or share major common interests, while the sites in Group 1 share relatively few common interests. This is particularly interesting in advertisement or advocation scenarios where certain cohesive focus groups need to be mined to target high return value.

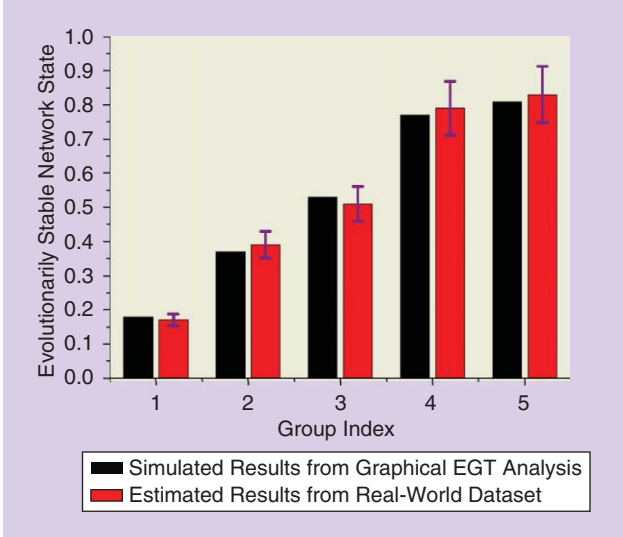
SEQUENTIAL USER BEHAVIOR: THE CHINESE-RESTAURANT-GAME FRAMEWORK

Another distinguishing feature of social systems is that users often participate sequentially in their own time and space. For example, users sequentially Q&A sites like Yahoo! Answers and Stack Overflow and decide whether to provide an answer, to vote on an existing answer, or not to participate. Other examples include online reviews, where customers write reviews for the product they purchase, and social news sites where online users post and promote stories under various categories.

The existence of network externality [5] in a social group dictates that users' actions and decisions influence each other. The network externality can be either positive or negative. When it is positive, users will have higher utilities when making the same decisions. On the contrary, when negative, users tend to make different decisions from others to achieve higher utilities. To achieve better performance, users should take into account the effect of network externality when making decisions.

On the other hand, users' decisions are also influenced by their knowledge of the system. In general, a user's knowledge of the system may be very limited because of the uncertainty in observations. This limitation reduces the accuracy of the user's decision and, thus, the overall system performance. The phenomenon of limited knowledge can be overcome through learning [27]–[30]. Users can learn from their previous experiences through machine-learning techniques and/or from other users' decisions and observations through social learning. All such information can help users to construct a belief, which can be probabilistic, on the unknown system states. In most cases, the accuracy of users' decisions can be greatly enhanced by taking the belief into account.

Therefore, to achieve the best utilities, users need to consider the effects of both learning and network externality when making decisions. While there are some existing works on combining positive network externality with learning [31]–[33], few works have been done on combining negative network externality with learning in the literature, mainly due to the difficulty of the problem, where a user has to consider the previous users' decisions and predict those of the subsequent users' decisions. Furthermore, the information leaked by a user's decision may eventually impair the utility the user can obtain. However, in practice, negative network externality commonly exists in social systems where users share and/or compete for resources and contents. To address this issue, we have developed a joint learning-decision-making framework, called the *Chinese-restaurant game* [17], [18], to study users' sequential learning and decision-making behavior in social systems.



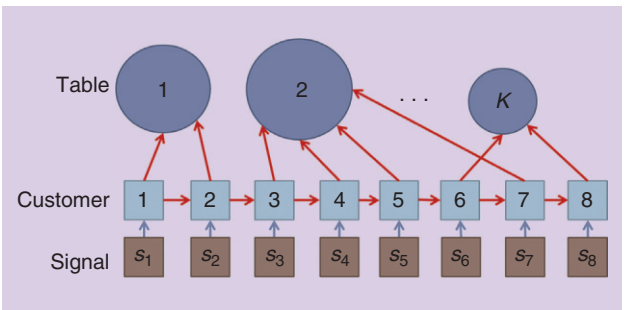
[FIG9] Experimental results of the evolutionarily stable strategy.

CHINESE-RESTAURANT-GAME FRAMEWORK

The well-known Chinese-restaurant process has been used in various fields, including machine learning, speech recognition, text modeling, and object detection in images and biological data clustering [34]. It offers an ideal structure to jointly formulate the decision-making problems with negative network externality. The Chinese-restaurant process is a nonparametric learning method for an unbounded number of objects in machine learning. In a Chinese-restaurant process, a restaurant has an infinite number of tables and customers arrive the restaurant sequentially. When a customer enters, he/she either joins one of the existing tables or requests a new table with a predetermined probability. However, there is not yet any notion of strategic decision making in the Chinese-restaurant process.

By introducing the strategic behavior into the nonstrategic Chinese-restaurant process, we proposed a new framework, the Chinese-restaurant game [17], [18], to study the learning and decision-making problem with negative network externality. To illustrate the framework, let us start with a Chinese restaurant that has a fixed number of tables, and customers sequentially come in requesting seats at these tables. Each customer may request a table. Since tables are available to all customers, there may be multiple customers requesting to sit at the same table, which, thereafter, incurs the negative network externality. We can imagine that the more personal space a customer has, the more comfortable the dining experience. Moreover, when the table sizes are unknown to the customers (before arriving at the restaurant), each of them may resort to some signals (e.g., through advertisements or discussions with previous customers) about the table sizes. By observing previous actions or signals, a user can exercise a learning process to make up the shortcoming of limited knowledge. With the proposed Chinese-restaurant game, we are able to develop an analytical framework involving the learning and decision making with negative network externality.

As shown in Figure 10, in the Chinese-restaurant game, there is a Chinese restaurant with K tables numbered $1, 2, ..., K$ and N customers labeled $1, 2, ..., N$. The table sizes are determined by the restaurant state $\theta \in \Theta$ and the table size functions $\{R_1(\theta), R_2(\theta), ..., R_K(\theta)\}$. When customer i arrives, he or she receives a signal s_i about the state θ and makes a decision about which table to choose such that he or she can maximize his or her utility, based on what was observed and the prediction of future customers' decisions. The prior distribution of the state



[FIG10] A system model of the Chinese-restaurant game.

information is assumed to be known by all customers. The signal is generated from a predefined distribution. Since there are uncertainties on the table sizes, customers who arrive first may not choose the right tables, and, consequently, their utilities may be lower. On the other hand, customers who arrive later may eventually have better chances to get the better tables since they can collect more information to make the right decisions. In other words, when signals are not perfect, learning can help to result in higher utilities for customers choosing later. Therefore, there is a tradeoff between more choices when playing first and more accurate signals when playing later. To study this tradeoff, some questions need to be answered: How can customers learn from their own signals and the information revealed by other customers? How can customers predict the decisions of future customers? And what are the best strategies for the customers?

To study how customers learn from the revealed information from others and their own signals, we first introduce the concept of belief to describe customers' uncertainty about the system state. One customer's belief on the system state is the conditional probability of the system state, given all the information observed by the customer, as follows:

$$\mathbf{g}_i = \{g_{i,l} \,|\, g_{i,l} = P(\theta = l \,|\, \mathbf{h}_i, s_i, \mathbf{g}_0), \forall l \in \Theta\}, \tag{6}$$

where $\mathbf{h}_i = \{s_1, s_2, ..., s_{i-1}\}$ is the signals observed by customer i, and $\mathbf{g}_0 = \{g_{0,l} \,|\, g_{0,l} = P(\theta = l), \forall l \in \Theta\}$ is the prior distribution.

With Bayesian learning [29], rational customers use Bayes' rule to find the optimal estimate about the system state and update their belief on the system state as follows:

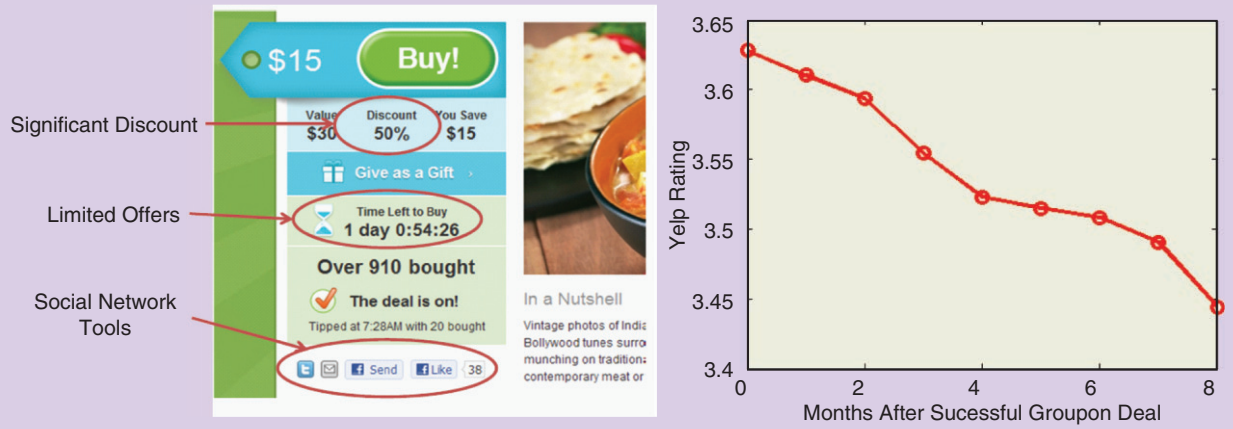
$$g_{i,l} = \frac{g_{0,l} P(\mathbf{h}_i, s_i \,|\, \theta = l)}{\sum_{l' \in \Theta} g_{0,l'} P(\mathbf{h}_i, s_i \,|\, \theta = l')}. \tag{7}$$

Because of their rationality and selfish nature, customers will choose their strategies to maximize their own utilities. In such a case, considering the incomplete information about the future customers, the best response of a customer is to maximize his or her expected utility based on all the observed information as follows:

$$\mathrm{BE}_i(\mathbf{n}_i, s_i, \mathbf{h}_i) = \arg\max_j E[U_i(R_j(\theta), n_j^*) \,|\, \mathbf{n}_i, s_i, \mathbf{h}_i, x_i = j], \tag{8}$$

where n_j^* is the final number of customers choosing table j; $U_i(R_j(\theta), n_j^*)$ is the utility of customer i choosing table j; $n_i = \{\mathbf{n}_{i1}, n_{i2}, ..., n_{iK}\}$ is the grouping observed by customer i, with n_{ik} being the number of customers choosing table k before customer i; and x_i is the action of customer i.

Note that the best response is determined by the final grouping, which depends on the subsequent customers' decisions. Since the decisions of subsequent customers are unknown to a customer when the customer is making the decision, a closed-form solution to the best response function is generally impossible and impractical. To find the best response for each customer, a recursive method based on backward induction is designed [18]. The key idea is to use the next customer's best response $\mathrm{BE}_{i+1}(\mathbf{n}_{i+1}, s_{i+1}, \mathbf{h}_{i+1})$ to derive the current customer's best response $\mathrm{BE}_i(\mathbf{n}_i, s_i, \mathbf{h}_i)$.



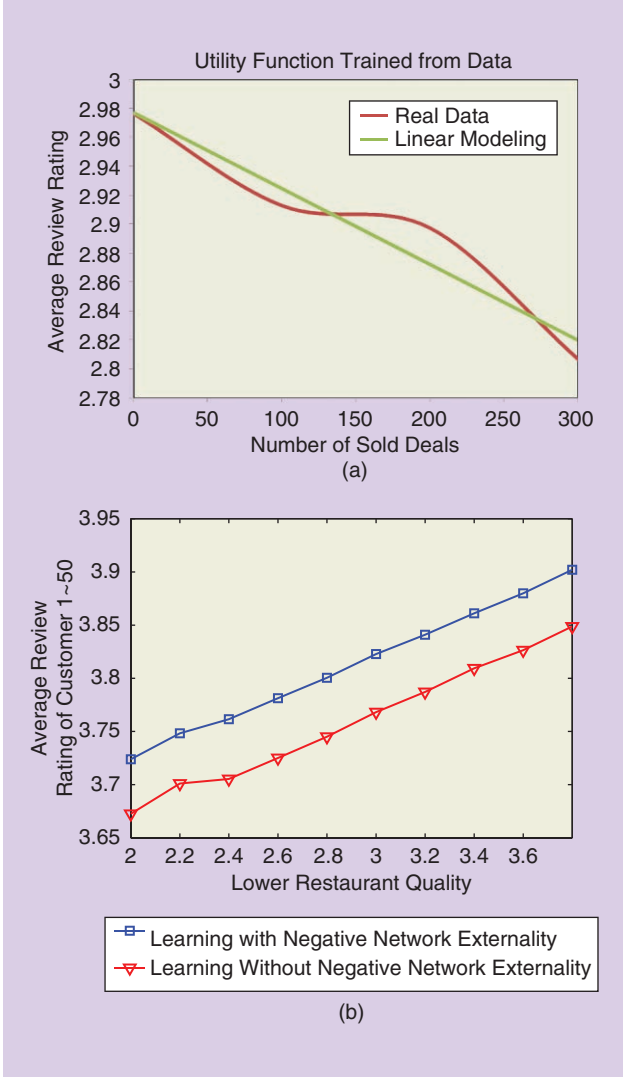
[FIG11] **Yelp star rating declines after a successful Groupon deal.**

With the best-response function, a customer's optimal decision is purely determined by the received signal given the grouping, i.e., the number of customers at each table and the information revealed by other customers. Therefore, we can partition the signal space into subspaces where within each subspace the customer will choose a specific table. By integrating the signal over each subspace, we can derive a recursive form of the probability mass function for the final grouping, i.e., the final number of customers at each table. With the recursive form of the final grouping, the expected utility of each customer can be computed, and the best response of all customers using backward induction can be derived.

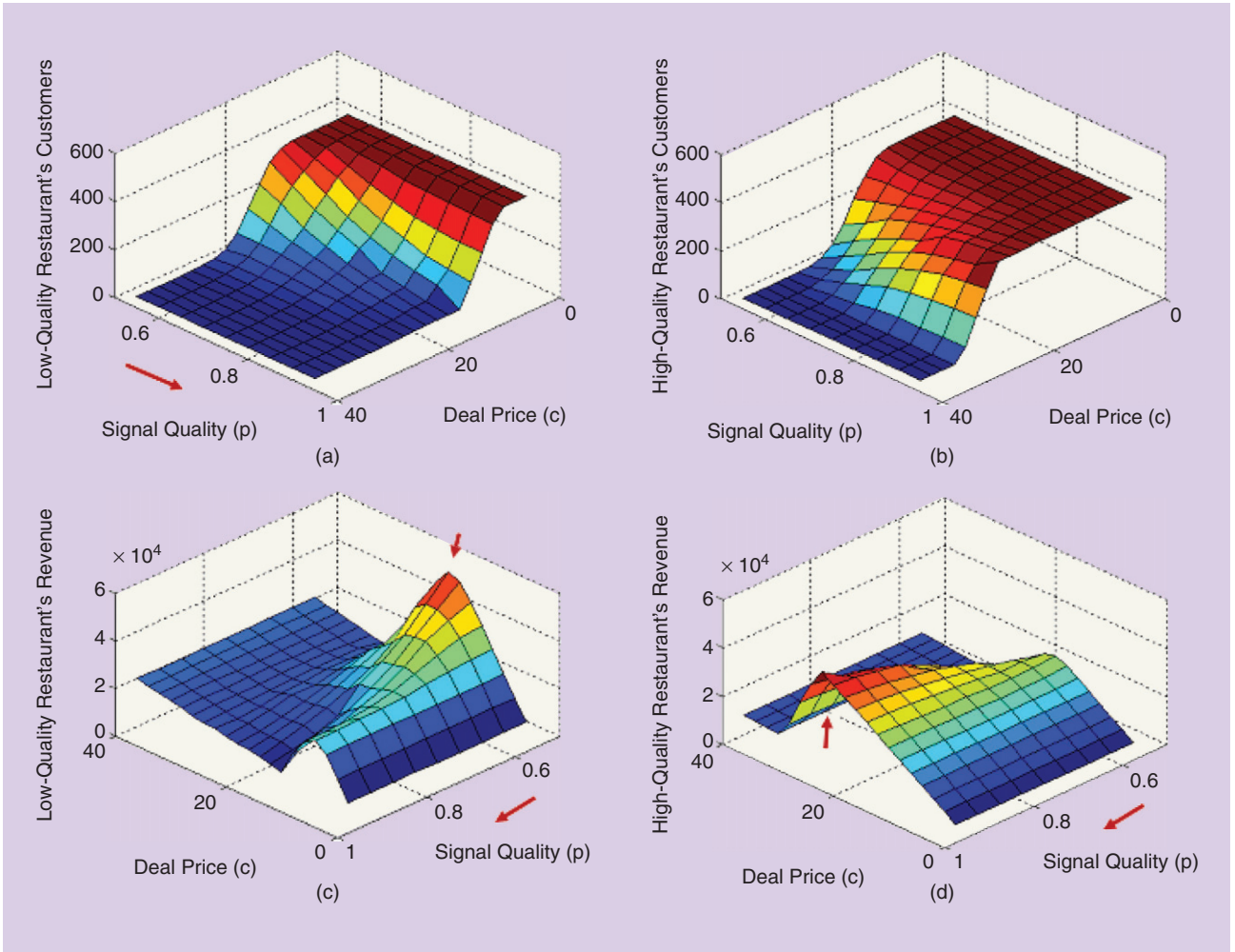
From previous discussions, we can see that the learning and decision making in the Chinese-restaurant-game framework are interweaved. On one hand, customers learn the system state from the information revealed by previous customers for better decision making. On the other hand, the decisions and information revealed by the customers will affect subsequent customers' learning and decision-making processes. Moreover, before any decision making, the utility function in (8) needs to be learned from the real data.

EXPERIMENTS WITH REAL DATA FROM SOCIAL SYSTEMS

We use deal selection on Groupon as an example to illustrate the Chinese-restaurant-game framework. Many have the experiences that some deals on Groupon look good but eventually turn out to have poor quality because of the overwhelming number of customers showing up at the same time, i.e., the negative network externality is at work. By collecting the data on Groupon and Yelp around the Washington, D.C., area for eight months, we indeed observe the decline of Yelp review ratings after some successful Groupon deals, as depicted in Figure 11. One can see a nonlinear decline function in review ratings. Let us use the real Yelp rating data to train the utility function of customers by approximating it as a linear model, as shown in Figure 12(a). Then, based on (8), we evaluate the



[FIG12] **(a) Utility function modeling using real data from Groupon and Yelp and (b) the performance comparison of our method with the learning method without negative network externality.**



[FIG13] A new restaurant's strategy: the number of customers choosing the restaurant with (a) low quality and (b) high quality and the result of the revenue of the (a) low-quality and (b) high-quality restaurants.

average utility, which is the average review rating customers can obtain. A comparison was made between the decision-learning method, denoted as *learning with negative network externality,* and that which does not consider negative network externality, denoted as *learning without negative network externality*. Note that the former considers the interplay between the learning and decision making, while the latter only considers the learning of system state but totally ignores the influence among customers' decision making. The results are shown in Figure 12(b). One can see that by combing learning with negative network externality, the proposed method can achieve much better utility for customers.

We further study the best pricing and promotion strategy of a new restaurant under the Chinese-restaurant-game framework. Let us consider two restaurants. One is always of high quality, and the other is a new restaurant, which could be of low or high quality. The same utility function trained from the real data in the above experiment is used to infer the strategy. The results are shown in Figure 13. One can see that if the new restaurant is of low quality, then the number of customers choosing the new restaurant decreases as signal quality increases and vice versa. One can also see that the optimal deal price of the high-quality restaurant is higher than that of the low-quality restaurant. Therefore, the high-quality restaurant should make every effort to increase the signal quality, while the low-quality restaurant should hide the quality information and use a low deal price to attract customers to increase the revenue. This offers a vivid example of using data to learn and come out with an optimal strategy.

EXTENSION TO THE CHINESE-RESTAURANT-GAME FAMILY

We have discussed the Chinese-restaurant game under a fixed population setting, i.e., there are a finite number of customers choosing the tables sequentially. However, in some applications, customers may arrive and leave the restaurant at any time, which results in the dynamic population setting. Examples include cloud storage service selection, deal selection on Groupon, and Wi-Fi access point selection in a conference hall [35]. In such a case, the utilities of customers will change from time to time because of the dynamic number of customers on each

table. To tackle this challenge, we have extended the Chinese-restaurant game to the dynamic population setting [36], [37], where we consider the scenario that customers may arrive and leave the restaurant with, e.g., a Poisson process. With such a dynamic population setting, each newly arriving customer not only learns the system state according to the information received and revealed by former customers, but also predicts the future customers' decisions to maximize the utility.

The Chinese-restaurant game is proposed by introducing strategic decision-making into the Chinese-restaurant process, where each customer can choose one table to maximize the utility. However, in some applications, users may want to simultaneously choose multiple resources. For example, mobile terminals may access multiple channels, cloud users may have multiple cloud storage services, and students may take multiple online courses. To further generalize the setting, we have introduced strategic decision making into another well-known random process, the Indian-buffet process [38], and develop a new framework, called the *Indian-buffet game*, to study the learning and decision-making problem with negative network externality under the scenario that customers can have multiple choices [39]. In the Indian-buffet game framework, we also consider multislot interactions where customers can interact and make decisions repeatedly, and only partial information is revealed, i.e., customers only reveal beliefs instead of full signals to others. We use the non-Bayesian social learning to learn from each other to improve the knowledge of the system and thus make better decisions. Similar extension can be applied to multiarmed-bandit problems by introducing decision-making processing into their formulation.

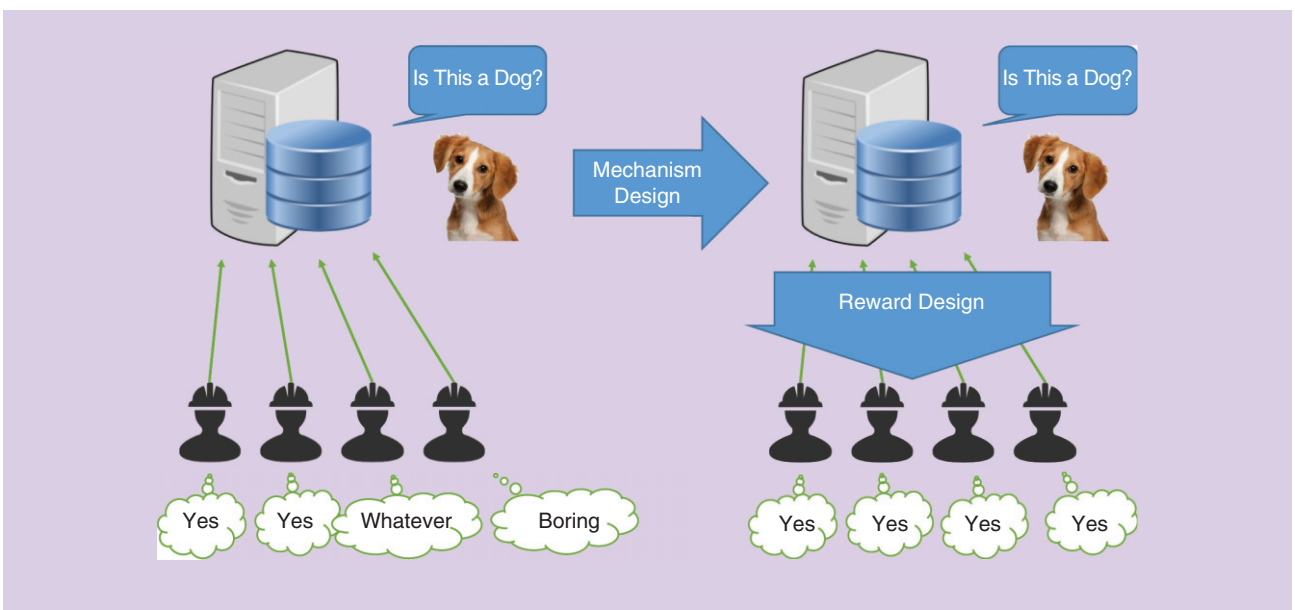
ONE KEY FACTOR FOR THE SUCCESS OF SUPERVISED AND SEMISUPERVISED LEARNING IS THE LARGE-SCALE LABELED DATA SET.

MECHANISM DESIGN IN DECISION LEARNING

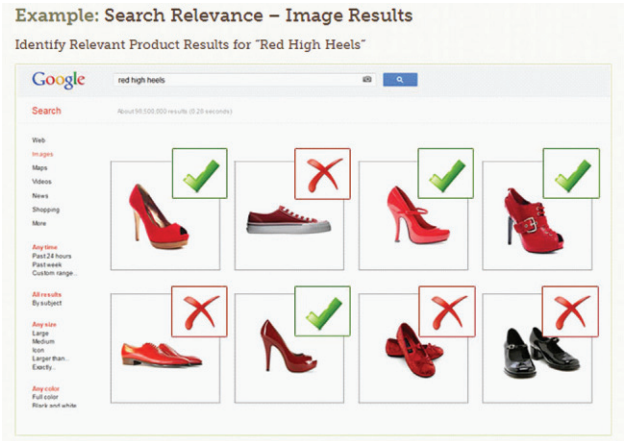
In this section, we will address decision learning from the system point of view. Can we design mechanisms for users to learn the desired behavior and thus achieve the goals of the system designer, as shown in Figure 14? In the following, we will use microtask crowdsourcing to illustrate how to design a mechanism to obtain high-quality data for data analytics.

One key factor for the success of supervised and semisupervised learning is a large-scale labeled data set [1], [2]. In general, a larger-scale data set will lead to a more accurate model and, thus, better performance. However, large-scale annotation is very expensive, which often becomes one of the bottlenecks of supervised and semisupervised learning. To address this challenge, microtask crowdsourcing, with access to a large and relatively cheap online labor pool, is a promising method since it can generate large volume of labeled data in a short time at a much lower price compared with traditional in-house solutions. An example of microtask crowdsourcing is illustrated in Figure 15.

On the other hand, because of the lack of proper incentives, microtask crowdsourcing suffers from quality issues. Since workers are paid a fixed amount of money per task they complete, it is profitable for them to provide random or bad-quality solutions to increase the number of submissions within a certain amount of time or with the least effort. It has been reported that most workers on Mturk, a leading marketplace for microtask crowdsourcing, do not contribute high-quality work [40]. To address this issue, a common machine-learning solution is to either add a data-curation



[FIG14] An illustration of mechanism design. Users may behave randomly without a well-designed mechanism, which can greatly degrade the system performance. Through mechanism design, the system designer can guide users to learn the desired behavior and thus achieve the expected goals at the system level.

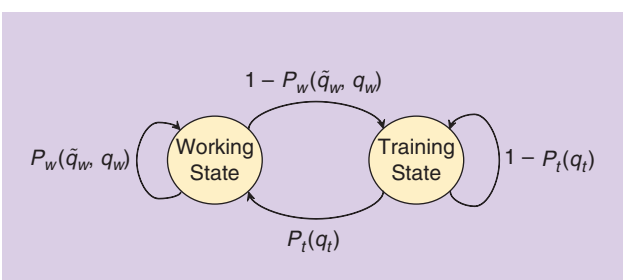


[FIG15] An example of microtask crowdsourcing.

phase to filter out low-quality data or to modify the learning algorithm to accept noisy labels [41]–[45].

In contrast to existing machine-learning solutions, we tackle such a problem by incentivizing the high-quality data in the first place [19], e.g., from the workers. This problem is challenging because of the inherent conflict between incentivizing high-quality solutions from workers and maintaining the low-cost advantage of microtask crowdsourcing for requesters. On one hand, requesters typically have a very low budget for each task in microtask crowdsourcing. On the other hand, the implementation of incentive mechanisms is costly as the operation of verifying the quality of submitted solutions is expensive [41]. Such a conflict makes it extremely challenging to design proper incentives for microtask crowdsourcing and motivates us to ask the following question: What incentive mechanisms should requesters employ to collect high-quality solutions in a cost-effective way? In a general sense, the core problem is how to design mechanism for obtaining good data.

To answer the question, we first study and model the behavior of workers. Specifically, let us consider a model with strategic workers, where the action of a worker is the quality of the solution $q \in [0, 1]$ and the primary objective of a worker is to maximize his or her own utility, defined as the reward he or she will receive minus the cost of producing solutions of a certain quality $c(q)$. Based on this model, we analyze two basic mechanisms that are widely adopted in existing microtask crowdsourcing applications: reward consensus mechanism M_c and reward accuracy mechanism M_a [19].



[FIG16] The state transition diagram of the incentive mechanism M_t.

REWARD CONSENSUS MECHANISM M_c

With this mechanism, a task is assigned to multiple workers. Only the same answer that is submitted by the majority of workers will be chosen as the correct solution and the workers whose solution agrees with the correct one will receive a positive reward. Through analyzing this mechanism, we find that there exists a minimum mechanism cost per task to obtain high-quality solutions [19]

$$C^*_{M_c} = 3c'(1), \tag{9}$$

where $c'(1)$ is the first-order derivative of the cost function $c(q)$ evaluating at the desired solution quality $q = 1$.

REWARD ACCURACY MECHANISM M_a

This mechanism assigns each task only to one worker. The requester evaluates with a certain probability the quality of submitted solutions directly, where each validation incurs a constant cost d. The validation can be erroneous with a probability of ε. The workers whose solutions are not evaluated or evaluated and confirmed as correct will receive a positive reward. Through analyzing this mechanism, we again find that there exists a minimum mechanism cost per task to obtain high-quality solutions [19]

$$C^*_{M_a} = \begin{cases} 2\sqrt{\dfrac{c'(1)d}{1-2\varepsilon}} - \varepsilon\dfrac{c'(1)}{1-2\varepsilon}, & \text{if } d \geq \dfrac{c'(1)}{1-2\varepsilon}; \\ \dfrac{c'(1)(1-\varepsilon)}{1-2\varepsilon} + d, & \text{otherwise.} \end{cases} \tag{10}$$

From (9) and (10), we can see that to obtain high-quality solutions using the two basic mechanisms (M_c and M_a), the unit cost incurred by requesters per task is subject to a lower bound constraint, which is beyond the control of requesters. In case that the budget of the requester is lower than the minimum cost constraint, it becomes impossible for the requester to achieve the desired quality solutions with these two basic mechanisms. In other words, neither of these two basic mechanisms is cost effective.

INCENTIVE MECHANISM VIA TRAINING M_t

To tackle this challenge, we design a cost-effective mechanism by employing quality-aware worker training as a tool to stimulate workers to provide high-quality solutions [19]. Different from current microtask crowdsourcing applications where training tasks are usually assigned to workers at the very beginning and are irrelevant to the quality of submitted solutions, we use the training tasks in a more effective way by assigning them to workers when they perform poorly. That is, when a worker performs poorly, he/she will be required to enter a training session without a reward to regain accreditation to be able to go back to perform in the regular session with a reward.

With the introduction of quality-aware training tasks, there will be two system states in our proposed mechanism: the working state and the training state. The working state is for production purposes, where workers work on standard tasks in return for rewards, while the training state is an auxiliary state, where workers do a set of training tasks to gain qualifications for the

working state. The state transition diagram is shown in Figure 16, where $P_w(\tilde{q}_w, q_w)$ represents the probability of a solution with quality q_w being accepted in the working state when other submitted solutions from the working state are of quality $\tilde{q}_w$, and $P_t(q_t)$ is the probability of a worker who produces solutions of quality q_t at the training state being allowed to enter the working state next time.

A COMMON MACHINE-LEARNING SOLUTION IS TO EITHER ADD A DATA-CURATION PHASE TO FILTER OUT LOW-QUALITY DATA OR TO MODIFY THE LEARNING ALGORITHM TO ACCEPT NOISY LABELS.

From Figure 16, we can see that the current action of a worker will affect the future system state of the worker. In other words, the quality of a worker's solution to one task will affect not only the worker's immediate utility but also his future utility due to the possible change of the system state. Such a dependence provides requesters with an extra degree of freedom in designing incentive mechanisms and thus enables them to collect high-quality solutions while still having control over their incurred costs.

To find the optimal action, each worker must solve a Markov decision process (MDP), according to the state transition diagram shown in Figure 16; the MDP faced by each worker also depends on other workers' actions. In essence, this is a challenging game-theoretic MDP problem [19]. Through analyzing the incentive mechanism M_t, we find that, as long as the number of training tasks is large enough, there always exists a desirable equilibrium where workers submit high-quality solutions at the working state. In other words, given any parameters in the working state, one can always guarantee the existence of a desirable equilibrium through the design of the training state. When the desirable equilibrium is adopted by all workers by following a certain design procedure, the minimum mechanism cost is theoretically proved to be zero [19], i.e.,

$$C^*_{M_t} = 0, \tag{11}$$

which means that one can collect high-quality solutions with an arbitrarily low cost. In other words, given any predetermined budget, the incentive mechanism M_t enables the requester to collect high-quality solutions while still staying within the budget.

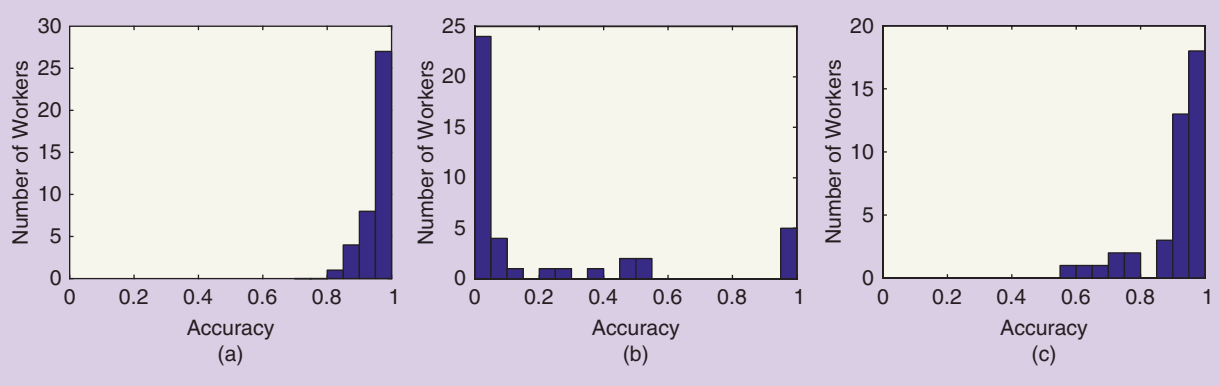
Notice that one can easily achieve better learning purposes with the high-quality data collected by the incentive mechanism M_t. Therefore, through modeling and analyzing users' strategic decision-making processes, one can design mechanisms from the system point of view to steer users' strategic behaviors to obtain better-quality data for better learning.

REAL BEHAVIORAL EXPERIMENTS

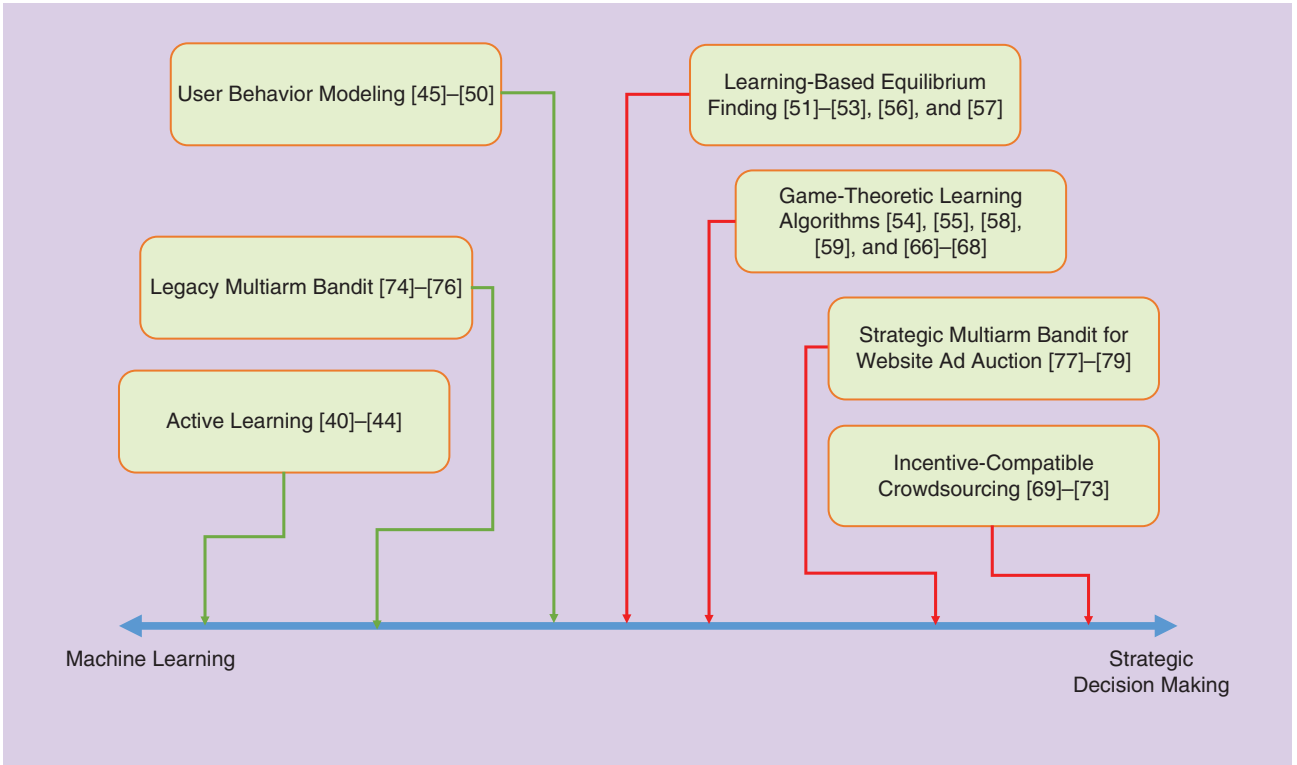
A set of behavioral experiments are conducted to test the incentive mechanism M_t in practice. We evaluate the performance of participants on a set of simple computational tasks under different incentive mechanisms. We compare the incentive mechanism M_t with the reward accuracy mechanism M_a, where the quality of submitted solutions is evaluated with a certain probability.

There are 41 participants in our experiments, most of whom are engineering graduate students. We use the accuracy of each participant as an indicator to the effectiveness of the incentive mechanisms, and the results are shown in Figure 17, where (a) shows the results of the reward accuracy mechanism M_a with sampling probability 1, i.e., every submitted solution is evaluated; (b) shows the results of the reward accuracy mechanism M_a with sampling probability 0.3, i.e., 30% of the submitted solutions are evaluated; and (c) shows the results using the incentive mechanism M_t with the sampling probability as that in (b).

As shown in Figure 17(a), with the highest sampling probability, most participants respond positively by submitting solutions with very high qualities. There is only one participant who had relatively low accuracy compared with others in that he/she



[FIG17] Histogram of accuracy: (a) the results using the reward accuracy mechanism M_a with sampling probability 1; (b) the results using the reward accuracy mechanism M_a with sampling probability 0.3; and (c) the results using the incentive mechanism M_t with sampling probability 0.3.



[FIG18] **A summary of related works.**

was playing the strategy of avoiding difficult tasks according to our exit survey. When a much lower sampling probability of 0.3 is used, it becomes profitable to increase the number of submissions by submitting lower-quality solutions, as most errors will simply not be detected. This explains why the majority of participants had very low accuracies, as shown in Figure 17(b). Notably, a few workers, five of 41, still exhibited very high accuracies in this case. Our exit survey suggests that their behaviors are influenced by a sense of work ethics, which prevents them from playing strategically to exploit the mechanism vulnerability. With the incentive mechanism M_t, as the introduction of training tasks makes it more costly to submit wrong solutions, participants need to reevaluate their strategies to achieve a good tradeoff between accuracy and the number of submitted tasks. From Figure 17(c), we can see that the accuracy of the participants with the incentive mechanism M_t has a very similar distribution to that of the group using the reward accuracy mechanism M_a with the highest sampling probability. Therefore, through the use of quality-aware worker training, the incentive mechanism M_t can greatly improve the effectiveness of the basic reward accuracy mechanism M_a with a low sampling probability to a level that is comparable to the one that has the highest sampling probability.

RELATED WORKS

Although not referred to specifically as decision learning, there has been a growing body of literature in recent years on the intersection of learning and strategic decision making, as summarized in Figure 18. One class of related works is learning to understand how human beings make strategic decisions from real data. For example, classical machine-learning techniques are used in [46] to predict how people make and respond to offers during negotiations and how they reveal information and their response to potential revelation actions by others. Their results showed that the strategies derived from machine-learning algorithms, even when not optimal, can beat real human beings [46]. The year-long study of empirical data shows that an experienced human being in a repeated game will be more cooperative but turn the tables more definitely when he or she is betrayed by the opponent [47]. Additionally, the study in [48] shows that human beings have very limited memory space and computation capability, which limits the optimality of their decisions. It has also been shown in [49] that a dynamic belief model, by ignoring the older signals in constructing the belief, works best in predicting human decisions. Through empirically analyzing the purchase history on Taobao, a large-scale online shopping social network, Guo et al. revealed that a real human values purchase experiences shared by his or her friend and would be willing to pay a higher price for trustworthy vendors [50]. Nevertheless, in such a complicated system, it is still difficult to predict the purchase decisions with more than 50% accuracy using traditional machine-learning algorithms [50]. In [51], how users make decisions on social computing systems is learned from real data and used to guide the design of mechanisms for the systems. The aforementioned works establish a solid foundation of decision learning. These

works reveal the real nature of people in making decisions and suggest that people usually behave in a rational way with several limitations such as short memory. These evidences become a strong support to the rationality assumption in decision learning.

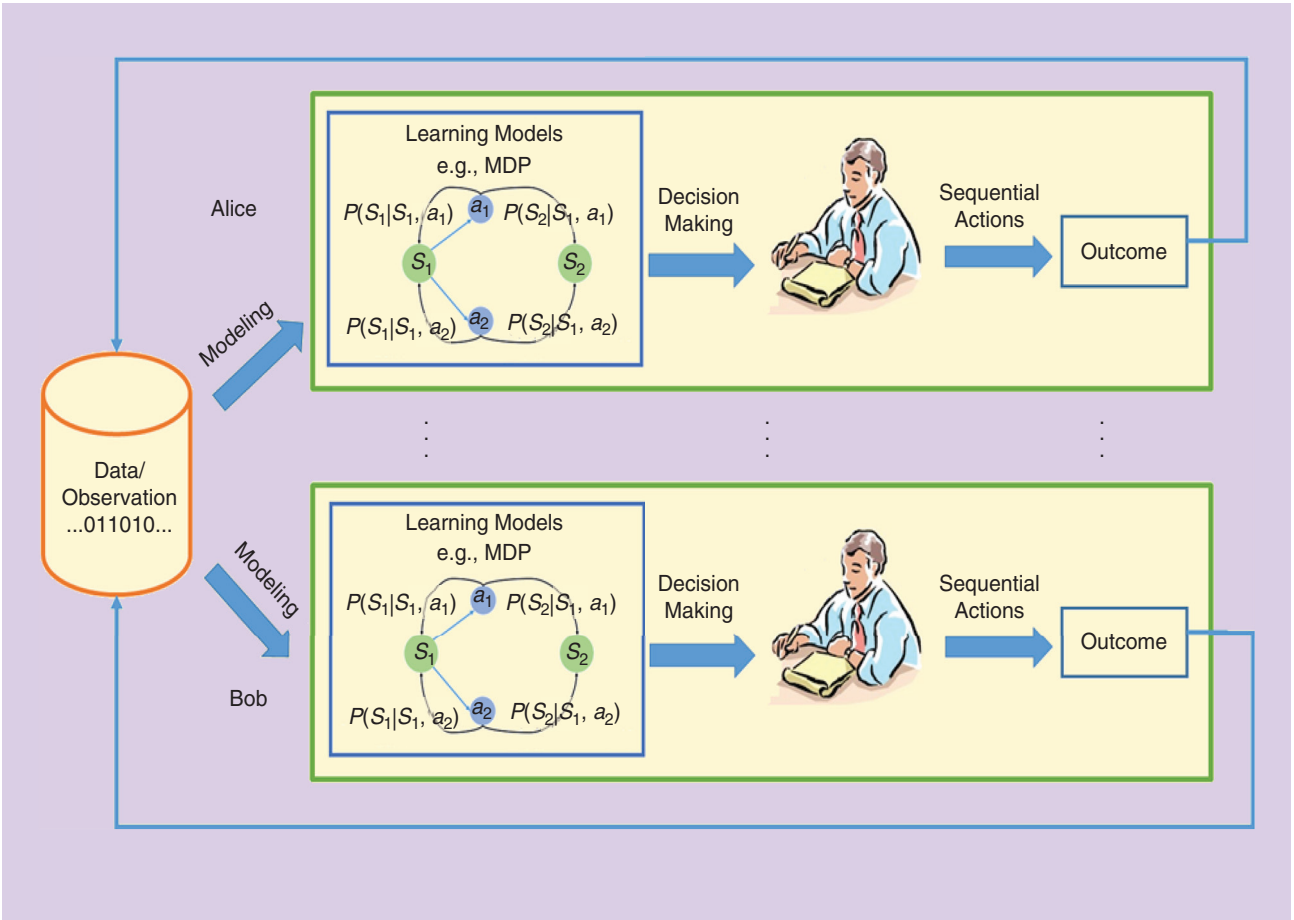
DECISION LEARNING IS AN EMERGING RESEARCH AREA TO BRIDGE LEARNING FROM LARGE VOLUMES OF DATA WITH STRATEGIC DECISION MAKING THAT UNDERSTANDS AND MODELS USER BEHAVIORS.

Another class of related works is finding equilibrium through learning. Finding Nash equilibrium is critical yet challenging in most game models since the difficulty has been shown to be polynomial parity arguments on directed graphs in general settings or even nondeterministic polynomial time-complete in specific problems [52]. Given that a general and exact solution is intractable, it is a natural choice to design proper learning algorithms to find the solutions. No-regret learning, for instance, has been shown to be a practical candidate. It has been applied in extensive-form games to reduce the number of subgame trees to explore [53]. The sufficient conditions for such type of learning algorithms to converge in the selfish routing problem [54] are also theoretically studied. Reinforcement learning is another candidate since its action-reward structure naturally forms the best response dynamic in game theory. Since traditional Q-learning may fail to converge if directly applied in a game, especially when the Nash equilibrium is not unique, maxmin Q-learning is proposed in [55] to find the Nash equilibrium in a two-player zero-sum game. The objective of maxmin Q-learning is modified from pure reward maximization into a max–min problem with an opponent's actions in mind. Nash Q-learning, a more general Q-learning algorithm, is proposed in [56] to handle multiplayer game with non-zero sum. The objective of Nash Q-learning is replaced with equilibrium conditions defined in game theory. The experiment results show that Nash Q-learning can help identify better Nash equilibrium than the traditional Q-learning algorithm. Learning has also been used to reduce the complexity in finding the subgame-perfect Nash equilibrium in a sequential game [57], which is PSPACE-hard in general. In [58], MDP and Monte Carlo simulation is used to reduce the complexity in identifying the optimal bidding strategy in sequential auctions. Finding the equilibrium point is also a critical part in decision learning since the performance of the decision-learning method relies on the equilibrium point. The aforementioned works contribute useful methods to help find the equilibrium point efficiently.

There have also been some related works that formulate the training problem in machine learning as a game. For instance, it was shown in [59] that a class of online learning algorithms can be modeled as a drifting game with both the trainer and the system as players. The learning algorithm in such a formation becomes the best response of the trainer to the system's reply to each training problem. Another application is maintaining fairness in a multiagent sequential decision problem. Given that the objective of the system is max–min fairness, one may model the learning model as a two-player game, where the first player aims at maximizing the utility of the target agent who is chosen by second player, while the second player chooses the agent with lowest utility as the target agent [60]. Decision learning not only inherits the spirit of the previously mentioned works but also makes one critical extension to address the competition effect in real world. That is, in decision learning, we introduce the game-theoretic competition effect into not only the system-versus-user relation but also user-versus-user relations.

Active learning is another related field [61]. Through actively choosing which data to learn from, active learning algorithms have the potential to greatly reduce the amount of labeling effort in machine-learning algorithms. Active learning with an explicit labeling cost has been widely studied [62]–[65]. It is known that active learning algorithms degrade quickly as the noise rate of labels increases [66]. To address the quality issue in labeled data collection, a variety of approaches have been proposed to filter low-quality labels and to increase the robustness of machine-learning algorithms [41]–[45]. In [67], a game-theoretic dynamics was proposed to approximately denoise the data to exploit the power of active learning. Incentive mechanisms have also been used to improve the quality of collected data. In [11], [68], and [69], all-pay auctions are applied to incentify high-quality user contributions. In [70], Shaw et al. conducted an experiment to compare the effectiveness of a collection of social and financial incentive mechanisms. A reputation-based incentive mechanism was proposed and analyzed for microtask crowdsourcing in [71]. In [72] and [73], Singer and Mittal proposed an online mechanism for microtask crowdsourcing where tasks are dynamically priced and allocated to workers based on their bids. In [74], Singla and Krause proposed a posted price scheme where workers are offered a take-it-or-leave-it price offer. The earlier works are closely related to decision learning since they also focus on the rationality and incentives of users in the learning process. Crowdsourcing is one of the main applications of decision learning, of which we also gave a detailed illustration previously.

A nonstrategic but related problem is the multiarmed-bandit problem [75]. In the multiarmed-bandit problem, a bandit with multiple arms is provided to a gambler. The gambler may have different levels of rewards by playing different arms each time. Thus, the gambler may try and learn in each play to maximize his collected rewards. Liu and Zhao extended this model by considering multiple agents and including the network externality in [76]. They studied how agents learn the expected payoff and the other agent's choice by estimating the regrets after choosing different arms based on his or her current belief. A multiarmed-bandit problem with costs in observations is discussed in [77]. Nevertheless, traditional studies in the multiarmed-bandit problem are generally nonstrategic. They assume agents will always follow the



[FIG19] In the coming big data tsunami, when a large volume of data is available, users can learn better models to improve their own decision making. On the other hand, their actions result in changes of the data pool, which consequently affects the models learned by the users. In summary, users' decisions and actions affect each other in an ever-changing fashion for user-generated data applications. Decision learning is an emerging research area to bridge learning from large volumes of data with strategic decision making that understands and models user behaviors.

learning rule designed by the system designer. Combining strategic thinking with the multiarmed-bandit problem has gained more attention recently mainly because of a popular and practical application: website ad auctions. The ad slot on a website is usually sold through auction. The value of the ad depends on two factors: the value of the product in the ad and the expected number of clicks on the ad. The former is known by the advertiser and can be collected through truthful auction such as a Vickery auction. Nevertheless, the expected number of clicks, or the click-through-rate (CTR) of the ad, is unknown to both the website owner and the advertiser. The no-regret algorithm in the multiarmed-bandit problem can be used to learn CTR while maintaining the truthfulness of the auction [78], [79]. In [80], it is shown that increasing number of explore stages will push the buyers to reveal their true valuation more, with fewer exploit stages for sellers to gain extra revenue from the learned valuation in return. The strategic multiarm-bandit problem is the one receiving the most attention in the transition from traditional machine learning to decision learning because of its structure and potentials in real world applications such as ad auctions. The great success also shows the potential of decision learning in real-world applications, especially those involving competitions.

CONCLUSIONS AND FINAL THOUGHTS

Decision learning is learning with strategic decision making that can analyze users' optimal behaviors from users' perspectives while designing optimal mechanisms from system designers' perspectives. In this article, we have used three social media examples to highlight the concepts of decision learning. Specifically, information diffusion over online social networks was used to illustrate how to learn users' utility functions from real data for understanding and modeling strategic decision making. Deal selection on Groupon with Yelp data was used to discuss how users learn from each other's interactions for better strategic decision making. Microtask crowdsourcing was used to discuss how to design mechanisms to steer users' strategic behaviors to obtain better-quality data for better learning. Besides the three examples discussed in this article, there can be other forms of joint learning with strategic decision making, including those discussed previously. In essence, in the coming big data tsunami, when a large

volume of data is available, users can learn better models to improve their own decision making, as depicted in Figure 19. On the other hand, their actions result in changes of the data pool, which consequently affects the models learned by the users.

In summary, users' decisions and actions affect each other in an ever-changing fashion for user-generated data applications. Decision learning is an emerging research area to bridge learning from large volumes of data with strategic decision making that understands and models user behaviors.

AUTHORS

Yan Chen (yan@umd.edu) received the Ph.D. degree from the University of Maryland, College Park, in 2011. A founding member and a principal technologist, he joined Origin Wireless, Inc., in 2013. He is currently a professor at the University of Electronic Science and Technology of China. His research interests are in data science, game theory, signal processing, and wireless communications. He is the recipient of the Best Paper Award from IEEE GLOBECOM in 2013, the Future Faculty Fellowship and Distinguished Dissertation Fellowship Honorable Mention in 2010 and 2011, respectively, and the Chinese Government Award for outstanding students abroad in 2011.

Chunxiao Jiang (chx.jiang@gmail.com) received the B.S. degree from Beihang University in 2008 and the Ph.D. degree from Tsinghua University (THU), Beijing, in 2013. From 2011 to 2014, he was a visiting researcher with the Signals and Information Group at the University of Maryland, College Park, with Prof. K.J. Ray Liu. He currently holds a postdoctoral position with the Department of Electrical Engineering, THU. His research interests include wireless communication and networking and social networks. He was a recipient of the Best Paper Award from IEEE GLOBECOM in 2013, the Beijing Distinguished Graduate Student Award, the Chinese National Fellowship, and the Tsinghua Outstanding Distinguished Doctoral Dissertation award in 2013.

Chih-Yu Wang (cywang@citi.sinica.edu.tw) received the B.S. degree in electrical engineering from National Taiwan University in 2007. He received the Ph.D. degree from the Graduate Institute of Communication Engineering, National Taiwan University. He was a visiting student at the University of Maryland, College Park, in 2011. He is currently an assistant research fellow at the Research Center for Information Technology Innovation, Academia Sinica. His research interests include game theory, wireless communications, social networks, and data science.

Yang Gao (yanggao.umd@gmail.com) received the B.S. degree from Tsinghua University, Beijing, China, in 2009 and the Ph.D. degree from the University of Maryland, College Park, in 2014, both in electrical engineering. Currently, he is a research scientist with Facebook, Inc. His research interests are in strategic behavior analysis and incentive mechanism design for crowdsourcing and social computing systems. He is a recipient of the silver medal of the 21st National Chinese Physics Olympiad, the honor of Excellent Graduate of Tsinghua University in 2009, the University of Maryland Future Faculty Fellowship in 2012, and the IEEE GLOBECOM 2013 Best Paper Award.

K.J. Ray Liu (kjrliu@umd.edu) was named a Distinguished Scholar–Teacher of the University of Maryland, College Park, in 2007, where he is Christine Kim Eminent Professor of Information Technology. He was a recipient of the 2016 IEEE Leon K. Kirchmayer Technical Field Award on graduate teaching and mentoring, IEEE Signal Processing Society 2014 Society Award, and IEEE Signal Processing Society 2009 Technical Achievement Award. Recognized by Thomson Reuters as a highly cited researcher, he is a Fellow of the IEEE and AAAS. He is on the IEEE Board of Directors. He was the 2012–2013 president of the IEEE Signal Processing Society.

REFERENCES

[1] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[2] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, U.K.: MIT Press, 2012.

[3] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, U.K.: MIT Press, 1991.

[4] M. J. Osborne, *An Introduction to Game Theory*. Oxford, U.K.: Oxford Univ. Press, 2004.

[5] M. Katz and C. Shapiro, "Technology adoption in the presence of network externalities," *J. Political Econ.*, vol. 94, no. 4, pp.822–841, 1986.

[6] Y. Chen and K. J. R. Liu, "Understanding microeconomic behaviors in social networking: An engineering view," *IEEE Signal Processing Mag.*, vol. 29, no. 2, pp. 53–64, 2012.

[7] H. V. Zhao, W. Lin, and K. J. R. Liu, *Behavior Dynamics in Media-Sharing Social Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[8] M. Burke, C. Marlow, and T. Lento, "Feed me: Motivating newcomer contribution in social network sites," in *Proc. SIGCHI Conf. Human Factors in Computing Systems*, 2009.

[9] K. Ling, G. Beenen, X. Wang, K. Chang, D. Frankowski, P. Resnick, and R. Kraut, "Using social psychology to motivate contributions to online communities," *J. Comput. Mediate. Commun.*, vol. 10, no. 4, 2005.

[10] J. Yang, L. A. Adamic, and M. S. Ackerman, "Crowdsourcing and knowledge sharing: Strategic user behavior on task CN," in *Proc. 9th ACM Conf. Electronic Commerce (EC)*, 2008.

[11] D. DiPalantino and M. Vojnovic, "Crowdsourcing and all-pay auctions," in *Proc. 10th ACM Conf. Electronic Commerce (EC)*, 2009.

[12] Y. R. Tausczik and J. W. Pennebaker, "Participation in an online mathematics community: Differentiating motivations to add," in *Proc. ACM 2012 Conf. Computer Supported Cooperative Work (CSCW)*, 2012.

[13] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Steering user behavior with badges," in *Proc. 22nd Int. Conf. World Wide Web (WWW)*, 2013.

[14] B. Wang, Y. Wu, and K. J. R. Liu, "Game theory for cognitive radio networks: An overview," *Comput. Netw.*, vol. 54, no. 14, pp. 2537–2561, 2010.

[15] C. Jiang, Y. Chen, and K. J. R. Liu, "Evolutionary dynamics of information diffusion over social networks," *IEEE Trans. Signal Processing*, vol. 62, no. 17, pp. 4573–4586, Sept. 2014.

[16] C. Jiang, Y. Chen, and K. J. R. Liu, "Graphical evolutionary game for information diffusion over social networks," *IEEE J. Sel. Top. Signal Processing.*, vol. 8, no. 4, pp. 524–536, Aug. 2014.

[17] C. Y. Wang, Y. Chen, and K. J. R. Liu, "Chinese restaurant game," *IEEE Signal Processing. Lett.*, vol. 19, no. 12, pp. 898–901, 2012.

[18] C. Y. Wang, Y. Chen, and K. J. R. Liu, "Sequential Chinese restaurant game," *IEEE Trans. Signal Processing*, vol. 61, no. 3, pp. 571–584, 2013.

[19] Y. Gao, Y. Chen, and K. J. R. Liu, "On cost-effective incentive mechanisms in microtask crowdsourcing," *IEEE Trans. Computat. Intell. AI Games*, vol. 7, no. 1, pp. 3–15, 2015.

[20] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proc. Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2009, pp. 497–506.

[21] J. M. Smith, *Evolutionary and the Theory of Games*. Cambridge, U.K.: Cambridge Univ. Press, 1982.

[22] M. A. Nowak and K. Sigmund, "Evolutionary dynamics of biological games," *Science*, vol. 303, no. 5659, pp. 793–799, 2004.

[23] H. Ohtsuki, M. A. Nowak, and J. M. Pacheco, "Breaking the symmetry between interaction and replacement in evolutionary dynamics on graphs," *Phys. Rev. Lett.*, vol. 98, no. 10, p. 108106, 2007.

[24] P. Shakarian, P. Roos, and A. Johnson, "A review of evolutionary graph theory with applications to game theory," *Biosystems*, vol. 107, no. 2, pp. 66–80, 2012.

[25] H. Ohtsukia and M. A. Nowak, "The replicator equation on graphs," *J. Theor. Biol.*, vol. 243, no. 1, pp. 86–97, 7 Nov. 2006.

[26] J. Leskovec. Stanford large network data set collection. [Online]. Available: http://snap.stanford.edu/data

[27] V. Bala and S. Goyal, "Learning from neighbours," *Rev. Econ. Stud.*, vol. 65, no. 3, pp. 595–621, 1998.

[28] B. Golub and M. O. Jackson, "Naive learning in social networks and the wisdom of crowds," *Amer. Econom. J.: Microecon.*, vol. 2, no. 1, pp. 112–149, 2010.

[29] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *Rev. Econ. Stud.*, vol. 78, no. 4, pp. 1201–1236, 2011.

[30] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dyn. Games Appl.*, vol. 1, no. 1, pp. 3–49, 2011.

[31] J. Wit, "Social learning in a common interest voting game," *Games Econ. Behav.*, vol. 26, no. 1, pp. 131–156, 1999.

[32] M. Battaglini, "Sequential voting with abstention," *Games Econ. Behav.*, vol. 51, no. 2, pp. 445–463, 2005.

[33] A. Dasgupta, "Coordination and delay in global games," *J. Econ. Theory*, vol. 134, no. 1, pp. 195–225, 2007.

[34] J. Pitman, *Combinatorial Stochastic Processes*. Lecture Notes for St. Flour Summer School. New Springer-Verlag, 2002.

[35] Y. H. Yang, Y. Chen, C. X. Jiang, C. Y. Wang, and K. J. R. Liu, "Wireless access network selection game with negative network externality," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5048–5060, 2013.

[36] C. Jiang, Y. Chen, Y.-H. Yang, and K. J. R. Liu, "Dynamic chinese restaurant game: Theory and application to cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1960–1973, Apr. 2014.

[37] C. Y. Wang, Y. Chen, H. Y. Wei, and K. J. R. Liu, "Optimal pricing in stochastic scalable video coding multicasting system," in *Proc. IEEE INFOCOM Mini-Conf.*, Apr. 2013.

[38] T. L. Griffiths and Z. Ghahramani, "The Indian buffet process: An introduction and review," *J. Mach. Learn. Res.*, vol. 12, pp. 1185–1224, Feb. 2011.

[39] C. Jiang, Y. Chen, Y. Gao, and K. J. R. Liu, "Indian buffet game with negative network externality and non-Bayesian social learning," *IEEE Trans. Syst., Man Cyber.: Syst.*, vol. 45, no. 4, pp. 609–623, Apr. 2015.

[40] P. Wais, S. Lingamnei, D. Cook, J. Fennell, B. Goldenberg, D. Lubarov, D. Marin, and H. Simons, "Towards building a high-quality workforce with mechanical turk," in *NIPS Workshop on Computational Social Science and the Wisdom of Crowds*, 2010.

[41] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality management on Amazon mechanical turk," in *Proc. ACM SIGKDD Workshop on Human Computation, HCOMP 10,* 2010, pp. 64–67.

[42] J. Whitehill, P. Ruvolo, T. Wu, J. Bergsma, and J. Movellan, "Whose vote should count more: Optimal integration of labels from labelers of unknown expertise," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2009, pp. 2035–2043.

[43] P. Welinder, S. Branson, S. Belongie, and P. Perona, "The multidimensional wisdom of crowds," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2010.

[44] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy, "Learning from crowds," *J. Mach. Learn. Res.*, vol. 11, pp. 1297–1322, Apr. 2010.

[45] V. C. Raykar and S. Yu, "Eliminating spammers and ranking annotators for crowdsourced labeling tasks," *J. Mach. Learn. Res.*, vol. 13, pp. 491–518, Feb. 2012.

[46] N. Peled, Y. K. Gal, and S. Kraus, "An agent design for repeated negotiation and information revelation with people," in *Proc. 27th AAAI Conf. Artificial Intelligence*, 2013.

[47] W. Mason, S. Suri, and D. J. Watts, "Long-run learning in games of cooperation," in *Proc. 15th ACM Conf. Economics and Computation (EC)*, June 2014, pp. 821–838.

[48] D. Fudenberg and A. Peysakhovich, "Recency, records and recaps: Learning and non-equilibrium behavior in a simple decision problem," in *Proc. 15th ACM Conf. Economics and Computation (EC)*, June 2014, pp. 971–986.

[49] S. Zhang and A. J. Yu, "Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2013, pp. 2607–2615.

[50] S. Guo, M. Wang, and J. Leskovec, "The role of social networks in online shopping: Information passing, price of trust, and consumer choice," in *Proc. 12th ACM Conf. Electronic Commerce (EC)*, 2011, pp. 157–166.

[51] Y. Gao, Y. Chen, and K. J. R. Liu, "Understanding sequential user behavior in social computing: To answer or to vote?," *IEEE Trans. Netw. Sci. Eng.*, vol. 2, no. 3, pp. 112–126, Aug. 2015.

[52] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a Nash equilibrium," *SIAM J. Comput.*, vol. 39, no. 1, pp. 195–259, 2009.

[53] N. Burch, M. Lanctot, D. Szafron, and R. G. Gibson, "Efficient Monte Carlo counterfactual regret minimization in games with many player actions," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2012.

[54] W. Krichene, B. Drighes, and A. Bayen, "On the convergence of no-regret learning in selfish routing," in *Proc. 31st Int. Conf. Machine Learning (ICML)*, 2014.

[55] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Proc. 11th Int. Conf. Machine Learning (ICML)*, 1994, pp. 157–163.

[56] J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games," *J. Mach. Learn. Res.*, vol. 4, pp. 1039–1069, Dec. 2003.

[57] T. Kalinowski, N. Narodytska, T. Walsh, and L. Xia, "Strategic behavior when allocating indivisible goods sequentially," in *Proc. 27th AAAI Conf. Artificial Intelligence*, 2013.

[58] A. Greenwald, J. Li, and E. Sodomka, "Approximating equilibria in sequential auctions with incomplete information and multi-unit demand," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2012.

[59] H. Luo and R. E. Schapire, "A drifting-games analysis for online learning and applications to boosting," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2014.

[60] C. Zhang and J. A. Shah, "Fairness in multi-agent sequential decision-making," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2014.

[61] B. Settles, "Active learning literature survey," Tech. Rep., Univ. Wisconsin-Madison, 2010.

[62] D. D. Margineantu, "Active cost-sensitive learning," in *Proc. IJCAI*, 2007.

[63] B. Settles, M. Craven, and L. Friedlan, "Active learning with real annotation costs," in *Proc. NIPS Workshop on Cost-Sensitive Learning*, 2008.

[64] S. Sabato, A. D. Sarwate, and N. Srebro, "Auditing: Active learning with outcome-dependent query costs," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2013.

[65] L. Yang and J. Carbonell, "Buy-in-bulk active learning," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2013.

[66] A. Beygelzimer, S. Dasgupta, and J. Langford, "Importance weighted active learning," in *Proc. ICML*, 2009.

[67] M. F. Balcan, C. Berlind, A. Blum, E. Cohen, K. Patnaik, and L. Song, "Active learning and best-response dynamics," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2014.

[68] S. Chawla, J. D. Hartline, and B. Sivan, "Optimal crowdsourcing contests," in *Proc. ACM-SIAM Symp. Discrete Algorithms(SODA)*, 2012.

[69] R. Cavallo and S. Jain, "Efficient crowdsourcing contests," in *Proc. Int. Conf. Autonomous Agents and Multiagent Systems (AAMAS)*, 2012.

[70] A. Shaw, J. Horton, and D. Chen, "Designing incentives for inexpert human raters," in *Proc. ACM Conf. Computer Supported Cooperative Work (CSCW)*, 2011.

[71] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications," in *Proc. IEEE INFOCOM*, 2012.

[72] Y. Singer and M. Mittal, "Pricing tasks in online labor markets," in *Proc. 3rd Workshop on Human Computation, (HCOMP) 11*, 2011.

[73] Y. Singer and M. Mittal, "Pricing mechanisms for crowdsourcing markets," in *Proc. 22nd Int. Conf. World Wide Web (WWW)*, 2013.

[74] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *Proc. 22nd Int. Conf. World Wide Web (WWW)*, 2013.

[75] D. Berry and B. Fristedt, *Bandit Problems: Sequential Allocation of Experiments*. London: Chapman and Hall, 1985.

[76] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Trans. Signal Processing*, vol. 58, no. 11, pp. 5667–5681, 2010.

[77] Y. Seldin, P. Bartlett, K. Crammer, and Y. Abbasi-Yadkori, "Prediction with limited advice and multiarmed bandits with paid observations," in *Proc. 31st Int. Conf. Machine Learning (ICML)*, 2014.

[78] N. R. Devanur and S. M. Kakade, "The price of truthfulness for pay-per-click auctions," *Proc. 10th ACM Conf. Electronic Commerce (EC)*, 2009.

[79] N. Gatti, A. Lazaric, and F. Trovò, "A truthful learning mechanism for contextual multi-slot sponsored search auctions with externalities," in *Proc. 13th ACM Conf. Electronic Commerce (EC)*, 2012, pp. 605–622.

[80] K. Amin, A. Rostamizadeh, and U. Syed, "Learning prices for repeated auctions with strategic buyers," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2013, pp. 1169–1177.

[SP]

[Mingyi Hong, Meisam Razaviyayn, Zhi-Quan Luo, and Jong-Shi Pang]

A Unified Algorithmic Framework for Block-Structured Optimization Involving Big Data

[With applications in machine learning and signal processing]

This article presents a powerful algorithmic framework for big data optimization, called the *block successive upper-bound minimization* (*BSUM*). The BSUM includes as special cases many well-known methods for analyzing massive data sets, such as the block coordinate descent (BCD) method, the convex-concave procedure (CCCP) method, the block coordinate proximal gradient (BCPG) method, the nonnegative matrix factorization (NMF) method, the expectation maximization (EM) method, etc. In this article, various features and properties of the BSUM are discussed from the viewpoint of design flexibility, computational efficiency, parallel/distributed implementation, and the required communication overhead. Illustrative examples from networking, signal processing, and machine learning are presented to demonstrate the practical performance of the BSUM framework.

Digital Object Identifier 10.1109/MSP.2015.2481563
Date of publication: 29 December 2015

INTRODUCTION

OVERVIEW OF OPTIMIZATION FOR BIG DATA

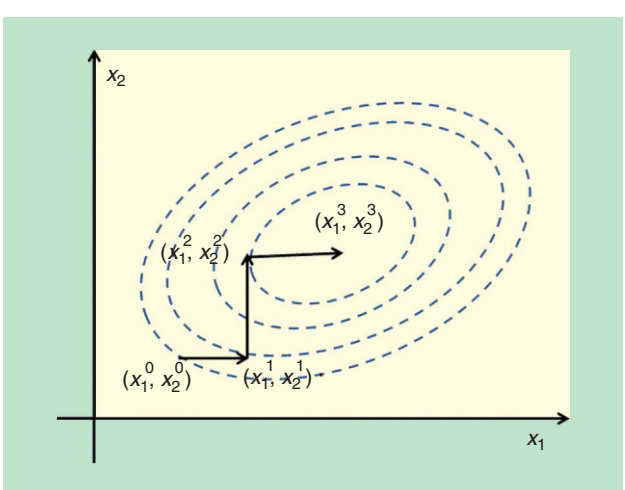
With advances in sensor, communication, and storage technologies, data acquisition is now more ubiquitous than at any other time. This has made available big data sets in many areas of engineering, biological, social, and physical sciences. While the proper modeling and analysis of such data sets can yield valuable information for inference, estimation, tracking, learning, and decision making, their size and complexity present great challenges in algorithm design and implementation.

Due to its central role in big data analytics, large-scale optimization has recently attracted significant attention not only from the optimization community, but also from the machine-learning, statistics, and signal processing communities. For example, emerging problems in image processing, social network analysis, and computational biology can easily exceed millions or billions of variables, and significant research is underway to enable fast accurate solutions to these problems [1]–[4]. Also, problems related to the design and provision of large-scale smart infrastructures such as wireless communication networks require real-time efficient resource allocation decisions to ensure optimal network performance. Traditional general-purpose optimization tools are inadequate for these problems due to the complexity of the model, the heterogeneity of the data, and, most importantly, the sheer data size [5]–[9]. Modern large-scale optimization algorithms, especially those that are capable of exploiting problem structures; dealing with distributed, time-varying, and incomplete data sets; and utilizing massively parallel computing and storage infrastructures, have become the workhorse in the big data era.

To be efficient for big data applications, optimization algorithms must have certain properties:

1) Each of their computational steps must be simple and easy to perform.
2) The intermediate results are easily stored.
3) They can be implemented in a distributed and/or parallel manner so as to exploit the modern multicore and cluster computing architectures.
4) A high-quality solution can be found using a small number of iterations.

These requirements preclude the use of high-order information about the problem (i.e., the Hessian matrix of the objective), which is usually too expensive to obtain, even for modest-sized problems.



[FIG1] The BCD method for a two-dimensional problem. The dashed curves represent the contours of the objective function, the solid lines represent the progress of the algorithm.

THE BCD METHOD

A very popular family of optimization algorithms that satisfies most of the aforementioned properties is the BCD method, sometimes known as the *alternating minimization/maximization* (*AM*) algorithm. The basic steps of the BCD are simple: 1) partition the entire optimization variables into small blocks and 2) optimize one block variable (or few blocks of variables) at each iteration while holding the remaining variables fixed. More specifically, consider the following block-structured optimization problem

$$\underset{x}{\text{minimize}} \quad f(x_1, x_2, \ldots, x_n), \quad \text{s.t. } x_i \in \mathcal{X}_i, i = 1, 2, ..., n, \quad (1)$$

where $f(\cdot)$ is a continuous function (possibly nonconvex, nonsmooth), each $\mathcal{X}_i \subseteq \mathbb{R}^{m_i}$ is a closed convex set, and each x_i is a block variable, $i = 1, 2, \ldots, n$. Define $x := (x_1, \cdots, x_n) \in \mathbb{R}^m$, and let $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \subseteq \mathbb{R}^m$. When applying the classical BCD method to solve (1), at every iteration r, a single block of variables, say $i = (r \bmod n) + 1$, is optimized by solving the following problem

$$x_i^r \in \underset{x_i \in \mathcal{X}_i}{\text{argmin}} \quad f(x_i, x_{-i}^{r-1}), \quad (2)$$

where we have defined $x_{-i}^{r-1} := (x_1^{r-1}, \cdots, x_{i-1}^{r-1}, x_{i+1}^{r-1}, \cdots, x_n^{r-1})$; for the rest of the variables $j \neq i$, let $x_j^r = x_j^{r-1}$. See Figure 1 for a graphical illustration of the algorithm.

The BCD algorithm is intuitively appealing and very simple to implement, yet it is extremely powerful. It enjoys tremendous popularity in a wide range of applications from signal processing communications to machine learning. Representative examples include image deblurring [10], statistical learning [11], wireless communications [12], etc. In recent years, there has been a surge of renewed interests to extend and generalize the BCD type of algorithms due to its applications in modern big data optimization problems. Theoretically, the BCD algorithm and its variants have been generalized significantly to accommodate various coordinate update rules [5], [13], [14] and have been made suitable for implementation on modern parallel processing architecture [15]–[19]. It can handle a wide range of nonsmooth, nonconvex cost functions [20]–[22]. Practically, it has been used in emerging large-scale signal processing and machine-learning applications, such as compressive sensing/sparse signal recovery [23], [24], matrix completion [25], [26], and tensor decomposition [27], [28], to name just a few. A recent survey of this algorithm can be found in [29].

THE BSUM METHOD

In this article, we introduce a unifying framework, the BSUM method, which generalizes the BCD type of algorithms [21]. Simply put, the BSUM framework includes algorithms that successively optimize certain upper bounds or surrogate functions of the original objectives, possibly in a block-by-block manner. The BSUM framework significantly expands the application domain of the traditional BCD algorithms. For example, it covers many classical statistics and machine-learning algorithms such as the EM method [30], the CCCP [31], and the multiplicative NMF [32]. It also includes as special cases many well-known signal processing algorithms such as the family of interference pricing algorithms [33], [34] and the weighted minimum mean square error algorithms [12], [35] for interference management in large-scale wireless systems.

The generality and flexibility of the BSUM offers an ideal platform to explore optimization algorithms for big data. Through the lens of the BSUM, one obtains not only a thorough understanding of the variety of algorithms and applications that are being covered, but more importantly the principle for designing a good algorithm suitable for big data. To this end, this article will first provide a concise overview of a few key theoretical characterizations of the algorithms that fall under the BSUM framework, BCD included. Our emphasis will be given on providing intuitive understanding as to when and where the BSUM framework should (or should not) work, and how its performance can be characterized. The second part of this article offers a detailed account of many existing large-scale optimization algorithms that fall under the BSUM framework, together with a few big data related applications that are of significant interests to the signal processing community. The last part of the article outlines some interesting extensions of the BSUM that further help expand its application domains. Throughout this article, special emphasis will be given on computational issues arising from big data optimization problems such as algorithm design for parallel and distributed computation, algorithm implementation on multicore computing platforms, and distributed storage of the data. In particular, we will discuss how these issues can be addressed in the BSUM framework by providing theoretical insights and examples from real practical problems.

[TABLE 1] A PSEUDOCODE OF THE BSUM ALGORITHM.

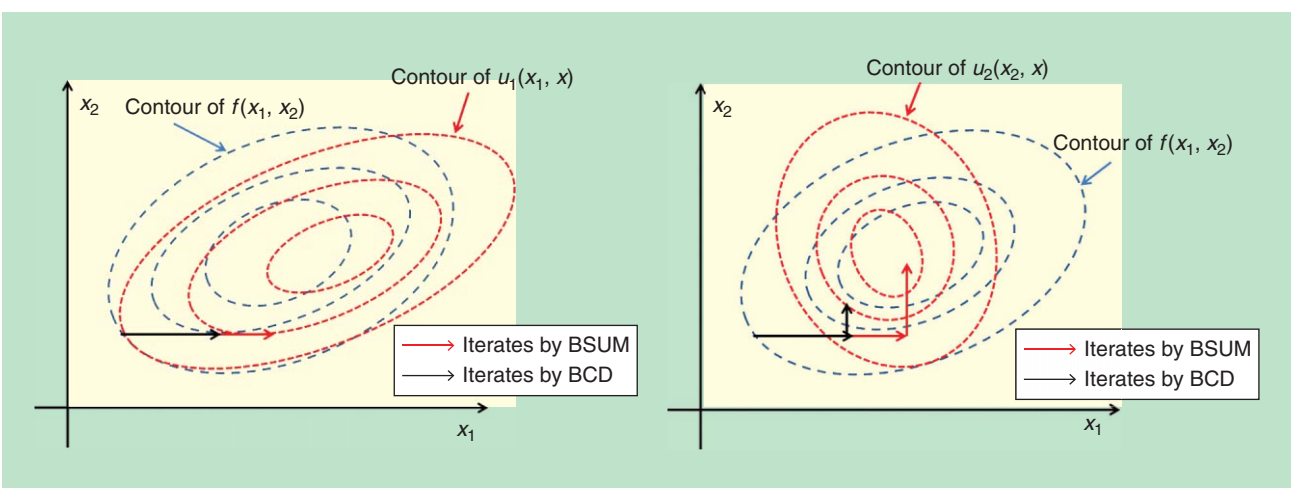
```
1 FIND A FEASIBLE POINT x^0 ∈ 𝒳 AND SET r = 0
2 REPEAT
3    PICK INDEX SET ℐ^r
4    LET x_i^r ∈ argmin_{x_i ∈ 𝒳_i} u_i(x_i, x^{r-1}),  ∀ i ∈ ℐ^r
5    SET x_k^r = x_k^{r-1},  ∀ k ∉ ℐ^r
6    r = r + 1,
7 UNTIL SOME CONVERGENCE CRITERION IS MET
```

BSUM AND ITS THEORETICAL PROPERTIES

THE MAIN IDEA

We start by providing a high-level description of the BSUM method and some of its theoretical properties. In practice, one of the main problems of directly applying BCD to solve (1) is that each of its subproblems (2) is often difficult to solve exactly, especially when $f(x)$ is nonconvex. Moreover, even such exact minimization can be performed, the BCD may not necessarily converge. One of the key insights to be offered by the BSUM framework is that, for both practical and theoretical considerations, obtaining an approximate solution of (2) is good enough to keep the algorithm going. To be more specific, let us introduce $u_i(x_i, z): \mathcal{X}_i \to \mathbb{R}$ as an approximation function of $f(x_i, z_{-i})$ for each coordinate i at a given feasible point $z \in \mathcal{X}$. Let us define a set $\mathcal{I}^r$ (possibly with $|\mathcal{I}^r| > 1$) as the block-variable indices to be picked at iteration r. Then at each iteration r, the BSUM method performs the following simple update

$$\begin{cases} x_i^r \in \operatorname{argmin}_{x_i \in \mathcal{X}_i} u_i(x_i, x^{r-1}), \forall\, i \in \mathcal{I}^r \\ x_k^r = x_k^{r-1}, \forall\, k \notin \mathcal{I}^r \end{cases}. \tag{3}$$



[FIG2] A comparison of BSUM and BCD methods for solving a two-dimensional problem. Each time a single coordinate is picked for update. The blue dashed curves represent the contours of the objective function $f(x_1, x_2)$, the red dotted curves represent the contours of the upper-bound functions $u_1(x_1, x)$ and $u_2(x_2, x)$; the black (resp. red) solid lines represent the progress of the BCD (resp. BSUM).

[TABLE 2] THE COMMONLY USED COORDINATE SELECTION RULES.

AT EACH ITERATION r, DEFINE A SET OF AUXILIARY VARIABLES $\{\hat{x}_i^r\}_{i=1}^n$ AS:

$$\hat{x}_i^r \in \operatorname*{argmin}_{x_i \in \mathcal{X}_i} u_i(x_i, x^{r-1}), i = 1, \cdots, n.$$

THEN WE HAVE THE FOLLOWING COMMONLY USED COORDINATE SELECTION RULES.

- *CYCLIC RULE*: THE COORDINATES ARE CHOSEN CYCLICALLY, I.E., IN THE ORDER OF $1, 2, \cdots, n, 1, 2, \cdots$.
- *ESSENTIALLY CYCLIC (E-C) RULE*: THERE EXISTS A GIVEN PERIOD $T \geq 1$ DURING WHICH EACH BLOCK IS UPDATED AT LEAST ONCE, I.E.,

$$\bigcup_{i=1}^{T} \mathcal{I}^{r+i} = \{1, \cdots, n\}, \forall\, r.$$

- *GAUSS-SOUTHWELL (G-SO) RULE*: AT EACH ITERATION r, $\mathcal{I}^r$ CONTAINS A SINGLE INDEX i^* THAT SATISFIES:

$$i^* \in \left\{ i \,\middle|\, \|\hat{x}_i^r - x_i^{r-1}\| \geq q \max_j \|\hat{x}_j^r - x_j^{r-1}\| \right\}$$

  FOR SOME CONSTANT $q \in (0, 1]$.
- *MAXIMUM BLOCK IMPROVEMENT (MBI) RULE*: AT EACH ITERATION r, $\mathcal{I}^r$ CONTAINS A SINGLE INDEX i^* THAT ACHIEVES THE BEST OBJECTIVE: $i^* \in \operatorname{argmin}_i f(\hat{x}_i^r, x_{-i}^{r-1})$.
- *RANDOMIZED RULE*: LET $p_{\min} \in (0, 1)$ BE A CONSTANT. AT EACH ITERATION r, THERE EXISTS A PROBABILITY VECTOR $p^r = (p_1^r, \ldots, p_n^r) \in \mathbb{R}^n$ SATISFYING $\sum_{i=1}^n p_i^r = 1$ AND $p_i^r > p_{\min}, \forall i$, WITH $\mathcal{I}^r$ CONTAINING A SINGLE RANDOM INDEX i_r^* DETERMINED BY

$$\Pr(i \in \mathcal{I}^r \,|\, x^{r-1}, x^{r-2}, \cdots, x^0) = p_i^r, \forall\, i.$$

The complete description of the BSUM is given in Table 1. See Figure 2 for a graphical comparison of the iterates generated by BSUM and BCD for a two-dimensional problem. It should be clear at this point that when $\mathcal{I}^r = \{(r \bmod n) + 1\}$ and no approximation is used [i.e., $u_i(x_i, z) = f(x_i, z_{-i})$] and at each iteration a single coordinate is selected), then the BSUM reduces to the classical cyclic BCD method. In Table 2, we also present several index selection rules that are covered by the BSUM framework. For simplicity of presentation, we will use the classical cyclic index selection rule, where $\mathcal{I}^r = \{(r \bmod n) + 1\}$, in the remainder of this article unless otherwise noted.

Next, we introduce the precise definition of the approximation function. The main idea is that, for each i, the approximation $u_i(x_i, x^r)$ should be an upper bound of the original objective function at the point of x^r (hence, the BSUM name of the framework); see Figure 3 for an illustration of the upper-bound minimization process. Intuitively, picking an upper-bound approximation function is reasonable because optimizing it at least should guarantee some descent of the original objective f; see Figure 3(c).

To be more precise, let us first define the directional derivative of a given function $f(x): \mathcal{X} \to \mathbb{R}$ at a point $x \in \mathcal{X}$ in direction d:

$$f'(x; d) \triangleq \liminf_{\lambda \downarrow 0} \frac{f(x + \lambda d) - f(x)}{\lambda}.$$

Using this definition, we make the following assumptions on the u_is.

Assumption A

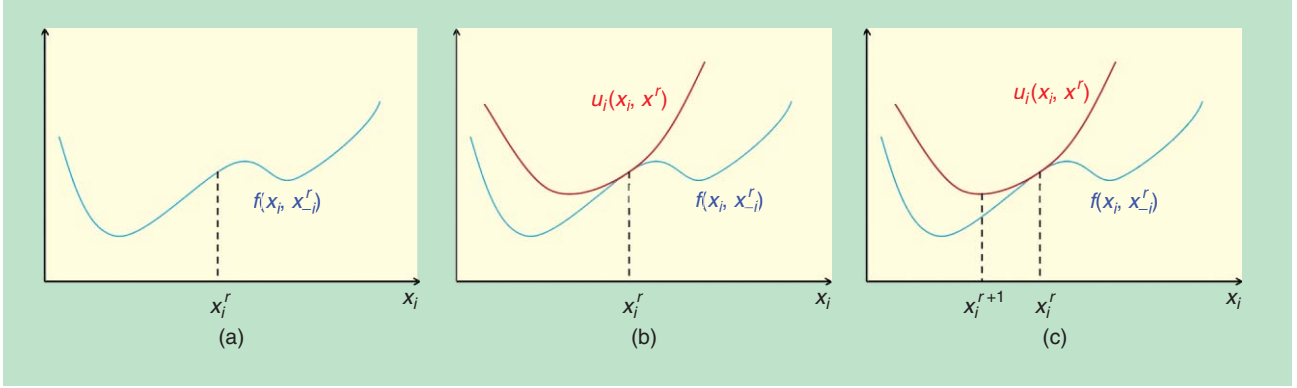
$$u_i(x_i, x) = f(x), \quad \forall x \in \mathcal{X}, \forall i \tag{A1}$$

$$u_i(x_i, z) \geq f(x_i, z_{-i}), \quad \forall x_i \in \mathcal{X}_i, \forall z \in \mathcal{X}, \forall i \tag{A2}$$

$$u_i'(x_i, z; d_i) \,|_{\,x_i = z_i} = f'(z; d), \quad \forall d = (0, \ldots, d_i, \ldots, 0) \text{ s.t. } z_i + d_i \in \mathcal{X}_i, \forall i \tag{A3}$$

$$u_i(x_i, z) \text{ is continuous in } (x_i, z), \quad \forall i. \tag{A4}$$

Intuitively, Assumptions (A1) and (A2) imply that the approximation function is a global upper bound of $f(x)$; while the assumption (A3) guarantees that the first-order behaviors of the



[FIG3] The upper-bound minimization step of the BSUM method is shown. Here we assume that coordinate i is updated at iteration $r + 1$. It is clear from the figure that after solving the BSUM subproblem (3), $f(x_i^{r+1}, x_{-i}^r) < f(x_i^r, x_{-i}^r)$, that is, the objective function is strictly decreased.

[TABLE 3] COMMONLY USED UPPER BOUNDS SATISFYING ASSUMPTION A.

- *PROXIMAL UPPER BOUND*: GIVEN A CONSTANT $\gamma > 0$, ONE CAN CONSTRUCT A BOUND BY ADDING A QUADRATIC PENALIZATION (I.E., THE PROXIMAL TERM)
$$u_i(x_i, z) := f(x_i, z_{-i}) + \frac{\gamma}{2}\|x_i - z_i\|^2.$$
- *QUADRATIC UPPER BOUND*: SUPPOSE $f(x) = g(x_1, \cdots, x_n) + h(x_1, \cdots, x_n)$, WHERE g IS SMOOTH WITH $\mathbf{H}_i$ AS THE HESSIAN MATRIX FOR THE iTH BLOCK. THEN ONE CAN CONSTRUCT THE FOLLOWING BOUND
$$u_i(x_i, z) := g(z_i, z_{-i}) + h(x_i, z_{-i}) + \langle \nabla_i g(z_i, z_{-i}), x_i - z_i \rangle + \frac{1}{2}(x_i - z_i)^T \Phi_i (x_i - z_i),$$
WHERE BOTH Φ_i AND $\Phi_i - \mathbf{H}_i$ ARE POSITIVE SEMIDEFINITE MATRICES.
- *LINEAR UPPER BOUND*: SUPPOSE f IS DIFFERENTIABLE AND CONCAVE, THEN ONE CAN CONSTRUCT
$$u_i(x_i, z) := f(z_i, z_{-i}) + \langle \nabla_i f(z_i, z_{-i}), x_i - z_i \rangle.$$
- *JENSEN'S UPPER BOUND*: SUPPOSE $f(x) := f(a_1^T x_1, \cdots, a_n^T x_n)$, WHERE $a_i \in \mathbb{R}^{m_i}$ IS A COEFFICIENT VECTOR, AND f IS CONVEX WITH RESPECT TO EACH $a_i^T x_i$. LET $w_i \in \mathbb{R}_+^{m_i}$ DENOTE A WEIGHT VECTOR WITH $\|w_i\|_1 = 1$. THEN ONE CAN USE JENSEN'S INEQUALITY AND CONSTRUCT
$$u_i(x_i, z) := \sum_{j=1}^{m_i} w_i(j) f\Big(\frac{a_i(j)}{w_i(j)}(x_i(j) - z_i(j)) + a_i^T z_i, z_{-i}\Big),$$
WHERE $w_i(j)$ REPRESENTS THE jTH ELEMENT IN VECTOR w_i.

objective function and the approximation function are the same at the point of approximation (cf. Figure 3). In Table 3, we provide a few commonly used upper bounds that satisfy Assumption A; see also [21] and [22] for additional examples. More discussion will be given in subsequent sections on how these bounds are used in practice.

For a popular subclass of problem (1), Assumption A can be further simplified; see the following example [21, Proposition 2].

Example 1: Consider the following special form of problem (1):

$$\begin{aligned} &\min f(x) := g(x_1, \cdots, x_n) \\ &\quad + h(x_1, \cdots, x_n), \quad \text{s.t. } x_i \in \mathcal{X}_i, i = 1, \cdots, n, \end{aligned} \tag{4}$$

where $g: \mathcal{X} \to \mathbb{R}$ is a smooth function and $h: \mathcal{X} \to \mathbb{R}$ is a possibly nonsmooth function whose directional derivative exists at every point $x \in \mathcal{X}$. Consider $u_i(x_i, z) = \hat{u}_i(x_i, z) + h(x)$, where $\hat{u}_i(x_i, z)$ approximates the smooth function g in the objective. Then assumption (A3) is implied by (A1) and (A2) and is therefore no longer needed.

Now that we have seen the main steps of the algorithm, we describe its theoretical properties. We address the following questions related to the convergence of the BSUM: When does the BSUM converge? How fast does it converge? When does the BSUM fail and why? The answers to these questions will be instrumental in understanding the framework as well as evaluating the performance of various related algorithms.

WHEN DOES THE BSUM CONVERGE?

To discuss the convergence property of the algorithm, we first investigate the optimality conditions for (1), which characterize the set of solutions that we would like our algorithm to reach. To this end, we introduce two related notions, one is called the *stationary solution* and the other is the *coordinatewise minimum solution*; see Table 4 for their precise definitions. Roughly speaking, the coordinatewise minimum $\hat{x}$ is the point where no single block $\hat{x}_i$, $i = 1, \cdots, n$ has a better direction to move to, while at a stationary point x^* the entire vector cannot move to a better direction. Further, a stationary solution must be a coordinatewise minimum, but the reverse direction is generally not true; see the example next.

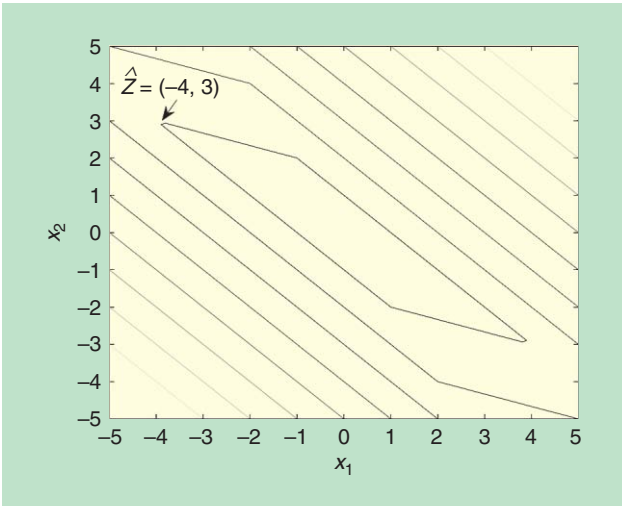
Example 2: Consider the convex function $f(z) = \|Az\|_1$, where $A = [3\ 4; 2\ 1] \in \mathbb{R}^{2\times 2}$. This function has a unique stationary solution $(z_1, z_2) = (0, 0)$, which is also the global optimal solution. Further, the point $\hat{z} = (-4, 3)$ is a coordinatewise minimum with respect to the two standard coordinates since $f'(\hat{z}; d) \geq 0, \forall d \in \{(d_1, d_2) \in \mathbb{R}^2 \,|\, d_1 d_2 = 0\}$; but this point is not a stationary solution as $f'(\hat{z}; \hat{d}) < 0$ for $\hat{d} = (4, -3)$. This fact can also be observed in the contour plot of the function in Figure 4.

This example confirms that the coordinatewise minimum can be much inferior to the stationary solution. Therefore in the subsequent discussion we will mainly focus on finding the stationary solutions rather than the coordinatewise minimum. An immediate question is: can one easily distinguish these two types of solutions, or for that matter, when does a coordinatewise minimum become a stationary solution? Let us define a regular point $x \in \mathcal{X}$ as a point that satisfies the following statement: if x is coordinatewise minimum, then it is a stationary solution. For a large and popular subclass of (1) expressed below in (5), where the nonsmooth function is separable across the blocks, all feasible points $x \in \mathcal{X}$ are regular

$$\min f(x) := g(x_1, \cdots, x_n) + \sum_{i=1}^{n} h_i(x_i), \quad \text{s.t. } x_i \in \mathcal{X}_i, i = 1, \cdots, n. \tag{5}$$

[TABLE 4] OPTIMALITY CONDITIONS.

- *STATIONARY SOLUTIONS*: THE POINT x^* IS A STATIONARY SOLUTION OF (1) IF $f'(x^*; d) \geq 0$ FOR ALL d SUCH THAT $x + d \in \mathcal{X}$. LET $\mathcal{X}^*$ DENOTE THE SET OF STATIONARY SOLUTIONS.
- *COORDINATEWISE MINIMUM SOLUTIONS*: $\hat{x} \in \mathcal{X}$ IS COORDINATEWISE MINIMUM OF PROBLEM (1) WITH RESPECT TO THE COORDINATES IN $\mathbb{R}^{m_1}, \mathbb{R}^{m_2}, \ldots, \mathbb{R}^{m_n}$, IF
$f(\hat{x} + d_k^0) \geq f(\hat{x}), \quad \forall d_k \in \mathbb{R}^{m_k}$ with $\hat{x} + d_k^0 \in \mathcal{X}, \forall k = 1, 2, \ldots, n,$
WHERE $d_k^0 = (0, \ldots, d_k, \ldots, 0)$.



[FIG4] The contour plot of the function $f(z) = \|Az\|_1$, where $A = [3\ 4; 2\ 1] \in \mathbb{R}^{2\times 2}$. The point $\hat{z} = (-4, 3)$ is a coordinatewise minimum but not a stationary solution [21]. (Figure reprinted with permission from the Society for Industrial and Applied Mathematics.)

The main convergence result for the BSUM method is given below, which is adapted from [21, Th. 2]

Theorem 1: Suppose the cyclic coordinate selection rule is chosen, i.e., $\mathcal{I}^r = \{(r \bmod n) + 1\}$. Let $\{x^r\}_{r=1}^{\infty}$ be a sequence generated by the BSUM algorithm. Suppose Assumption A holds, and that each x^r is regular. Then the following is true:

a) *Suppose that the function $u_i(x_i, y)$ is quasi-convex in x_i for $i = 1, \ldots, n$. Furthermore, assume that the subproblem (3) has a unique solution for any point $x^{r-1} \in \mathcal{X}$. Then every limit point x^* of $\{x^r\}$ is a stationary point of (1).*

b) *Suppose the level set $\mathcal{X}^0 = \{x | f(x) \le f(x^0)\}$ is compact. Furthermore, assume that the subproblem (3) has a unique solution for any point $x^{r-1} \in \mathcal{X}$, $r \ge 1$ for at least $n-1$ blocks. Then $\{x^r\}$ converge to the set of stationary points, i.e.,*

$$\lim_{r\to\infty} \ d(x^r, \mathcal{X}^*) = 0,$$

where $d(x, \mathcal{X}^*) \triangleq \min_{x^* \in \mathcal{X}^*} \|x - x^*\|$ and $\mathcal{X}^*$ is the set of stationary points.

Here, the first part of the result deals with the possibility of an unbounded sequence, whereas the second part assumes that the sequence lies in a compact set, therefore it has a slightly stronger claim. Note that Theorem 1 can be easily extended to all the coordinate selection rules given in Table 2, and for the randomized version the convergence is with probability 1.

Special attention should be given to the conditions set forth by Theorem 1. First, it says that the upper-bound function needs to be picked carefully to satisfy both Assumption A and the requirement that at least $n-1$ subproblems (3) have a unique solution. However, when the objective function $f(x)$ is convex, the uniqueness requirement of the per-block solution can be dropped; see [36]. Also, Theorem 1 requires the problem to be well-defined so that coordinatewise optimal solutions are equivalent to the stationary solution (which precludes objective function $\|Ax\|_1$ in Example 2). Next we will demonstrate, through a couple of concrete examples, that relaxing some of these conditions indeed leads to the divergence of the BSUM.

WHEN DOES BSUM FAIL?

Here we provide a few examples in which BSUM fails to converge to any stationary solutions. The examples in this section serve as reminders that in practice, to avoid those pitfalls, extreme care must be exercised when designing large-scale optimization algorithms.

Our first example comes from Example 2. It demonstrates the consequence of lacking the regularity condition.

Example 3: Consider the following unconstrained convex optimization problem $\min_x \left\|\sum_{i=1}^2 A_i x_i\right\|_1$, where $A_1 = [3\ 2]^T$ and $A_2 = [4\ 1]^T$. Clearly by defining $A = [3\ 4; 2\ 1]$, the objective function can be rewritten as $\|Ax\|_1$, which is not regular at point $(x_1, x_2) = (-4, 3)$ (cf. Example 2). It follows that, by setting $u_i(x_i, z) = \left\|\sum_{i=1}^2 A_i x_i\right\|_1$ (no approximation is used) and letting $(x_1^0, x_2^0) = (-4, 3)$, the BSUM algorithm will not be able to move forward for either x_1 or x_2, thus it will be stuck at the noninteresting point $(-4, 3)$, far away from the only stationary solution $(0, 0)$.

The next example shows that BSUM fails to converge if the feasible set $\mathcal{X}$ is no longer a Cartesian product of feasible sets $\mathcal{X}_1, \cdots, \mathcal{X}_n$, a fact that we have taken for granted so far.

Example 4: Consider the following simple quadratic problem:

$$\min \quad x_1^2 + x_2^2, \quad \text{s.t.} \quad x_1 + x_2 = 2.$$

The optimal objective value is two. Consider the BSUM algorithm with an arbitrary approximation function satisfying Assumption A, but initiated at the point $(x_1^0, x_2^0) = (0, 2)$. Then the BSUM method will be stuck at this noninteresting point without making any progress because it is not possible to change a single block without violating the coupling constraint.

Our next example shows that the BSUM could diverge if the approximation function u_i violates Assumption A.

Example 5: Consider the simple quadratic problem

$$\min \quad x_1^2 + x_2^2 + 2x_1x_2, \quad \text{s.t.} \quad -1 \le x_1, x_2 \le 1.$$

The optimal objective value is 0, with $(x_1^*, x_2^*) = (0, 0)$. Consider using the BSUM algorithm with a linear approximation function, which violates Assumption (A2). More specifically, for a given feasible tuple $(x_1, x_2) = (\hat{x}_1, \hat{x}_2)$, the x_1's subproblem becomes

$$\min_{x_1} \quad \langle \hat{x}_1 + \hat{x}_2, x_1 \rangle, \quad \text{s.t.} \quad -1 \le x_1 \le 1.$$

Clearly, the optimal solution is either $x_1 = -1$ or $x_1 = 1$. The same happens when solving the subproblem for x_2. Therefore the algorithm will never reach the optimal solution $(x_1^*, x_2^*) = (0, 0)$. Further, if the feasible sets of x_1 and x_2 are unbounded, then the BSUM can diverge.

Our last example is taken from Powell [37]. It shows that without the uniqueness assumption of each subproblem (3), the BSUM algorithm is not necessarily convergent.

Example 6: Consider the following unconstrained problem

$$\min \quad f(x) := -x_1x_2 - x_2x_3 - x_3x_1 + (x_1-1)_+^2 + (-x_1-1)_+^2 + (x_2-1)_+^2 + (-x_2-1)_+^2 + (x_3-1)_+^2 + (-x_3-1)_+^2,$$

where the notation $(z)_+^2$ *means* $(\max\{0, z\})^2$. *In this case, fixing* (x_2, x_3) *and optimizing over* x_1 *yields the following solution*

$$x_1 = \begin{cases} (1+\frac{1}{2}|x_2+x_3|)\operatorname{sign}(x_2+x_3), & \text{if } x_2+x_3 \neq 0, \\ [-1,1], & \text{otherwise} \end{cases}. \quad (6)$$

A similar solution can be obtained for x_2 *and* x_3 *as well. Suppose we set* $u_i(x_i, z) = f(x)$ *for all* i *(no approximation is used) and let* $(x_1^0, x_2^0, x_3^0) = (-1-\epsilon, 1+1/2\epsilon, -1-1/4\epsilon)$ *for some* $\epsilon > 0$. *Then it can be shown the applying the cyclic version of the BSUM algorithm, the iterates will be cycling around six points* $(1,1,-1)$, $(1,-1,-1)$, $(1,-1,1)$, $(-1,-1,1)$, $(-1,1,1)$, $(-1,1,-1)$, *and none of these six points is a stationary solution of the original problem. The reason for such pathological behavior is that, in any one of the six limit points above, there are at least two subproblems that have multiple optimal solutions. For example, at* $(1,1,-1)$, *and fixing* x_2, x_3 *(resp.* x_1, x_3*), the optimal solution for* x_1 *(resp.* x_2*) is any element in the interval* $[-1,1]$; *cf. (6).*

A natural question at this point is, can we make the BSUM work for these examples? The answer is affirmative, but how this can be done requires a case by case study. To handle the first two examples (i.e., Example 3 and 4), a generalized version of BSUM is needed, which will be discussed in the "Extensions" section. For the third example, one can simply pick a better upper bound to guarantee convergence. For example, if we pick the proximal upper bound (cf. Table 3): $u_i(x_i, z) = f(x) + \gamma/2\|x_i - z_i\|^2$, then each subproblem will have a unique solution, and the algorithm will not be trapped by the noninteresting solutions given in Example 6. Notice that the use of $\gamma/2\|x_i - z_i\|^2$ inhibits the move of x_i from its current position z_i. Thus, the main message here is that when optimizing each block, it is beneficial, at least theoretically, to be less greedy so that a conservative step is taken towards reducing the objective. The extent of the "conservativeness" for the per block update is then naturally characterized by the chosen upper bounds. Quite interestingly, the difficulty with the nonunique subproblem solution can also be resolved by using randomization. Formally, we have the following corollary to Theorem 1 [38].

Corollary 1: Suppose the level set $\mathcal{X}^0 = \{x | f(x) \leq f(x^0)\}$ *is compact. Then, under the randomized block selection rule, the iterates generated by the BSUM algorithm converge to the set of stationary points almost surely, i.e.,*

$$\lim_{r\to\infty} d(x^r, \mathcal{X}^*) = 0, \quad \text{almost surely.}$$

HOW FAST DOES THE BSUM CONVERGE?

Now that we have examined the convergence of the BSUM, let us proceed next to characterize the convergence speed of the algorithm. There is no doubt that this is an important issue, especially so for big data optimization—the sheer size of the data and limited computational resource makes it difficult to optimize a problem to global optimality. Consequently, we are generally satisfied with high-quality suboptimal solutions and are mostly concerned with how quickly such solutions can be obtained.

Recently, extensive research efforts have been devoted to analyzing the convergence rate for various BSUM-type algorithms, most of which use randomized coordinate selection rules and/or quadratic upper-bound functions (cf. Table 3) to solve convex problems; for example; see [5], [8], and [39]–[43]. Since it is not possible to go over all the details of these different variations of BSUM, here we present, at a high level, a fairly general result that covers a large family of upper-bound functions satisfying Assumption A, as well as all coordinate selection rules outlined in Table 2.

To this end, let us make the following additional assumptions.

Assumption B

B1) $f(x) := g(x) + \sum_{i=1}^n h_i(x_i)$, where $g(x)$ is a smooth convex function with Lipschitz continuous gradient, i.e., there exists a constant L such that $\|\nabla g(x) - \nabla g(y)\| \leq L\|x-y\|$, $\forall\, x, y \in \mathcal{X}$. Further both g and h_is are convex functions.

B2) The level set $\{x | f(x) \leq f(x^0), x \in \mathcal{X}\}$ is compact.

B3) Each upper-bound function $u_i(x_i, z)$ is strongly convex with respect to x_i.

An ϵ-optimal solution $x^\epsilon \in \mathcal{X}$ is defined as an $x^\epsilon \in \{x | x \in \mathcal{X}, f(x) - f(x^*) \leq \epsilon\}$, where $f(x^*)$ is the globally optimal objective value of problem (5). Suppose both Assumptions A and B are satisfied. Then it is shown in [38] and [42] that BSUM takes at most c/ϵ iterations to find an ϵ-optimal solution, for all coordinate rules specified in Table 2, where $c > 0$ is a constant only related to the description of the problem. Such a type of convergence rate is known as *sublinear convergence*. Here, the main message is that under Assumptions A and B, the algorithm generally converges sublinearly in the order of $1/\epsilon$. Further, for different special forms of BSUM, the constant c in front of $1/\epsilon$ can be significantly refined so that it is independent of problem dimension; see [5] and [8]. It is also interesting to note here that when the objective f is either strongly convex or convex but with certain special structure, the BSUM achieves the linear rate of convergence. That is, BSUM takes at most $O(\log(c/\epsilon))$ iterations to find an ϵ-optimal solution, which is much faster than the sublinear rate; see, e.g., [44] and [45] for the related discussions.

Finally, we briefly mention that it is possible to relax certain conditions in Assumption B to obtain refined rates. For example, [42] shows that dropping the per-block strong convexity assumption in (B3) still achieves an $O(1/\epsilon)$ sublinear convergence. In [38], [46] it is shown that, when removing the convexity Assumption (B1), it is also possible to characterize the convergence rate to stationary solutions. In [42] it is shown that when there are two blocks of variables, the cyclic version of the BSUM can be accelerated to achieve an improved $O(1/\sqrt{\epsilon})$ complexity. In a few recent works [47], [48], it is shown that when randomized block selection and the quadratic upper bound are used, it is possible to accelerate the BSUM with any $n > 2$ blocks.

ALGORITHMS COVERED BY THE BSUM FRAMEWORK

Now that we have a good understanding of the theoretical properties of the family of BSUM algorithms, we demonstrate in this section the wide applicability of the BSUM framework by revealing its connection to a few well-known algorithms for high-dimensional massive data analysis. For each of the examples presented below, we pay special attention to the design of the upper-bound functions.

THE BCD ALGORITHM

The first algorithm that the BSUM covers is obviously the classic BCD described in the "The BCD Method" section. Here the upper-bound function is simply the original objective itself, i.e., $u_i(x_i, z) := f(x_i, z_{-i}), \ \forall\, x_i \in \mathcal{X}_i, z \in \mathcal{X}, \forall\, i$. We would like to mention that all the convergence and rate of convergence analysis of BSUM carries over to the BCD method. Specifically, the result in the section "How Fast Does the BSUM Converge?" implies that the BCD method (with coordinate update rules specified in Table 2) converges in a sublinear manner whenever Assumptions (B1) and (B2) are satisfied. This result by itself is interesting, as there has been limited theoretical analysis for the general form of a BCD algorithm when applied to problems satisfying Assumptions (B1) and (B2), but not (B3).

THE CCCP

Consider the following unconstrained nonconvex problem, known as the *difference of convex* (*DC*) program: $\min f(x) := g_1(x) - g_2(x)$ where both g_1 and g_2 are convex functions. The well-known CCCP algorithm [31] generates a sequence $\{x^r\}$ by solving

$$x^{r+1} = \arg\min_x u(x, x^r),$$

where $u(x, x^r) := g_1(x) - \langle x - x^r, \nabla g_2(x^r)\rangle - g_2(x^r)$. Clearly, the linear upper bound in Table 3 is used here, therefore CCCP is a special case of the BSUM algorithm with a single block of variables. Furthermore, the updates can also be done in a block coordinate manner.

THE MAJORIZATION-MINIMIZATION ALGORITHM

The majorization-minimization (MM) algorithm, which has been widely used in statistics [49], can be viewed as the single block version of BSUM. Consider the problem of $\min_{x\in\mathcal{X}} f(x)$ where $f(x)$ is a smooth function. The basic idea of the MM algorithm is to first construct a "majorization" function $u(x, z)$ for the original objective $f(z)$, such that

$$u(x, z) \geq f(z), \forall\, x, z \in \mathcal{X}, \quad u(x, x) = f(x), \forall\, x \in \mathcal{X}. \tag{7}$$

Then the algorithm generates a sequence of iterates by successively minimizing $u(x, x^r)$. This algorithm represents a slight generalization of the CCCP discussed previously, but nevertheless still falls in the framework of BSUM.

As another concrete example of the MM algorithm, let us consider the celebrated EM algorithm [30]. Let w be a vector observation used for estimating the parameter x. The maximum likelihood estimate of x is given as (where $p(w\,|\,x)$ denotes the conditional probability of w given x)

$$\hat{x}_{\mathrm{ML}} = \arg\max_x \ln p(w\,|\,x). \tag{8}$$

Let the random vector z be some hidden/unobserved variable. The EM algorithm generates a sequence $\{x^r\}$ by iteratively performing the following two steps 1) E-Step: Calculate $g(x, x^r) := \mathbb{E}_{z|w,x^r}\{\ln p(w, z\,|\,x)\}$ and 2) M-Step: $x^{r+1} = \arg\max_x g(x, x^r)$. To see why the EM algorithm is a special case of MM (hence a special case of BSUM) let us rewrite (8) as: $\min_x -\ln p(w\,|\,x)$, the objective of which can be bounded by

$$\begin{aligned}
-\ln p(w\,|\,x) &= -\ln \mathbb{E}_{z|x}\, p(w\,|\,z, x) \\
&= -\ln \mathbb{E}_{z|x}\left[\frac{p(z\,|\,w, x^r)\,p(w\,|\,z, x)}{p(z\,|\,w, x^r)}\right] \\
&= -\ln \mathbb{E}_{z|w,x^r}\left[\frac{p(z\,|\,x)\,p(w\,|\,z, x)}{p(z\,|\,w, x^r)}\right] \\
&\leq -\mathbb{E}_{z|w,x^r} \ln\left[\frac{p(z\,|\,x)\,p(w\,|\,z, x)}{p(z\,|\,w, x^r)}\right] \\
&= -\mathbb{E}_{z|w,x^r} \ln p(w, z\,|\,x) + \mathbb{E}_{z|w,x^r} \ln p(z\,|\,w, x^r) \\
&:= u(x, x^r),
\end{aligned}$$

where the inequality is due to the fact that a convex function f must satisfy $\mathbb{E}[f(x)] \geq f(\mathbb{E}[x])$ (by the Jensen's inequality). Since $\mathbb{E}_{z|w,x^r} \ln p(z\,|\,w, x^r)$ is not a function of x, the M-step can be written as

$$x^{r+1} = \arg\min_x u(x, x^r).$$

Furthermore, it is not hard to see that $u(x^r, x^r) = -\ln p(w\,|\,x^r)$, therefore, both conditions in (7) are satisfied. Similarly as in the previous case, one can extend the EM to a block coordinate version; see [21] for a detailed discussion.

THE PROXIMAL POINT ALGORITHM (

The classical proximal point algorithm (PPA) (see, e.g., [50, Sec. 3.4.3]) obtains a solution of the problem $\min_{x\in\mathcal{X}} f(x)$ by solving the following equivalent problem

$$\min_{x\in\mathcal{X},\, y\in\mathcal{X}} f(x) + \frac{\gamma}{2}\|x - y\|^2, \tag{9}$$

where $f(\bullet)$ is a convex function, $\mathcal{X}$ is a closed convex set, and $\gamma > 0$ is a coefficient. Clearly (9) is strongly convex in both x and y so long as $f(x)$ is convex [but not jointly strongly convex in (x, y)]. This problem can be solved by performing the following two steps alternatingly

$$x^r = \arg\min_{x\in\mathcal{X}} \left\{f(x) + \frac{1}{2c}\|x - y^{r-1}\|^2\right\}, \quad y^r = x^{r+1}. \tag{10}$$

Equivalently, the PPA algorithm can be viewed as successively minimizing the single block version of the proximal upper bound $u(x; x^r)$ given in Table 3. Note that for a problem with a

single block of variables, if $x \in \mathcal{X}$ is coordinatewise minimum, then it must be a global minimum solution. Therefore, every feasible solution $x \in \mathcal{X}$ is regular, and the convergence of PPA is covered by BSUM.

Furthermore, the PPA can be generalized to solve the multiblock problem (1), where $f(\cdot)$ is convex in each of its block components, but not necessarily strictly convex. Directly applying BCD may fail to find a stationary solution for this problem, as the per-block subproblems can contain multiple solutions (cf. Example 6). Alternatively, we can consider an alternating PPA [51], which successively solves the following:

$$\min_{x_i} \quad f(x_i, x^r_{-i}) + \frac{\gamma}{2}\|x_i - x_i^r\|^2, \quad \text{s.t.} \quad x_i \in \mathcal{X}_i.$$

Clearly this algorithm is a special form of BSUM with strongly convex proximal upper bound (cf. Table 3). It follows that each subproblem has a unique optimal solution, and by Theorem 1 it must converge to a stationary solution.

THE FORWARD-BACKWARD SPLITTING ALGORITHM

The forward-backward splitting (FBS) algorithm (also known as the *proximal splitting algorithm*; see, e.g., [52] and the references therein) for nonsmooth optimization solves the composite problem (4) with a single block of variables (i.e., $n = 1$), where h is convex and lower semicontinuous; g is convex and has Lipschitz continuous gradient, i.e., $\|\nabla g(x) - \nabla g(y)\| \le L\|x - y\|$, $\forall\, x, y \in \mathcal{X}$ and for some $L > 0$.

Define the proximity operator $\text{prox}_h: \mathcal{X} \to \mathcal{X}$ as

$$\text{prox}_h(x) = \arg\min_{y \in \mathcal{X}} h(y) + \frac{1}{2}\|x - y\|^2.$$

The FBS iteration is given below [52]:

$$x^{r+1} = \underbrace{\text{prox}_{\beta h}}_{\text{backward step}} \underbrace{(x^r - \beta \nabla g(x^r))}_{\text{forward step}}, \tag{11}$$

where $\beta \in (0, 1/L]$. Define

$$u(x, x^r) := h(x) + \frac{1}{2\beta}\|x - x^r\|^2 + \langle x - x^r, \nabla g(x^r)\rangle + g(x^r), \tag{12}$$

which is the quadratic upper bound in Table 3, with $\Phi_1 := 1/\beta \mathbf{I}$. It is easy to see that the iteration (11) is equivalent to the following iteration $x^{r+1} = \arg\min_{x \in \mathcal{X}} u(x, x^r)$, therefore, it again falls under the BSUM framework.

Similar to the previous example, we can generalize the FBS algorithm to solve multiple block problems of the form (5). The resulting algorithm, sometimes also known as the *BCPG method*, has received significant attention recently due to its efficiency for solving certain big data optimization problem such as the least absolute shrinkage and selection operator (LASSO) [11]. For recent developments and applications for BCPG, see [5], [39], and [53].

Here we make a special note that, by appealing to the general convergence rate result in the section "How Fast Does the BSUM Converge?" the BCPG method with any coordinate selection rules in Table 2 gives a sublinear convergence rate, when it is used to solve (5) that satisfies Assumption B.

THE NMF ALGORITHM

Consider the following NMF problem:

$$\min_{W \in \mathbb{R}^{M \times K}, H \in \mathbb{R}^{K \times M}} f(W, H) := \frac{1}{2}\|V - WH\|_F^2, \quad \text{s.t.}\ W \ge 0, H \ge 0, \tag{13}$$

where $V \in \mathbb{R}_+^{M \times N}$ is given. The problem has been extensively studied since Lee and Seung's seminal work [32], and it has wide applications in factor analysis, dictionary learning, speech analysis and so on [54]. In [32], a simple and efficient multiplicative algorithm is proposed:

$$[H^{r+1}]_{j,i} = [H^r]_{j,i} \frac{[(W^r)^T V]_{j,i}}{[(W^r)^T W^r H^r]_{j,i}}, \quad j = 1, \cdots, K, i = 1, \cdots, N \tag{14a}$$

$$[W^{r+1}]_{j,i} = [W^r]_{j,i} \frac{[V(H^{r+1})^T]_{j,i}}{[W^r H^{r+1}(H^{r+1})^T]_{j,i}}, \quad j = 1, \cdots, M, i = 1, \cdots, K. \tag{14b}$$

Here, $[H^{r+1}]_{j,i}$ means the (j, i)th component of matrix H^{r+1}. In the following we show that when the iterates are well-defined (i.e., $[W_{ij}^r] > 0$ and $[H_{ij}^r] > 0$), the NMF iteration (14a) and (14b) is also covered by BSUM [55].

Let H_i and V_i represent the ith column of H and V, respectively. Then, at a given iterate $\{W^r, H^r\}$, the subproblem for optimizing H_i is given by

$$\min_{H_i \ge 0} f(H_i, \{W^r, H^r\}) := \frac{1}{2}\|V_i - W^r H_i\|_F^2 + \frac{1}{2}\sum_{j \ne i}\|V_j - W^r H_j^r\|_F^2. \tag{15}$$

Define the upper-bound function $u_i(H_i, \{W^r, H^r\})$ as

$$u_i(H_i, \{W^r, H^r\}) := f(H_i^r, \{W^r, H^r\}) + (H_i - H_i^r)^T \nabla_{H_i} f(H_i, \{W^r, H^r\}) + \frac{1}{2}(H_i - H_i^r)^T \Phi_i(W^r, H_i^r)(H_i - H_i^r),$$

where $\Phi_i(W^r, H_i^r)$ is a diagonal matrix given by

$$\Phi_i(W^r, H_i^r) := \text{Diag}\left(\frac{[(W^r)^T W^r H_i^r]_1}{[H_i^r]_1}, \cdots, \frac{[(W^r)^T W^r H_i^r]_K}{[H_i^r]_K}\right).$$

Clearly, $\Phi_i(W^r, H_i^r) \succ 0$, and it is easy to show that $\Phi_i(W^r, H_i^r) \succ (W^r)^T W^r$, where $(W^r)^T W^r$ is the Hessian of the objective of (15) [32]. This implies that $u_i(H_i, \{W^r, H^r\})$ is the quadratic upper bound given in Table 3 of $f(H_i, \{W^r, H^r\})$. Further, one can check that the subproblem that minimizes $u_i(H_i, \{W^r, H^r\})$ has a unique solution, given by (14a). Similar analysis can be established for the W-block update rule as well. Therefore, we conclude that the iterates (14a) and (14b) are a special case of BSUM. Finally, we note that it is also possible to use different upper-bound functions to derive more efficient update rules for the NMF problem (13); see, e.g., [56], where both the concave upper bound and the Jensen's upper bound (cf. Table 3) are used.

THE ITERATIVE REWEIGHTED LEAST SQUARES METHOD

The iterative reweighted least squares (IRLS) method is a popular algorithm used for solving big data problems such as sparse recovery [57]. Consider

$$\min_{x} \quad h(x) + \sum_{j=1}^{\ell} (|A_j x + b_j|_2, \quad \text{s.t.} \quad x \in \mathbb{R}^m, \tag{16}$$

where $A_j \in R^{k_i \times m}$, $b_j \in \mathbb{R}^{k_i}$, and $h(x)$ is some convex function not necessarily smooth. For a set of applications for this model, see [41, Sec. 4]. Consider the following smooth approximation of (16):

$$\min_{x} h(x) + g(x) := h(x) + \sum_{j=1}^{\ell} \sqrt{\|A_j x + b_j\|^2 + \eta^2}, \quad \text{s.t.} \quad x \in \mathbb{R}^m, \tag{17}$$

where η is some small constant and $g(x)$ denotes the smooth part of the objective. The IRLS algorithm solves (17) by performing the following iteration:

$$x^{r+1} = \arg\min_{x \in \mathbb{R}^m} \left\{ h(x) + \frac{1}{2} \sum_{j=1}^{\ell} \frac{\|A_j x + b_j\|^2 + \eta^2}{\sqrt{\|A_j x^r + b_j\|^2 + \eta^2}} \right\}.$$

Define the following function for $g(x)$:

$$u(x, x^r) = \frac{1}{2} \left(\sum_{j=1}^{\ell} \frac{\|A_j x + b_j\|^2 + \eta^2}{\sqrt{\|A_j x^r + b_j\|^2 + \eta^2}} + \sqrt{\|A_j x^r + b_j\|^2 + \eta^2} \right). \tag{18}$$

It is clear that $g(x^r) = u(x^r, x^r)$, so Assumption (A1) is satisfied. To verify Assumption (A2), we apply the arithmetic-geometric inequality, and have

$$u(x, x^r) = \frac{1}{2} \left(\sum_{j=1}^{\ell} \frac{\|A_j x + b_j\|^2 + \eta^2}{\sqrt{\|A_j x^r + b_j\|^2 + \eta^2}} + \sqrt{\|A_j x^r + b_j\|^2 + \eta^2} \right)$$

$$\geq \sum_{j=1}^{\ell} \sqrt{\|A_j x + b_j\|^2 + \eta^2} = g(x), \forall\, x \in \mathbb{R}^m.$$

Then according to Example 1, Assumption (A3) is automatically true, therefore we have verified that $u(x, x^r)$ defined in (18) is indeed an upper-bound function for the smooth function $g(x)$. It follows that the IRLS algorithm corresponds to a single-block BSUM algorithm. Notice that using the BSUM framework we can easily generalize the IRLS to the multiblock scenario.

APPLICATIONS OF THE BSUM FRAMEWORK

Here we briefly review a few applications of the BSUM framework in wireless communication, bioinformatics, signal processing, and machine learning.

WIRELESS COMMUNICATION AND TRANSCEIVER DESIGN

Consider a multiple-input, multiple-output interference channel with K transmitter-receiver pairs. Let M (resp. N) be the number of antennas at each transmitter (resp. receiver) and each transmitter k, $k = 1, 2, \ldots, K$, is interested in transmitting one data stream to its own receiver. Let $\mathrm{x}_k \in \mathbb{C}^M$ be the transmitted signal of user k; assuming linear channel model, the received signal of user k can be written as

$$\mathrm{y}_k = \underbrace{\mathrm{H}_{kk}\mathrm{x}_k}_{\text{desired signal}} + \underbrace{\sum_{j \neq k} \mathrm{H}_{kj}\mathrm{x}_j}_{\text{multiuser interference}} + \underbrace{\mathrm{n}_k}_{\text{noise}},$$

where $\mathbf{H}_{kj} \in \mathbb{C}^{N \times M}$ is the channel from transmitter j to receiver k and $\mathbf{n}_k \in \mathbb{C}^N$ denotes the additive white Gaussian noise at the receiver k with distribution $\mathcal{CN}(0, \sigma^2 \mathbf{I})$.

When linear beamformers are employed at the transmitters and receivers, the transmitted signal and the estimated received data stream can be respectively written as

$$\mathbf{x}_k = \mathbf{v}_k s_k$$

and

$$\hat{s}_k = \mathbf{u}_k^H \mathbf{y}_k,$$

where $\mathbf{v}_k \in \mathbb{C}^M$ and $u_k \in \mathbb{C}^N$ are, respectively, the transmit and receive beamformers. Here the transmitted data stream and the estimated data stream at the receiver are denoted by $s_k \in \mathbb{C}$ and $\hat{s}_k \in \mathbb{C}$, respectively.

A crucial task in modern wireless networks is to design the transmit and receive beamformers $\mathbf{v}_k$ and $\mathbf{u}_k$ to maximize a given utility of the system. Here, for simplicity of presentation, we consider the sum rate utility function as our objective. Therefore, our goal is to solve the following optimization problem:

$$\begin{aligned} \max_{\mathbf{u},\mathbf{v}} \quad & \sum_{k=1}^{K} R_k(\mathbf{u}, \mathbf{v}) \\ \text{s.t.} \quad & \|v_k\|^2 \leq P_k, \quad \forall k = 1, 2, \ldots, K, \end{aligned} \tag{19}$$

where P_k is the total power budget of user k and $R_k(\mathbf{u}, \mathbf{v})$, which is the communication rate of user k, is given by

$$R_k(\mathbf{u}, \mathbf{v}) = \log\left(1 + \frac{|\mathbf{u}_k^H \mathbf{H}_{kk} \mathbf{v}_k|^2}{\sigma^2 \|u_k\|^2 + \sum_{j \neq k} |\mathbf{u}_k^H \mathbf{H}_{kj} \mathbf{v}_j|^2}\right).$$

Problem (19) is nonconvex and known to be NP-hard [58]. Using the well-known relation between the signal-to-interference-plus-noise ratio (SINR) and the mean square error (MSE) value, one can rewrite (19) as [12], [59]:

$$\begin{aligned} \min_{\mathbf{v},\mathbf{u}} \quad & \sum_{k=1}^{K} \log(e_k(\mathbf{u}, \mathbf{v})) \\ \text{s.t.} \quad & \|\mathbf{v}_k\|^2 \leq P_k, \quad \forall k = 1, 2, \ldots, K, \end{aligned} \tag{20}$$

where $e_k(\mathbf{u}, \mathbf{v})$ is the MSE value and is given by

$$e_k(\mathbf{u}, \mathbf{v}) = |\mathbf{u}_k^H \mathbf{H}_{kk} \mathbf{v}_k - 1|^2 + \sum_{j \neq k} |\mathbf{u}_k^H \mathbf{H}_{kj} \mathbf{v}_j|^2 + \sigma^2.$$

Since the $\log(\cdot)$ function is concave, it is upper bounded by its first-order approximation (i.e., the linear upper bound in Table 3). Therefore, we can define the function

$$u(\mathbf{x}, \mathbf{x}^r) = \sum_{k=1}^{K} (\log(e_k(\mathbf{x}^r)) + (e_k(\mathbf{x}^r))^{-1} (e_k(\mathbf{x}) - e_k(\mathbf{x}^r))), \tag{21}$$

where $x \triangleq (\mathbf{u}, \mathbf{v})$ is the optimization variable and $\mathbf{x}^r \triangleq (\mathbf{u}^r, \mathbf{v}^r)$ denotes the beamformer at iteration r. It is not hard to see that the approximation function in (21) is a valid upper bound in the

BSUM framework and at each iteration r, this choice of approximation function leads to a quadratic programming problem which has closed-form solutions. The resulting algorithm, dubbed WMMSE, converges to a stationary point of the problem and, in practice, it typically converges in a few iterations [12] even for larger-size problems [60].

For more details of the algorithm and its extensions to various beamformer design scenarios and different utility functions, refer to [12], [20], [59], [61], and [62]. It is also worth noting that many other interesting transceiver design algorithms also fall into the BSUM framework; see [33] and [63]–[65] for more details.

BIOINFORMATICS AND SIGNAL PROCESSING

Here we briefly outline two interesting big data applications of the BSUM framework in bioinformatics and signal processing.

ABUNDANCE ESTIMATION IN MODERN HIGH-THROUGHPUT SEQUENCING TECHNOLOGIES

An essential step in the analysis of modern high throughput sequencing technologies of biological data is to estimate the abundance level of each transcript in the experiment. Mathematically, this problem can be stated as follows. Consider M transcript sequences $s_1,\cdots,s_M \in \{A,C,G,T\}^L$ with the corresponding abundance levels $\rho_1,\cdots,\rho_M$ such that $\sum_{m=1}^{M}\rho_m = 1$. Let $R_1,\cdots,R_N$ be noisy sequencing reads originated from the transcript sequences, where each read R_n, $n=1,\cdots,N$, is originated from only one of the transcript sequences $s_1,\cdots,s_M$. Given the observed reads, the likelihood of the abundance levels $\rho_1,\cdots,\rho_M$ can be written as

$$\begin{aligned}
&\Pr(R_1,\ldots,R_N;\rho_1,\ldots,\rho_M)\\
&\quad=\prod_{n=1}^{N}\Pr(R_n;\rho_1\ldots\rho_M)\\
&\quad=\prod_{n=1}^{N}\left(\sum_{m=1}^{M}\Pr(R_n\,|\,\text{read } R_n \text{ from sequence } s_m)\Pr(s_m)\right)\\
&\quad=\prod_{n=1}^{N}\left(\sum_{m=1}^{M}\alpha_{nm}\rho_m\right),
\end{aligned}$$

where $\alpha_{nm} \triangleq \Pr(R_n\,|\,\text{read } R_n \text{ from sequence } s_m)$ can be obtained efficiently using an alignment algorithm such as the ones based on the Burrows-Wheeler transform; see, e.g., [66] and [67]. Therefore, given $\{\alpha_{nm}\}_{n,m}$, the maximum likelihood estimation of the abundance levels can be stated as

$$\hat{\rho}_{ML} = \arg\min_{\rho} \; -\sum_{n=1}^{N}\log\left(\sum_{m=1}^{M}\alpha_{nm}\rho_m\right)$$

$$\text{s.t.}\quad \sum_{m=1}^{M}\rho_m = 1, \quad \text{and} \quad \rho_m \geq 0, \forall m = 1,\ldots,M. \tag{22}$$

As a special case of the EM algorithm, a popular approach for solving this optimization problem is to successively minimize a local tight upper bound of the objective function. In particular, the eXpress software [4] solves the following optimization problem at the rth iteration of the algorithm:

$$\begin{aligned}
\rho^{r+1} = \arg\min_{\rho}\;&-\sum_{n=1}^{N}\left(\sum_{m=1}^{M}\left(\frac{\alpha_{nm}\rho_m^r}{\sum_{m'=1}^{M}\alpha_{nm'}\rho_{m'}^r}\log\left(\frac{\rho_m}{\rho_m^r}\right)\right)\right.\\
&\left.+\log\left(\sum_{m=1}^{M}\alpha_{nm}\rho_m^r\right)\right)\\
\text{s.t.}\;\sum_{m=1}^{M}\rho_m = 1, \quad &\text{and} \quad \rho_m \geq 0, \forall m = 1,\ldots,M.
\end{aligned} \tag{23}$$

Using Jensen's inequality, it is not hard to check that (23) is a valid upper bound of (22) in the BSUM framework. Moreover, (23) has a closed-form solution obtained by

$$\rho_m^{r+1} = \frac{1}{N}\sum_{n=1}^{N}\frac{\alpha_{nm}\rho_m^r}{\sum_{m'=1}^{M}\alpha_{nm'}\rho_{m'}^r}, \quad \forall m = 1,\ldots,M,$$

which makes the algorithm computationally efficient at each step. In practice, the above algorithm for abundance estimation converges in a few iterations. Moreover, this algorithm is perfectly suitable for distributed storage and multicore machines. In particular, since the number of reads N is much larger than the number of sequences M, one can store the reads $R_1,\ldots,R_N$ in n_p different processing units. Hence, at each iteration r, the processing unit p, $p=1,\ldots,n_p$, can compute the local value

$$\hat{\rho}_{m,p}^{r+1} = \frac{1}{N}\sum_{n\in\mathcal{N}_p}\frac{\alpha_{nm}\rho_m^r}{\sum_{m'=1}^{M}\alpha_{nm'}\rho_{m'}^r}, \quad \forall m = 1,\ldots,M,$$

where $\mathcal{N}_p$ is the set of reads stored at processor p with $\cup_{p=1}^{n_p}\mathcal{N}_p = \{1,2,\ldots,N\}$. Then, all processors update their global abundance estimate through the consensus procedure

$$\rho_m^{r+1} = \sum_{p=1}^{n_p}\hat{\rho}_{m,p}^{r+1}, \quad \forall m = 1,\ldots,M.$$

For a very recent application of BSUM algorithm in gene RNA-seq abundance estimation, see [68].

TENSOR DECOMPOSITION

The CANDECOMP/PARAFAC (CP) decomposition has applications in different areas such as clustering [69] and compression [70]. For ease of presentation, here we only consider third-order tensors. Given a third-order tensor $\mathcal{X} \in \mathbb{R}^{m_1\times m_2\times m_3}$, its rank R CP decomposition is given by $\mathcal{X} = \sum_{r=1}^{R}\mathbf{a}_r\circ\mathbf{b}_r\circ\mathbf{c}_r$, where $\mathbf{a}_r\in\mathbb{R}^{m_1}$, $b_r\in\mathbb{R}^{m_2}$, $c_r\in\mathbb{R}^{m_3}$; and the notation "$\circ$" stands for the outer product operator.

In general, finding the CP decomposition of a given tensor is NP-hard [71]. A well-known algorithm for finding the CP decomposition is the alternating least squares (ALS) algorithm proposed in [72] and [73]. This algorithm is, in essence, the BCD algorithm on the following optimization problem:

$$\min_{\{\mathbf{a}_r,\mathbf{b}_r,\mathbf{c}_r\}_{r=1}^{R}}\left\|\mathcal{X}-\sum_{r=1}^{R}\mathbf{a}_r\circ\mathbf{b}_r\circ\mathbf{c}_r\right\|_F^2. \tag{24}$$

In the ALS algorithm, we consider three blocks of variables: $\{\mathbf{a}_r\}_{r=1}^{R}$, $\{\mathbf{b}_r\}_{r=1}^{R}$, and $\{\mathbf{c}_r\}_{r=1}^{R}$. At each iteration of the algorithm, two blocks are held fixed and only one block is updated by solving (24). The

block selection rule is cyclic and, therefore, one needs the uniqueness of the minimizer assumption at each iteration for theoretical convergence guarantee. Clearly, this assumption does not hold for (24), therefore convergence is not always guaranteed. In addition, another well-known drawback of the ALS algorithm is the "swamp" effect where the objective remains almost constant for many iterations and then starts decreasing again. It has been observed in the literature that the employment of proximal upper bound (see Table 3) could help reduce the swamp effect [74]. It is also suggested in [74] that decreasing the proximal coefficient (γ in Table 3) during the ALS algorithm could further improve the performance of the algorithm. Notice that these modifications in the algorithm makes the algorithm a special case of BSUM framework. Consequently, its theoretical convergence is also guaranteed by Theorem 1.

Figure 5 compares the performance of the naive ALS algorithm with the one using proximal upper bound. The figure shows that the proximal ALS algorithm has less swamp effect as compared to the naive ALS method. For more details of the algorithm, refer to [21] and [74]; and to [75] for the application of BSUM and CP decomposition in gene expression and brain imaging.

MACHINE LEARNING: SPARSE DICTIONARY LEARNING AND SPARSE LINEAR DISCRIMINANT ANALYSIS

DICTIONARY LEARNING FOR SPARSE REPRESENTATION

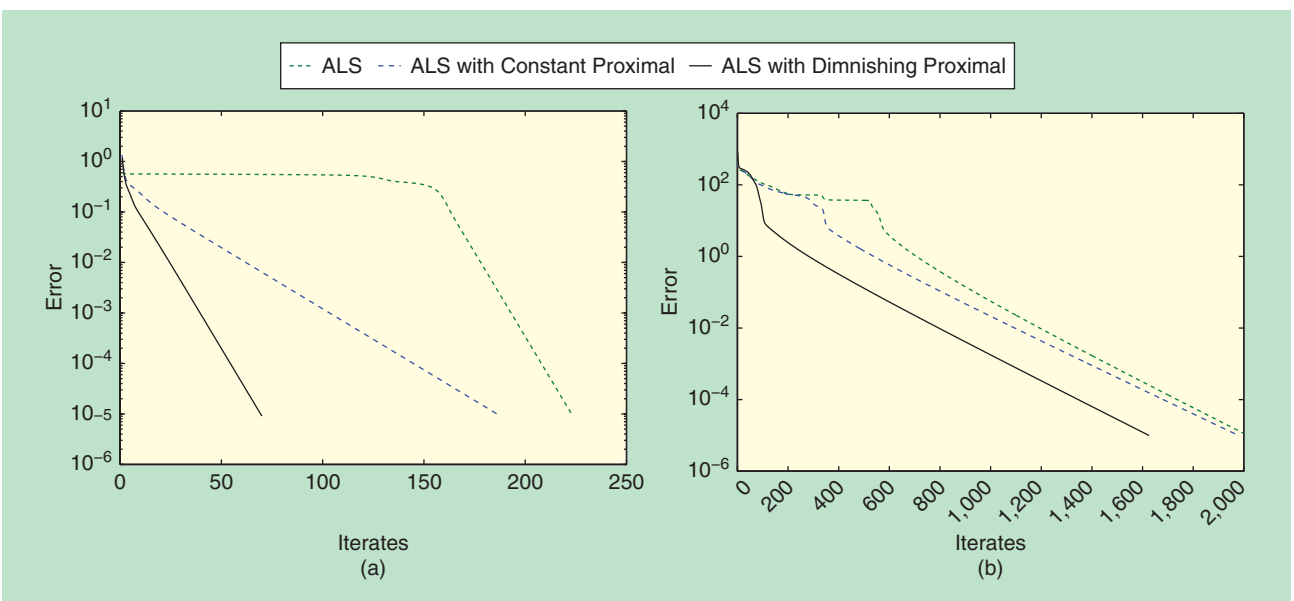
In compressive sensing [76], [77] problems, a given data signal is represented by sparse linear combination of the signals in a given set called a *dictionary*. In many applications even the dictionary is not known a priori, therefore, it should be learned from the data. More precisely, given a set of training signals $\{\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^n\}$, the dictionary-learning task is to find a dictionary set $\{\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^n\}$ that can sparsely represent the signals in the training set. Defining the matrices $\mathbf{Y} \triangleq [\mathbf{y}_1, \ldots, \mathbf{y}_N]$, $\mathbf{A} \triangleq [\mathbf{a}_1, \ldots, \mathbf{a}_k]$, and $\mathbf{X} \triangleq [\mathbf{x}_1, \ldots, \mathbf{x}_N]$, the dictionary-learning problem can be written as [78], [79]

$$\begin{aligned} \min_{\mathbf{A},\mathbf{X}} \quad & d(\mathbf{Y}, \mathbf{A}, \mathbf{X}) \\ \text{s.t.} \quad & \mathbf{A} \in \mathcal{A}, \mathbf{X} \in \mathcal{X}, \end{aligned} \tag{25}$$

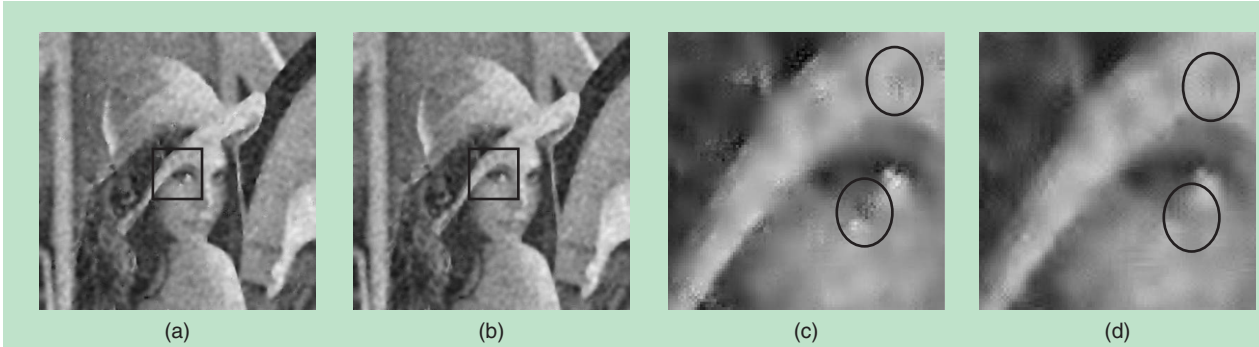
where the sets $\mathcal{X}$ and $\mathcal{A}$ are given based on the prior knowledge on the data. The function $d(\cdot, \cdot, \cdot)$ measures the goodness-of-fit of the model. For example, a popular choice of the function $d(\cdot, \cdot, \cdot)$ and the set $\mathcal{A}$ leads to the following optimization problem [79]:

$$\begin{aligned} \min_{\mathbf{A},\mathbf{X}} \quad & \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|^2 + \lambda \|\mathbf{X}\|_1 \\ \text{s.t.} \quad & \|\mathbf{a}_i\|_F^2 \le \beta_i, \end{aligned}$$

where the first term in the objective keeps our estimated signals close to the training set and the second term forces the representation to be sparse. One popular approach in the dictionary-learning algorithm is to alternatingly update the dictionary $\mathbf{A}$ and the coefficients $\mathbf{X}$ [80]. However, naively updating these two variables to its global optimum requires solving a sparse recovery problem at each iteration, which is costly for larger-sized problems. Motivated by the idea of inexact steps in the BSUM framework, one can iteratively replace the objective by a locally tight upper bound, which is easier to minimize at each iteration and, hence, leads to computationally cheaper steps in the algorithm. It is not hard to see that utilizing the quadratic upper bound in Table 3 with diagonal matrices Φ_i leads to closed-form updates at each step [78]. Unlike many existing algorithms in the literature [79], [80], the resulting algorithm is guaranteed to converge theoretically to the set of stationary solutions as the result of Theorem 1.



[FIG5] A comparison of the ALS and proximal ALS algorithm [21]. The proximal ALS algorithm is the BSUM approach using the proximal upper bound; see Table 3. The "ALS with diminishing proximal" algorithm utilizes a decreasing proximal coefficient during the iterates of the algorithm. In the example $m_1 = m_2 = m_3 = R = 100$. (Figure reprinted with permission from the Society for Industrial and Applied Mathematics.)



[FIG6] **Sample denoised images ($\sigma = 100$). (a) K-SVD, (b) proposed algorithm, (c) K–SVD (zoomed in), and (d) proposed algorithm (zoomed in) [78].**

Figure 6 and Table 5 show the performance of the resulting algorithm for dictionary learning in an image denoising problem. The denoising is performed on the Lena image corrupted by additive Gaussian noise with various variances σ^2. As can be seen from Table 5, the proposed algorithm results in larger PSNR values than the K-SVD method [80] when the noise level is large. Moreover, the proposed algorithm contains less visual artifacts. Furthermore, each step of the proposed algorithm is in closed form and is computationally favorable, while each step of the K-SVD method requires an inner iterative method.

[TABLE 5] THE IMAGE DENOISING RESULT COMPARISON ON THE "LENA IMAGE" FOR DIFFERENT NOISE LEVELS. VALUES ARE AVERAGED OVER TEN MONTE CARLO SIMULATIONS [78].

σ /PSNR	DCT	K-SVD	PROPOSED ALGORITHM
20/22.11	32	32.38	30.88
60/12.57	26.59	26.86	26.37
100/8.132	24.42	24.45	24.46
140/5.208	22.96	22.93	23.11
180/3.025	21.73	21.69	21.96

SPARSE LINEAR DISCRIMINANT ANALYSIS

The linear discriminant analysis (LDA), which is closely related to analysis of variance (ANOVA) and regression analysis, is widely used in machine learning and statistics for classification and dimensionality reduction purposes; see, e.g., [81]. Let us, for the ease of presentation, focus only on the binary classification problem: Let $\mathbf{x}_i \in \mathbb{R}^p$, $i = 1, 2, \ldots, N$, denote the zero-centered observations, where each observation x_i belongs to one and only one of the two classes C_0 and C_1. Given the binary classes, the standard within-class covariance estimate can be calculated by

$$\hat{\Sigma}_w = \frac{1}{N} \sum_{k \in \{0,1\}} \sum_{i \in C_k} (\mathbf{x}_i - \hat{\mu}_k)(\mathbf{x}_i - \hat{\mu}_k)^T,$$

where $\hat{\mu}_k = 1/N \sum_{i \in C_k} \mathbf{x}_i$ is observations mean in class C_k. Similarly, the standard between-class covariance estimate is given by

$$\hat{\Sigma}_b = \frac{1}{N} (N_0 \hat{\mu}_0 \hat{\mu}_0^T + N_1 \hat{\mu}_1 \hat{\mu}_1^T),$$

with N_0 (resp. N_1) being the cardinality of the set C_0 (resp. C_1). The goal of LDA is to find a lower-dimensional subspace so that the projection of the observations onto the selected subspace leads to well-separated classes. In other words, the task is to project data points into a subspace with large between-class variance relative to the within-class variance. For simplicity, consider projection onto one-dimensional subspace defined by the vector $\beta \in \mathbb{R}^p$; see [82] for details on projection to larger-dimensional subspaces. Then the inner product $\langle \beta, x \rangle$ is the projection of the observation $\mathbf{x}$ onto the selected subspace; and the within-class variance of the projected data points is given by $\hat{\sigma}_w = \beta^T \hat{\Sigma}_w \beta$; while the between class variance can be written as $\hat{\sigma}_b = \beta^T \hat{\Sigma}_b \beta$. Therefore, in the LDA problem, we are interested in solving

$$\begin{aligned} \max_{\beta} \quad & \beta^T \hat{\Sigma}_b \beta \\ \text{s.t.} \quad & \beta^T \hat{\Sigma}_w \beta \le 1. \end{aligned} \tag{25}$$

Unfortunately, when the number of features is large relative to N, the matrix $\hat{\Sigma}_w$ is rank deficient and therefore (25) is ill posed. To resolve this issue and to have a small generalization error, [82] suggests to regularize the optimization problem with a convex penalty function $P(\cdot)$; and solve

$$\begin{aligned} \max_{\beta} \quad & \beta^T \hat{\Sigma}_b \beta - P(\beta) \\ \text{s.t.} \quad & \beta^T \hat{\Sigma}_w \beta \le 1. \end{aligned} \tag{26}$$

Clearly, this optimization problem is nonconvex. As suggested in [82], one can linearize the first part of the objective in (26) iteratively to obtain a tight upper bound of the objective. It is not hard to see that the algorithm used in [82] is BSUM with the linear upper bound given in Table 3.

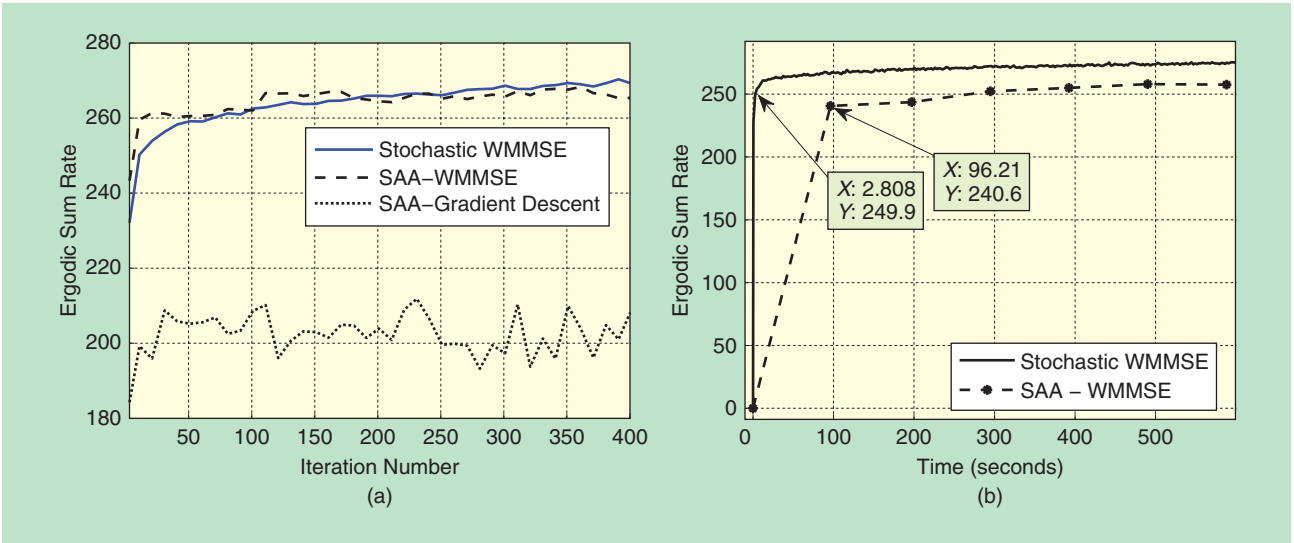
EXTENSIONS

In this section, we discuss extensions and generalizations of the BSUM framework in various settings.

STOCHASTIC OPTIMIZATION

Consider the following stochastic optimization problem:

$$\begin{aligned} \min_{x} \quad & f(x) \triangleq \mathbb{E}_{\xi}[g(x, \xi)] \\ \text{s.t.} \quad & x \in \mathcal{X}, \end{aligned} \tag{27}$$



[FIG7] **An iteration and running time comparison of SSUM versus SAA [20].**

where $\mathcal{X}$ is a closed convex set and ξ is a random variable modeling the uncertainty in our optimization problem. A standard classical approach for solving (27) is the sample average approximation (SAA) method; see [83] and the references therein. At iteration r of the SAA method, given a new realization ξ^r of the random variable ξ, the SAA method generates a new iterate x^r by solving a problem with the sample average $1/r\sum_{i=1}^{r} g(x,\xi^i)$ as its objective, where $\xi^1, \xi^2, \ldots, \xi^r$ are independent identically distributed realizations of the random variable ξ.

A major drawback of the SAA method is that each of its iteration can be computationally very expensive. The computational inefficiency arises from either the nonconvexity of the objective, or not having closed-form solutions at each iteration.

Motivated by the BSUM framework, the authors of [20] and [84] suggest using an inexact version of the SAA method, in which a sequence of upper bounds of the objective are minimized. In particular, at each iteration r, the optimization variable is updated by

$$x^r \in \arg\min_{x} \quad \frac{1}{r}\sum_{i=1}^{r} \hat{g}(x, x^{i-1}, \xi^i) \quad \text{s.t.} \quad x \in \mathcal{X},$$

where $\hat{g}(\cdot, x^{i-1}, \xi^i)$ is an upper bound of the function $g(\cdot, \xi^i)$ around the point x^{i-1}. The approximation function $\hat{g}(\cdot, x^{i-1}, \xi^i)$ is assumed to be in the form of the BSUM approximation. The resulting algorithm, named *stochastic successive upper-bound minimization* (SSUM), is guaranteed to converge to the set of stationary solutions almost surely; see [20] for more details. Further, it is shown to be capable of dealing with various practical problems in signal processing and machine learning. For example, as we will see shortly, the authors of [85] apply the SSUM framework to cope with uncertainties in channel estimation for a wireless beamformer design problem. As another example, the online sparse dictionary-learning algorithm proposed in [86] is a special case of SSUM.

The stochastic optimization framework is well suited for many modern big data applications, especially when the entire data set is not available initially and the data points are made available over time. These problems can be considered as the previously mentioned general stochastic optimization problem; see also [87]. In addition, many statistical model fitting problems, such as the simple classical regression problem, can be cast as minimizing the following sum-cost function $\sum_{\ell=1}^{L} g(x,\xi_\ell)$. Typically, the number of data points L is very large, making it difficult for batch processing. Therefore, it is desirable to implement algorithms working with only one (or a few) data point(s) at each step. In these scenarios, the stochastic optimization framework is useful since the sum-cost minimization problem can be viewed as a stochastic optimization problem $\min \mathbb{E}[g(x,\xi)]$ with ξ being uniformly drawn from the set $\{\xi_1, \ldots, \xi_L\}$.

As an example of the SSUM method, consider the wireless transceiver design problem discussed in the "Wireless Communication and Transceiver Design" section, where the channel coefficients $\{\mathbf{H}_{ij}\}_{i,j}$ are not exactly known. In this scenario, we can consider the channel coefficients as random variables and solve the following stochastic optimization problem:

$$\max_{\mathbf{u},\mathbf{v}} \quad \sum_{k=1}^{K} \mathbb{E}[R_k(\mathbf{u},\mathbf{v})] \quad \text{s.t.} \quad \|\mathbf{v}_k\|^2 \le P_k, \quad \forall k = 1,2,\ldots,K, \tag{28}$$

which is the stochastic counterpart of the optimization problem (19). Utilizing the upper bound (21) in the SSUM algorithm leads to the stochastic WMMSE algorithm [85]. Figure 7 illustrates the numerical performance of the SSUM methods as compared with SAA. At each iteration of the SAA procedure, one should solve a nonconvex optimization problem. Two different methods are considered: the gradient descent method with random initialization and the WMMSE algorithm, which is known to converge in few iterations for this problem. As Figure 7 illustrates, the running time of the SAA algorithm is much longer than that of the SSUM.

COUPLING CONSTRAINTS

So far in this article, we have assumed that the constraints in the optimization problem is separable and convex. In other words, the constraint set $\mathcal{X}$ in (1) is of the form $\mathcal{X} = \mathcal{X}_1 \times \ldots \mathcal{X}_n$ with each $\mathcal{X}_i$ being convex. A natural extension of the BSUM framework is to modify it to deal with coupling and nonconvex constraints.

LINEAR COUPLING

Consider the following convex problem with linear coupling constraints

$$\begin{aligned} \min_{(x_1,\ldots,x_n)} \quad & f(x_1,\ldots,x_n) \\ \text{s.t.} \quad & \sum_{i=1}^{n} A_i x_i = b \\ & x_i \in \mathcal{X}_i, \ \forall i = 1,2,\ldots,n, \end{aligned} \tag{29}$$

where $x_i \in \mathbb{R}^{m_i}$, $A_i \in \mathbb{R}^{k \times m_i}$, $b \in \mathbb{R}^k$ and $f(\cdot)$ is a convex function. As seen in Example 4, the direct extension of the BCD/BSUM approach does not work for this type problem. A popular approach for solving the above optimization problem is the alternating direction method of multipliers (ADMM) [88], [89]. This approach is based on finding a saddle point of the augmented Lagrangian function

$$L(x_1,\ldots,x_n;\lambda) = f(x_1,\ldots,x_n) + \left\langle \lambda, \sum_{i=1}^{n} A_i x_i - b \right\rangle + \frac{\rho}{2} \left\| \sum_{i=1}^{n} A_i x_i - b \right\|^2,$$

where $\lambda \in \mathbb{R}^k$ is the Lagrange multiplier corresponding to the linear constraint; $\rho > 0$ is the augmented Lagrangian coefficient; and $\langle \cdot, \cdot \rangle$ denotes the inner product operator.

At each iteration of the ADMM method, either a primal block variable x_i is updated according to

$$x_i^{r+1} \leftarrow \arg\min_{x_i \in \mathcal{X}_i} \ L(x_i, x_{-i}^r; \lambda^r),$$

or the dual Lagrange multiplier λ is updated according to the gradient ascent rule

$$\lambda^{r+1} \leftarrow \lambda^r + \alpha^r \left(\sum_{i=1}^{n} A_i x_i^r - b \right),$$

where α^r is the dual step size at iteration r. The update orders for the primal and dual variables could be either cyclic or randomized.

Similar to the BSUM framework, one can replace the augmented Lagrangian function $L(\cdot, x_{-i}^r; \lambda^r)$ with its tight upper bound $\tilde{L}_i(\cdot, x^r; \lambda^r)$ at iteration $r+1$, where

$$\tilde{L}_i(x_i, x^r, \lambda^r) = u_i(x_i, x^r) + \left\langle \lambda^r, A_i x_i + \sum_{j \neq i} A_j x_j^r - b \right\rangle + \frac{\rho}{2} \left\| A_i x_i + \sum_{j \neq i} A_j x_j^r - b \right\|^2$$

with $u_i(\cdot, x^r)$ being a locally tight approximation of the function $f(\cdot, x_{-i}^r)$ around the point x_i^r satisfying Assumption A. The resulting algorithm, named the *BSUM method of multipliers* (*BSUMM*) [44], is guaranteed to converge to the global optimal of (29) under some regularity assumptions [44]. For extensions to nonconvex problems, see the recent work [90] and [91]. There are a few other interesting techniques that deal with linearly coupling constraint. For example, [92] and [93] propose to randomly pick two blocks of variables to update at each iteration, and [94] proposes new algorithms based on minimizing the augmented Lagrangian function. We refer the readers to these papers for more related works in this direction.

Let us illustrate an application of BSUMM to a multicommodity routing problem, which arises in the design of next-generation cloud-based communication networks [95]. Consider a connected wireline network $\mathcal{N} = (\mathcal{V}, \mathcal{L})$ that is controlled by $K+1$ network controllers (NCs) as illustrated in Figure 8. Let $\mathcal{V}$ denote the set of network nodes, which is partitioned into K subsets, i.e., $\mathcal{V} = \cup_{i=1}^{K} \mathcal{V}^i$, $\mathcal{V}^i \cap \mathcal{V}^j = \emptyset$, $\forall i \neq j$. The set of directed links that connect nodes of $\mathcal{V}$ is denoted as $\mathcal{L} \triangleq \{l = (s_l, d_l) | \ \forall s_l, d_l \in \mathcal{V}\}$, where $l = (s_l, d_l)$ denotes the directed link from node s_l to node d_l. Each NC i controls $\mathcal{V}^i$ and the links connecting these nodes, i.e., $\mathcal{L}^i \triangleq \{l = (s_l, d_l) \in \mathcal{L} | \ \forall s_l, d_l \in \mathcal{V}^i\}$, (cf. Figure 8). The network $\mathcal{N}^i \triangleq (\mathcal{V}^i, \mathcal{L}^i)$ is called the subnetwork i.

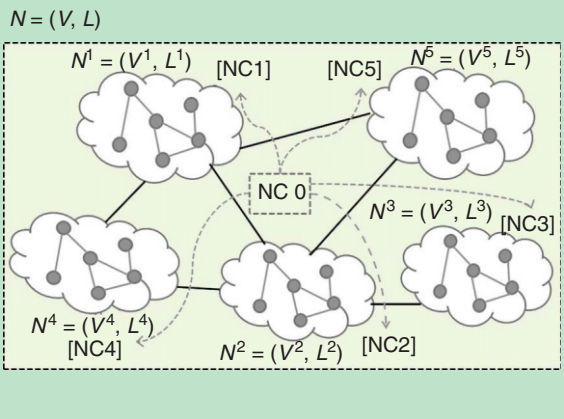
Our objective is to transport M data flows over the network, with each flow m being routed from a source node $s_m \in \mathcal{V}$ to the sink node $d_m \in \mathcal{V}$. We use $r_m \geq 0$ to denote flow m's rate, and use $f_{l,m} \geq 0$ to denote its rate on link $l \in \mathcal{L}$. We also assume that a master node exists, which controls the data flow rates $\{r_m\}_{m=1}^{M}$. The central NC 0 controls the subnetwork $\mathcal{N}^0$, consisting of the master node and the links connecting different subnetworks, i.e., $\mathcal{L}^0 = \cup_{i \neq j} \mathcal{L}_{ij}^0$.

We consider two types of network constraints:

1) *Link capacity constraints*. Assume each link $l \in \mathcal{L}$ has a fixed capacity denoted as C_l. The total flow rate on link l is constrained by

$$1^T \mathbf{f}_l \leq C_l, \ \forall l \in \mathcal{L}, \tag{30}$$

where 1 is the all-one vector and $\mathbf{f}_l \triangleq [f_{l,1}, \ldots, f_{l,M}]^T$.



[FIG8] A wireline network consists of five subnetworks. Each of them is controlled by an NC, and these NCs are coordinated globally by a central NC 0 [95].

[TABLE 6] A COMPARISON OF DIFFERENT ALGORITHMS. $M = 200$.

APPROACHES	TIME	NUMBER OF ITERATIONS
GUROBI	11.4690 s	N/A
SYNCHRONOUS BSUMM (TEN NCs, ONE CORE)	18.0679 s	264.71
SYNCHRONOUS BSUMM (TEN NCs, TEN CORE)	5.80 s	264.71
ASYNCHRONOUS BSUMM (TEN NCs, TEN CORE)	4.23 s	2109.07

2) *Flow conservation constraints*. For any node $v \in \mathcal{V}$ and data flow m, the total incoming flow should be equal to the total outgoing flow:

$$\sum_{l \in \text{In}(v)} f_{l,m} + 1_{v=s(m)} r_m = \sum_{l \in \text{Out}(v)} f_{l,m} + 1_{v=d(m)} r_m, \quad m = 1, \cdots, M, \ \forall v \in \mathcal{V}, \tag{31}$$

where $\text{In}(v) \triangleq \{l \in \mathcal{L} | d_l = v\}$ and $\text{Out}(v) \triangleq \{l \in \mathcal{L} | s_l = v\}$ denote the set of links going into and coming out of a node v respectively; $1_{v=x} = 1$ if $v = x$ otherwise $1_{v=x} = 0$.

To provide fairness to the users, we maximize the minimum rate of all data flows. The problem can be formulated as the following linear program (LP)

$$\max_{\mathbf{f}, r} \ r_{\min} \quad \text{s.t.} \ \ \mathbf{f} \geq 0, \ r_m \geq r_{\min}, \ m = 1, \cdots, M \tag{32a}$$

$$(30) \text{ and } (31), \tag{32b}$$

where $\mathbf{f} \triangleq \{f_l | l \in \mathcal{L}\}$ and $r \triangleq \{r_{\min}, r_m | m = 1 \sim M\}$. Obviously, one can use off-the-shelf optimization packages such as Gurobi [96] to solve the LP (32), but this is only viable in a centralized setting where all the flows are managed by a single controller.

To enable distributed/parallel network management across the NCs, we need to allow each NC i to independently optimize the variables belonging to the subnetwork $\mathcal{N}^i$. However, this task is difficult because the optimization variables of (32) is coupled (indeed each flow rate f_m appears in exactly two flow conservation constraints). To address this problem, we introduce a few sets of new variables to decouple the flow conservation constraints across different subnetworks (see [95] for the detailed reformulation). The reformulated problem (32) is given by

$$\begin{aligned} &\max_{\mathbf{x}} \ r_{\min} \\ &\text{s.t.} \{r_{\min}, \mathbf{x}_{02}\} \in \mathcal{X}_0, \ \{\mathbf{x}_{i1}, \mathbf{x}_{i2}\} \in \mathcal{X}_{i1}, \ x_{i3} \in \mathcal{X}_{i2}, \\ &\underbrace{x^i_{01} = x^i_{02}}_{\text{in}\mathcal{N}^0}, \ \underbrace{x_{i1} = x^i_{01}}_{\text{in}\mathcal{N}^i \text{and} \mathcal{N}^0}, \ \underbrace{x_{i2} = x_{i3}}_{\text{in}\mathcal{N}^i}, \ i = 1 \cdots K, \end{aligned}$$

where $\mathcal{X}_0$, $\mathcal{X}_{i1}$, and $\mathcal{X}_{i2}$ are some feasible sets, and $\{r_{\min}, \mathbf{x}_{02}, \mathbf{x}_{01}, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \mathbf{x}_{i3}\}$ are the block variables. By applying the BSUMM, we can obtain a parallel/distributed algorithm. A few remarks about the implementation of this algorithm are:

1) The replication of link/flow variables for links across different subnetworks allows each subnetwork to be considered separately and independently. This feature makes the BSUMM subproblems solvable in parallel. The requirement of the replicated variables being the same as the original variables is enforced by the linear coupling constraints, and they can be satisfied asymptotically as the BSUMM algorithm converges.

2) The subproblems of the proposed BSUMM-based algorithm can be solved by each NC very efficiently. For example, the update of $\{r_{\min}, \mathbf{x}_{02}\}$ can be performed by each NC in closed form; the update of $\{\mathbf{x}_i, \mathbf{x}^i_{01}\}$ can be performed by running the well-known RELAX code [97].

3) A careful implementation of the BSUMM allows the NCs to act asynchronously, in the sense that they do not need to coordinate with each other for computation. Such asynchronous implementation has the potential of greatly improving the computational efficiency.

We illustrate the BSUMM implementation over a mesh wireline network with 126 nodes, which is randomly partitioned into nine subnetworks with 306 directed links within these subnetworks and 100 directed links connecting the subnetworks. The capacities for the links within (resp. between) the subnetworks are uniformly distributed in [50,100] megabits/second (resp. [20,50] megabits/second). All simulation results are averaged over 200 randomly selected data flow pairs and link capacity.

To demonstrate the benefit of parallelization, we also utilize a high-performance computing cluster, and make each computing node to be an NC. We compare a few different approaches for solving (32):

1) use Gurobi [96], a centralized LP solver

2) apply the synchronous BSUMM algorithm, with $K = 10$ NCs; the computation is done by either a single or by ten distributed computing cores

3) apply the asynchronous BSUMM with $K = 10$ NCs; the computation is done in ten distributed computing cores.

Note that the asynchrony in the network arises naturally from the per-node computational delay and network communication delay. In Table 6, we demonstrate the performance of various algorithms when $M = 200$. Clearly, the asynchronous BSUMM with a small number of NCs outperforms all the rest of the algorithms.

The numerical results suggest that appropriate network decomposition and asynchronous implementation are both critical for the fast convergence of BSUMM. In practice, we observe that the network should be decomposed following a few guidelines:

- the computation burden across the subnetworks is well balanced
- the subroutine within the network can achieve its maximum efficiency
- the total number of replicated auxiliary variables is small.

NONCONVEX CONSTRAINTS

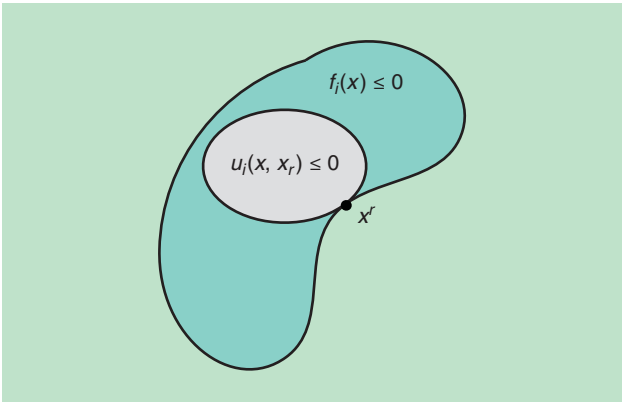
The BSUM idea can be straightforwardly extended to a nonconvex constraint scenario for single block optimization problems. To proceed, consider the optimization problem:

$$\begin{aligned} \min_{x} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, \quad \forall i = 1, 2, \ldots, \ell, \end{aligned} \tag{33}$$

where the functions $f_i(\cdot)$ are not necessarily convex. Since dealing with nonconvex constraints is often not easy, one popular approach is to replace the functions $f_i(x), i = 1, 2, \ldots, \ell$, with their locally tight upper bound $u_i(x, x^r)$ iteratively. In other words, the update rule of the iterates is given by [98]

$$\begin{aligned} x^{r+1} \leftarrow \arg\min_{x} \quad & u_0(x, x^r) \\ \text{s.t.} \quad & u_i(x, x^r) \le 0, \quad \forall i = 1, 2, \ldots, \ell. \end{aligned} \tag{34}$$

As illustrated in Figure 9, the iterative approximation of the constraints is a restriction of the constraints and hence the iterates remain feasible during the algorithm. If, in addition, some constraint qualification conditions are satisfied, the resulting algorithm is guaranteed to converge to the set of stationary solutions of (33); see [38, Th. 1] for detailed conditions and analysis.



[FIG9] **An illustration of constraint convexification.**

PARALLEL VERSION AND EXTENSIONS TO GAME THEORY

With the recent advances in multicore and cluster computational platforms, it is desirable to design "parallel" algorithms for multiblock optimization problem where multiple cores update the block variables in parallel to optimize the objective function. A naive parallel extension of the BCD approach for solving (1) is to update all blocks (or a subset of them) in parallel by solving

$$x_i^{r+1} \leftarrow \arg\min_{x_i \in \mathcal{X}_i} f(x_i, x_{-i}^r), \quad \forall i = 1, \ldots, n.$$

Unfortunately, this naive extension of the BCD algorithm does not converge, in general, and might result in a zigzag/oscillating or divergent behavior. As an example, consider the problem

$$\begin{aligned} \min_{(x_1, x_2)} \quad & (x_1 - x_2)^2 \\ \text{s.t.} \quad & -1 \le x_1, x_2 \le 1. \end{aligned}$$

Clearly, this problem is convex with bounded feasible set and its optimal value is zero. However, the above naive parallel extension of the algorithm leads to the following iteration path:

$$\begin{aligned} (x_1^0, x_2^0) = (1, -1) \rightarrow (x_1^1, x_2^1) = (-1, 1) \rightarrow (x_1^2, x_2^2) \\ = (1, -1) \rightarrow \cdots, \end{aligned}$$

which is clearly not convergent. This is caused by aggressive steps used in the algorithm. To make the algorithm convergent, it is then necessary to employ controlled steps that are also small enough. Furthermore, in the case of nonconvex objective function $f(\cdot)$ in (1), the approximation functions could be again used to obtain computationally efficient update rules. The resulting algorithm, dubbed *parallel successive convex approximation* (*PSCA*), is summarized in Table 7. To see the convergence analysis of this algorithm and other related algorithms such as the flexible parallel algorithm, refer to [18], [99], [100], and the references therein.

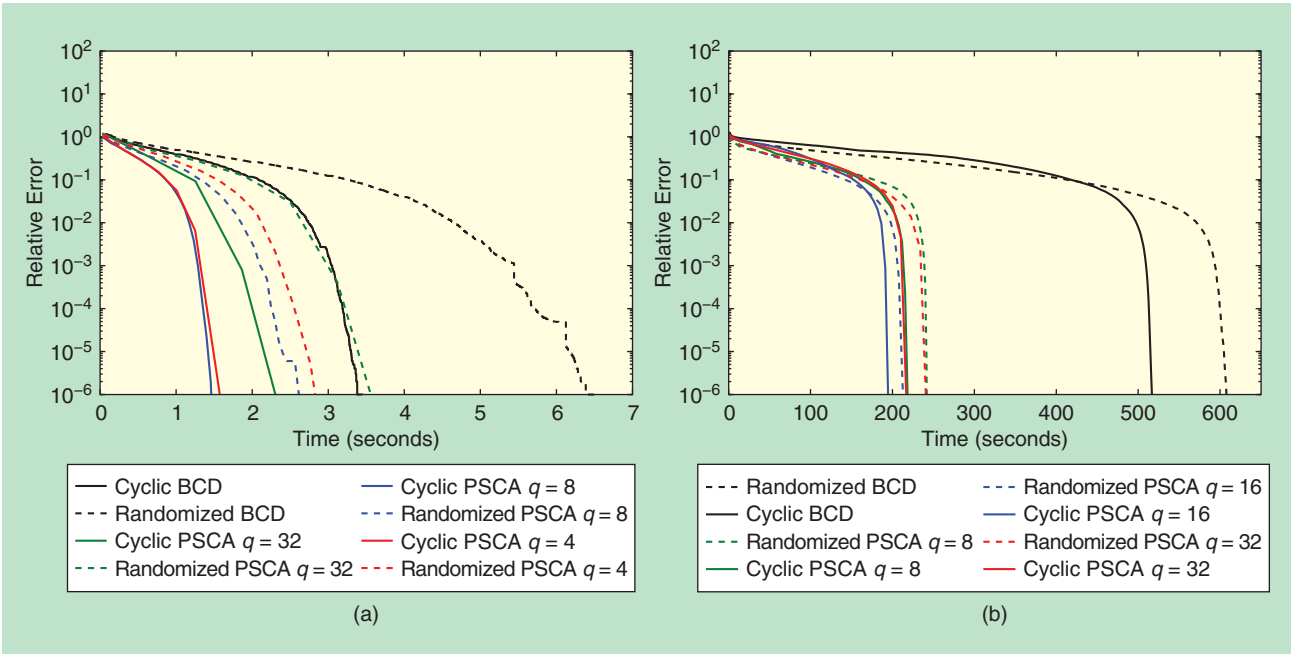
Notice that PSCA can be viewed as a way of solving a multiagent optimization problem where multiple agents/users try to optimize a common objective by updating their own variable iteratively. Furthermore, it can be used in a game-theoretic setting where each player in the game utilizes the best response strategy by optimizing a locally tight upper bound of its own utility function. This algorithm is guaranteed to converge for some particular class of games under some regularity assumptions on the players' utility functions. The convergence analysis presented in [101] is based on certain contraction approach as well as monotone convergence for potential games.

Figure 10 illustrates the behavior of the cyclic and randomized parallel PSCA method as compared with their serial counterparts (i.e., the "cyclic BCD" and the "randomized BCD") applied to the LASSO problem. The performance of the PSCA method is also illustrated for a different number of processors and various block selection rules. In Figure 10(a) and (b), parallelization can result in more efficient algorithm; however, the convergence speed does not grow linearly with the number of processing cores. Moreover, increasing the number of processors beyond certain point results in slower convergence, which can be attributed to the increased communication overhead among the nodes.

Note that the parallel update rule is very useful in dealing with distributed data sets. Consider solving the LASSO problem with the following objective: $\|Ax - b\|^2 + \lambda\|x\|_1$. Assume we have q processing cores each having their own memory. Let us partition the matrix A and vector x into q blocks: $A = [A_1, \ldots, A_q]$ and $x = [x_1^T, \cdots, x_q^T]^T$. If PSCA is implemented in a way that each core j is only responsible for updating

[TABLE 7] A PSEUDOCODE OF THE PSCA ALGORITHM.

1 FIND A FEASIBLE POINT $x^0 \in \mathcal{X}$; SET $r = 0$, AND CHOOSE A STEP-SIZE SEQUENCE $\{\gamma^r\}$
2 **REPEAT**
3 PICK INDEX SET $\mathcal{I}^r$
4 LET $\hat{x}_i^r = \arg\min_{x_i \in \mathbb{X}_i} u_i(x_i, x^{r-1}), \ \forall\, i \in \mathcal{I}^r$
5 SET $x_k^r = x_k^{r-1}, \ \forall k \notin \mathcal{I}^r$
6 SET $x_i^r = x_i^{r-1} + \gamma^r(\hat{x}_i^r - x_i^{r-1}), \ \forall i \in \mathcal{I}^r$
7 $r = r + 1$,
8 **UNTIL** SOME CONVERGENCE CRITERION IS MET



[FIG10] A comparison of the serial BCD with PSCA method for solving the LASSO problem: $\min_x \|Ax-b\|^2+\lambda\|x\|_1$. The matrix A and vector b are generated according to [102]; and q denotes the number of processors used in each experiment. The dimension of A is $2,000\times 10,000$ for (a) and $1,000\times 100,000$ for (b). The experiments are done over a computer cluster using the Message Passing Interface in C [18].

block j, then at each iteration $r+1$, core j's problem of interest can be written as

$$\min \ \|A_j x_j - b_j'\|^2+\lambda\|x_j\|_1,$$

where $b_j' \triangleq b-\sum_{i\neq j} A_i x_i^r$. Notice that the value of b_j' can be calculated by letting each node i compute the value of $A_i x_i$ and broadcast it to other nodes. Under this architecture, each node does not need to know the complete matrix A and only local information is enough for a distributed implementation of the PSCA method.

PRACTICAL CONSIDERATIONS

There are a number of factors that we need to consider when using the BSUM framework. The first consideration is about the choice of the upper bound. What is a good bound for a given application? The answer is generally problem dependent, as we have already seen in a few examples. The general guideline is that a good upper bound should be able to ensure algorithm convergence, best exploit the problem structure, and make the subproblems easily solvable (preferably in closed form). For example, a simple proximal upper bound is not likely to perform well for the transceiver design problem discussed in the section "Wireless Communication and Transceiver Design" as the resulting subproblems will not decompose over the variables.

The second consideration is about the choice of the block update rules. As we have seen in the section "How Fast Does the BSUM Converge?" different update rules lead to quite distinct convergence behavior. For convex problems, deterministic rules such as the cyclic rule promise the worst-case convergence rates, while the randomized rule ensures convergence rate in either averaged or high-probability sense. Further, there is barely any theoretical rate analysis for nonconvex problems, regardless of the block selection rules. Therefore, the best strategy in practice is to perform an extensive numerical study and pick the best rule for the application at hand. For example, researchers have found that the MBI rule is effective for certain tensor decomposition problems [13]; the cyclic rule can be superior to the randomized rule for certain LASSO problems, and certain G-So rules can outperform the randomized rule [43], [14].

The third consideration is about the choice of the parallelization schemes. There has been extensive research on parallelizing various special cases and variations of the BSUM type algorithm; see [17], [19], [99], [103]–[106], and the references therein. These algorithms differ in a number of implementation details and in their applicability. For example, most of the implementations use randomized block selection rules to pick the variable blocks, while [99] and [105] additionally use certain variations of the G-So rule. The majority of the schemes only work for convex problems, with the exception of [99] and [103], which work for general nonsmooth and nonconvex problems in the form of (5). When assessing whether a given problem is suitable for parallelization, it is important to know that oftentimes the number of blocks that can be updated in parallel is data dependent. For example, when solving LASSO problems, it is shown in [105] and [106] that the degree of parallelization is dependent on the maximum eigenvalue of certain submatrices of the data matrix. Some recent results [107] show that for a certain randomized coordinate descent method, such dependency could be mild. For solving general

convex nonsmooth problems, [104] shows that the step size should be carefully selected based on both the "separability" of the problem (or the sparsity of the data matrix) as well as the degree of parallelization. If the application at hand does not satisfy these conditions, the alternatives usually are: 1) exploit the problem structure and pick a good upper bound, so that the subproblems are decomposable, leading to parallel and step-size-free updates; see, for example, the NMF problem in the section "The NMF Algorithm" and the WMMSE algorithm in the section "Wireless Communication and Transceiver Design"; 2) to use the diminishing stepsizes for updating the blocks; see, for example, [19], [99], and [103].

ISSUES AND OPEN RESEARCH PROBLEMS

This article presents a comprehensive algorithmic framework, BSUM, for block-structured large-scale optimization. The main strength of the BSUM framework is its strong theoretical convergence guarantee and its flexibility. As demonstrated in this article, the BSUM framework covers a number of well-known but seemingly unrelated algorithms as well as their new extensions. Moreover, it is amenable to a number of different data models as well as to parallel implementation on modern multicore computing platforms.

To close, we briefly highlight a couple of issues and open research topics related to the BSUM framework.

- *Communication delay and overhead in parallel implementations*: As discussed in the section "Parallel Version and Extensions to Game Theory," the convergence speed of the parallel version of the BSUM framework does not increase linearly with the number of computational nodes. In fact, after a point, increasing the number of computational nodes can lead to a slower convergence speed. As mentioned previously, this is due to the delay caused by communication among the nodes. This observation gives rise to two important research questions: First, given the maximum allowable number of computation nodes and the communication overhead of the nodes, what is the optimum choice of the number of cores for solving a given optimization problem? Answering this question requires computation/communication tradeoff analysis of the proposed optimization approach. Second, can the BSUM framework be extended and implemented in a (semi)asynchronous manner? If this is possible, then the communication overhead can be reduced significantly since the nodes are not required to wait for each other before updating the variables, making the algorithm lock-free. For recent efforts on this research direction see [17].
- *Nonlinear coupling constraints:* As we observe in the "Extensions" section, the BSUM framework can also be used in the presence of linear coupling or nonconvex decoupled constraints. How can the BSUM framework be generalized to problems with nonlinear coupling constraints? More precisely, can the BSUM framework with block-wise update rules be applied to the optimization problem of the following form?

$$\min_{x} \quad f_0(x_1, \ldots, x_n)$$

$$\text{s.t.} \quad f_i(x_1, \ldots, x_n) \leq 0, \forall i = 1, 2, \ldots, n.$$

Example 4 shows that the naive extension of the BCD approach fails to find the optimal solution even in the convex setting. A popular approach to tackle the aforementioned problem is to place the constraints in the objective using Lagrange multipliers and update the multipliers iteratively. However, this approach typically leads to double-loop algorithms and requires subgradient steps in the dual space, which is known to be slow.

ACKNOWLEDGMENTS

M. Hong and M. Razaviyayn contributed equally to this work. M. Hong is supported by the National Science Foundation, grant CCF-1526078. Z.-Q. Luo is supported by the National Science Foundation, grant CCF-1526434.

AUTHORS

Mingyi Hong (mingyi@iastate.edu) received his B.E. degree in communications engineering from Zhejiang University, China, in 2005; his M.S. degree in electrical engineering from Stony Brook University, New York, in 2007; and his Ph.D. degree in systems engineering from the University of Virginia in 2011. He was previously a research assistant professor with the Department of Electrical and Computer Engineering, University of Minnesota, Twin Cities. He is currently an assistant professor with the Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames. His research interests include signal processing, wireless communications, large-scale optimization and its applications in compressive sensing, complex networks, and high-dimensional data analysis.

Meisam Razaviyayn (meisamr@stanford.edu) received his B.S. degree in electrical engineering from Isfahan University of Technology, Iran, in 2008; his M.S. degrees in mathematics and electrical engineering from the University of Minnesota in 2013; and his Ph.D. degree in electrical engineering from the University of Minnesota in 2014. He is currently a postdoctoral research fellow in the Electrical Engineering Department at Stanford University, California. His research interests include large-scale optimization, machine learning and statistics signal processing, and wireless communication and its computational issues.

Zhi-Quan Luo (luozq@umn.edu) received his B.S. degree in applied mathematics in 1984 from Peking University, Beijing, China. He received his Ph.D. degree in operations research in 1989 from the Massachusetts Institute of Technology. From 1989 to 2003, he held a faculty position with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, where he eventually became the department head and held a Canada research chair in information processing. Since 2003, he has been with the Department of Electrical and Computer Engineering at the University of Minnesota, Twin Cities as a full professor and holds an endowed ADC chair in digital technology. Since 2014, he has been with the Chinese University of Hong Kong (Shenzhen) as the vice president (academic). His research interests include optimization algorithms, signal processing, and digital communication.

Jong-Shi Pang (jongship@usc.edu) received his B.S. degree in mathematics from National Taiwan University in 1973 and his

Ph.D. degree in operations research from Stanford University, California, in 1977. He is the Epstein family chair professor in the Department of Industrial and Systems Engineering, University of Southern California. He was the head and Caterpillar professor in the Department of Industrial and Enterprise Systems Engineering at the University of Illinois at Urbana-Champaign. He was previously the Margaret A. Darrin Distinguished Professor in applied mathematics at Rensselaer Polytechnic Institute in Troy, New York. He has received several awards and honors. He was the member of the inaugural 2009 class of the SIAM fellows. He is an ISI highly cited author in the mathematics category. His research interests are in continuous optimization and equilibrium programming and their applications in engineering, economics, and finance.

REFERENCES

[1] V. Krishnamurthy, O. Namvar, and M. Hamdi, "Signal processing of social networks: Interactive sensing and learning," *Found. Trends Signal Process.*, vol. 7, nos. 1–2. pp. 1–19, 2014.

[2] *Special Issue on Signal Processing for Big Data, IEEE Signal Processing Mag.*, vol. 31, no. 5, Sept. 2014.

[3] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning*. Cambridge, MA: MIT Press, 2012.

[4] A. Roberts and L. Pachter, "Streaming fragment assignment for real-time analysis of sequencing experiments," *Nat. Methods*, vol. 10, no. 1, pp. 71–73, 2013.

[5] Y. Nesterov, "Efficiency of coordinate descent methods on huge-scale optimization problems," *SIAM J. Optim.*, vol. 22, no. 2, pp. 341–362, 2012.

[6] V. Cevher, S. Becker, and M. Schmidt, "Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 32–43, Sept. 2014.

[7] N. S. Aybat, "First order methods for large-scale sparse optimization," Ph.D. dissertation, Columbia Univ., 2014.

[8] P. Richtárik and M. Takáč, "Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function," *Math. Program.*, vol. 144, no. 1, pp. 1–38, 2014.

[9] S. Cui, Z.-Q. Luo, A. Hero, and J. Moura, *Big Data over Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[10] Y.-L. You and M. Kaveh, "A regularization approach to joint blur identification and image restoration," *IEEE Trans. Image Processing*, vol. 5, no. 3, pp. 416–428, 1996.

[11] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *J. Stat. Softw.*, vol. 33, no. 1, pp. 1–22, 2010.

[12] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.

[13] B. Chen, S. He, Z. Li, and S. Zhang, "Maximum block improvement and polynomial optimization," *SIAM J. Optim.*, vol. 22, no. 1, pp. 87–107, 2012.

[14] J. Nutini, M. Schmidt, I. H. Laradji, M. Friedlander, and H. Koepke, "Coordinate descent converges faster with the Gauss-Southwell rule than random selection," in *Proc. 30th Int. Conf. Machine Learning (ICML)*, 2015, pp. 1632–1641.

[15] P. Richtárik and M. Takáč, "Parallel coordinate descent methods for big data optimization," *Math. Program.*, to be published.

[16] F. Facchinei, S. Sagratella, and G. Scutari, "Flexible parallel algorithms for big data optimization," in *Proc. 2014 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 7208–7212.

[17] J. Liu, C. Re S. J. Wright, and V. Bittorf, "An asynchronous parallel stochastic coordinate descent algorithm," in *Proc. Int. Conf. Machine Learning (ICML)*, 2014.

[18] M. Razaviyayn, M. Hong, Z.-Q. Luo, and J. S. Pang, "Parallel successive convex approximation for nonsmooth nonconvex optimization," in *Proc. Neural Information Processing (NIPS)*, 2014, pp. 1440–1448.

[19] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *IEEE Trans. Signal Processing*, vol. 63, no. 7, pp. 1874–1889, 2015.

[20] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," preprint, 2013.

[21] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[22] J. Mairal, "Optimization with first-order surrogate functions," in *Proc. Int. Conf. Machine Learning (ICML)*, 2013, pp. 783–791.

[23] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[24] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Top. Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.

[25] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, 2012.

[26] R. Sun and Z.-Q. Luo, "Guaranteed matrix completion via nonconvex factorization," preprint, 2014.

[27] J. Kim, Y. He, and H. Park, "Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework," *J. Glob Optim.*, vol. 58, no. 2, pp. 285–319, Feb. 2014.

[28] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[29] S. J. Wright, "Coordinate descent algorithms," *Math. Program.*, vol. 151, no. 1, pp. 3–34, 2015.

[30] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[31] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Comput.*, vol. 15, pp. 915–936, 2003.

[32] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Neural Information Processing Systems (NIPS)*, 2000, pp. 556–562.

[33] C. Shi, R. A. Berry, and M. L. Honig, "Monotonic convergence of distributed interference pricing in wireless networks," in *Proc. IEEE Int. Conf. Symp. Information Theory*, 2009, pp. 1619–1623.

[34] M. Hong and Z.-Q. Luo, "Signal processing and optimal resource allocation for the interference channel," in *Academic Press Library in Signal Processing*. New York: Academic, 2013.

[35] M. Hong, Q. Li, and Y.-F. Liu, "Decomposition by successive convex approximation: A unifying approach for linear transceiver design in interfering heterogeneous networks," *IEEE Trans. Wireless Commun.*, to be published.

[36] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 103, no. 9, pp. 475–494, 2001.

[37] M. J. D. Powell, "On search directions for minimization algorithms," *Math. Program.*, vol. 4, no. 1, pp. 193–201, 1973.

[38] M. Razaviyayn, "Successive convex approximation: Analysis and applications," Ph.D. dissertation, Univ. of Minnesota, 2014.

[39] Z. Lu and L. Xiao, "On the complexity analysis of randomized block-coordinate descent methods," *Math. Program.*, vol. 152, no. 1, pp. 615–642, Aug. 2015.

[40] A. Beck and L. Tetruashvili, "On the convergence of block coordinate descent type methods," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2037–2060, 2013.

[41] A. Beck, "On the convergence of alternating minimization with applications to iteratively reweighted least squares and decomposition schemes," *SIAM J. Optim.*, vol. 25, no. 1, pp. 185–209, 2015.

[42] M. Hong, X. Wang, M. Razaviyayn, and Z.-Q. Luo, "Iteration complexity analysis of block coordinate descent methods," preprint, 2013.

[43] A. Saha and A. Tewari, "On the nonasymptotic convergence of cyclic coordinate descent method," *SIAM J. Optim.*, vol. 23, no. 1, pp. 576–601, 2013.

[44] M. Hong, T.-H. Chang, X. Wang, M. Razaviyayn, S. Ma, and Z.-Q. Luo, "A block successive upper bound minimization method of multipliers for linearly constrained convex optimization," preprint, 2013.

[45] M. Kadkhodaei, M. Sanjabi, and Z.-Q. Luo, "On the linear convergence of approximate proximal splitting methods for non-smooth convex minimization," *J. Oper. Res. Soc. China*, vol. 2, no. 2, pp. 123–141, 2014.

[46] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[47] Q. Lin, Z. Lu, and L. Xiao, "An accelerated proximal coordinate gradient method and its application to regularized empirical risk minimization," *SIAM J. Optim.*, to be published.

[48] O. Fercoq and P. Richtárik, "Accelerated, parallel and proximal coordinate descent," *SIAM J. Optim.*, vol. 25, no. 4, pp. 1927–2023, 2015.

[49] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Am. Stat.*, vol. 58, no. 1, pp. 30–37, 2004.

[50] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, 2nd ed. Belmont, MA: Athena-Scientific, 1999.

[51] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints," *Oper. Res. Lett.*, vol. 26, no. 3, pp. 127–136, 2000.

[52] P. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer Optimization and Its Applications*. New York: Springer, 2011, pp. 185–212.

[53] Z. Lu and L. Xiao, "Randomized block coordinate non-monotone gradient method for a class of nonlinear programming," preprint, 2013.

[54] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Machine Learning (ICML)*, 2009, pp. 689–696.

[55] Q. Li, "On the convergence of Lee-Seung's multiplicative update for NMF," 2012, Tech. Note.

[56] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the beta-divergence," *Neural Comput.*, vol. 23, no. 3, pp. 1–24, 2011.

[57] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.

[58] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Top. Signal Process*sing, vol. 2, no. 1, pp. 57–73, 2008.

[59] H. Baligh, M. Hong, W.-C. Liao, Z.-Q. Luo, M. Razaviyayn, M. Sanjabi, and R. Sun, "Cross-layer provision of future cellular networks: A WMMSE-based approach," *IEEE Signal Processing Mag.*, vol. 31, no. 6, pp. 56–68, Nov. 2014.

[60] M. Razaviyayn, H. Baligh, A. Callard, and Z.-Q. Luo, "Joint user grouping and transceiver design in a mimo interfering broadcast channel," *IEEE Trans. Signal Processing*, vol. 62, no. 1, pp. 85–94, Jan. 2014.

[61] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.

[62] M. Hong, R. Sun, H. Baligh, and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogenous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226–240, 2013.

[63] C. Shi, R. A. Berry, and M. L. Honig, "Distributed interference pricing with MISO channels," in *Proc. 46th Annual Allerton Conf. Communication, Control, and Computing* 2008, Sept. 2008, pp. 539–546.

[64] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO Ad Hoc Cognitive Radio Networks," *IEEE Trans. Inform. Theory*, vol. 57, no. 5, pp. 3117–3131, 2011.

[65] C. T. K. Ng and H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1446–1454, 2010.

[66] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, "Ultrafast and memory-efficient alignment of short dna sequences to the human genome," *Genome Biol.*, vol. 10, no. 3, pp. R25, 2009.

[67] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows–Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009.

[68] N. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal RNA-Seq quantification," preprint, 2015.

[69] A. Shashua and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," in *Proc. 22nd Int. Conf. Machine Learning (ICML)*, 2005, pp. 792–799.

[70] I. Ibraghimov, "Application of the three-way decomposition for matrix compression," *Numer. Linear Algebra Appl.*, vol. 9, no. 6–7, pp. 551–565, 2002.

[71] J. Håstad, "Tensor rank is NP-complete," *J. Algorithms*, vol. 11, no. 4, pp. 644–654, 1990.

[72] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[73] R. A. Harshman, "Foundations of the parafac procedure: Models and conditions for an "explanatory" multimodal factor analysis," *UCLA Working Papers in Phonetics, 16*, pp. 1–84, 1970.

[74] C. Navasca, L. De Lathauwer, and S. Kindermann, "Swamp reducing technique for tensor decomposition," in *Proc. 16th European Signal Processing Conf. (EUSIPCO)*, Aug. 2008, pp. 1–5.

[75] J. A. Bazerque, G. Mateos, and G. B. Giannakis, "Rank regularization and Bayesian inference for tensor completion and extrapolation," *IEEE Trans. Signal Process*sing, vol. 61, no. 21–24, pp. 5689–5703, 2013.

[76] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[77] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[78] M. Razaviyayn, H.-W. Tseng, and Z.-Q. Luo, "Dictionary learning for sparse representation: Complexity and algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 5247–5251.

[79] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2006, pp. 801–808.

[80] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[81] G. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, vol. 544. Hoboken, NJ: Wiley, 2004.

[82] D. M. Witten and R. Tibshirani, "Penalized classification using Fisher's linear discriminant," *J. R. Stat. Soc. Series B (Stat. Methodol.)*, vol. 73, no. 5, pp. 753–772, 2011.

[83] S. Kim, R. Pasupathy, and S. G. Henderson, "A guide to sample average approximation," in *Handbook of Simulation Optimization*, M. C. Fu, Ed. vol. 216: International Series in Operations Research and Management Science. New York: Springer, 2015, pp. 207–243.

[84] J. Mairal, "Stochastic majorization-minimization algorithms for large-scale optimization," in *Proc. Advances in Neural Information Processing Systems*, 2013, pp. 2283–2291.

[85] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic weighted MMSE approach to sum rate maximization for a MIMO interference channel," in *Proc. IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, IEEE, 2013, pp. 325–329.

[86] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Jan. 2010.

[87] K. Slavakis, S.-J. Kim, G. Mateos, and G. B. Giannakis, "Stochastic approximation vis-a-vis online learning for big data analytics," *IEEE Signal Processing Mag.*, vol. 31, no. 6, pp. 124–129, 2014.

[88] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[89] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*. New York: Springer-Verlag, 1984.

[90] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," submitted for publication.

[91] B. Ames and M. Hong, "Alternating directions method of multipliers for l1-penalized zero variance discriminant analysis and principal component analysis," submitted for publication.

[92] I. Necoara, Y. Nestrov, and F. Glineur, "A random coordinate descent method on large-scale optimization problems with linear constraints," Univ. Politehnica Bucharest, Tech. Rep, 2011.

[93] S. Reddi, A. Hefny, C. Downey, A. Dubey, and S. Sra, "Large-scale randomized-coordinate descent methods with non-separable linear constraints," preprint, 2014.

[94] N. S. Aybat, G. Iyengar, and Z. Wang, "An asynchronous distributed proximal method for composite convex optimization," in *Proc. 32th Int. Conf. Machine Learning (ICML)*, 2015, pp. 2454–2462.

[95] W.-C. Liao, M. Hong, H. Farmanbar, and Z.-Qu. Luo, "Semi-asynchronous routing for large-scale hierarchical networks," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2015, pp. 2894–2898.

[96] Gurobi Optimization, Inc. *Gurobi Optimizer Reference Manual*, 2013. [Online]. Available: http://www.gurobi.com/documentation/6.0/refman/

[97] D. P. Bertsekas, P. Hosein, and P. Tseng, "Relaxation methods for network flow problems with convex arc costs," *SIAM J. Control Optim.*, vol. 25, no. 5, pp. 1219–1243, Sept. 1987.

[98] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, 1978.

[99] F. Facchinei, S. Sagratella, and G. Scutari, "Flexible parallel algorithms for big data optimization," preprint, 2013.

[100] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Trans. Signal Processing*, vol. 63, no. 3, pp. 641–656, 2014.

[101] J.-S. Pang and M. Razaviyayn, "A unified distributed algorithm for non-cooperative games with non-convex and non-differentiable objectives," preprint, 2014.

[102] Yu. Nesterov, "Gradient methods for minimizing composite functions," *Math. Program.*, vol. 140, no. 1, pp. 125–161, 2013.

[103] M. Razaviyayn, M. Hong, Z.-Q. Luo, and J.-S. Pang, "Parallel successive convex approximation for nonsmooth nonconvex optimization," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2014, pp. 1440–1448.

[104] P. Richtárik and M. Takáč, "Distributed coordinate descent method for learning with big data," preprint, 2013.

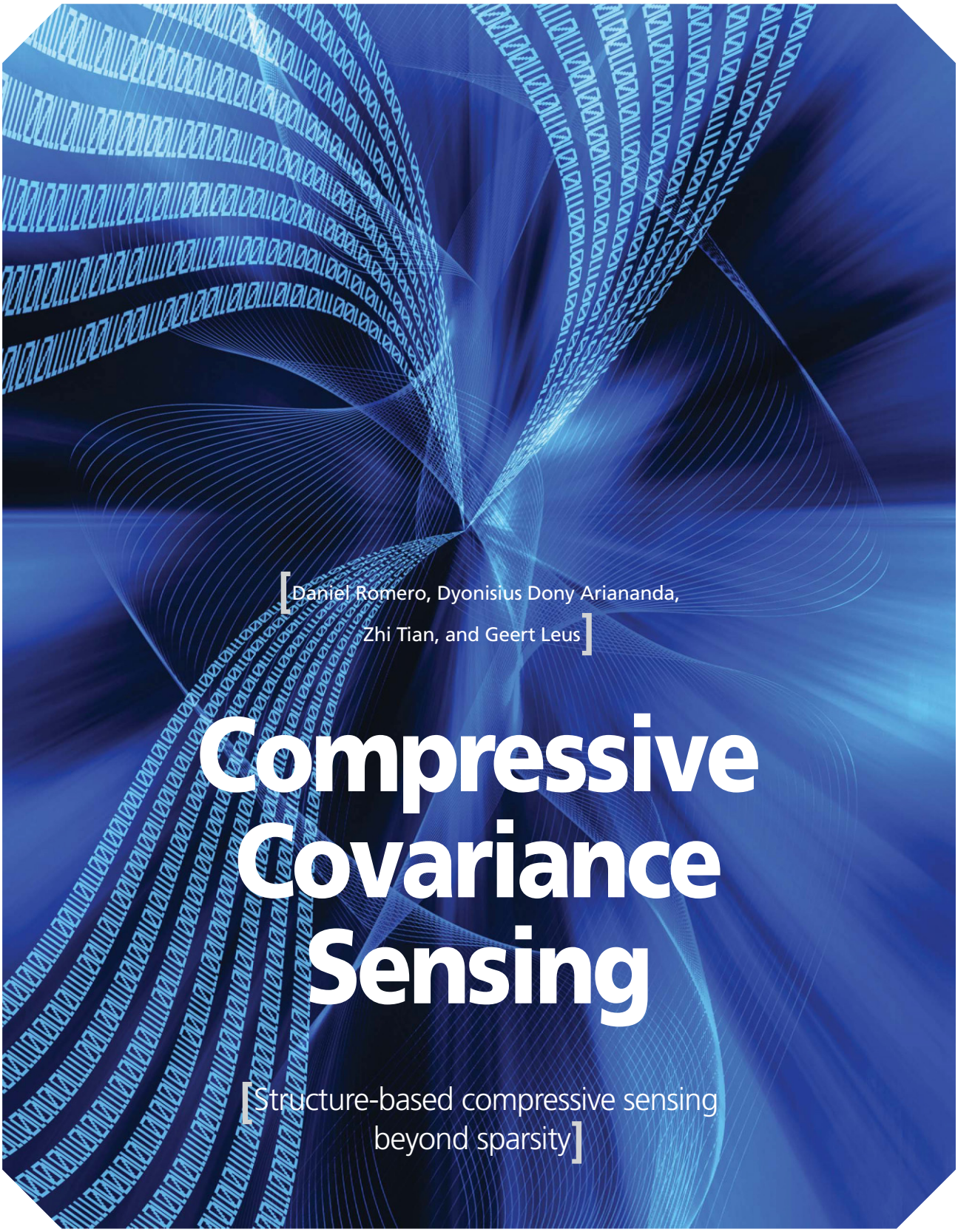
[105] Z. Peng, M. Yan, and W. Yin, "Parallel and distributed sparse optimization," in *Proc. IEEE 2013 Asilomar Conf. Signals, Systems and Computers*, 2013, pp. 659–646.

[106] C. Scherrer, A. Tewari, M. Halapanavar, and D. J. Haglin, "Feature clustering for accelerating parallel coordinate descent," in *Proc. Advanced in Neural Information Processing Systems (NIPS)*, 2012, pp. 28–36.

[107] O. Fercoq, Z. Qu, P. Richtárik, and M. Takáč, "Fast distributed coordinate descent for non-strongly convex losses," in *Proc. IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP)*, 2014, pp. 1–6.

[SP]

IEEE SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®



[Daniel Romero, Dyonisius Dony Arianandaa,
Zhi Tian, and Geert Leus]

Compressive Covariance Sensing

[Structure-based compressive sensing beyond sparsity]

IMAGE LICENSED BY INGRAM PUBLISHING

Digital Object Identifier 10.1109/MSP.2015.2486805
Date of publication: 29 December 2015

1053-5888/16©2016IEEE

IEEE SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®

Compressed sensing deals with the reconstruction of signals from sub-Nyquist samples by exploiting the sparsity of their projections onto known subspaces. In contrast, this article is concerned with the reconstruction of second-order statistics, such as covariance and power spectrum, even in the absence of sparsity priors. The framework described here leverages the statistical structure of random processes to enable signal compression and offers an alternative perspective at sparsity-agnostic inference. Capitalizing on parsimonious representations, we illustrate how compression and reconstruction tasks can be addressed in popular applications such as power-spectrum estimation, incoherent imaging, direction-of-arrival estimation, frequency estimation, and wideband spectrum sensing.

INTRODUCTION

The incessantly growing size of sensing problems has spurred an increasing interest in simultaneous data acquisition and compression techniques that limit sensing, storage, and communication costs. Notable examples include compressed sensing [1], support recovery [2], sub-Nyquist sampling of multiband or multitone signals [3]–[5], and array design for aperture synthesis imaging [6]–[8]. The overarching paradigm of sub-Nyquist sampling can impact a broad swath of resource-constrained applications arising in data sciences, broadband communications, large-scale sensor networks, bioinformatics, and medical imaging, to name a few.

The aforementioned techniques rely on parsimonious models that capture relevant information and enable compression. In compressed sensing, for example, signals can be reconstructed from sub-Nyquist samples provided that they admit a sparse representation in a known transformed domain. Whereas this form of structure arises naturally in many applications, it is often the case that either the underlying signal is not sparse or the sparsifying transformation is difficult to model or manipulate. Those scenarios call for alternative approaches to allow compression by capturing other forms of structure.

A prominent example is the family of methods exploiting structural information in the statistical domain, which includes those intended to reconstruct the second-order statistics of wide-sense stationary (WSS) signals, such as power, autocorrelation, or power-spectral density. It is widely accepted that statistics of this class play a central role in a multitude of applications comprising audio and voice processing, communications, passive sonar, passive radar, radioastronomy, and seismology, for example [9]. Although reconstruction of second-order statistics from compressed observations dates back several decades (see, e.g., [6] and the references therein), the recent interest in compressive sensing and reconstruction has propelled numerous advances in this context.

The purpose of this article is to provide a fresh look at the recent contributions in this exciting area, which is referred to as *compressive covariance sensing* (*CCS*). Admittedly, a straightforward approach to reconstruct second-order statistics is to apply an estimation method over a waveform uncompressed via a non-CCS procedure. However, it is not difficult to see that such a two-step approach incurs large computational complexity and heavily limits the compression ratio. CCS methods, on the other hand, proceed in a single step by directly recovering relevant second-order statistics from the compressed samples, thus allowing a more efficient exploitation of the statistical structure.

SAMPLING SECOND-ORDER STATISTICS

To introduce the basic notions of CCS, consider the problem of measuring the fine variations of a spatial field to achieve a high angular resolution in source localization. Since the large sensor arrays required in the absence of compression incur prohibitive hardware costs, multiple acquisition schemes have been devised to reduce the number of sensors without sacrificing resolution.

A WARM-UP EXAMPLE

Suppose that a uniform linear array (ULA) with L antennas, such as the one in Figure 1(a), observes T snapshots of a zero-mean spatial signal whose complex baseband representation is given by $\boldsymbol{x}_\tau \in \mathbb{C}^L$, $\tau = 0, 1, \ldots T-1$. Many array processing algorithms rely on estimates of the so-called spatial covariance matrix $\boldsymbol{\Sigma}_x := E\{\boldsymbol{x}_\tau \boldsymbol{x}_\tau^H\}$ to form images or to obtain information such as the bearing of certain sources [6], [9]. A straightforward estimate of $\boldsymbol{\Sigma}_x$ is the sample covariance matrix, given by

$$\hat{\boldsymbol{\Sigma}}_x = \frac{1}{T}\sum_{\tau=0}^{T-1} \boldsymbol{x}_\tau \boldsymbol{x}_\tau^H. \tag{1}$$

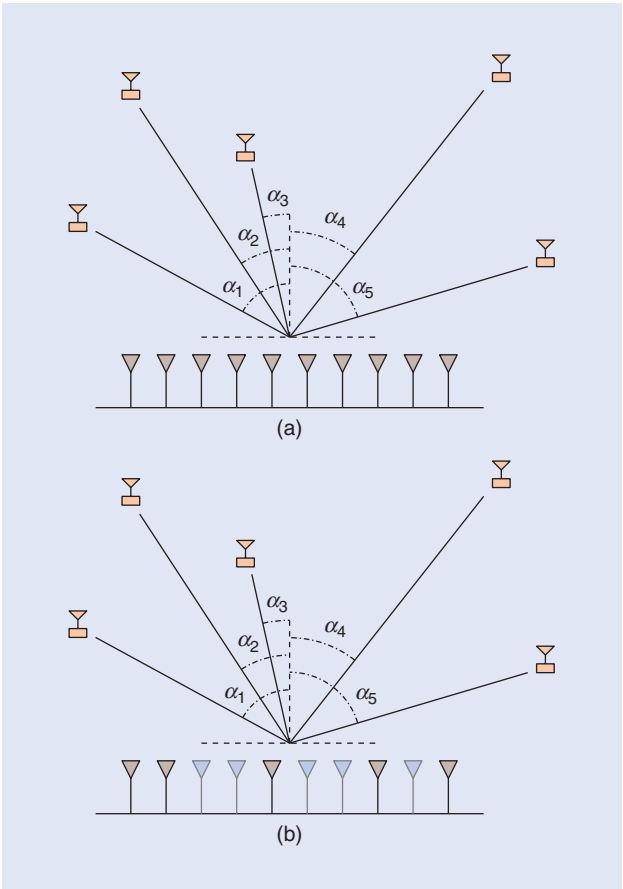
If the impinging signals are generated by uncorrelated point sources in the far field [see Figure 1(a)], the matrix $\boldsymbol{\Sigma}_x$ exhibits a Toeplitz structure (see the section "Modal Analysis"), meaning that the coefficients are constant along the diagonals. Thus, one may represent the (m,n)th entry of $\boldsymbol{\Sigma}_x$ by

$$\sigma[m-n] = E\{x_\tau[m]x_\tau^*[n]\}, \tag{2}$$

where $x_\tau[m]$ represents the mth entry of $\boldsymbol{x}_\tau$. Noting that $\boldsymbol{\Sigma}_x$ is also Hermitian reveals that all the information is contained in the coefficients $\sigma[l]$, $l = 0, \ldots, L-1$. These observations suggest the possibility of constructing estimators with improved performance [10], [11]; simply consider replacing the elements on each diagonal of $\hat{\boldsymbol{\Sigma}}_x$ with their arithmetic mean. This operation renders a more satisfactory estimate than the sample covariance matrix in (1) because it utilizes the underlying Toeplitz structure.

Let us now adopt a different standpoint. Instead of attempting to improve the estimation performance, the described structure can also be exploited to reduce the number of antennas required to estimate $\boldsymbol{\Sigma}_x$ (see, e.g., [6]–[8]). Suppose, in particular, that only a subset of the antennas in the ULA is used to sample the spatial field of interest, the others being disconnected [see Figure 1(b)].

Let the set $\mathcal{K} := \{k_0, \ldots, k_{K-1}\}$ collect the indices of the K active antennas. The vector signal received by this subarray, which can be thought of as a compressed observation, is given by $\boldsymbol{y}_\tau = [x_\tau[k_0], \ldots, x_\tau[k_{K-1}]]^T$. The (i,j)th entry (we adopt the convention that the first row/column of any vector/matrix is associated with the index 0, the second with the index 1, and so on) of the covariance matrix $\boldsymbol{\Sigma}_y := E\{\boldsymbol{y}_\tau \boldsymbol{y}_\tau^H\}$ is, therefore,



[FIG1] (a) An uncompressed ULA with ten antennas receiving the signals from five sources in the far field. (b) A compressed array with five antennas. The five antennas marked in light gray were removed, but the achievable spatial resolution remains the same.

$$\mathrm{E}\{y_\tau[i]y_\tau^*[j]\} = E\{x_\tau[k_i]x_\tau^*[k_j]\} = \sigma[k_i - k_j]. \qquad (3)$$

Thus, $\mathbf{\Sigma}_y$ is made up of a subset of the entries of $\mathbf{\Sigma}_x$. It is clear that $\mathbf{\Sigma}_x$ can be reconstructed from a sample estimate of $\mathbf{\Sigma}_y$ if all the entries of the former show up at least once in the latter. From (3), this means that, for every $l = 0, \ldots, L-1$, there must exist at least one pair of elements k, k' in $\mathcal{K}$ satisfying $k - k' = l$. Sets $\mathcal{K}$ of this nature are called *sparse rulers*, and, if they contain a minimum number of elements, they are termed *minimal sparse rulers*, as explained in "Linear Sparse Rulers." In Figure 1(b), for example, only the antennas at positions $\mathcal{K} = \{0, 1, 4, 7, 9\}$ are operative, but the array can reconstruct the same spatial covariance matrix as the array in Figure 1(a).

Mathematically, the problem of constructing sparse rulers is interesting on its own and has been extensively analyzed (see [12] and the references therein). Since finding minimal sparse rulers is a combinatorial problem with no closed-form solution, devising structured yet suboptimal designs has received much attention (see, e.g., [12] and [13]).

An intimately related concept is the minimum-redundancy array [7], [9], well known within the array processing community. A minimum-redundancy array is a minimal linear sparse ruler whose length is maximum given its number of marks. For example, $\mathcal{K}_1 = \{0, 1, 2, 3, 7\}$, $\mathcal{K}_2 = \{0, 1, 2, 5, 8\}$, and $\mathcal{K}_3 = \{0, 1, 2, 6, 9\}$ are minimal sparse rulers of length 7, 8, and 9, respectively. However, $\mathcal{K}_1$ and $\mathcal{K}_2$ are not minimum-redundancy arrays, since a minimal sparse ruler of greater length can be found with the same number of marks, an example being $\mathcal{K}_3$.

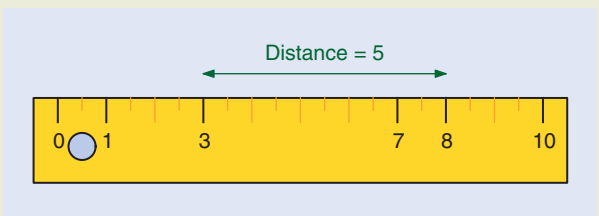
Deploying a smaller number of antennas allows cost savings beyond the costs associated with the antennas themselves: radio-frequency (RF) equipment, such as filters, mixers, and analog-to-digital converters (ADCs), needs to be deployed only for the active antennas. Moreover, the fact that the endpoints 0 and $L-1$ are always in $\mathcal{K}$ for any length-$(L-1)$ linear sparse ruler $\mathcal{K}$ means that the aperture of the subarray equals the aperture of the uncompressed array. Therefore, this antenna reduction comes at no cost in angular resolution. The price to be paid is, however, slower convergence of the estimates; generally, the smaller the $|\mathcal{K}|$, the larger the number of snapshots T required to attain a target performance. Hence, for signals defined in the spatial domain, this kind of compression is convenient when hardware savings make up for an increase in the acquisition time, as is usually the case in array processing.

LINEAR SPARSE RULERS

A set $\mathcal{K} \subset \{0, \ldots, L-1\}$ is a length-$(L-1)$ (linear) sparse ruler if for every $l = 0, \ldots, L-1$, there exists at least one pair of elements k, k' in $\mathcal{K}$ satisfying $k - k' = l$. Two examples of length 10 are $\mathcal{K} = \{0, 1, 2, 5, 7, 10\}$ and $\mathcal{K} = \{0, 1, 3, 7, 8, 10\}$.

The name *sparse ruler* stems from the geometric interpretation of $\mathcal{K}$ as a physical ruler where all but the marks with indices in $\mathcal{K}$ have been erased. Despite lacking part of its marks, a sparse ruler is still able to measure any integer distance between zero and $L-1$. In Figure S1, we observe that the ruler $\mathcal{K} = \{0, 1, 3, 7, 8, 10\}$ is capable of measuring any object of length five by using the marks three and eight.

A length-$(L-1)$ minimal sparse ruler is a length-$(L-1)$ sparse ruler $\mathcal{K}$ with minimum number of elements $|\mathcal{K}|$. The set $\mathcal{K} = \{0, 1, 3, 7, 8, 10\}$ is a length-10 minimal linear sparse ruler since it has six elements and there exists no length-10 sparse ruler with five or fewer elements.



[FIGS1] A sparse ruler can be thought of as a ruler with a part of its marks erased, but the remaining marks allow all integer distances between zero and its length to be measured.

IMPORTANCE OF COVARIANCE STRUCTURES

In the previous example, the Hermitian Toeplitz structure of $\mathbf{\Sigma}_x$ allowed us to recover the second-order statistics of $\boldsymbol{x}_\tau$ from those of its compressed version $\boldsymbol{y}_\tau$. More generally, it is expected that our ability to compress a signal while preserving the second-order statistical information depends on the structure of $\mathbf{\Sigma}_x$. In other words, we expect that the more structured $\mathbf{\Sigma}_x$ is, the stronger the compression on $\boldsymbol{x}_\tau$ it may induce.

CCS METHODS, ON THE OTHER HAND, PROCEED IN A SINGLE STEP BY DIRECTLY RECOVERING RELEVANT SECOND-ORDER STATISTICS FROM THE COMPRESSED SAMPLES, THUS ALLOWING A MORE EFFICIENT EXPLOITATION OF THE STATISTICAL STRUCTURE.

In certain applications, such as power-spectrum estimation for communication signals, the covariance matrix is known to be circulant [14]–[17]. Recall that a circulant matrix is a special type of Toeplitz matrix where each row is the result of applying a circular shift to the previous one. For this reason, it can be seen that $\sigma[l] = \sigma[l-L]$. This increased structure relaxes the requirements on $\mathcal{K}$, which is no longer required to be a linear sparse ruler but a circular one; see "Circular Sparse Rulers" for a definition.

Because of their ability to measure two different distances using each pair of marks, circular sparse rulers lead to a greater compression than their linear counterparts. In other words, $\mathcal{K}$ needs fewer elements to be a length-$(L-1)$ circular sparse ruler than to be a length-$(L-1)$ linear sparse ruler.

Circular sparse rulers can be designed in several ways. For certain values of L, minimal rulers can be obtained in closed form [18]. Other cases may require exhaustive search, which motivates suboptimal designs. Immediate choices are length-$(L-1)$ or length-$\lfloor L/2 \rfloor$ minimal linear sparse rulers [19]. In fact, the latter provide optimal solutions for most values of L below 60 [20].

Aside from Toeplitz and circulant, another common structure is the one present in those applications where the covariance matrix is known to be banded [19]. A type of Toeplitz matrix, d-banded matrices satisfy $\sigma[l] = 0$ for all $l > d$ and arise in those cases where we sample a WSS time signal whose autocorrelation sequence $\sigma[l]$ vanishes after d lags. Sampling patterns for banded matrices are discussed in [20], which suggests that the achievable compression is dependent on the parameter d. These designs also hold for certain situations where we are only interested in the first d correlation lags [21].

These typical covariance structures, including Toeplitz, circulant, and banded, are illustrated in Figure 2, along with their most popular applications. Generally speaking, in many cases including the previous ones, prior knowledge constrains covariance matrices to be linear combinations of certain known matrices, say $\{\mathbf{\Sigma}_i\}_i$. In other words, there must exist coefficients α_i such that

$$\mathbf{\Sigma}_x = \sum_{i=0}^{S-1} \alpha_i \mathbf{\Sigma}_i. \tag{4}$$

Without any loss of generality, we may assume that the scalars α_i are real [20] and the matrices $\mathbf{\Sigma}_i$ are linearly independent. Thus, $\mathcal{S} = \{\mathbf{\Sigma}_0, \ldots, \mathbf{\Sigma}_{S-1}\}$ is a basis, and S represents the dimension of the model. This expansion encompasses all the previous examples as particular cases as long as the right set of matrices $\mathbf{\Sigma}_i$ is chosen. It can be seen that $S = 2L-1$ for Toeplitz matrices, $S = L$ for circulant matrices, and $S = 2d-1$ for d-banded matrices (see Figure 2). As we will see in the section "Optimal Designs," S is related to how compressible $\mathbf{\Sigma}_x$ is.

The problem of estimating the coefficients α_i is known as *structured covariance estimation* or *covariance matching* [10], [22] and has a strong connection with CCS. Nonetheless, this line of work flourished before the surge of compressed sensing in signal processing, when the main goal was to design robust and performance-enhanced estimators with a small sample size. CCS offers a new way to exploit covariance structures for joint signal acquisition and compression.

COMPRESSION

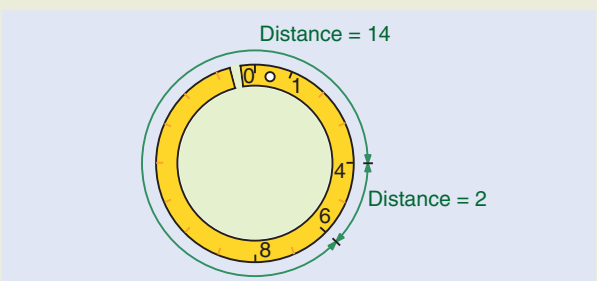
The previous array processing example describes how compression can be exerted for signals acquired in the spatial

CIRCULAR SPARSE RULERS

A set $\mathcal{K} \subset \{0, \ldots, L-1\}$ is a length-$(L-1)$ circular sparse ruler if for every $l = 0, \ldots, L-1$, there exists at least one pair of elements $k, k' \in \mathcal{K}$ satisfying $(k-k') \bmod L = l$. An example of a length-15 circular sparse ruler is $\mathcal{K} = \{0, 1, 4, 6, 8\}$. It can be seen that any length-l linear sparse ruler, with $L/2 \le l \le L-1$, is also an example of a length-L circular sparse ruler.

A circular sparse ruler can be thought of as the result of wrapping around a linear ruler. This operation allows us to measure two different distances using each pair of marks (see Figure S2).

A length-$(L-1)$ circular sparse ruler is minimal if there exists no length-$(L-1)$ circular sparse ruler with fewer elements.



[FIGS2] **Generally, in a circular sparse ruler, each pair of marks allows two distances to be measured.**

domain—only a subset $\mathcal{K}$ of antennas was used to estimate $\mathbf{\Sigma}_x$; the remaining antennas can be disconnected, or, more simply, they need not be deployed. Broadly, acquisition hardware represents the bottleneck of many current signal processing systems, whose designs aim at meeting an ever-increasing demand for processing rapidly changing signals. In practice, Nyquist acquisition of wideband signals becomes prohibitive in many applications since the sampling rate drastically affects power consumption and hardware complexity. The ambition to break this bandwidth barrier has prompted a growing interest in innovative acquisition hardware architectures that replace traditional equipment, such as the slow and power-hungry ADCs. In this section, we delve into compression methods that can be applied not only for compressive acquisition of spatial signals but also for time signals and more general classes of signals.

ACQUISITION HARDWARE REPRESENTS THE BOTTLENECK OF MANY CURRENT SIGNAL PROCESSING SYSTEMS, WHOSE DESIGNS AIM AT MEETING AN EVER-INCREASING DEMAND FOR PROCESSING RAPIDLY CHANGING SIGNALS.

In particular, suppose that we are interested in estimating the second-order statistics of $x(t)$, indexed by the continuous-time index t. A traditional ADC ideally produces the sequence

$$x[l] = x(lT_s), \quad l = 0, \ldots, L-1, \tag{5}$$

where $1/T_s$ is the sampling rate, a number that must exceed the Nyquist rate of $x(t)$ to avoid aliasing. Unfortunately, power consumption, amplitude resolution, and other parameters dictated by the application establish stringent upper bounds on the values that the sampling rate can take on. These limitations conflict with the constantly increasing need for larger bandwidths and, hence, higher Nyquist rates.

A compression approach similar to the one described for the spatial domain may potentially alleviate these limitations by reducing the average sampling rate. Generally known as *nonuniform sampling*, this approach advocates the acquisition of a small number of samples indexed by a subset of the Nyquist grid:

$$y[i] = x(k_i T_s), \quad \mathcal{K} = \{k_0, \ldots, k_{K-1}\}. \tag{6}$$

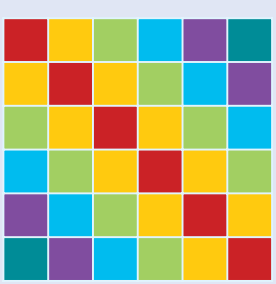
As we will soon see, this average rate reduction has led to the technology of compressive ADCs (C-ADCs), conceived to circumvent the aforementioned hardware tradeoffs. Before exploring this topic, let us expand the families of samplers we are about to consider.

By forming $\boldsymbol{x} = [x[0], \ldots, x[L-1]]^T$ and $\boldsymbol{y} = [y[0], \ldots, y[K-1]]^T$, the operation in (6) can be equivalently represented as a row-selection operation

$$\boldsymbol{y} = \bar{\mathbf{\Phi}}\boldsymbol{x}. \tag{7}$$

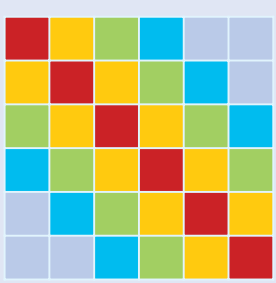
The matrix $\bar{\mathbf{\Phi}} \in \mathbb{C}^{K \times L}$, which contains ones at the positions (i, k_i) and zeros elsewhere, is, therefore, a sparse matrix with at most one nonzero entry at each row or column. Rather than restricting ourselves to matrices of this form, there are certain applications where the usage of dense compression matrices has proven to be successful, both in the time domain (see, e.g., [4] and [5]) and in the spatial domain (see, e.g., [23]). In correspondence with this terminology, we talk about dense samplers when $\bar{\mathbf{\Phi}}$ is dense and about sparse samplers when $\bar{\mathbf{\Phi}}$ is sparse.

As opposed to most applications in array processing, it is common in time-domain applications to observe just a single realization of the signal of interest, i.e., $T = 1$. This is why we dropped the subscript τ from $\boldsymbol{x}$ and $\boldsymbol{y}$ in (7) when compared to $\boldsymbol{x}_\tau$ and $\boldsymbol{y}_\tau$ in the previous section. For simplicity, we omit this subscript throughout when possible, keeping in mind that several snapshots may be available.



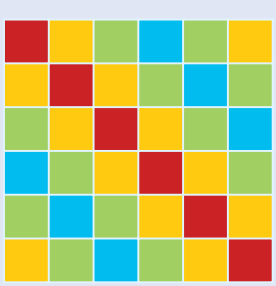
$S = 2L - 1$ Real Unknowns
WSS Processes
Direction-of-Arrival Estimation
Incoherent Imaging

(a)



$S = 2d - 1$ Real Unknowns
WSS with d Limited Lags
MA(d) Time Series

(b)



$S = L$ Real Unknowns
Orthogonal Frequency-Division Multiplexing Signals
Multiband Signals
Incoherent Imaging

(c)

[FIG2] Some common covariance structures, along with their main applications: (a) Toeplitz, (b) d-banded, and (c) circulant.

[TABLE 1] OPTIMAL DESIGNS AND COMPRESSION RATIOS.

	SPARSE SAMPLERS		DENSE SAMPLERS	
	DESIGN	RATIO	DESIGN	RATIO
TOEPLITZ	$\mathcal{K}$ IS A LENGTH-$(N-1)$ LINEAR SPARSE RULER	$\frac{1}{N}\sqrt{2.435(N-1)} \le \eta_{\min} \le \frac{1}{N}\left\lfloor\sqrt{3(N-1)}\right\rfloor$	$\Phi \in \mathbb{C}^{M\times N}$ IS DRAWN FROM A CONTINUOUS DISTRIBUTION WITH $M \ge \sqrt{\frac{2NB-1}{2B-1}}$	$\eta_{\min} \approx \sqrt{\frac{2NB-1}{(2B-1)N^2}}$
CIRCULANT	$\mathcal{K}$ IS A LENGTH-$(N-1)$ CIRCULAR SPARSE RULER	$\frac{2+\sqrt{4N-3}}{2N} \le \eta_{\min} \le \frac{1}{N}\left\lfloor\sqrt{3\left\lfloor\frac{N}{2}\right\rfloor}\right\rfloor$	$\Phi \in \mathbb{C}^{M\times N}$ IS DRAWN FROM A CONTINUOUS DISTRIBUTION WITH $M \ge \sqrt{\frac{NB}{2B-1}}$	$\eta_{\min} \approx \sqrt{\frac{B}{(2B-1)N}}$
d-BANDED	CASE $N \le d \le N(B-1)$:		$\Phi \in \mathbb{C}^{M\times N}$ IS DRAWN FROM A CONTINUOUS DISTRIBUTION WITH $M \ge \sqrt{\frac{2d+1}{2B-1}}$	$\eta_{\min} \approx \sqrt{\frac{2d+1}{(2B-1)N^2}}$
	$\mathcal{K}$ IS A LENGTH-$(N-1)$ CIRCULAR SPARSE RULER	$\frac{2+\sqrt{4N-3}}{2N} \le \eta_{\min} \le \frac{1}{N}\left\lfloor\sqrt{3\left\lfloor\frac{N}{2}\right\rfloor}\right\rfloor$		

When observation windows for time signals are long, hardware design considerations make it convenient to split a sampling pattern into shorter pieces that are repeated periodically. This amounts to grouping data samples in blocks that are acquired using the same pattern. Likewise, the usage of periodic arrays in the spatial domain may also present advantages [16].

In these cases, the uncompressed observations $\boldsymbol{x}$ are divided into B blocks of size $N = L/B$ as $\boldsymbol{x} = [\boldsymbol{x}^T[0], \ldots, \boldsymbol{x}^T[B-1]]^T$, and each block is compressed individually to produce an output block of size M:

$$\boldsymbol{y}[b] = \Phi \boldsymbol{x}[b]. \tag{8}$$

It is clear that one can assemble the vector of compressed observations as $\boldsymbol{y} = [\boldsymbol{y}^T[0], \ldots, \boldsymbol{y}^T[B-1]]^T$ and the matrix $\bar{\Phi}$ from (7) as $\bar{\Phi} = \boldsymbol{I}_B \otimes \Phi$, where $\otimes$ represents the Kronecker product.

In the case of sparse samplers, the block-by-block operation means that the pattern $\mathcal{K}$ can be written as

$$\mathcal{K} = \{m + bN: \quad m \in \mathcal{M}, b = 0, \ldots, B-1\}, \tag{9}$$

where $\mathcal{M} \subset \{0, \ldots, N-1\}$ is the sampling pattern used at each block. For example, $\mathcal{M}$ can be a length-$(N-1)$ linear sparse ruler. Thus, $\mathcal{M}$ can be thought of as the period of $\mathcal{K}$, or we may alternatively say that $\mathcal{K}$ is the result of a B-fold concatenation of $\mathcal{M}$. In sparse sampling schemes of this form, known in the literature as *multicoset samplers* [3], the matrix Φ is the result of selecting the rows of $\boldsymbol{I}_N$ indexed by $\mathcal{M}$.

OPTIMAL DESIGNS

One critical problem in CCS is to design a sampler $\bar{\Phi}$ that preserves the second-order statistical information, in the sense that it allows reconstruction of the uncompressed covariance matrix from the compressed observations.

Design techniques for sparse and dense samplers hinge on different basic principles. Whereas sparse samplers are designed based on discrete mathematics (as explained previously), existing designs for dense samplers rely on probabilistic arguments. Inspired by compressed sensing techniques, these designs generate sampling matrices at random and provide probabilistic guarantees on their admissibility.

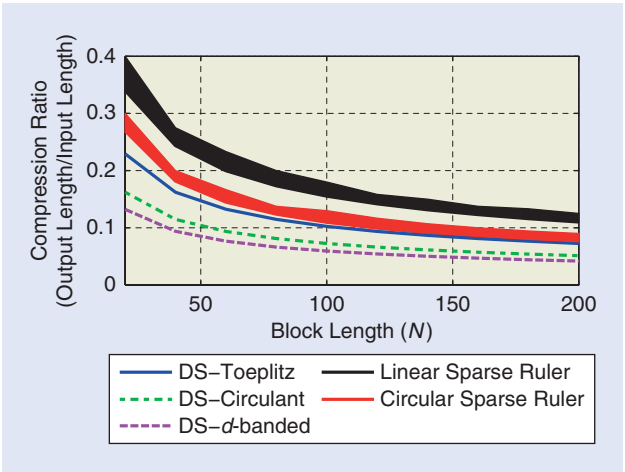
Optimal rates for dense samplers are known in closed form for linear covariance parameterizations such as Toeplitz, circulant, and banded [20]. On the other hand, their evaluation for sparse samplers requires solving combinatorial problems such as the minimal sparse ruler problem. Table 1 summarizes the optimum designs, along with the maximum compression ratios, for the aforementioned parameterizations [20]. The compression ratio is defined as

$$\eta := \frac{|\mathcal{K}|}{L} = \frac{|\mathcal{M}|}{N} \tag{10}$$

and satisfies $0 \le \eta \le 1$. Note that the stronger the compression, the smaller η. It can also be interpreted as the reduction in the average sampling rate: if $x[l]$ represents the sample sequence acquired at the Nyquist rate $1/T_s$, for instance, then the compressed sequence $y[k]$ corresponds to an average sampling rate of η/T_s.

The designs from Table 1 are compared in Figure 3. The vertical axis depicts the reduction in the sampling rate that can be achieved in applications such as compressive wideband spectrum sensing [24]–[26], where the spectrum occupancy over a very wide band is decided via power-spectrum estimation. Note that the sampling rate can be reduced considerably, even in the absence of sparsity. For instance, even for a moderate block length of $N = 50$, the minimum sampling rate is less than one-fourth of the Nyquist rate in all cases. Asymptotically for increasing N, sampling rate savings are proportional to $1/\sqrt{N}$ (cf. Table 1).

The superior efficiency of dense samplers over their sparse counterpart also manifests itself in Figure 3. In fact, it can be



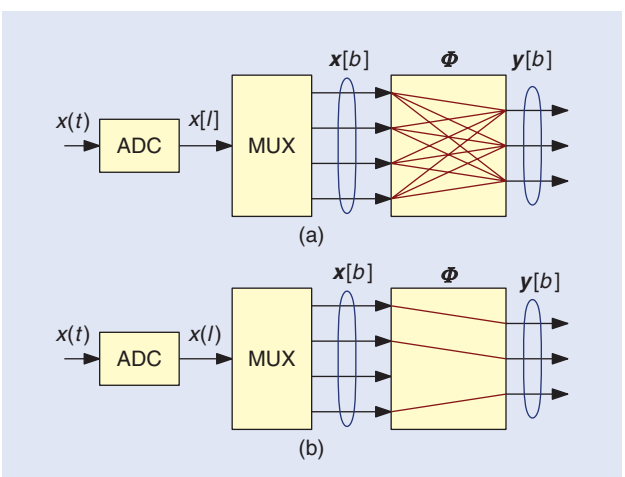
[FIG3] The optimum compression ratios when $B = 10$ and $d = BN/3$ using dense samplers (DSs) for Toeplitz, circulant, and d-banded matrices, and using a linear sparse ruler and a circular sparse ruler. Moderate block lengths yield strong compression.

shown that certain random designs for dense samplers achieve optimal compression ratios with probability one, in the sense that no other sampler (either dense or sparse) can achieve a lower ratio.

TECHNOLOGIES

The acquisition systems that can be used to implement the previously described sampling schemes, which are essentially the same as those used by many other sub-Nyquist acquisition techniques, have recently experienced an intense development.

For example, time signals can be compressively acquired using C-ADCs such as interleaved ADCs [27], random demodulators [5], modulated wideband converters [4], and random modulators pre-integrators [28]. If $\boldsymbol{x}$ contains the Nyquist samples of $x(t)$, their operation can be described by (7) (see Figure 4). Note, however, that no C-ADC internally acquires Nyquist samples since this would entail precisely the disadvantages of conventional ADCs that they attempt to avoid. Nonetheless, they represent a convenient mathematical abstraction.



[FIG4] The mathematical model for the operation of a C-ADC: (a) dense sampler and (b) sparse sampler.

As for spatial signals, sparse samplers can be easily implemented by removing unused antennas. On the other hand, dense samplers require analog combining (see, e.g., [23]).

MAIN APPLICATIONS

The problems that can be formulated in CCS terms are those relying exclusively on the second-order moments of a certain signal $\boldsymbol{x}$. In this section, we elaborate on the mathematical formulation of the signal processing problems involved in some of the main applications. In each case, we indicate the set of basis matrices $\mathcal{S} = \{\boldsymbol{\Sigma}_0, \ldots, \boldsymbol{\Sigma}_{S-1}\}$ to be used [see (4)].

APPLICATIONS IN THE TIME DOMAIN

CCS is especially convenient to acquire wideband signals, whose rapid variations cannot be easily captured by conventional ADCs. As described in the section "Technologies," this difficulty motivates the usage of C-ADCs, whose operation can be described by (7). Their usage in CCS has been considered in a number of applications where acquisition designs and reconstruction algorithms have been proposed. Some of them are detailed next.

- *Compressive power-spectrum estimation:* The goal is to estimate $\boldsymbol{\Sigma}_x$ from $\boldsymbol{y}$ subject to the constraint that $\boldsymbol{\Sigma}_x$ is Hermitian Toeplitz and positive semidefinite. This means that the matrices in $\mathcal{S}$ span the subspace of Hermitian Toeplitz matrices. If the length L (in samples) of the acquisition window is greater than the length of the autocorrelation sequence $\sigma[m]$, then $\mathcal{S}$ can be set to a basis of the subspace of d-banded matrices [19]. Other approaches in the literature follow from the consideration of bases for the subspace of circulant matrices, which arise by stating the problem in the frequency domain [14], [15]. The positive (semi)definiteness of $\boldsymbol{\Sigma}_x$ can be ignored to obtain simple estimators, or it can be enforced using methods like those in [10].
- *Wideband spectrum sensing:* Applications such as dynamic spectrum sharing in cognitive radio networks [29] require monitoring the power of different transmitters operating on wide frequency bands. Suppose that a spectrum sensor is receiving the signal $\boldsymbol{x} = \sum_i \sqrt{\alpha_i}\boldsymbol{x}^{(i)}$, where the component $\sqrt{\alpha_i}\boldsymbol{x}^{(i)}$ contains the Nyquist samples of the signal received from the ith transmitter. If $\boldsymbol{x}^{(i)}$ is power normalized, then α_i is the power received from the ith transmitter. Since the second-order statistics of $\boldsymbol{x}^{(i)}$, collected in $\boldsymbol{\Sigma}_i = E\{\boldsymbol{x}^{(i)}(\boldsymbol{x}^{(i)})^H\}$, are typically known [25], [30], [31], estimating the power of each transmitter amounts to estimating the α_i's in the expansion (4).

  CCS is of special relevance in this application since the typically large number of transmitters means that $x(t)$ is wideband, which motivates the usage of C-ADCs. Various estimation algorithms have been proposed on these grounds in [30].
- *Frequency estimation:* C-ADCs can be used to identify sinusoids in wideband signals [32]. If R denotes the number of sinusoids, the uncompressed signal samples can be modeled as $x[l] = \sum_{i=0}^{R-1} s_i a^{(i)}[l] + w[l]$, where $s_i \in \mathbb{C}$ is random, $w[l]$ is noise, and $a^{(i)}[l] = e^{\jmath\omega_i l}$ is a complex exponential whose frequency ω_i is to be estimated, possibly

along with the variance of s_i. This is the problem of estimating a sparse power spectrum [5].

Many algorithms for estimating these frequencies rely on estimates of the covariance matrix $\mathbf{\Sigma}_x = E\{\boldsymbol{x}\boldsymbol{x}^H\}$, which is known to be Hermitian Toeplitz and positive semidefinite [11]. From the observations provided by a C-ADC, one can first reconstruct $\mathbf{\Sigma}_x$ and subsequently apply one of the existing techniques that take $\mathbf{\Sigma}_x$ as the input. To accomplish such reconstruction, one can use (4), with $\mathcal{S}$ being a set spanning the subspace of Hermitian Toeplitz matrices.

APPLICATIONS IN THE SPATIAL DOMAIN

In applications requiring estimating the so-called angular spectrum (e.g., sonar, radar, astronomy, localization), introducing compression may considerably decrease hardware costs. In schemes using sparse sampling (see, e.g., [6]–[8], [13], and [33]), only the antennas corresponding to the nonnull columns of $\bar{\mathbf{\Phi}}$ need to be physically deployed, whereas in schemes employing dense sampling [23], the number of antennas is preserved after introducing compression, but the number of RF chains is reduced.

In applications employing CCS, the received signal is typically modeled as a sum of incoherent planar waves emitted by a collection of sources in the far field. The spatial field produced by each source results in a Toeplitz spatial covariance matrix which depends on the angle of arrival of that source. The sum of all contributions and noise, assumed white for simplicity, therefore produces a Toeplitz $\mathbf{\Sigma}_x$.

Two problems are usually considered:

- *Incoherent imaging:* If a continuous source distribution is assumed, then the angular spectrum is dense. The problem can be formulated as described previously for compressive power-spectrum estimation, since the only structure present is that $\mathbf{\Sigma}_x$ is Hermitian Toeplitz and positive semidefinite [8]. However, recent works show that the problem can also be stated using circulant covariance matrices [16], [17].
- *Direction-of-arrival estimation:* The goal is to estimate the angles of arrival of a finite number of sources. A broad family of methods exists to this end (see, e.g., [8], [13], [33], and [34]), most of them following the same principles as described previously for frequency estimation, since both problems admit the formulation of sparse power-spectrum estimation.

Most applications listed in this section have been covered with the two compression methods introduced in previous sections, i.e., sparse and dense sampling, either periodic or nonperiodic. For time signals, periodicity typically arises because of the block-by-block operation of C-ADCs (see, e.g., [19], [30], and [35]); for spatial signals, by consideration of periodic arrays [16], [17].

ESTIMATION AND DETECTION

Having described the model and compression schemes for CCS, we turn our attention to the reconstruction problem. For estimation, it boils down to recovering $\mathbf{\Sigma}_x$ in (4) from the compressive measurements $\boldsymbol{y}$.

Since $\boldsymbol{y} = \bar{\mathbf{\Phi}}\boldsymbol{x}$, it follows that $\mathbf{\Sigma}_y = \bar{\mathbf{\Phi}}\mathbf{\Sigma}_x\bar{\mathbf{\Phi}}^H$. If $\mathbf{\Sigma}_x$ is given by (4), $\mathbf{\Sigma}_y$ can be similarly represented as

$$\mathbf{\Sigma}_y = \sum_{i=0}^{S-1} \alpha_i \bar{\mathbf{\Sigma}}_i, \quad \alpha_i \in \mathbb{R}, \tag{11}$$

where $\bar{\mathbf{\Sigma}}_i = \bar{\mathbf{\Phi}}\mathbf{\Sigma}_i\bar{\mathbf{\Phi}}^H$. This means that $\mathbf{\Sigma}_y$ and $\mathbf{\Sigma}_x$ share the coordinates α_i. If the compression is accomplished properly, for example using the designs discussed in previous sections, these coordinates are identifiable and can be estimated from the observations of $\boldsymbol{y}$.

MAXIMUM LIKELIHOOD

If the probability distribution of the observations is known, one may resort to a maximum-likelihood estimate of $\mathbf{\Sigma}_y$. For example, if $\boldsymbol{y}$ is zero-mean Gaussian and

$$\hat{\mathbf{\Sigma}}_y = \frac{1}{T}\sum_{\tau=0}^{T-1} \boldsymbol{y}_\tau \boldsymbol{y}_\tau^H, \tag{12}$$

is the sample covariance matrix of the compressed observations, the maximization of the log-likelihood leads to the following problem:

$$\underset{\{\alpha_i\}_i}{\text{minimize}} \quad \log|\mathbf{\Sigma}_y| + \text{Tr}(\mathbf{\Sigma}_y^{-1}\hat{\mathbf{\Sigma}}_y) \tag{13}$$

subject to (11). Numerous algorithms have been proposed to solve this nonconvex problem (see, e.g., [10], [30], and [36]).

LEAST SQUARES

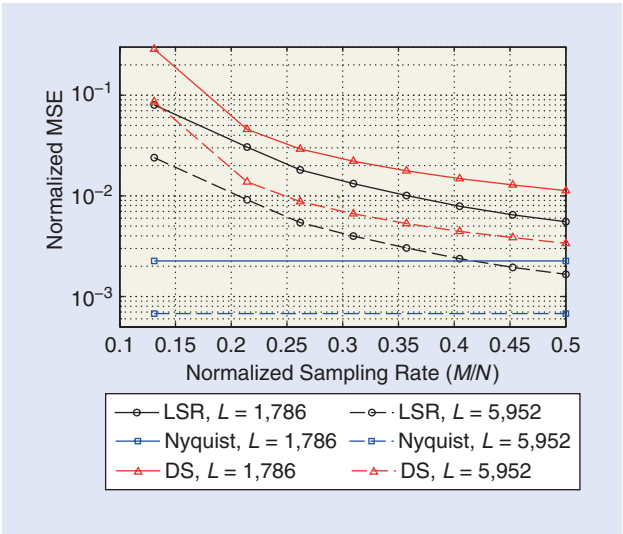
The maximum-likelihood approach involves high computational costs and requires an accurate statistical characterization of the observations. For these reasons, it is customary to rely on geometrical considerations and project the sample covariance matrix onto the span of $\mathcal{S}$.

From $\mathbf{\Sigma}_y = \bar{\mathbf{\Phi}}\mathbf{\Sigma}_x\bar{\mathbf{\Phi}}^H$, it follows that $\boldsymbol{\sigma}_y = (\bar{\mathbf{\Phi}}^* \otimes \bar{\mathbf{\Phi}})\boldsymbol{\sigma}_x$, where $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_x$ are, respectively, the vectorizations of $\mathbf{\Sigma}_y$ and $\mathbf{\Sigma}_x$. Vectorizing (4) yields $\boldsymbol{\sigma}_x = \sum_{i=0}^{S-1} \alpha_i \boldsymbol{\sigma}_i$ or, in matrix form, $\boldsymbol{\sigma}_x = S\boldsymbol{\alpha}$, where we have arranged the vectors $\boldsymbol{\sigma}_i$ as columns of the matrix S and the coordinates α_i as elements of the vector $\boldsymbol{\alpha}$. This results in the relation $\boldsymbol{\sigma}_y = (\bar{\mathbf{\Phi}}^* \otimes \bar{\mathbf{\Phi}})S\boldsymbol{\alpha}$. If the $M^2B^2 \times S$ matrix $(\bar{\mathbf{\Phi}}^* \otimes \bar{\mathbf{\Phi}})S \in \mathbb{C}$ is full-column rank, substituting $\boldsymbol{\sigma}_y$ by a sample estimate $\hat{\boldsymbol{\sigma}}_y$ produces an overdetermined system $\hat{\boldsymbol{\sigma}}_y = (\bar{\mathbf{\Phi}}^* \otimes \bar{\mathbf{\Phi}})S\hat{\boldsymbol{\sigma}}$, whose solution via least squares yields the desired estimate in closed form [14], [19], [30]:

$$\hat{\mathbf{\Sigma}}_x^{\text{LS}} = \text{vec}^{-1}\{S[(\bar{\mathbf{\Phi}}^* \otimes \bar{\mathbf{\Phi}})S]^\dagger \hat{\boldsymbol{\sigma}}_y\}. \tag{14}$$

Here, the operator $\text{vec}^{-1}\{\cdot\}$ restacks a vector into a square matrix.

Figure 5 illustrates the performance of this technique when $\mathbf{\Sigma}_x$ is 168-banded (see [19] for more details) and several sampling designs are used. Clearly, the mean squared error of the estimate is larger when compression is introduced since it reduces the total number of samples. This effect is not exclusive to least-squares estimation—it negatively affects any estimator. For this reason, including compression usually requires longer observation time if a certain target performance metric is to be achieved. This does not conflict with the ultimate purpose of



[FIG5] The mean-squared error of the estimate of the least-squares algorithm when Σ_x is 168-banded. (Figure adapted from [19].)

compression, which is to reduce the average sampling rate—a parameter that critically affects the hardware cost.

However, note that this approach does not exploit the fact that Σ_x is positive semidefinite. This constraint can be enforced to improve the estimation performance at the expense of greater complexity. For instance, one may attempt to minimize the least squares cost $\| \hat{\sigma}_y - (\bar{\Phi}^* \otimes \bar{\Phi}) S \hat{\sigma} \|^2$ subject to the constraint $\Sigma_x \geq 0$, which is a convex problem. Other constraints can also be imposed if more prior information is available. For instance, the elements of $\hat{\alpha}$ might be nonnegative [30], in which case one would introduce the constraint $\hat{\alpha} \geq 0$. It can be known that $\hat{\alpha}$ is sparse either by itself or on a linearly transformed domain, in which case one may impose the constraint $\| F_s \hat{\alpha} \|_0 \leq S_0$, where S_0 is the number of nonzero entries and F_s takes $\hat{\alpha}$ to the domain where it is sparse. For instance, the elements of $F_s \hat{\alpha}$ may be samples of the power spectrum [37]. Since the zero-norm in this constraint is not convex, it is typically relaxed to an ℓ_1-norm. For example, an ℓ_1-norm regularized least-squares formulation can be adopted as follows:

$$\underset{\alpha}{\text{minimize}} \ \| \hat{\sigma}_y - (\bar{\Phi}^* \otimes \bar{\Phi}) S \hat{\alpha} \|^2 + \lambda \| F_s \hat{\alpha} \|_1. \tag{15}$$

In (15), signal compression is induced by the statistical structure of Σ_x beyond sparsity, while the additional sparsity structure can lead to stronger compression at the expense of increased computational complexity compared to the closed-form solution in (14).

DETECTION

In detection theory, we are interested in deciding whether a signal of interest is present or not. This operation is typically hindered by the presence of noise and other waveforms, such as clutter in radar or interference in communications.

In many cases, this problem can be stated in terms of the second-order statistics of the signals involved, so the goal is to decide one of the following hypotheses:

$$\begin{aligned} \mathcal{H}_0&: \ \Sigma_x = \Sigma_w \\ \mathcal{H}_1&: \ \Sigma_x = \Sigma_r + \Sigma_w, \end{aligned} \tag{16}$$

where Σ_r and Σ_w, respectively, collect the second-order statistics of the signal of interest and noise/interference. Our decision must be based on the observation of the compressed samples $y = \bar{\Phi} x$, whose covariance matrix Σ_y is given by $\bar{\Phi} \Sigma_w \bar{\Phi}^H$ under $\mathcal{H}_0$ and by $\bar{\Phi} (\Sigma_r + \Sigma_w) \bar{\Phi}^H$ under $\mathcal{H}_1$. A most powerful detection rule exists for this simple setting and can be found using the Neyman–Pearson lemma [11]. If $p(y; \mathcal{H}_i)$ denotes the density under hypothesis $\mathcal{H}_i$, this rule decides $\mathcal{H}_1$ when the ratio $p(y; \mathcal{H}_1)/p(y; \mathcal{H}_0)$ exceeds a certain threshold set to achieve a target probability of false alarm [11].

More general problems arise by considering basis expansions like the one in (4). In this case, the goal may be to decide whether one of the α_i, say α_0, is positive or zero, while the others are unknown and treated as nuisance parameters [30]. Since in these cases no uniformly most-powerful test exists, one must resort to other classes of detectors, such as the generalized likelihood ratio test, which makes a decision by comparing $p(y; \hat{\alpha}_{\mathcal{H}_1})/p(y; \hat{\alpha}_{\mathcal{H}_0})$ against a threshold, where $\hat{\alpha}_{\mathcal{H}_i}$ is the maximum-likelihood estimate of α under hypothesis $\mathcal{H}_i$ [30].

MODAL ANALYSIS

As mentioned in the section "Main Applications," the problem of estimating the frequency of a number of noise-corrupted sinusoids and the problem of estimating the direction of arrival of a number of sources in the far field are instances of the class of sparse spectrum estimation problems, which allow a common formulation as modal analysis [11].

Suppose that the observations are given by

$$x = \sum_{i=0}^{R-1} s_i a^{(i)} + w = As + w, \tag{17}$$

where $a^{(i)} = [1, e^{j\omega_i}, \ldots, e^{j\omega_i (L-1)}]^T$ are the so-called steering vectors, $A = [a^{(0)}, \ldots, a^{(R-1)}]$ is the manifold matrix, w is noise and the coefficients s_i, collected in the vector s, are uncorrelated random variables. The structure of $a^{(i)}$ stems from the fact that each antenna receives the signal s_i with a different phase shift. Because the antennas are uniformly spaced in a ULA, the relative phase shift between each pair of antennas is an integer multiple of a normalized quantity ω_i, which is a function of the angle of arrival.

The covariance matrix of x is given by

$$\Sigma_x = A \Sigma_s A^H + \sigma_w^2 I_L, \tag{18}$$

where σ_w^2 is the power of the noise process, assumed white for simplicity, and Σ_s is the covariance matrix of s, which is diagonal since the sources are uncorrelated. Note that these assumptions result in Σ_x having a Toeplitz structure.

The compressed observations can be written as $y = \bar{\Phi} x = \bar{A} s$, where $\bar{A} = \bar{\Phi} A$, and have covariance matrix

$$\boldsymbol{\Sigma}_y = \bar{\boldsymbol{A}}\boldsymbol{\Sigma}_s\bar{\boldsymbol{A}}^H + \sigma_w^2\bar{\boldsymbol{\Phi}}\bar{\boldsymbol{\Phi}}^H. \quad (19)$$

The parameters ω_i can be estimated from $\boldsymbol{\Sigma}_y$ using adaptations of traditional techniques such as multiple signal classification [35] and the minimum variance distortionless response algorithm [38].

THE MEAN SQUARED ERROR OF THE ESTIMATE IS LARGER WHEN COMPRESSION IS INTRODUCED SINCE IT REDUCES THE TOTAL NUMBER OF SAMPLES.

Alternative approaches are based on the observation that the vectorization of (19) can be written in terms of the Khatri–Rao product, defined as the columnwise application of the Kronecker product, as

$$\operatorname{vec}(\boldsymbol{\Sigma}_y) = (\bar{\boldsymbol{A}}^* \odot \bar{\boldsymbol{A}})\operatorname{diag}\{\boldsymbol{\Sigma}_s\} + \sigma_w^2\operatorname{vec}(\bar{\boldsymbol{\Phi}}\bar{\boldsymbol{\Phi}}^H). \quad (20)$$

The matrix $\bar{\boldsymbol{A}}^* \odot \bar{\boldsymbol{A}}$ can be thought of as a virtual manifold matrix, since this expression has the same structure as (17) [13], [39]. An especially convenient structure is when $\bar{\boldsymbol{A}}^* \odot \bar{\boldsymbol{A}}$ contains all the rows in the manifold matrix of a ULA [40]. To obtain this structure, array geometries like two-level nested arrays [13], coprime arrays [32], and linear sparse rulers [34] can be used.

Other approaches stem from the idea of gridding. One can construct the matrix $\bar{\boldsymbol{A}}$ using a fine grid of angles ω_i and then estimate $\boldsymbol{s}$ from $\boldsymbol{y} = \bar{\boldsymbol{A}}\boldsymbol{s}$ exploiting the idea that most of its components will be zero since, for a grid fine enough, most of the columns of $\bar{\boldsymbol{A}}$ will correspond to angles where there are no sources. In other words, $\boldsymbol{s}$ is sparse, which means that the techniques from [5] and [41] can be applied to recover this vector. This technique does not have to rely on second-order statistics, but similar grid-based approaches can be devised that operate on (20) instead [42].

PREPROCESSING

Most of the methods described in this article make use of the sample covariance matrix of $\boldsymbol{y}$, defined in (12). Under general conditions, the average $T^{-1}\sum_\tau \boldsymbol{y}_\tau\boldsymbol{y}_\tau^H$ converges to the true $\boldsymbol{\Sigma}_y$ as T becomes large. If compression does not destroy relevant second-order statistical information, the matrix $\boldsymbol{\Sigma}_y$ contains all the information required to identify all entries of $\boldsymbol{\Sigma}_x$, but a considerably large number T of snaspshots will be required for $\hat{\boldsymbol{\Sigma}}_y$ to be close to $\boldsymbol{\Sigma}_y$, which is necessary to obtain a reasonable estimate of $\boldsymbol{\Sigma}_x$.

Typically, in those applications involving spatial signals, the outputs of all antennas are synchronously sampled. If $\boldsymbol{y}_\tau$ collects the samples acquired at time instant τ, it is clear that multiple observations of $\boldsymbol{y}$ can be obtained by considering successive snapshots $\tau = 0,1,\ldots,T-1$. This means that, whereas $\boldsymbol{y}$ contains samples across space, the different snapshots are acquired along the time dimension. Conversely, in applications involving time-domain signals, $\boldsymbol{y}$ contains samples acquired over time. A possible means to observe multiple realizations is by considering the vectors $\boldsymbol{y}_\tau$ observed at different locations $\tau = 0, 1, \ldots, T-1$. In this case, while $\boldsymbol{y}$ contains time samples, τ ranges across space. This establishes a duality relation between the space and the time domains: when the observed signals are defined on one domain, multiple observations can be acquired over the other.

Unfortunately, many applications do not allow averaging over the dual domain, and one must cope with a single observation, say $\boldsymbol{y}_0$, producing the estimate $\hat{\boldsymbol{\Sigma}}_y = \boldsymbol{y}_0\boldsymbol{y}_0^H$. This matrix is not a satisfactory estimate of $\boldsymbol{\Sigma}_y$ since it is always rank one and is not Toeplitz. For this reason, an estimation/detection method working on this kind of estimate may exhibit a poor performance.

The key observation in this case is that, although multiple realizations cannot be acquired, sometimes it is possible to gather a large number of samples in the domain where the signal is defined. One can therefore exploit the Toeplitz structure of $\boldsymbol{\Sigma}_x$ to obtain a more convenient estimate [30]. In particular, because of the block-by-block operation described by (8), the fact that $\boldsymbol{\Sigma}_x$ is Toeplitz means that $\boldsymbol{\Sigma}_y$ is block Toeplitz; that is, it can be written as

$$\boldsymbol{\Sigma}_y = \begin{bmatrix} \boldsymbol{\Sigma}_y[0] & \boldsymbol{\Sigma}_y[-1] & \cdots & \boldsymbol{\Sigma}_y[-B+1] \\ \boldsymbol{\Sigma}_y[1] & \boldsymbol{\Sigma}_y[0] & \cdots & \boldsymbol{\Sigma}_y[-B+2] \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_y[B-1] & \boldsymbol{\Sigma}_y[B-2] & \cdots & \boldsymbol{\Sigma}_y[0] \end{bmatrix}, \quad (21)$$

where the (nonnecessarily Toeplitz) $M \times M$ blocks $\boldsymbol{\Sigma}_y[k]$ are given by

$$\boldsymbol{\Sigma}_y[k] = E\{\boldsymbol{y}[b]\boldsymbol{y}^H[b-k]\}, \quad \forall b. \quad (22)$$

This suggests the estimate

$$\hat{\boldsymbol{\Sigma}}_y[k] = \frac{1}{\text{number of terms}}\sum_b \boldsymbol{y}[b]\boldsymbol{y}^H[b-k]. \quad (23)$$

Moreover, since $\boldsymbol{\Sigma}_y$ is Hermitian, this computation needs only to be carried out for $k = 0,\ldots,B-1$. More sophisticated estimates exhibiting different properties were analyzed in [30].

Another observation is that the smaller k, the higher the quality of the estimates of $\boldsymbol{\Sigma}_y[k]$. The reason is that the number of averaging terms in (23) is larger for blocks lying close to the main diagonal than for distant ones. Thus, it seems reasonable to operate on a cropped covariance matrix

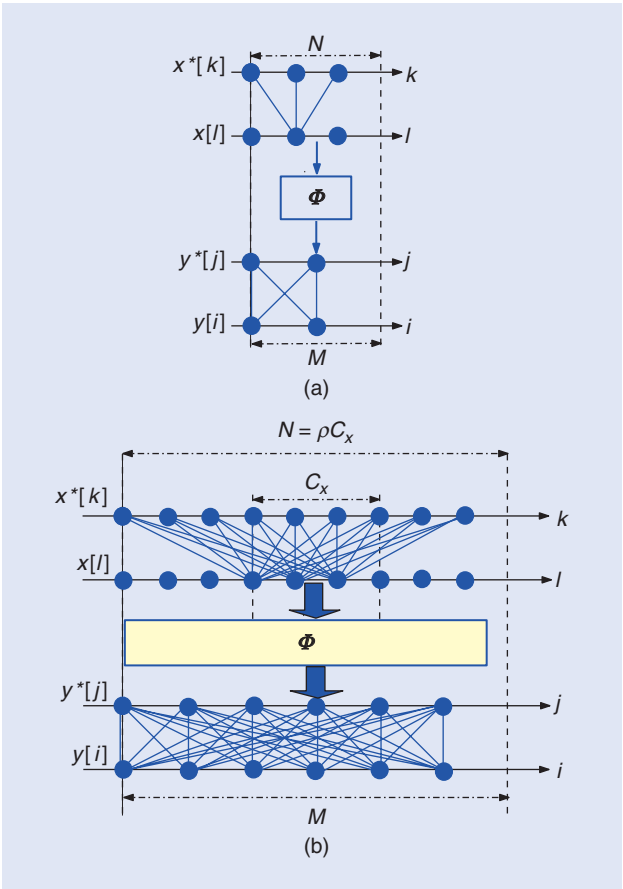
$$\boldsymbol{\Sigma}_y = \begin{bmatrix} \boldsymbol{\Sigma}_y[0] & \boldsymbol{\Sigma}_y[-1] & \cdots & \boldsymbol{\Sigma}_y[-\tilde{B}+1] \\ \boldsymbol{\Sigma}_y[1] & \boldsymbol{\Sigma}_y[0] & \cdots & \boldsymbol{\Sigma}_y[-\tilde{B}+2] \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_y[\tilde{B}-1] & \boldsymbol{\Sigma}_y[\tilde{B}-2] & \cdots & \boldsymbol{\Sigma}_y[0] \end{bmatrix}, \quad (24)$$

where $\tilde{B} < B$. Note that, in this case, the dimension of the cropped matrix is less than the length of the observation vector $\boldsymbol{y}$.

In certain cases, this technique leads to important computational savings at a small performance loss since the terms being retained are those of the highest quality [30].

ADVANCED TECHNIQUES

Having explained the basic principles of CCS, we now illustrate the broad applications of CCS by considering other forms of second-order statistics as well as implementation issues in practical systems.



[FIG6] (a) A compression of a block of $N = 3$ samples of a WSS signal using a 2×3 matrix Φ, which produces compressed blocks of $M = 2$ samples. (b) A compression of a block of $N = 9$ samples of a cyclostationary signal, which produces compressed blocks of $M = 6$ samples.

CYCLOSTATIONARITY

Cyclostationarity is exhibited in many man-made signals with inherent periodicity, which is a useful feature for estimation, detection, and classification of digital communication signals [26]. Although there are several methods to reconstruct the second-order statistics of a cyclostationary signal from compressed observations (see, e.g., [26], [43], and [44]), in this section, we only illustrate the main principles underlying these techniques using a simple model.

We say that a signal is cyclostationary if its time-varying covariance function is periodic. Formally, the time-varying covariance function of a zero-mean process $x[l]$ is defined as $\sigma[l, k] = E\{x[l]x^*[l-k]\}$, and it is said to be periodic when there exists an integer C_x, called the *cyclic period*, such that $\sigma[l + n_c C_x, k] = \sigma[l, k]$ for any integer n_c [26], [43]. Although other forms of cyclostationarity exist, we confine ourselves to this one for simplicity. Note that cyclostationary signals generalize WSS signals, since the latter may be viewed as a particular case of the former with $C_x = 1$.

Suppose that the length of the sampling block is an integer multiple of the cyclic period, that is, $N = \rho C_x$ for some integer ρ. Then, the vector $\boldsymbol{x}[b]$ can be divided into ρ subblocks of length C_x as

$$\boldsymbol{x}[b] = [\tilde{\boldsymbol{x}}^T[b\rho], \tilde{\boldsymbol{x}}^T[b\rho + 1], \ldots, \tilde{\boldsymbol{x}}^T[b\rho + \rho - 1]]^T. \quad (25)$$

The fact that $\sigma[l, k]$ is periodic along l means that $\boldsymbol{\Sigma}_x$ is block Toeplitz with $C_x \times C_x$ blocks. By defining an $N \times N$ matrix $\boldsymbol{\Sigma}_x[b] = E\{\boldsymbol{x}[b']\boldsymbol{x}^H[b' - b]\}$, we can write

$$\boldsymbol{\Sigma}_x = \begin{bmatrix} \boldsymbol{\Sigma}_x[0] & \boldsymbol{\Sigma}_x[-1] & \cdots & \boldsymbol{\Sigma}_x[-B+1] \\ \boldsymbol{\Sigma}_x[1] & \boldsymbol{\Sigma}_x[0] & \cdots & \boldsymbol{\Sigma}_x[-B+2] \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_x[B-1] & \boldsymbol{\Sigma}_x[B-2] & \cdots & \boldsymbol{\Sigma}_x[0] \end{bmatrix}, \quad (26)$$

where the blocks $\boldsymbol{\Sigma}_x[b]$ also have a block Toeplitz structure with blocks $\boldsymbol{\Sigma}_{\tilde{x}}[\varrho] = E\{\tilde{\boldsymbol{x}}[\varrho']\tilde{\boldsymbol{x}}^H[\varrho' - \varrho]\}$:

$$\boldsymbol{\Sigma}_x[b] = \begin{bmatrix} \boldsymbol{\Sigma}_{\tilde{x}}[b\rho] & \cdots & \boldsymbol{\Sigma}_{\tilde{x}}[b\rho - \rho + 1] \\ \boldsymbol{\Sigma}_{\tilde{x}}[b\rho + 1] & \cdots & \boldsymbol{\Sigma}_{\tilde{x}}[b\rho - \rho + 2] \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{\tilde{x}}[b\rho + \rho - 1] & \cdots & \boldsymbol{\Sigma}_{\tilde{x}}[b\rho] \end{bmatrix}. \quad (27)$$

Cyclostationarity provides an alternative perspective to understand compression of second-order statistics, even for WSS sequences. The main idea is that the resulting sequence $y[i]$ of compressed observations is cyclostationary with cyclic period M, which is larger than that of the original signal C_x.

Figure 6(a) intuitively explains this effect for a WSS signal ($C_x = 1$) satisfying $\sigma[l] = 0$ for $|l| > 1$. In that figure, the dots on the l-axis represent a block of $N = 3$ samples of the WSS sequence $x[l]$, and the dots on the k-axis represent their complex conjugates. The three lines connecting the dots in both axes represent the (possibly) different values of correlation between samples. Note that no extra lines need to be drawn since only $\sigma[-1]$, $\sigma[0]$, and $\sigma[1]$ are allowed to be different from zero. Since the correlation of a WSS signal is determined by the time-lags independent of the time origin, only one representative dot along $x[l]$ is chosen as the time origin. A similar representation is provided at the bottom of Figure 6(a) for the compressed sequence $y[i]$, which can be seen to be cyclostationary with cyclic period $C_y = M$ [just apply the above considerations to (21)]. Note that the four line segments effectively capture all the different correlation values between samples of $y[i]$. Here, $y[i]$ is no longer WSS due to the compression process, and, hence, all time origins along $y[i]$ within a block are selected to depict the correlations.

Observe that, although the number of samples in each block was reduced from three to two after compression, the number of different correlation values has increased from three to four. This means that, whereas one cannot reconstruct the samples of $x[l]$ from $y[i]$ without further assumptions, there is a chance of reconstructing the second-order statistics of $x[l]$ from those of $y[i]$. In fact, if Φ satisfies certain conditions, one can, for instance, estimate $\boldsymbol{\Sigma}_y$ from $y[i]$ using sample statistics and

obtain an estimate of $\mathbf{\Sigma}_x$ via least squares, as described in the section "Least Squares."

Now assume that $x[l]$ is a cyclostationary signal of cyclic period $C_x = 3$ and that $\sigma[l, k]$ is such that each subblock of C_x samples is only correlated with the neighboring subblocks. Figure 6(b) illustrates a case where a block of $N = \rho C_x = 9$ samples is compressed to produce a block of $M = 6$ samples. As before, all (possibly) distinct correlation values have been represented with the corresponding line segments. Observe that, although the number of output samples is lower than the number of input samples, it may be possible to use the $M^2 = 36$ correlation values at the output to reconstruct the $\rho C_x^2 = 27$ correlation values at the input.

We next describe a reconstruction method based on least squares. Note from (8) and (22) that the $M \times M$ blocks of $\mathbf{\Sigma}_y$ [c.f. (21)] can be written as

$$\mathbf{\Sigma}_y[b] = \mathbf{\Phi}\mathbf{\Sigma}_x[b]\mathbf{\Phi}^H. \tag{28}$$

To exploit the block Toeplitz structure of $\mathbf{\Sigma}_x[b]$ [see (27)], we first vectorize both sides of (28) and apply the properties of the Kronecker product to obtain

$$\mathrm{vec}(\mathbf{\Sigma}_y[b]) = (\mathbf{\Phi}^* \otimes \mathbf{\Phi})\,\mathrm{vec}(\mathbf{\Sigma}_x[b]). \tag{29}$$

Now we rewrite the rightmost vector of this expression as

$$\mathrm{vec}(\mathbf{\Sigma}_x[b]) = \mathbf{T}\boldsymbol{\beta}_x[b], \tag{30}$$

where

$$\begin{aligned}\boldsymbol{\beta}_x[b] = [&\mathrm{vec}^T(\mathbf{\Sigma}_{\tilde{x}}[b\rho]), \mathrm{vec}^T(\mathbf{\Sigma}_{\tilde{x}}[b\rho+1]), \ldots,\\ &\mathrm{vec}^T(\mathbf{\Sigma}_{\tilde{x}}[b\rho+\rho-1]), \mathrm{vec}^T(\mathbf{\Sigma}_{\tilde{x}}[b\rho-\rho+1]),\\ &\ldots, \mathrm{vec}^T(\mathbf{\Sigma}_{\tilde{x}}[b\rho-1])]^T\end{aligned} \tag{31}$$

is a $(2\rho-1)C_x^2 \times 1$ vector containing all the possibly distinct entries of $\mathbf{\Sigma}_x[b]$, and where $\mathbf{T}$ is the $N^2 \times (2\rho-1)C_x^2$ repetition matrix, which maps (and repeats) the elements of $\boldsymbol{\beta}_x[b]$ into $\mathrm{vec}(\mathbf{\Sigma}_x[b])$. Substituting (30) in (29) yields

$$\mathrm{vec}(\mathbf{\Sigma}_y[b]) = (\mathbf{\Phi}^* \otimes \mathbf{\Phi})\mathbf{T}\boldsymbol{\beta}_x[b]. \tag{32}$$

Note that, while $\mathbf{\Phi}$ generally has more columns than rows (as $M < N$), the $M^2 \times (2\rho-1)C_x^2$ matrix $(\mathbf{\Phi}^* \otimes \mathbf{\Phi})\mathbf{T}$ can have more rows than columns. Hence, under certain conditions, (32) is an overdetermined system for each $b = -B+1, \ldots, 0, 1, \ldots, B-1$. Substituting $\mathbf{\Sigma}_y[b]$ by a sample estimate, one can obtain an estimate of $\boldsymbol{\beta}_x[b]$ as the least-squares solution of that system and obtain an estimate of $\mathbf{\Sigma}_x$ by plugging the result in (30).

This approach has been proposed in [43] using dense samplers. A more specific case is discussed in [44], which specifically proposes the usage of a sparse matrix $\mathbf{\Phi}$ with a block diagonal structure.

DYNAMIC SAMPLING

There are situations where the signal itself does not possess evident covariance structure, but we can effect compression by means of dynamic sampling.

Let us go back to the array processing example in the section "A Warm-Up Example," where the Toeplitz structure of $\mathbf{\Sigma}_x$ allowed us to estimate $\mathbf{\Sigma}_x$ using $M < N$ antennas. This structure relies on the assumption that the sources are uncorrelated. If this is not the case, then the only structure present in $\mathbf{\Sigma}_x$ is Hermitian and positive semidefinite, which means that $\mathbf{\Sigma}_x$ cannot be estimated with fewer than N antennas.

A possible way to circumvent this problem is to adopt a dynamic scheme where a full array of N antennas (the uncompressed array) is deployed but only a certain subset of antennas is activated at each time slot [40]. The activation pattern may change periodically over time, which allows computing sample statistics for every activation pattern. With this technique, only a small number of RF chains need to be deployed. This is illustrated in Figure 7, where only $K = 4$ out of the $L = 7$ physical antennas are active at each time slot. The antenna selection may be implemented using analog circuitry. Note that a similar scheme could be used relying on dense samplers. Alternative settings include [45], where different arrays are obtained by sampling different frequencies.

To estimate $\mathbf{\Sigma}_x$, the least-squares method from previous sections can be used. Let $\bar{\mathbf{\Phi}}_g$ denote the $K \times L$ compression matrix used during the gth time slot. The covariance matrix of the compressed observations at time slot g is given by

$$\mathbf{\Sigma}_{y_g} = \bar{\mathbf{\Phi}}_g \mathbf{\Sigma}_x \bar{\mathbf{\Phi}}_g^H. \tag{33}$$

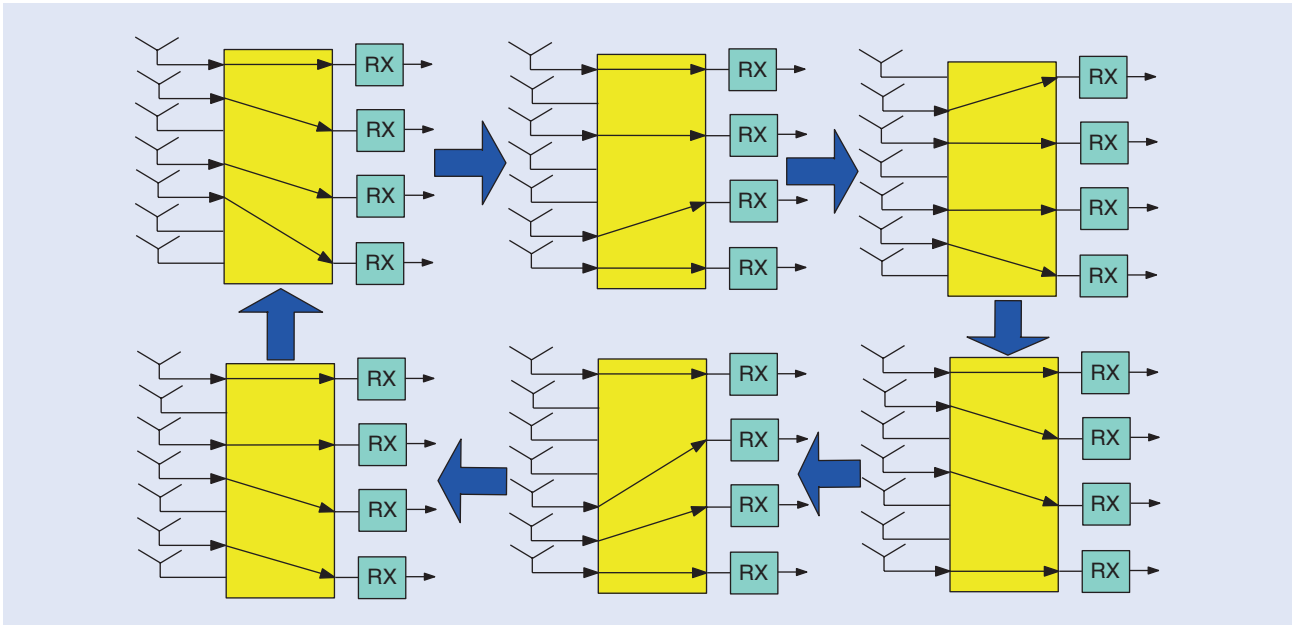
Vectorizing both sides and combining the result for the G time slots in each period yields

$$\begin{bmatrix}\mathrm{vec}(\mathbf{\Sigma}_{y_0})\\ \mathrm{vec}(\mathbf{\Sigma}_{y_1})\\ \vdots\\ \mathrm{vec}(\mathbf{\Sigma}_{y_{G-1}})\end{bmatrix} = \begin{bmatrix}\bar{\mathbf{\Phi}}_0^* \otimes \bar{\mathbf{\Phi}}_0\\ \bar{\mathbf{\Phi}}_1^* \otimes \bar{\mathbf{\Phi}}_1\\ \vdots\\ \bar{\mathbf{\Phi}}_{G-1}^* \otimes \bar{\mathbf{\Phi}}_{G-1}\end{bmatrix} \mathrm{vec}(\mathbf{\Sigma}_x) = \mathbf{\Psi}\mathrm{vec}(\mathbf{\Sigma}_x). \tag{34}$$

If the $GK^2 \times L^2$ matrix $\mathbf{\Psi}$ has full column rank, then it is possible to estimate $\mathbf{\Sigma}_{y_g}$, $g = 0, \ldots, G-1$ using sample statistics and then obtain an estimate of $\mathbf{\Sigma}_x$ as the least-squares solution of (34). It can be shown that this full rank condition is satisfied if every pair of antennas is simultaneously active in at least one time slot per scanning period [40]. To estimate $\mathbf{\Sigma}_{y_g}$ via sample statistics, one may simply average over the observations in the gth time slot of each period.

COMPRESSIVE COVARIANCE ESTIMATION OF MULTIBAND SIGNALS

When uncorrelated signal sources are concerned, a multiband signal structure arises in many applications [14], [35], [46]. Suppose that our goal is to estimate the second-order statistics, e.g., the power spectrum, of a time-domain (spatial-domain) signal which



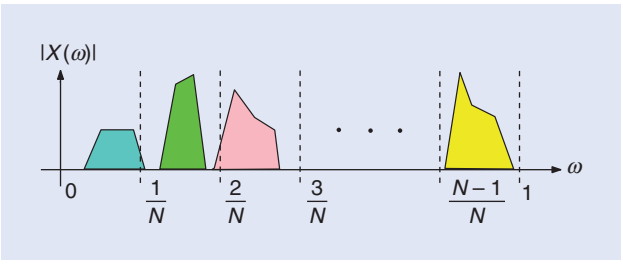
[FIG7] **Implementation of dynamic spatial sampling using antenna switching.**

has a multiband structure in the frequency (angular) domain (see Figure 8) [14], [35], [46]. For simplicity, consider a time-domain signal $x(t)$, although the discussion immediately carries over to the spatial domain [46]. We show how this problem can be cast as the problem of compressing a circulant covariance matrix (see the section "Importance of Covariance Structures").

The trick is to reformulate the problem in the frequency domain. Let $X(\omega)$ denote the discrete-time Fourier transform (DTFT) at digital frequency $\omega \in [0, 1)$ of the sequence $x[l]$, $l = 0, \ldots, L-1$. Let us also split the frequency axis $\omega \in [0, 1)$ into N bins of size $1/N$ (see Figure 8) and introduce, for $\omega \in [0, 1/N)$, the $N \times 1$ vector $\boldsymbol{x}(\omega) = [X(\omega), X(\omega + 1/N), \ldots, X(\omega + N - 1/N)]^T$.

Now, suppose that instead of concatenating the vectors $\boldsymbol{x}[b]$ vertically to form $\boldsymbol{x}$ (see the section "Compression"), we arrange them as columns of the $N \times B$ matrix $\boldsymbol{X}$. Repeating the same operation for the compressed samples in $\boldsymbol{y}$ produces the $M \times B$ matrix $\boldsymbol{Y}$. Clearly, since $\bar{\Phi} = \boldsymbol{I}_B \otimes \Phi$, it follows that the compression model of (7) can be rewritten as

$$\boldsymbol{Y} = \Phi \boldsymbol{X}. \tag{35}$$



[FIG8] **An example of a signal with a multiband structure. Here, the digital frequency axis ω is split into N uniform bins.**

Let us form the $N \times 1$ vector $\bar{\boldsymbol{x}}(\omega)$, whose nth entry contains the DTFT of the nth row of $\boldsymbol{X}$. Note that the collection of samples in each row of $\boldsymbol{X}$ is the result of downsampling $x[l]$ by a factor of N. This operation produces N aliases in the frequency domain, which means that the spectrum has period $1/N$. Thus, it suffices to consider $\bar{\boldsymbol{x}}(\omega)$ in the frequency interval $\omega \in [0, 1/N)$. Likewise, define the $M \times 1$ vector $\bar{\boldsymbol{y}}(\omega)$, $\omega \in [0, 1/N)$, as the vector containing the DTFTs of the rows of $\boldsymbol{Y}$. Clearly, (35) can then be expressed in the frequency domain using these vectors:

$$\bar{\boldsymbol{y}}(\omega) = \Phi \bar{\boldsymbol{x}}(\omega). \tag{36}$$

The relationship between $\boldsymbol{x}(\omega)$ and $\bar{\boldsymbol{x}}(\omega)$ can be shown to be given by [14], [35], [46]

$$\bar{\boldsymbol{x}}(\omega) = \frac{1}{N} \boldsymbol{F}_N^H \boldsymbol{x}(\omega), \quad \omega \in [0, 1/N), \tag{37}$$

where $\boldsymbol{F}_N$ is the $N \times N$ discrete Fourier transform (DFT) matrix. From (36) and (37), it follows that

$$\boldsymbol{\Sigma}_{\bar{y}}(\omega) = E[\bar{\boldsymbol{y}}(\omega)\bar{\boldsymbol{y}}^H(\omega)] = \Phi \boldsymbol{\Sigma}_{\bar{x}}(\omega) \Phi^H, \tag{38}$$

and

$$\boldsymbol{\Sigma}_{\bar{x}}(\omega) = E[\bar{\boldsymbol{x}}(\omega)\bar{\boldsymbol{x}}^H(\omega)] = \frac{1}{N^2} \boldsymbol{F}_N^H \boldsymbol{\Sigma}_x(\omega) \boldsymbol{F}_N, \tag{39}$$

where $\boldsymbol{\Sigma}_x(\omega) = E[\boldsymbol{x}(\omega)\boldsymbol{x}^H(\omega)]$. If the frequency bands are uncorrelated, for instance, because they were produced by different sources, and if the width of each band is less than $1/N$, which

is the width of the bin, then $\mathbf{\Sigma}_x(\omega)$ in (39) is a diagonal matrix for all $\omega \in [0,1/N)$ [46]. Such a diagonal structure is characteristic of multiband signals, which enables compression beyond sparsity. Likewise, since F_N is a DFT matrix, it implies a circulant structure in $\mathbf{\Sigma}_{\bar{x}}(\omega)$.

Compare (38) with the expression $\mathbf{\Sigma}_y = \bar{\mathbf{\Phi}}\mathbf{\Sigma}_x\bar{\mathbf{\Phi}}^H$ from previous sections. We observe that $\mathbf{\Sigma}_{\bar{y}}(\omega)$ is the result of compressing the circulant matrix $\mathbf{\Sigma}_{\bar{x}}(\omega)$. A possible means of estimating the second-order statistics of $x[l]$ is, for example, by using sample statistics to estimate $\mathbf{\Sigma}_{\bar{y}}(\omega)$, reconstructing $\mathbf{\Sigma}_{\bar{x}}(\omega)$ using least squares, and finally recovering $\mathbf{\Sigma}_x(\omega)$ from (39) [46].

CCS IS ESPECIALLY CONVENIENT TO ACQUIRE WIDEBAND SIGNALS, WHOSE RAPID VARIATIONS CANNOT BE EASILY CAPTURED BY CONVENTIONAL ADC.

COOPERATIVE CCS

As mentioned in the section "Preprocessing," in certain cases, multiple sensors are used to observe a time signal in multiple spatial locations, which can result in improved convergence of the sample statistics [47]. Here, we show that this setting can also be used to introduce strong compression.

Suppose that a collection of sensors are deployed across a certain area to estimate the second-order statistics of a certain WSS time signal $x(t)$. Although different sensors observe different signal values, we can assume that the second-order statistics of the received signals are approximately the same for all sensors. This is the case, for example, if the channels from each signal source to all sensors (possibly after passing through an automatic gain control) have approximately the same statistics [21]. As before, let us collect those statistics in the Toeplitz covariance matrix $\mathbf{\Sigma}_x$.

We now describe a particularly interesting case where the sensors use multicoset sampling. To do so, recall from the section "A Warm-Up Example" that, in the single-sensor case, $\mathbf{\Sigma}_x$ can be reconstructed from the covariance matrix of the compressed observations $\mathbf{\Sigma}_y$ if all the entries of $\mathbf{\Sigma}_x$ show up at least once in $\mathbf{\Sigma}_y$. In the cooperative scenario, a milder condition may be imposed by capitalizing on the availability of multiple sensors.

Let us form Z groups of sensors by arranging together all the sensors that share the same multicoset sampling pattern. The sought condition can be given in terms of the matrices $\mathbf{\Sigma}_{y,z}$, $z = 0,\ldots,Z-1$, where $\mathbf{\Sigma}_{y,z}$ represents the covariance matrix of the compressed observations at the sensors within the zth group. The requirement now is that, to reconstruct $\mathbf{\Sigma}_x$, every entry of $\mathbf{\Sigma}_x$ is only required to show up in at least one of the matrices $\{\mathbf{\Sigma}_{y,z}\}_{z=0}^{Z-1}$. This observation yields great compression improvements per sensor, as the sampling burden is now distributed across sensors.

To illustrate this effect, suppose that $\mathbf{\Sigma}_x$ is such that, in the noncooperative scenario, the optimum compression pattern $\mathcal{M}$ for each block is a circular sparse ruler (see Table 1). In the cooperative setting, let $\mathcal{M}_z$ denote the multicoset sampling pattern used by all sensors in group z, and let $\Omega(\mathcal{M}_z)$ represent the set containing all modular differences between elements of $\mathcal{M}_z$:

$$\Omega(\mathcal{M}_z) = \{(m-m') \bmod N : m, m' \in \mathcal{M}_z\}. \tag{40}$$

It can be shown that a collection of sampling patterns $\{\mathcal{M}_z\}_{z=0}^{Z-1}$ ensures the identifiability of $\mathbf{\Sigma}_x$ if and only if [21]

$$\bigcup_{z=0}^{Z-1} \Omega(\mathcal{M}_z) = \{0,1,\ldots,N-1\}. \tag{41}$$

Clearly, for $Z = 1$, this condition reduces to the noncooperative condition, which requires $\mathcal{M}_0$ to be a circular sparse ruler. Each $\mathcal{M}_z$, $z = 0,\ldots,Z-1$, is called an *incomplete circular sparse ruler* since it does not contain all possible differences between 0 and $N-1$ (see "Circular Sparse Rulers"). However, (41) clearly implies that, for every given integer modular distance $n \in \{0,1,\ldots,N-1\}$, at least one of those incomplete circular sparse rulers can measure n. An example of collection of incomplete circular sparse rulers is the one composed of the sets $\mathcal{M}_0 = \{0,1,6\}$, $\mathcal{M}_1 = \{0,2,10\}$, and $\mathcal{M}_2 = \{0,3,7\}$, represented geometrically in Figure 9. Observe that, as in the case of circular sparse rulers, each mark provides two distances, one clockwise and the other counterclockwise.

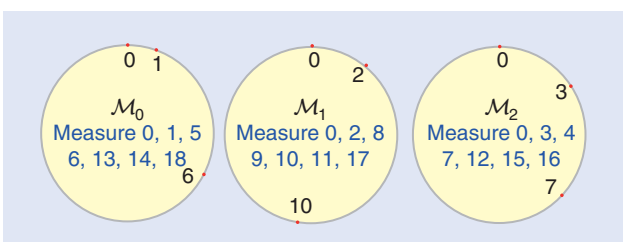
The next question is how to minimize the overall compression ratio. The idea is to minimize the number of marks in each ruler while satisfying (41). This task is intimately connected to the so-called nonoverlapping circular Golomb rulers [21].

Alternative schemes for cooperative CCS include [48], which exploits the cross-correlation between observations at different sensors, and [37], where the observations are not only linearly compressed but also quantized to a single bit.

OPEN QUESTIONS

Despite the long history of structured covariance estimation and recent excitement on compressed sensing of sparse signals, research on CCS is still at an early stage. Extensive work is required to improve its applicability and theoretical understanding. Some possible future directions are listed in this section.

As for sampler design, its most existing schemes rely on identifiability criteria [8], [20], but other criteria are yet to be explored. For instance, it is important to find sampler designs minimizing the Cramér–Rao bound for unbiased estimation of the parameters of interest. Of special relevance are deterministic schemes maximizing



[FIG9] Incomplete circular sparse rulers used in a setting with $Z = 3$ groups of sensors. The correlation lags that the sensors in each group measure are listed inside each circumference.

compression for a target estimation performance. Other sampling schemes, like gridless or continuous irregular sampling, are yet to be investigated from a CCS perspective. The problem becomes reconstructing second-order statistics when the sample locations are not a subset of a regularly-spaced grid. This problem differs from the existing literature on gridless or continuous sparse reconstruction, which aims to accurately recover sparse input signals.

Cooperative schemes also deserve extensive research. For instance, distributed implementations and data fusion techniques for CCS with affordable communication overhead need to be revisited [37]. This includes schemes where sensors quantize their observations before reporting them to the fusion center. In this context, either the correlations or the raw data can be quantized. The latter is possible since under some conditions the correlation function of the original raw data can be computed from the correlation function of the quantized data.

CCS may also be of critical relevance in big data analytics because of its ability to meaningfully reduce the dimension of the data set. In this context, online, adaptive and distributed implementations are yet to be devised. Moreover, as more big-data applications employ a network of high-dimensional signals for data mining and exploration, it is an interesting new direction to see how the CCS framework benefits covariance estimation problems for data-starved inference networks. Such problems arise under the umbrella of probabilistic analysis for high-dimensional data sets with many variables and few samples. As a precursor, sparse (inverse) covariance estimation has already become a popular topic in statistical inference for analysis on graphs, where the sparsity of the (inverse) covariance matrix is exploited, in the context of correlation mining. When high-dimensional or wideband random processes are concerned, CCS has been applied for covariance estimation based on the exploitation of various structures in the data: Gaussianity, stationarity, and compression [49]. Fruitful exploration along this direction may lead to CCS for inference networks, which will find broad applications in analyzing astronomical data, network data, biomedical diagnostics, and video imaging, to name a few.

CYCLOSTATIONARITY PROVIDES AN ALTERNATIVE PERSPECTIVE TO UNDERSTAND COMPRESSION OF SECOND-ORDER STATISTICS, EVEN FOR WSS SEQUENCES.

Finally, we highlight the relevance of extending the reviewed techniques to nonstationary process analysis, for instance, exploiting the framework of underspread processes [50]. Future research may also consider nonlinear parameterizations as well as nonlinear compression.

CONCLUSIONS

This article presented a renewed perspective on a traditional topic in signal processing, which we dubbed *CCS*. We introduced a joint signal acquisition and compression framework for a number of applications and problems that deal with second-order statistics. The basic principle underlying CCS is that the desired signal statistics can be reconstructed directly from properly compressed observations without having to recover the original signal itself, which can be costly in terms of both computational and sensing resources. This standpoint entails multiple benefits, such as the possibility of introducing strong compression without need for sparsity, as required by compressed sensing.

ACKNOWLEDGMENTS

This work was supported by the Spanish Government and the European Regional Development Fund (ERDF) (project TEC2013-47020-C2-1-R COMPASS), the Galician Regional Government and ERDF (projects GRC2013/009, R2014/037, and AtlantTIC), the U.S. Army Research Office under grant W911NF-15-1-0492, 2015-2016, the National Science Foundation under grant EARS 1343248, and NWO-STW under the VICI program (10382).

AUTHORS

Daniel Romero (dromero@umn.edu) received his M.S. and Ph.D. degrees from the University of Vigo, Spain, in 2011 and 2015, respectively. He recently joined the Department of Electrical and Computer Engineering, University of Minnesota. His research interests are in the areas of statistical signal processing, communications, and big-data analytics.

Dyonisius Dony Arianandа (D.A.Dyonisius@tudelft.nl) received his bachelor's degree (2000) from Gadjah Mada University, Yogyakarta, Indonesia, and his M.S. (2009) and Ph.D. (2015) degrees from the Delft University of Technology (TU Delft), The Netherlands, all in electrical engineering. He was also granted the Australian Development Scholarship to complete his M.S. degree in internetworking at the University of Technology, Sydney, New South Wales, Australia, in 2005. He was an academic staff member with the Informatics Engineering Department, Sanata Dharma University, Yogyakarta, Indonesia, from 2000 to 2003 and with the Electrical Engineering Department, Atma Jaya University, Jakarta, Indonesia, from 2006 to 2007. He is now a postdoctoral researcher with TU Delft, working on compressive sampling.

Zhi Tian (ztian1@gmu.edu) is a professor with the Electrical and Computer Engineering Department, George Mason University, Virginia. She was previously on the faculty of Michigan Technological University and worked with the U.S. National Science Foundation. Her research interests lie in statistical signal processing theory and its broad applications. Her current research focuses on compressed sensing for random processes, statistical inference of large-scale data, cognitive radio networks, and distributed wireless sensor networks. She is a Fellow of the IEEE.

Geert Leus (G.J.T.Leus@tudelft.nl) is a full professor at the Delft University of Technology, The Netherlands. He received a 2002 Young Author Best Paper Award and a 2005 Best Paper Award from the IEEE Signal Processing Society (SPS). He was the chair of the IEEE Signal Processing for Communications and Networking Technical Committee and an associate editor of *IEEE Transactions on Signal Processing, IEEE Transactions on*

Wireless Communications, IEEE Signal Processing Letters, and *EURASIP Journal on Advances in Signal Processing (JASP)*. Currently, he is a member-at-large of the Board of Governors of the IEEE SPS and a member of the IEEE Sensor Array and Multichannel Technical Committee. He serves as the editor-in-chief of *EURASIP JASP*. He is a Fellow of the IEEE.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] Y. Jin and B. D. Rao, "Support recovery of sparse signals in the presence of multiple measurement vectors," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 3139–3157, May 2013.

[3] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 2173–2183, Sept. 2000.

[4] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Top. Signal Processing*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[5] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[6] R. T. Hoctor and S. A. Kassam, "The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging," *Proc. IEEE*, vol. 78, no. 4, pp. 735–752, Apr. 1990.

[7] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, Mar. 1968.

[8] S. U. Pillai, Y. Bar-Ness, and F. Haber, "A new approach to array geometry for improved spatial spectrum estimation," *Proc. IEEE*, vol. 73, no. 10, pp. 1522–1524, 1985.

[9] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing.* Hoboken, NJ: Wiley, 2004.

[10] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proc. IEEE*, vol. 70, no. 9, pp. 963–974, Sept. 1982.

[11] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, vol. 1. Reading, MA: Addison-Wesley, 1991.

[12] D. A. Linebarger, I. H. Sudborough, and I. G. Tollis, "Difference bases and sparse sensor arrays," *IEEE Trans. Inform. Theory*, vol. 39, no. 2, pp. 716–721, 1993.

[13] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[14] C. P. Yen, Y. Tsai, and X. Wang, "Wideband spectrum sensing based on sub-Nyquist sampling," *IEEE Trans. Signal Processing*, vol. 61, no. 12, pp. 3028–3040, 2013.

[15] M. A. Lexa, M. E. Davies, J. S. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2011, pp. 3884–3887.

[16] J. D. Krieger, Y. Kochman, and G. W. Wornell, "Design and analysis of multi-coset arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2013, pp. 3781–3785.

[17] D. D. Arianandа, D. Romero, and G. Leus, "Compressive angular and frequency periodogram reconstruction for multiband signals," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, San Martin, France, Dec. 2013, pp. 440–443.

[18] J. Singer, "A theorem in finite projective geometry and some applications to number theory," *Trans. Amer. Math. Soc.*, vol. 43, no. 3, pp. 377–385, 1938.

[19] D. D. Arianandа and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Processing*, vol. 60, no. 9, pp. 4775–4789, 2012.

[20] D. Romero, R. López-Valcarce, and G. Leus, "Compression limits for random vectors with linearly parameterized second-order statistics," *IEEE Trans. Inform. Theory*, vol. 61, no. 3, pp. 1410–1425, Mar. 2015.

[21] D. D. Arianandа, D. Romero, and G. Leus, "Cooperative compressive power spectrum estimation," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, A Coruna, Spain, June 2014, pp. 97–100.

[22] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digit. Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.

[23] V. Venkateswaran and A. J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4131–4143, 2010.

[24] G. Leus and D. D. Arianandа, "Power spectrum blind sampling," *IEEE Signal Processing Lett.*, vol. 18, no. 8, pp. 443–446, 2011.

[25] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 2007, vol. 4, pp. IV. 1357–1360.

[26] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-nyquist sampling for wideband spectrum sensing," *IEEE J. Sel. Top. Signal Processing*, vol. 6, no. 1, pp. 58–69, Feb. 2012.

[27] W. C. Black and D. Hodges, "Time interleaved converter arrays," *IEEE J. Solid-State Circuits*, vol. 15, no. 6, pp. 1022–1029, Dec. 1980.

[28] S. Becker, *Practical Compressed Sensing: Modern Data Acquisition and Signal Processing*, Ph.D. dissertation, California Institute of Technology, 2011.

[29] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Mag.*, vol. 24, no. 3, pp. 79–89, 2007.

[30] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *IEEE Trans. Signal Processing*, vol. 61, no. 24, pp. 6232–6246, 2013.

[31] G. Vázquez-Vilar and R. López-Valcarce, "Spectrum sensing exploiting guard bands and weak channels," *IEEE Trans. Signal Processing*, vol. 59, no. 12, pp. 6045–6057, 2011.

[32] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.

[33] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spencer, "Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays. Part I: Fully augmentable arrays," *IEEE Trans. Signal Processing*, vol. 46, no. 9, pp. 2458–2471, 1998.

[34] S. Shakeri, D. D. Arianandа, and G. Leus, "Direction of arrival estimation using sparse ruler array design," in *Proc. IEEE Int. Workshop on Signal Processing Advances in Wireless Communications*, June 2012, pp. 525–529.

[35] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

[36] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Processing*, vol. 92, no. 7, pp. 1580–1590, 2012.

[37] O. Mehanna and N. Sidiropoulos, "Frugal sensing: Wideband power spectrum sensing from few bits," *IEEE Trans. Signal Processing*, vol. 61, no. 10, pp. 2693–2703, May 2013.

[38] Y. Wang, A. Pandharipande, and G. Leus, "Compressive sampling based MVDR spectrum sensing," in *Proc. Cognitive Information Processing*, June 2010, pp. 333–337.

[39] W.-K. Ma, T.-H. Hsieh, and C.-Y. Chi, "DOA estimation of quasi-stationary signals via Khatri–Rao subspace," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2009, pp. 2165–2168.

[40] D. D. Arianandа and G. Leus, "Direction of arrival estimation for more correlated sources than active sensors," *Signal Processing*, vol. 93, no. 12, pp. 3435–3448, 2013.

[41] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[42] P. Pal and P.P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Trans. Signal Processing*, vol. 63, no. 3, pp. 711–726, Feb. 2015.

[43] G. Leus and Z. Tian, "Recovering second-order statistics from compressive measurements," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Dec. 2011, pp. 337–340.

[44] D. D. Arianandа and G. Leus, "Non-uniform sampling for compressive cyclic spectrum reconstruction," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2014, pp. 41–45.

[45] E. BouDaher, F. Ahmad, and M. G. Amin, "Sparse reconstruction for direction-of-arrival estimation using multi-frequency co-prime arrays," *EURASIP J. Adv. Signal Processing*, vol. 2014, no. 1, p. 168, 2014.

[46] D. D. Arianandа, D. Romero, and G. Leus, "Compressive periodogram reconstruction using uniform binning," *IEEE Trans. Signal Processing*, vol. 63, no. 16, pp. 4149–4164, Aug. 2015.

[47] R. J. Kozick and B. M. Sadler, "Source localization with distributed sensor arrays and partial spatial coherence," *IEEE Trans. Signal Processing*, vol. 52, no. 3, pp. 601–616, Mar. 2004.

[48] D. D. Arianandа and G. Leus, "Cooperative compressive wideband power spectrum sensing," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Nov. 2012, pp. 303–307.

[49] A. O. Nasif, Z. Tian, and Q. Ling, "High-dimensional sparse covariance estimation for random signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing,* IEEE, 2013, pp. 4658–4662.

[50] G. Matz and F. Hlawatsch, "Nonstationary spectral analysis based on time-frequency operator symbols and underspread approximations," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1067–1086, 2006.

[SP]

Game Theory for Networks

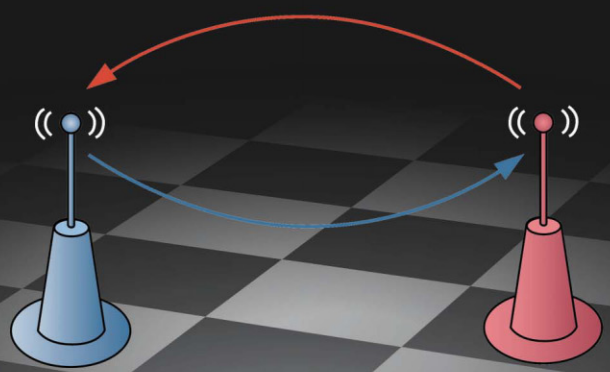


IMAGE LICENSED BY INGRAM PUBLISHING

[A tutorial on game-theoretic tools for emerging signal processing applications]

[Giacomo Bacci, Samson Lasaulce, Walid Saad, and Luca Sanguinetti]

The aim of this tutorial is to provide an overview, although necessarily incomplete, of game theory (GT) for signal processing (SP) in networks. One of the main features of this contribution is to gather in a single article some fundamental game-theoretic notions and tools that, over the past few years, have become widespread in the SP literature. In particular, both strategic-form and coalition-form games are described in detail, and the key connections and differences between them are outlined. Moreover, particular attention is also devoted to clarifying the connections between strategic-form games and distributed optimization and learning algorithms. Beyond an introduction to the basic concepts and main solution approaches, several carefully designed examples are provided to allow a better understanding of how to apply the described tools.

Digital Object Identifier 10.1109/MSP.2015.2451994
Date of publication: 29 December 2015

INTRODUCTION

GT is a branch of mathematics that enables the modeling and analysis of the interactions between several decision makers (called *players*) who can have conflicting or common objectives. A game is a situation in which the benefit or cost achieved by each player from an interactive situation depends not only on its own decisions but also on those taken by the other players. For example, the time a car driver needs to get home generally depends not only on the route he or she chooses but also on the decisions made by the other drivers. Therefore, in a game, the actions and objectives of the players are tightly coupled. Until very recently, GT

1053-5888/16©2016IEEE

has been used only marginally in SP, with notable examples being some applications in robust detection and estimation [1] as well as watermarking [2] (in which the watermarking problem is seen as a game between the data embedder and the attacker). However, the real catalyst of the application of GT to SP has been the blooming of all issues related to networking in general, and distributed networks, in particular. The interactions that take place in a network can often be modeled as a game, in which the network nodes are the players that compete or form coalitions to get some advantage and enhance their quality of service. The main motivation behind formulating a game in a network is the large interdependence between the actions of the network nodes due to factors such as the use of common resources (e.g., computational, storage, or spectral resources), with interference across wireless networks being an illustrative case study. Paradigmatic examples of this approach can be found in the broad field of SP for communication networks in which GT is used to address fundamental networking issues, such as controlling the power of radiated signals in wireless networks, with the line of research largely originated from the seminal work in [3]; beamforming for smart antennas [4]; precoding in multiantenna radio transmission systems [5]; data security [6]; spectrum sensing in cognitive radio (CR) [7]; spectrum and interference management [8]; multimedia resource management [9]; and image segmentation [10], [11].

Spurred and motivated by the well-established application to the listed fields, GT has also proliferated many other branches of SP and has very recently been used for modeling and analyzing the following "classical" SP problems: distributed estimation in sensor networks [12]; adaptive filtering [13]; waveform design for multiple-input, multiple-output (MIMO) radar estimation [14]; jamming of wireless communications [15] and MIMO radar applications [16]; and finding the position of network nodes [17]. In addition to the aforementioned examples, we must eventually point out the important connection that is building up between GT and SP through the fields of machine-learning algorithms [18] and distributed optimization [19]. As explained in the section "Learning Equilibria in Strategic-Form Games," there is a close relationship between game-theoretic concepts and learning-algorithm aspects. In this respect, one of the key messages of this contribution is that the solution of a game (often called an *equilibrium*, as discussed later) can often coincide with the convergence point that results from the interaction among several automata that implement iterative or learning algorithms. Therefore, there is an important synergy between GT and the broad field of multiagent learning.

Despite the clear intersection between GT, learning, and optimization, as corroborated by a significant number of SP papers that exploit GT, it is worth noting that games usually have some features that are not common in classical optimization problems. In this respect, GT possesses its own tools, approaches, and notions. For example, in contrast to a classical optimization problem, in which a certain function must be optimized under some constraints, the meaning of *optimal decision*, or, equivalently, *strategy*, is generally unclear in interactive situations involving several decision makers, since none of them controls all the variables of the problem and these players can also have different objectives. To address such situations, GT is enriched with concepts from different disciplines, such as economics and biology. This leads to notions that one does not encounter when studying, for instance, convex optimization. Examples of these notions are auctions, cooperative plans, punishments, rationality, risk aversion, trembling hand, and unbeatable strategies, to name a few. Remarkably, such concepts can actually be exploited to design algorithms. Although a player can be an automaton, a machine, a program, a person, an animal, a living cell, a molecule, or, more generally, any decision-making entity, it is essential to have in mind that a game is, first and foremost, a mathematical tool that aims at modeling and analyzing an interactive situation. Before delving into the specific details of the various game models, we first provide a detailed overview on the different game models available in the GT literature.

There are three dominant mathematical representations for a game: 1) the strategic form, 2) the extensive form, and 3) the coalition form. Other representations exist, e.g., the standard form, which is used in the theory of equilibrium selection [20], and the state-space representation [21], but their use is rather marginal. The extensive form, which is typically used to investigate dynamical situations in computer science, will not be discussed in this survey. The main reason is that the extensive form, although more general (see [22] and [23] and references therein for more details) than the strategic form, is often mathematically less tractable for typical SP problems. Defining the corresponding model and providing important results related to the strategic form is the purpose of the "Strategic-Form Games" section, whereas the "Learning Equilibria in Strategic-Form Games" section shows how some solution concepts that are inherent to the strategic form can be related to algorithmic aspects. The "Coalition-Form Games" section discusses the coalition form, which, unlike the strategic form, deals with options available to subsets of players (called *cooperative groups* or *coalitions*), what cooperative coalitions can form, and how the coalition utility is divided among its members. The algorithms that can be used to implement this approach are detailed in the section "Algorithms for Coalition-Form Games." Note that, as described throughout the article, for a given SP problem, the structure of the problem at hand and the practical constraints associated with it will determine whether the strategic or the coalition form is the most suitable representation. For example, it may occur that both forms are acceptable in terms of information assumptions, while complexity issues will lead to selecting one over the other.

To sum up, the main objectives of this tutorial are as follows. The primary goal of this survey is to provide a holistic reference on the use of GT in SP application domains. Some surveys have already been published in the SP literature [24] and communications literature [25], [26]. Our motivation is not only to provide a refined and updated view of GT with respect to these existing tutorials but also to establish explicit connections across the different tools of GT. This tutorial is intended for researchers and graduate students (with some expertise in networks and SP) interested in obtaining a comprehensive overview of game-theoretic concepts and distributed algorithm design, and it aims to

- give the reader a global—although necessarily partial—overview of GT highlighting connections and differences between strategic- and coalition-form games in a single article
- delineate differences and connections between GT and optimization
- explain the strong relationship between game-theoretic solution concepts, such as the Nash equilibrium (NE), and distributed SP algorithms
- provide many application examples to help the reader understand the way the described tools can be applied to different contexts.

For absolute beginners in GT, we refer readers to a recent lecture note [27], whereas we invite those interested in a thorough and textbook-oriented discussion on GT applied to wireless communications and SP to refer to the textbooks [22] and [23]. For the reader's convenience, Table 1 lists the acronyms for the game-theoretic terms used throughout the tutorial, and Figure 1 provides a reference for the structure of this tutorial, adopting the typical methodology used to address the game-theoretic problems and listing the topics described in each section.

STRATEGIC-FORM GAMES

DEFINITION

A game in strategic (equivalently, normal) form is represented by a family of multivariate functions $u_1, \ldots, u_K$; $K \geq 1$. The index set of this family, which is denoted here by $\mathcal{K} = \{1, \ldots, K\}$, is called the *set of players* and, for each $k \in \mathcal{K}$, u_k is commonly called the *utility* (equivalently, *payoff*) *function* of player k. The strategic form assumes that u_k can be any function of the following form:

$$\begin{aligned} u_k: \ & \mathcal{S}_1 \times \ldots \times \mathcal{S}_K \longrightarrow \mathbb{R} \\ & (s_1, \ldots, s_K) \quad \longmapsto u_k(s), \end{aligned} \tag{1}$$

where $\mathcal{S}_k$ is called the *set of strategies* of player k, s_k is the strategy of player k, $s = (s_1, \ldots, s_K) \in \mathcal{S}$ is the strategy profile, and $\mathcal{S} = \mathcal{S}_1 \times \ldots \times \mathcal{S}_K$. We refer to a strategic-form game by using the compact triplet notation $\mathcal{G} = (\mathcal{K}, (\mathcal{S}_k)_{k \in \mathcal{K}}, (u_k)_{k \in \mathcal{K}})$. The notation $s_{-k} = (s_1, \ldots, s_{k-1}, s_{k+1}, \ldots, s_K)$ is used to denote the strategies taken by all other players, except player k. With a slight abuse of notation, the whole strategy profile is denoted by $s = (s_k, s_{-k})$. The strategic-form representation may encompass a large number of situations in SP. To mention a few, players in a game can be radars competing to improve their performance in terms of the probability of false alarms or missed detections; sensors in a sensor network, which coordinate to estimate a field in a distributed way; base stations allocating the resources in a cellular network to optimize the system throughput; several digital signal processors, which have to compete for or manage computing resources; or a watermarking device or algorithm, which has to find a good strategy against potential attackers.

Formally, it is worth noting that, in its general formulation, the strategic form is characterized by the simultaneous presence of two key features:

- Each player k can have its own objective, which is captured by a per-player specific function $u_k(s)$.
- Each player k has partial control over the optimization variables as it can control its strategy $s_k \in \mathcal{S}_k$ only.

Although the first feature is tied with multiobjective optimization, a clear difference exists in the control of the optimization variables because, in multiobjective optimization, one has full control over all the variables. Additionally, quite often in multiobjective optimization problems (see, e.g., [28]), an aggregate objective must be defined. The second feature is closely related to the framework of distributed optimization, although a common objective function is usually considered in this context, i.e., $\forall k$ $u_k(s) = u(s)$. More importantly, the conventional assumption in distributed optimization is that the decision-making process is basically driven by a single designer (controller), which provides a set of strategies that the players strictly follow. Despite being a possible scenario (which might be very relevant for some algorithmic aspects), in GT, the players, in general, have the freedom to choose their strategies by themselves.

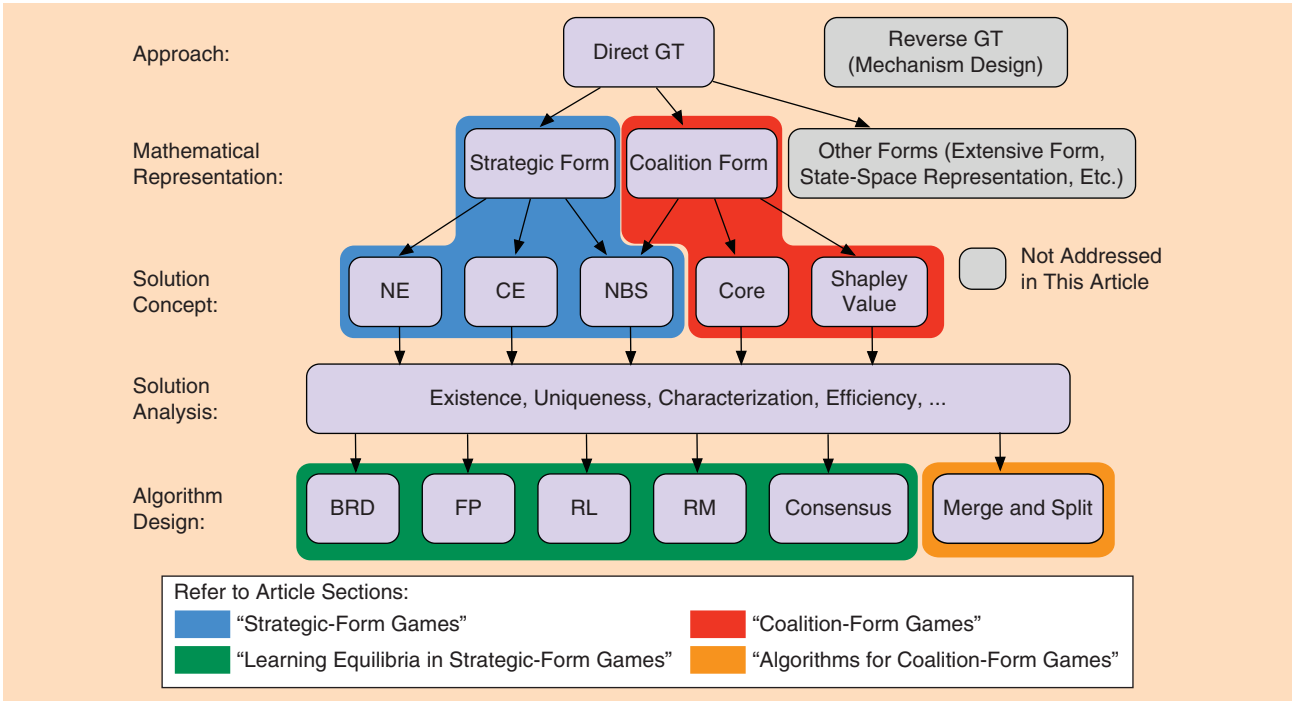
A central question is how to "solve" a strategic-form game. The very notion of optimality in this context is unclear since, as previously explained, we are in the presence of multiple objectives, and the variables, which impact the utility functions, cannot be controlled jointly. This is the reason why the problem needs to be defined before being solved and why there exists the need for introducing game-theoretic solution concepts.

SOLUTION CONCEPTS

The NE is a fundamental solution concept for a strategic-form game, on which many other concepts are built. This section is mostly dedicated to the NE and discusses more briefly other solution concepts that might also be considered. In [29], Nash proposed a simple but powerful solution concept, which is now known as an *NE* (equivalently, *Nash point*).

[TABLE 1] A LIST OF GT ACRONYMS USED THROUGHOUT THIS TUTORIAL.

BR	BEST RESPONSE	OCF	OVERLAPPING COALITION FORMATION
BRD	BEST-RESPONSE DYNAMICS	PF	PARTITION FUNCTION
CCE	COARSE CORRELATED EQUILIBRIUM	PO	PARETO OPTIMALITY
CE	CORRELATED EQUILIBRIUM	POA	PRICE OF ANARCHY
CF	CHARACTERISTIC FUNCTION	RL	REINFORCEMENT LEARNING
FP	FICTITIOUS PLAY	RM	REGRET MATCHING
NBS	NASH BARGAINING SOLUTION	SE	STRONG EQUILIBRIUM
NE	NASH EQUILIBRIUM	SO	SOCIAL OPTIMALITY
NTU	NONTRANSFERABLE UTILITY	TU	TRANSFERABLE UTILITY



[FIG1] **The logical structure of this tutorial.**

DEFINITION 1 (NE)

An NE of the game $\mathcal{G} = (\mathcal{K}, (\mathcal{S}_k)_{k\in\mathcal{K}}, (u_k)_{k\in\mathcal{K}})$ is a strategy profile $s^{\mathrm{NE}} = \left(s_1^{\mathrm{NE}}, \ldots, s_K^{\mathrm{NE}}\right) = \left(s_k^{\mathrm{NE}}, s_{-k}^{\mathrm{NE}}\right)$ such that

$$\forall k \in \mathcal{K}, \forall s_k \in \mathcal{S}_k, u_k\left(s_k^{\mathrm{NE}}, s_{-k}^{\mathrm{NE}}\right) \geq u_k\left(s_k, s_{-k}^{\mathrm{NE}}\right). \tag{2}$$

■

A simple instance of an NE in everyday life would be to say that if everyone drives on the right, no single driver has an incentive to drive on the left. As a more technical comment on the above definition, it can be seen that s^{NE} represents a strategy profile in the broad sense. For instance, it may be a vector of actions, a vector of probability distributions, or a vector of functions. Probability distributions naturally appear when considering an important extended version of the strategies of $\mathcal{G}$, i.e., mixed strategies. When $\mathcal{S}_k$ is finite (the continuous case is obtained by using an integral instead of a discrete sum in the definition), they are defined next.

DEFINITION 2 (MIXED STRATEGIES)

Let $\Delta(\mathcal{X})$ be the set of distribution probabilities over the generic set $\mathcal{X}$ (that is the unit simplex). Player k's mixed strategy $\pi_k \in \Delta(\mathcal{S}_k)$ is a distribution that assigns a probability $\pi_k(s_k)$ to each strategy s_k, such that $\sum_{s_k \in \mathcal{S}_k} \pi_k(s_k) = 1$. For mixed strategies, the (joint) probability distribution over the strategy profile s is, by definition, the product of the marginals π_k, $k \in \mathcal{K}$. ■

A mixed strategy consists of choosing a lottery over the available actions. In the case where a player has two possible choices, choosing a mixed strategy amounts to choosing a coin with a given probability of having heads (or tails): the player flips the coin to determine the action to be played. Using mixed strategies, each player can play a certain strategy s_k with probability $\pi_k(s_k)$. Note that the strategies considered so far, termed *pure strategies*, are simply a particular case of mixed strategies, in which probability one is assigned to one strategy, and zero to the others. The importance of mixed strategies, aside from being more general mathematically than pure strategies, comes, in part, from the availability of existence results for mixed NE. The latter is defined next.

DEFINITION 3 (MIXED NE)

A mixed strategy NE of the game $\mathcal{G} = (\mathcal{K}, (\mathcal{S}_k)_{k\in\mathcal{K}}, (u_k)_{k\in\mathcal{K}})$ is a mixed strategy profile $\pi^{\mathrm{NE}} = \left(\pi_1^{\mathrm{NE}}, \ldots, \pi_K^{\mathrm{NE}}\right) = \left(\pi_k^{\mathrm{NE}}, \pi_{-k}^{\mathrm{NE}}\right)$ such that

$$\forall k \in \mathcal{K}, \forall \pi_k \in \Delta(\mathcal{S}_k), \tilde{u}_k\left(\pi_k^{\mathrm{NE}}, \pi_{-k}^{\mathrm{NE}}\right) \geq \tilde{u}_k\left(\pi_k, \pi_{-k}^{\mathrm{NE}}\right), \tag{3}$$

where

$$\tilde{u}_k(\pi_k, \pi_{-k}) = \mathbb{E}(u_k) = \sum_{s\in\mathcal{S}} \left(\prod_{j\in\mathcal{K}} \pi_j(s_j)\right) u_k(s) \tag{4}$$

is the expected utility of player k when selecting the mixed strategy π_k, and $\mathcal{S} = \mathcal{S}_1 \times \ldots \times \mathcal{S}_K$. ■

By definition, an NE of $\mathcal{G}$ is a point such that, for every index k, the function u_k cannot be (strictly) increased by just changing the value of the variable s_k at the equilibrium. For this reason, an NE is said to be strategically stable to unilateral deviations. The NE has at least two other very attractive features.

- In its mixed version, its existence is guaranteed for a broad class of games.
- It may result from the repeated interaction among players, which are only partially informed about the problem.

In particular, some well-known distributed and/or learning algorithms may converge to an NE (see the "Learning Equilibria in Strategic-Form Games" section).

Elaborating more on the first feature, it should be stressed that existence is a fundamental issue in GT. In fact, one might think of various solution concepts for a game. For example, one might consider a point which is stable to K deviations rather than to a single one (with K being the number of players). This solution concept is known as a *strong equilibrium* (*SE*) (e.g., see [22] and [23]): an SE is a strategy profile from which no group of players (of any size) can deviate and improve the utility of every member of the group while the players outside the deviating group maintain their strategy to that of the equilibrium point. The SE is therefore stable to multiple deviations, and the number of deviations can be up to K. This is a strong requirement, which explains why it is quite rarely satisfied in a static game (see [22] for a static-game example where it is met). In fact, the SE is particularly relevant in infinitely repeated games. To better understand this, refer to the "Coalition-Form Games" section, where the notion of core is described; indeed, it turns out that a specific version of the core, the β-core, of a game coincides with the SE utilities in an infinite repetition of that game [30]. Considering the SE as a solution concept (in a context of purely selfish players of a static game) might be inappropriate since it will typically not exist; instead, the NE offers more positive results in terms of existence. Indeed, tackling the existence issue of an NE for a strategic-form game $\mathcal{G}$ needs further study on a fixed-point problem for which quite positive results can be obtained. To this end, the notion of best response (BR) for a player must be first introduced.

DEFINITION 4 (BR)

Player k's BR $\mathrm{BR}_k(s_{-k})$ to the vector of strategies s_{-k} is the set-valued function

$$\mathrm{BR}_k(s_{-k}) = \arg\max_{s_k \in \mathcal{S}_k} u_k(s_k, s_{-k}). \tag{5}$$

■

By introducing the auxiliary notion of composite (or, equivalently, global game's) BR

$$\begin{aligned} \mathrm{BR}:\ \mathcal{S} &\to \mathcal{S} \\ s &\mapsto \mathrm{BR}_1(s_{-1}) \times \ldots \times \mathrm{BR}_K(s_{-K}), \end{aligned} \tag{6}$$

we have the following characterization for an NE.

DEFINITION 5 (NE CHARACTERIZATION)

Let $\mathcal{G} = (\mathcal{K}, (\mathcal{S}_k)_{k\in\mathcal{K}}, (u_k)_{k\in\mathcal{K}})$ be a strategic-form game. A strategy profile s^{NE} is an NE if and only if

$$s^{\mathrm{NE}} \in \mathrm{BR}(s^{\mathrm{NE}}). \tag{7}$$

■

The characterization of an NE in terms of a fixed-point problem is due to Nash [29] and explains why common existence theorems are based on topological and geometrical assumptions such as compactness for the sets of strategies or continuity for the utility functions. The following two theorems explain why the NE is an attractive solution concept from the existence issue standpoint: they show that any finite game or compact continuous game possesses at least one mixed NE.

THEOREM 1 ([31])

In a strategic-form game $\mathcal{G} = (\mathcal{K}, (\mathcal{S}_k)_{k\in\mathcal{K}}, (u_k)_{k\in\mathcal{K}})$, if $\mathcal{K}$ is finite and $\mathcal{S}_k$ is finite for every k, then there exists at least one NE, possibly involving mixed strategies. ■

THEOREM 2 ([31])

In a strategic-form game $\mathcal{G} = (\mathcal{K}, (\mathcal{S}_k)_{k\in\mathcal{K}}, (u_k)_{k\in\mathcal{K}})$, if $\mathcal{S}_k$ is compact and u_k is continuous in $s \in \mathcal{S}$ for every $k \in \mathcal{K}$, then there exists at least one NE, possibly involving mixed strategies. ■

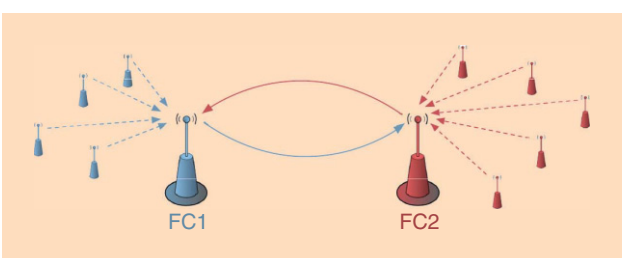
To better illustrate the meaning of the strategic-form representation and the notion of NE, let us consider a simple example, which is an instance of what is referred to as the *prisoner's dilemma* in the GT literature [32].

EXAMPLE 1 (THE WIRELESS SENSOR'S DILEMMA)

Consider the wireless sensor network sketched in Figure 2, which is populated by a number of wireless sensors sending their own measurements (e.g., target detection and temperature), to their fusion centers (FCs), labeled as FC1 and FC2. For the sake of graphical representation, sensors communicating with sensors and the FCs are represented with blue and red colors, respectively. Gathering information at each FC from a larger population of nodes (in this case, those covered by the other FC) helps improve its measurement accuracy. However, sharing data among different populations of nodes implies additional transmission of information across the FCs, which is, in general, costly due to energy expenditure. In this context, the two FCs can independently and simultaneously decide whether to share (i.e., relay) the information. Depending on both decisions, each FC gets a (dimensionless) utility in the form "accuracy minus spent energy," given according to Figure 3 (known as a *payoff matrix*), in which $0 \le e \le 1$ represents the cost incurred by an FC for relaying the measurements of the other. ■

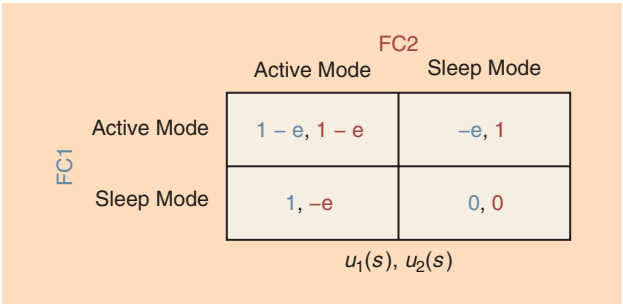
The communication problem corresponding to Example 1 can be modeled as a strategic-form game where the set of players is $\mathcal{K} = \{\mathrm{FC1}, \mathrm{FC2}\}$ and the action (strategy) sets are $\mathcal{S}_k = \{\text{sleep mode}, \text{active mode}\}$ for $k \in \{1, 2\}$. The utility function for FC1 (the one for FC2 follows by symmetry) is given by

$$u_1(s_1, s_2) = \begin{cases} -e & \text{if } (s_1, s_2) = (\text{active mode, sleep mode}) \\ 0 & \text{if } (s_1, s_2) = (\text{sleep mode, sleep mode}) \\ 1-e & \text{if } (s_1, s_2) = (\text{active mode, active mode}) \\ 1 & \text{if } (s_1, s_2) = (\text{sleep mode, active mode}). \end{cases} \tag{8}$$

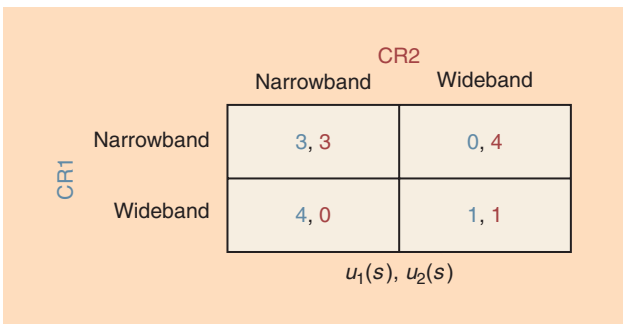


[FIG2] The wireless sensor's dilemma.

This game is said to be a *static* (equivalently, *one-shot*) *game* since each player takes a single action once and for all. Since this game is finite, it has at least one mixed NE (according to Theorem 1). To find these equilibria, let us denote by ρ_1 (respectively, ρ_2) the probability that FC1 (respectively, FC2) assigns to the action active mode. The mixed NE of the considered game can be found by computing the expected utilities. For player FCk with $k \in \{1,2\}$, it can be written as $\tilde{u}_k(\rho_1, \rho_2) = -e\rho_k + \rho_{-k}$. The BR of player k is given by: $\forall \rho_{-k} \in [0,1]$, $\overline{\text{BR}}_k(\rho_{-k}) = 0$. Since, by definition, Nash equilibria are intersection points of the BRs, the unique mixed NE is $(\rho_1^{\text{NE}}, \rho_2^{\text{NE}}) = (0,0)$, which is a pure NE consisting of the action profile (sleep mode, sleep mode).



[FIG3] **A wireless sensor's dilemma game under matrix form.**



[FIG4] **A CR's dilemma game under matrix form (utilities may be expressed in megabits/second).**

EXAMPLE 2 (THE COGNITIVE RADIO'S DILEMMA)

Observe that Example 1 is general enough to encompass many different applications., e.g., it can be used to model a cognitive network with two CRs, CR1 and CR2, which have to decide independently and simultaneously to transmit either over a narrow or a wide frequency band. In this case, the two corresponding actions are respectively denoted by *narrowband* and *wideband*. Depending on the CR's decisions, each CR transmits at a certain data rate (say, in megabits per second) accordingly to Figure 4. The first (second) component of each pair corresponds to the transmission rate (i.e., utility) of CR1 (CR2). For instance, if both use a wide band, their transmission rate is the same and equals 1 megabit/second. ■

The action (sleep mode) in Example 1 (or, in Example 2, wideband) is called a *strictly dominant action* for player k since, for any given action chosen by the other player, it provides a strictly higher utility than any other choice. At the equilibrium (sleep mode, sleep mode), the wireless sensors have a zero utility. We can see that there exists an action profile at which both players would gain a higher utility. The action profile (active mode, active mode) is said to *Pareto-dominate* the action profile (sleep mode, sleep mode). More generally, in any game, when there exists a strategy profile, which provides a utility for every player that is greater than the equilibrium utility, the equilibrium is said to be *Pareto-inefficient*. Inefficiency is generally a drawback of considering the NE as a solution concept. From an engineering point of view, it would be more desirable to find an equilibrium that is Pareto-efficient, i.e., a Pareto-optimal (PO) point.

DEFINITION 6 (PARETO-OPTIMAL PROFILE)

A strategy profile s^{PO} is a PO point if there exists no other strategy profile s, such that $u_k(s) \geq u_k(s^{\text{PO}})$ for all $k \in \mathcal{K}$, and $u_k(s) > u_k(s^{\text{PO}})$ for some $k \in \mathcal{K}$. ■

In addition to Pareto optimality, another widely used related concept is the weak PO point, defined next.

DEFINITION 7 (WEAK PARETO-OPTIMAL PROFILE)

A strategy profile s^{PO} is a weak PO point if there exists no other strategy profile s, such that $u_k(s) > u_k(s^{\text{PO}})$ for all $k \in \mathcal{K}$. ■

In other words, when operating at a PO strategy profile, it is not possible to increase the utility of one player without decreasing that of at least one other. In many occasions, beyond the concept of Pareto optimality, the performance (in terms of social efficiency) of an NE can be measured by comparing it to a socially optimal profile, which is defined as a maximizer of the social welfare (or, more properly, sum-utility) $\sum_{k \in \mathcal{K}} u_k(s)$. Observe that other global measures can be used to introduce some fairness (e.g., see [33]). For example, through Definition 12, the Nash product (defined later on in this section) is considered and can be shown to be proportionally fair (see [34]). Formally stated, a social-optimal (SO) point is defined next.

DEFINITION 8 (SOCIAL OPTIMUM)

A strategy profile s^{SO} is an SO point if it satisfies

$$s^{\text{SO}} \in \arg \max_{s \in \mathcal{S}} \sum_{k \in \mathcal{K}} u_k(s). \tag{9}$$

■

Both PO and SO points can be seen as possible solution concepts for a game. Often, implementing these solution concepts will require some coordination between the players and typically relies on the need for significant information and knowledge assumptions. In the framework of distributed networks, such coordination degree and/or knowledge might not be available or may be costly, and, thus, social and Pareto optimality can only be used to measure the performance loss induced by decentralization. There is a common and simple measure of efficiency, which allows us to quantify the gap between the performance of centralized (in some sense, classical) optimization and distributed optimization. Indeed, the efficiency of the Nash equilibria can be measured using the concept of *price of anarchy* (*PoA*) [35], which is defined next.

DEFINITION 9 (PRICE OF ANARCHY)

The PoA corresponds to

$$\text{PoA} = \frac{\max_{s\in\mathcal{S}} \sum_{k\in\mathcal{K}} u_k(s)}{\min_{s\in\mathcal{S}^{\text{NE}}} \sum_{k\in\mathcal{K}} u_k(s)}, \tag{10}$$

where $\mathcal{S}^{\text{NE}}$ denotes the set of all NE in a game. ■

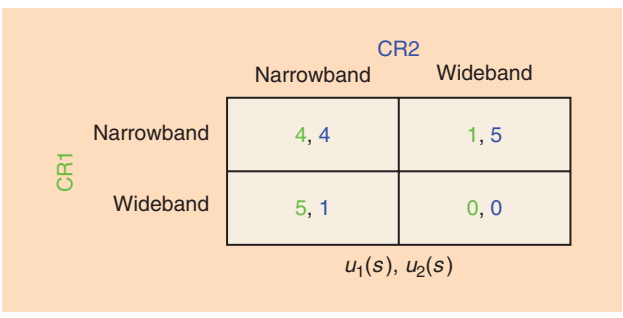
Otherwise stated, the PoA provides a measure of the performance loss (in terms of social welfare) of the "worst" NE compared to a socially optimal strategy. The closer the PoA is to 1, the higher the efficiency of the NE. One of the features of the PoA is that it can be upper-bounded in some important cases, e.g., in congestion games with monomial costs [36]; a congestion game is a special form of game in which the utility of a player depends on its own action and depends on others' action only through the way they distribute over the available actions (often called *edges* or *routes*). For instance, if the cost (the opposite of the utility) is linear, the PoA is upper-bounded by (4/3), showing that the price of decentralization is relatively small in this scenario.

To illustrate the notions of PoA, let us reconsider Example 2, where the four possible utility profiles are reported in Figure 4. The game has three Pareto optima: (4, 0), (0, 4) and (3, 3). Geometrically, a utility vector is PO if there is no point in the northeast orthant whose origin is located at the candidate point. In the considered game, there is a unique NE. Here, the PoA equals $(3+3)/(1+1) = 3$. If there is no means of coordinating the two CRs, which may happen when both transmitters have been designed independently or are owned by different economic players, the loss in terms of social efficiency has to be undergone. However, if there is a common designer as in the framework of distributed optimization, it may be possible to decrease the PoA.

Remark 1

One way to improve efficiency is to keep on considering an NE as the solution concept but to transform the game. The corresponding general framework is referred to as *mechanism design* [37]. Affine pricing is a very special instance of mechanism design: it consists of applying an affine transformation on the utility functions and tuning the introduced parameters to obtain an NE, which is more efficient than the one considered in the original game [3]. ■

Another possibility to improve efficiency is to keep the game unchanged but to modify the solution concept. This may be either a correlated equilibrium (CE) or a Nash bargaining solution (NBS). A CE is a joint distribution over the possible actions or pure strategy profiles of the game from which no player has interest in deviating unilaterally. More formally, we have the following definition.



[FIG5] A simple CR's coordination game, which exhibits nontrivial CE (utilities may be expressed in megabits/second).

DEFINITION 10 (CORRELATED EQUILIBRIUM)

A CE is a joint probability distribution $q^{\text{CE}} \in \Delta(\mathcal{S})$, which verifies

$$\forall k \in \mathcal{K}, \forall \sigma_k, \sum_{s\in\mathcal{S}} q^{\text{CE}}(s_k, s_{-k}) u_k(s_k, s_{-k}) \geq \sum_{s\in\mathcal{S}} q^{\text{CE}}(s_k, s_{-k}) u_k(\sigma_k(s_k), s_{-k}), \tag{11}$$

where σ_k: $\mathcal{S}_k \to \mathcal{S}_k$ can be any mapping, and $\mathcal{S}_{-k} = \mathcal{S}_1 \times \ldots \times \mathcal{S}_{k-1} \times \mathcal{S}_{k+1} \times \ldots \times \mathcal{S}_K$. ■

We know that a pure NE is a special case of mixed NE for which the individual probability distributions used by the players are on the vertices of the unit simplex. We see now that a mixed NE is a special case of a CE for which joint probability distributions over the action profiles factorizes as the product of its marginals. It is important to know how to obtain a CE in practice. Aumann showed that the availability of an "exogenous public signal" to players allows the game to reach new equilibria, which are in the convex hull of the set of mixed NE of the game [22]. The term *public signal* implies that every player can observe it; the adjective exogenous is added to explicitly indicate that the signal is not related to the player's actions. A simple example would be the realization of a Bernoulli random variable such as the outcome obtained by flipping a coin. Additionally, if exogenous private signals are allowed, new equilibria outside this hull can be reached and lead to better outcomes; by private, it is meant that each player observes the realizations of its own lottery. The obtained equilibria are precisely CE. Having a CE therefore means that the players have no interest in ignoring (public or private) signals, which would recommend them to play according to the realizations of a random lottery whose joint distribution corresponds to a CE q^{CE}. In the case of the wireless sensor's dilemma, it can be checked that the only CE boils down to the unique pure NE of the game, showing that sending a broadcast signal to the wireless sensors would not allow them to reach another equilibrium, which might be more efficient. To better illustrate the meaning of CE, consider the modified version of Example 2 shown in Figure 5 in matrix form. Observe that it no longer has the structure of a prisoner's dilemma (no strictly dominant strategy for the players exists). Figure 6 shows the set of CE of this game. In particular, it turns out that a public signal allows the CR to reach any CE in the convex hull of the points (5, 1), (1, 5) (pure NE) and [(5/2), (5/2)] (strict mixed NE). Private signals allow one to extend this region. The set of CE becomes the convex hull of the points (1, 5), (5, 1), [(5/2), (5/2)], and [(10/3), (10/3)].

Another notion of equilibrium derived from the notion of CE is the coarse correlated equilibrium (CCE). It is mathematically more general than the CE, and, hence, the set of CE is included in the set of CCE. One of the motivations for mentioning it here is that CCE can be learned by implementing simple algorithms such as

regret-matching (RM)-based learning schemes [18] (see the "Learning Equilibria in Strategic-Form Games" section for further details).

DEFINITION 11 (COARSE CORRELATED EQUILIBRIUM)

A CCE is a joint probability distribution $q^{\mathrm{CCE}} \in \Delta(\mathcal{S})$ which verifies

$$\forall k \in \mathcal{K}, \forall s_k \in \mathcal{S}_k, \sum_{s \in \mathcal{S}} q^{\mathrm{CCE}}(s) u_k(s) \geq \sum_{s_{-k} \in \mathcal{S}_{-k}} q^{\mathrm{CCE}}_{-k}(s_{-k}) u_k(s'_k, s_{-k}), \tag{12}$$

where $q^{\mathrm{CCE}}_{-k}(s_{-k}) = \sum_{s'_k \in \mathcal{S}_k} q^{\mathrm{CCE}}(s'_k, s_{-k})$. ■

A possible interpretation of this definition is as follows. Following the notion of CCE, players are assumed to decide, before receiving the recommendation associated with a public or private signal, whether to commit to it. At a CCE, all players are willing to follow the recommendation, given that all the others also choose to commit. That is, if a single player decides not to follow the recommendations, it experiences a lower (expected) utility. Based on this interpretation, the difference between the CCE and the CE is that in the latter, the players choose whether or not to follow a given recommendation after it has been received. Therefore, there is no a priori commitment.

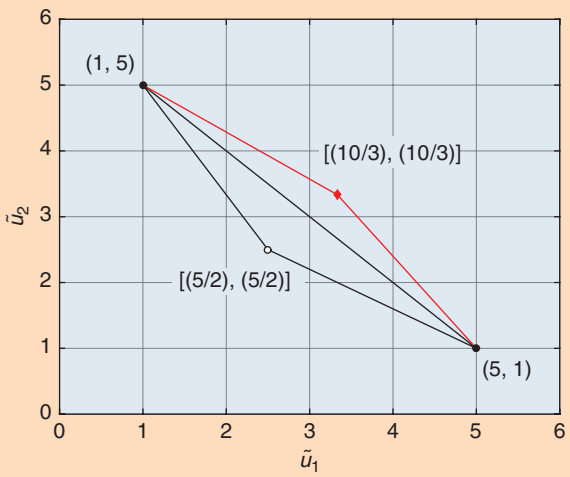
Another effective approach that can be taken to further improve the efficiency of the game solution while also addressing fairness issues is to seek alternative solution concepts. One example of such alternative solutions is the concept of NBS [38], which has been originally defined for two-player games. The implementation of the NBS typically requires some form of coordination or exchange of information among the players. As explained in the "Learning Equilibria in Strategic-Form Games" section, the NBS can be related to SP algorithms such as consensus algorithms. The NBS was used in the networking literature about 20 years ago to obtain fair solutions to flow-control problems in communication networks [33]. More recently, it has been exploited in different contexts, such as in [34] to solve bandwidth allocation problems, in [39] to achieve weighted proportional fairness in resource allocation in wireless networks, or in [4] to obtain cooperative beamforming strategies in interference networks where transmitters are equipped with multiple antennas. Another example can be found in [9], wherein the bargaining methodology is employed to address the problem of rate allocation for collaborative video users (see also [40]). Following [38], let us define the NBS for two-player games. For this, we denote by $\mathcal{U}$ the set of feasible utility points of the strategic-form game of interest and assume that $\mathcal{U}$ is a closed and convex set. Let us denote by (λ_1, λ_2) a given point in $\mathcal{U}$, which will be referred to as a *status quo* (or, equivalently, a *disagreement point*). The NBS is then defined next.

DEFINITION 12 (NBS)

The NBS is the unique PO profile, which is a solution of

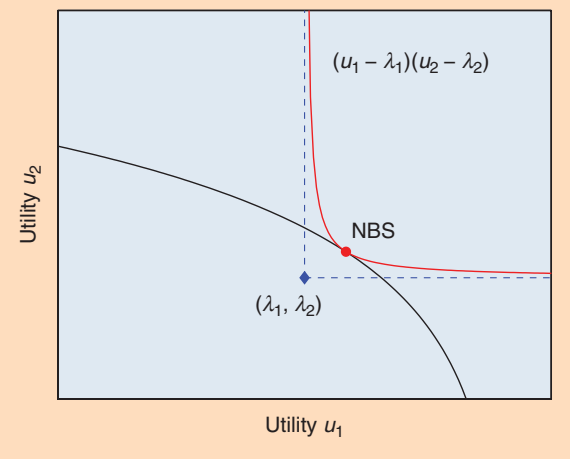
$$\max_{(u_1, u_2) \in \mathcal{U}} \quad (u_1 - \lambda_1)(u_2 - \lambda_2)$$
$$\text{subject to} \quad u_1 \geq \lambda_1, u_2 \geq \lambda_2, \tag{13}$$

where $\mathcal{U}$ is a convex set. ■



[FIG6] The set of CE of the game given in Figure 5 in the expected utility plane.

The graphical interpretation of the NBS is shown in Figure 7. The solution of (13) corresponds to the point of tangency between the Pareto boundary of $\mathcal{U}$ and the hyperbola $(u_1 - \lambda_1)(u_2 - \lambda_2) = \kappa$, where κ is properly chosen to ensure only one intersection between the two curves. The original definition of the NBS by Nash only concerns two-player games but it can be extended by considering K players. For this, the two-factor product above, which is called the *Nash product*, becomes $\prod_{k=1}^{K}(u_k - \lambda_k)$. However, when there are more than two parties involved in the bargaining, coalition forming is always possible, and this definition may need to be replaced by modified versions, such as the coalition NBS [41] (please refer to the "Coalition-Form Games" section for further details). We will conclude the discussion on the NBS by providing an example that is drawn from [42], i.e., a beamforming game for communications in the presence of interference.



[FIG7] The graphical interpretation of the NBS point (red circle) as the intersection between the Pareto boundary of $\mathcal{U}$ and the hyperbola $(u_1 - \lambda_1)(u_2 - \lambda_2) = \kappa$, where the status quo $\lambda = (\lambda_1, \lambda_2)$ is represented by the blue diamond.

EXAMPLE 3 (BEAMFORMING GAME [42])

Consider two N-antenna transmitters. Transmitter $i \in \{1,2\}$ has to choose a beamforming vector $w_i \in \mathbb{C}^N$, such that $w_i^H w_i = 1$ (where the superscript ^{H} stands for Hermitian transpose). The signal observed by the single-antenna receiver i is given by $y_i = h_{ii}^H w_i x_i + h_{ji}^H w_j x_j + z_i$, $j = -i$, $h_{ji} \in \mathbb{C}^N$ are fixed for all (i,j), $x_i \in \mathbb{C}$, and $z_i \sim \mathbb{C}\mathcal{N}(0,1)$ is complex white Gaussian noise. By choosing the utility function as $u_i = \log(1+\text{SINR}_i)$ with $\text{SINR}_i = (|h_{ii}^H w_i|^2 \mathbb{E}|x_i|^2)/(1+|h_{ji}^H w_j|^2 \mathbb{E}|x_j|^2)$, it can be shown that any point of the Pareto frontier can be reached by beamforming vectors, which linearly combine the zero-forcing (ZF) beamforming solution (w_i^{ZF}) and maximum ratio transmission (MRT) beamforming solution $\left(w_i^{\text{MRT}}\right)$ [42]. Therefore, finding the NBS amounts to finding the appropriate linear combination coefficient α_i, which is defined as $w_i = \alpha_i w_i^{\text{ZF}} + (1-\alpha_i) w_i^{\text{MRT}}$. The unique NE of the considered game corresponds to $(\alpha_1^{\text{NE}}, \alpha_2^{\text{NE}}) = (0,0)$, i.e., each transmitter uses ZF beamforming. By choosing the unique NE of the game under investigation to be the status quo point, i.e., $\lambda_i = u_i(\alpha_i^{\text{NE}}, \alpha_j^{\text{NE}})$, the NBS is then given by

$$(\alpha_1^{\text{NBS}}, \alpha_2^{\text{NBS}}) = \arg \max_{(\alpha_1,\alpha_2)\in[0,1]^2} [u_1(\alpha_1,\alpha_2) - u_1(0,0)] \times [u_2(\alpha_1,\alpha_2) - u_2(0,0)]. \quad (14)$$

By construction, the obtained solution is necessarily more Pareto-efficient than the NE. However, computing the NBS typically requires more channel state information than what is required by the NE [42]. ■

SPECIAL CLASSES OF STRATEGIC-FORM GAMES

In this section, we review some special classes of strategic-form games, which show a relevant share of the game-theoretic approaches available in the SP literature. For the sake of brevity, we list here only the distinguishing features of each class but also provide a (nonexhaustive) list of relevant references that can be used to gather more specific details on problem modeling and solution tools. For other interesting classes of games (not reported here due to space constraints), interested readers are referred to specific literature on the topic (e.g., [22], [23], and [31]).

ZERO-SUM GAMES

One of the most common types of strategic-form games is the two-player zero-sum game. A two-player zero-sum game is a game in which the sum of the utilities is zero or can be made zero by appropriate positive scaling and translation, which do not depend on the players' actions or strategies. In other words, it is a game such that $\mathcal{K} = \{1,2\}$, $u_1(s_1,s_2) + u_2(s_1,s_2) = 0$. In such a game, one player is a maximizer, i.e., aims to maximize its gain, while the other player is a minimizer, i.e., aims to minimize its losses (which are the gains of the other player). In SP, zero-sum games are especially popular when modeling security games involving an attacker and a defender. In such games, the attacker's gains are most often equal to the defender's losses, yielding a zero-sum situation. An example in this context can be found in [16], in which the interaction between a target and a MIMO radar—both smart—is modeled as a two-player zero-sum game since the target and the radar are completely hostile. The mutual information criterion is used in formulating the utility functions. In [43], the problem of polarimetric waveform design for distributed MIMO radar from a game-theoretic perspective is also formulated as a two-player zero-sum game played between an opponent and the radar design engineer. In [2], the authors use a two-player zero-sum game to model a watermarking problem where a source sequence (the cover text) needs to be copyright-protected before it is distributed to the public. Another example is given by a two-user communication channel (such as the Gaussian multiple access channel) with a constraint on the total sum-rate [44].

Despite being one of the most well-studied and analyzed classes of strategic-form games in GT (in part because many results can be derived), zero-sum games can be restrictive. In fact, the majority of the studied problems in SP are better modeled as nonzero-sum games.

CONTINUOUS QUASI-CONCAVE GAMES

A game is said to be continuous if, for all $k \in \mathcal{K}$, the utility function u_k is continuous in the strategy profile s. It is said to be quasi-concave if u_k is quasi-concave with respect to s_k for any fixed s_{-k} and $\mathcal{S}_k$ is a compact and convex set. For such games, we can take advantage of Theorem 2, which ensures the existence of at least one pure-strategy NE. A flurry of research activity on energy-efficient resource allocation in wireless communications or sensor networks makes use of quasi-concave utility functions, which aim at trading off the performance of network agents while saving as much energy as possible. Since the performance usually increases with the amount of resources employed, a useful modeling provides

$$u_k(s) = \frac{f_k\left(\frac{s_k}{1+\sum_{j\neq k} s_j}\right)}{s_k} \quad (15)$$

under the hypothesis of a one-dimensional strategy set $\mathcal{S}_k = [0, P^{\max}]$, with $P^{\max}$ being the maximum transmit power. As long as f_k shows some desirable properties (such as sigmoidness), which are often verified in many SP and communications scenarios, the ratio u_k proves to be quasi-concave with respect to s_k. This is the case, for instance, when $f(x) = (1-e^{-x})^M$, $M > 1$ or $f(x) = e^{-(a/x)}$, $a > 0$.

CONTINUOUS CONCAVE GAMES

The same assumptions as for the previous special class of games are made, except that u_k is now a concave function of s_k. The existence of a pure NE is guaranteed in such games since individual concavity implies individual quasi-concavity. Interestingly, if we make one more assumption, called the *diagonally strict condition* (*DSC*), the uniqueness of the NE can also be guaranteed. This is since sufficient conditions for ensuring uniqueness are quite rare in the GT literature. The DSC is met if there exists a vector of (strictly) positive components $r = (r_1, \ldots, r_K)$, such that

$$\forall (s,s') \in \mathcal{S}^2, s \neq s' : (s-s')(\gamma_r(s') - \gamma_r(s))^T > 0, \quad (16)$$

where $\gamma_r(s) = [r_1(\partial u_1(s)/\partial s_1), \ldots, r_K(\partial u_K(s)/\partial s_K)]$. An example of this game can be found in [45]. Therein, the scenario investigated is a set of multiantenna transmitters, which have to choose a precoding matrix to optimize their expected individual transmission rate between each of them and a common multiantenna receiver.

SUPERMODULAR GAMES

Supermodular games are thoroughly investigated in [46]. A strategic-form game is supermodular if, for all $k \in \mathcal{K}$, $\mathcal{S}_k$ is a compact subset of $\mathbb{R}$; u_k is upper semicontinuous in s; and $u_k(s_k, s_{-k}) - u_k(s_k, s'_{-k})$ is nondecreasing in s_k for all $k \in \mathcal{K}$, $\mathcal{S}_k$ and for all $s_{-k} \geq s'_{-k}$, where the inequality is intended to be component-wise. In the example of power control, this definition is very easy to understand. If all the transmitters, except k, increase their power level, then transmitter k has interest in increasing its own power as well. Two properties make supermodular games appealing in the SP community: 1) the set of pure-strategy NE is not empty; and 2) iterative distributed algorithms such as the best-response dynamics (BRD; see the "Learning Equilibria in Strategic-Form Games" section for more details) can be used to let the players converge to one NE of the game. As an example, we can perform an affine transformation of the utility functions in (15) such that they become

$$U_k(s) = \frac{f_k\left(\frac{s_k}{1+\sum_{j\neq k} s_j}\right)}{s_k} - c_k s_k, \tag{17}$$

with $c_k \geq 0$ being a parameter to be tuned. The latter parameter induces a penalty in terms of utility, which increases with the transmit power. The corresponding transformation is called *affine* or *linear pricing* and aims at improving (social) efficiency at the equilibrium. The corresponding game can be shown to be supermodular provided that the action space is reduced as detailed in [3]. Other examples of supermodular games can be found in the SP literature. For instance, in [17], the problem of time-of-arrival-based positioning is formulated as a supermodular game.

POTENTIAL GAMES

A strategic-form game is said to be potential if, for all $k \in \mathcal{K}$, s_k, $s'_k \in \mathcal{S}_k$ and all $s_{-k} \in \mathcal{S}\backslash\mathcal{S}_k$, the difference $u_k(s_k, s_{-k}) - u_k(s'_k, s_{-k})$ can be related to a global potential function $\Phi(s)$ that does not depend on the specific player k. There exist at least four types of potential games: weighted, exact, ordinal, and generalized, according to the relationship between the differences in utilities and potential functions [22]. For example, a game is an exact potential game if there exists a function Φ, such that $u_k(s_k, s_{-k}) - u_k(s'_k, s_{-k}) = \Phi(s_k, s_{-k}) - \Phi(s'_k, s_{-k})$. Similarly to supermodular games, the interest in potential games stems from the guarantee of the existence of pure-strategy NE, and from the study of a single function, which allows the application of theoretical tools borrowed from other disciplines, such as convex optimization [47]. For instance, a maximum point for Φ is an NE for $\mathcal{G}$. Similarly to supermodular games, convergence of iterative distributed algorithms such as the BRD algorithm is guaranteed in potential games. Examples of potential games can be found in [48] for a problem of power allocation, in [49] for radar networks, or in [50] for a problem of multiportfolio optimization. In [51], the authors make use of a potential game to study cooperative consensus problems for sensor deployment. Other simple examples of potential games are games with a common utility function or games for which each utility only depends on the individual action or strategy.

REPEATED GAMES

It is important to note that the definition of the strategic form does not require any particular assumption on the sets of strategies $\mathcal{S}_1, \ldots, \mathcal{S}_K$. In particular, as seen throughout this section, an example of $\mathcal{S}_k$ can be a discrete alphabet (as in the wireless sensor's dilemma), or an interval of $\mathbb{R}$ (as in the example of energy-efficient power control game). In the mentioned examples, the game is said to be static because each player takes a single action. It should be stressed, however, that the strategic form can also be used to model some dynamic games in which players have to take an action in a repeated manner and even in a continuous-time manner (e.g., in some differential games). In dynamic games, the sets of strategies become more complex. They can be sets of sequences of functions or sets of sequences of probability distributions. Due to space limitations in this article, we will only mention the case of repeated games, which will allow us to identify some differences in terms of modeling and analysis between static and repeated games.

A repeated game belongs to a subclass of dynamic games, in which the players face the same single-stage game, say, $\Gamma = (\mathcal{K}, (\mathcal{A}_k)_{k\in\mathcal{K}}, (\nu_k)_{k\in\mathcal{K}})$, where $\mathcal{A}_k$ is the set of possible actions for player k, and ν_k is its instantaneous utility function. The game is played over several stages. The number of stages can be either finite or infinite. The single-stage game is called, equivalently, the *constituent*, *component*, or *stage game*. When introducing the notion of time, the strategies s_k become complete plans of actions, which depend on the unfolding of the game through time. More precisely, a strategy in a repeated game typically corresponds to a sequence of maps or functions, which assign an action to a sequence of observations. Similarly, the utility functions of the repeated game are modified and correspond now to average or long-term utilities. Often, average utilities are of the form

$$u_k(s) = \sum_{t=1}^{+\infty} \theta_t \nu_k(a(t)), \tag{18}$$

where $(\theta_t)_{t\geq 1}$ represents a sequence of weights, which can model different aspects depending on the scenario under consideration (e.g., see [22]). Typical choices for $(\theta_t)_{t\geq 1}$ are as follows:

- $\forall t \in \{1, \ldots, T\}, \theta_t = (1/T)$, and $\forall t \geq T+1, \theta_t = 0$; this type of game is referred to as a *finitely repeated game*.
- $\forall t \geq 1, \theta_t = (1-\delta)\delta^t$, with $0 \leq \delta < 1$; this type of game is referred to as a *repeated game with discount*.
- When the limit exists, $\forall t \geq 1, \theta_t = (1/T)$; this type of game is called an *infinitely repeated game*.

The definition of the strategies $s_1, \ldots, s_K$ strongly depends on the observation assumptions made. For instance, in a repeated game with perfect monitoring and perfect recall, i.e., a game where every player observes all the past actions and is able to store them, the strategy of player $k \in \mathcal{K}$ is given by the following sequence of causal functions:

$$\forall t \geq 1, s_{k,t} : \begin{array}{ccc} \mathcal{A}^{t-1} & \rightarrow & \mathcal{A}_k \\ (a(1), \ldots, a(t-1)) & \mapsto & a_k(t), \end{array} \tag{19}$$

where $a(t) = (a_1(t), \ldots, a_K(t))$ is the profile of actions played at stage t and $\mathcal{A}^0 = \emptyset$ by convention. This strategy is called a *pure strategy*.

Even in the special case of repeated games just described, we can identify some important differences between static and repeated games in terms of equilibrium analysis. The existence issue is fundamental for the NE to be relevant as a solution concept for the problem of interest. Note that, while uniqueness is an important issue for static games, e.g., to be able to predict the convergence point of a distributed algorithm, it is generally much less relevant for a repeated game, since the number of equilibria can be large and even infinite. This is the reason why equilibria are not characterized in terms of equilibrium strategies, but rather in terms of equilibrium utilities. This characterization corresponds to a theorem called the *Folk theorem* [31]. We have seen that efficiency is an important issue for a static game. For a repeated game, due to the fact that players can observe the history of the actions played and therefore exchange information, there may exist efficient equilibria and those equilibria can be attained. For example, in the case of the wireless sensor's dilemma, the following strategies can be checked to be equilibrium strategies of an infinite repeated game with perfect observation:

$$\forall t \geq 2, \quad s^{\star}_{k,t} = \begin{cases} \text{narrowband} & \text{if } a_j(t-1) = \text{ narrowband}, j \in \{1, \ldots, K\} \\ \text{wideband} & \text{otherwise}, \end{cases} \tag{20}$$

with $a(1) = (\text{narrowband}, \ldots, \text{narrowband})$. By implementing these strategies, each player gets a utility that equals three, whereas it was one in the static game version. Therefore, repeating the game and considering long-term utilities allows one to reach more efficient points at every stage of the game. This can be interpreted as a form of cooperation among the players. Thus far, we have mentioned two forms of cooperation, i.e., through bargaining and cooperative plans in repeated games. In the "Coalition-Form Games" section, we will see that the coalition form offers another way of implementing cooperative solutions in games. From the above discussion, it follows that referring to strategic-form games as *noncooperative games* and to coalition games as *cooperative games* is questionable. Indeed, cooperation may exist in the former while players may still be selfish in the latter.

Remark 2

In general, extensive-form games group all situations in which the players are allowed to have a sequential interaction, meaning that the move of each player is conditioned by the previous moves of all players in the game. This class of games is termed *dynamic games*. Repeated games are a subclass of dynamic games, in which the players face the same single-stage (static) game every period. While extensive-form games are not treated because of the lack of space needed to address their general aspect, repeated games, which represent a notable example, are included in this tutorial, thanks to their broad field of application in the SP scenario. ■

BAYESIAN GAMES

When one wants to perform the direct maximization of a function while some of its parameters are unknown, a possible solution is to consider an expected version of the function of interest (e.g., think of the famous expectation-maximization algorithm). When solving a game, a similar approach can be adopted. In the presence of multiple decision-makers, the problem is however more difficult. To understand this, assume that each player chooses a prior distribution over the parameters it does not know (e.g., the overall channel state): this is its belief. However, a player also has to assume what it knows about the belief of the other players. Going further, a player needs to have a belief about the belief on the other players on its own belief. This leads to the quite complex notion of hierarchy of beliefs. This approach seems to be inapplicable in practice. Why should an automaton or a computer implement such an elaborate level of reasoning? An important result of practical interest is that a simpler model might capture the whole hierarchy of beliefs. This model is known as *Harsanyi's model* [52], and it is very close in spirit to what is done in estimation problems in the presence of uncertain parameters. Once the game is formulated as a strategic-form (Bayesian) game, standard tools can be exploited. Although it is exactly an NE in the presence of expected utilities, in this context, an NE is called a *Bayesian equilibrium*. Application examples of Bayesian games in the literature of SP for communications can be found in [53]. Therein, the unknown parameter is typically the communication channel state. In [54], the authors illustrate how Bayesian games are natural settings to analyze multiagent adaptive sensing systems.

LEARNING EQUILIBRIA IN STRATEGIC-FORM GAMES

To better understand the relationship between the solution concepts described in the "Strategic-Form Games" section and algorithmic aspects, we will first consider some experiments, which were conducted by the biologist David Harper [55]. These experiments are of interest to better understand how equilibria can be achieved (learned) by repeated interactions driven by simple decision-making rules. In winter 1979, Harper conducted experiments on a flock of 33 ducks on a lake in a botanical garden at Cambridge University in the United Kingdom. Two observers who were acting as bread tossers were located at two fixed points on the lake surface 20 m apart. Pieces of bread were thrown at regular intervals. For instance, one of the experiments assumes that the frequency of supply for one observer (called the *least-profitable site*) is 12 items/minute, whereas it was equal to 24 items/minute for the other observer. Figure 8 represents the number of ducks at the least-profitable site against time; the dots indicate the mean points, while the vertical segments represent the dispersion of the measures. After about a minute, the number of ducks at the least-profitable site stabilizes around 11, which means that 22 ducks are at the most-profitable site. The corresponding point is an NE: every duck that would switch to the other site in a

unilateral manner would get less food. Figure 8 shows that, at the beginning of the trial, each duck behaves like a conventional optimizer: most of the ducks go to the most-profitable site. This choice does not take into account that the site-selection problem a duck faces is not a conventional optimization problem but a game: what a duck gets does not only depend on its choice but also on others' choices. During the transient period, the ducks that switch to the other site realize they get more food at the least-profitable site. Other ducks do so as long as an equilibrium is reached. Quite likely, the ducks do not know their utility functions and, more generally, the parameters of the game they play. They may hardly be qualified as rational players as well. Nonetheless, some sort of iterative "auction" process (known as *tâtonnement*) has led them to an NE, showing that an NE can emerge as the result of repeated interactions between entities that have only partial information on the problem and only implement primitive decision-making or learning rules. The purpose of this section is to provide learning rules (or SP algorithms) among many others from the vast literature of multiagent learning, learning in games, or distributed optimization, which may lead to equilibria.

Although the remainder of this article only focuses on distributed optimization and multiagent learning algorithms as solution concepts for a certain static game, it may also be possible to interpret a multiagent-learning rule as a strategy for a certain dynamic game [22], showing also the existence of a relationship between learning and dynamic games.

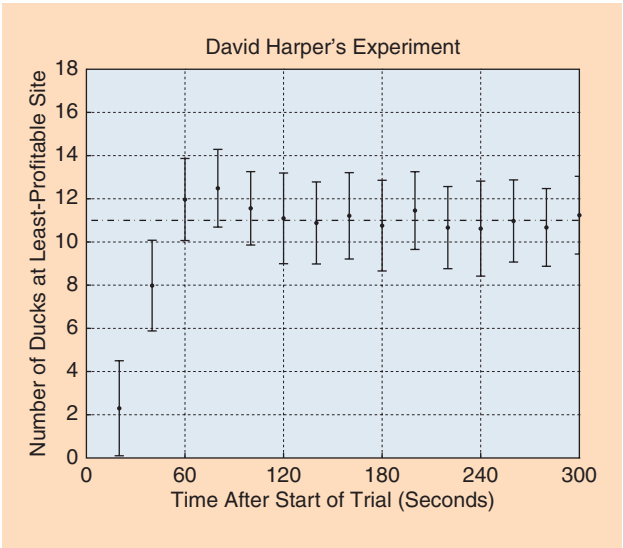
BEST-RESPONSE DYNAMICS

BRD is a popular and simple learning rule that may lead to equilibria. The BRD has been used in various disciplines, but, as its use is specialized, the different instances of it are not always recognized as the same algorithm. Two instances of it are the Gauss–Seidel method [56] and the Lloyd–Max algorithm [57]. The Gauss–Seidel method is an iterative algorithm that allows to numerically solve a linear system of equations. Let us review this method in the special case of two unknowns x_1, x_2 and two observations y_1, y_2. The goal is to solve the system

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \tag{21}$$

where the entries a_{kj} are assumed to be known and meet some classical conditions, which can be found in [56]. By denoting $(x_1(t), x_2(t))$, the value for the pair (x_1, x_2) at iteration t, x_1 is updated as $x_1(t+1)$, which is obtained by solving $a_{11}x_1(t+1) + a_{12}x_2(t) - y_1 = 0$. Then, $x_2(t+1)$ is obtained by solving $a_{21}x_1(t+1) + a_{22}x_2(t+1) - y_2 = 0$. This can be interpreted as a game with two players in which x_k is the action of player k and setting (or making close) to zero $a_{kk}x_k + a_{k,-k}x_{-k} - y_k$ is its objective or cost function. The Gauss–Seidel method precisely implements the sequential BRD of the latter game.

As observed in [58], another special instance of the BRD is the Lloyd–Max algorithm, which was originally used for scalar quantization and is now extensively used in data-compression techniques in information theory and SP. Designing a signal quantizer means choosing how to partition the source signal space into cells or regions and choosing a representative for each of them. It turns out

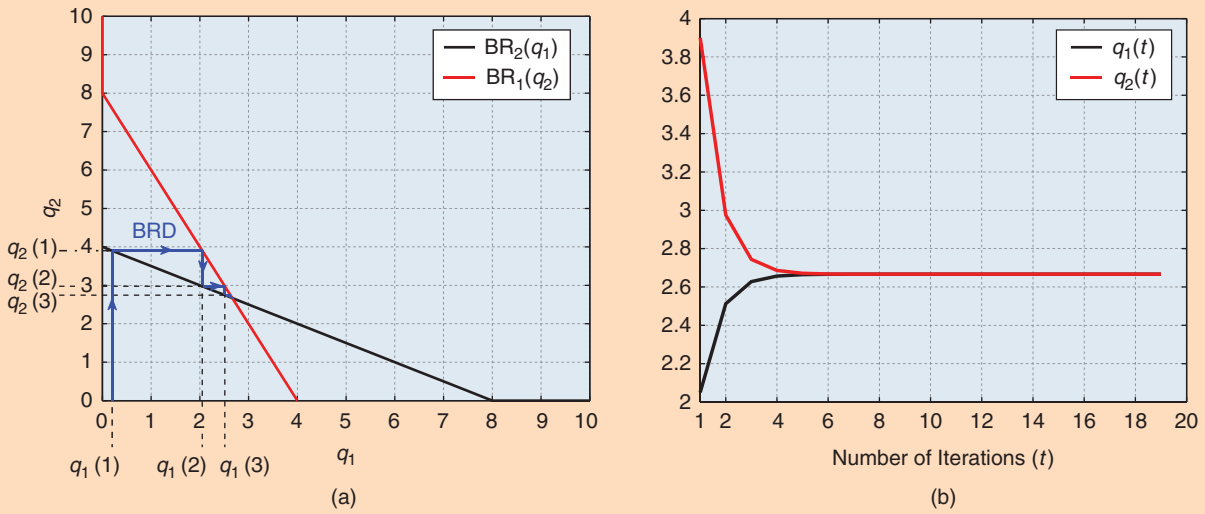


[FIG8] The ducks are given the choice between two bread tossers for which the frequency of supply of the most-profitable site is twice that of the least profitable. After switching a few times between the two sites, the ducks stick to a given choice. The corresponding point is an NE.

that finding in a joint manner the set of regions and the set of representatives which minimize the distortion (i.e., the quantization noise level) is a difficult problem in general. The Lloyd–Max algorithm is an iterative algorithm, in which each iteration comprises two steps. First, one fixes a set of regions and computes the best representatives in the sense of the distortion. Second, for these representatives, one updates the regions so that distortion is minimized. This procedure is repeated until convergence and corresponds to a special instance of the sequential BRD of a game with two players which have a common cost function. As shown in the "Special Classes of Strategic-Form Games" section, since the cost function is common, the game is potential. As explained later, convergence of the sequential BRD is guaranteed in such games.

EXAMPLE 4 (COURNOT TÂTONNEMENT)

Another well-known instance of the BRD is the Cournot tâtonnement. It was originally introduced by Cournot in 1838 to study an economic competition between two firms where each one has to decide the quantity of goods to produce. In particular, Cournot showed that the following dynamical procedure converges: firm 1 chooses a certain quantity of goods $q_1(1)$, firm 2 observes the quantity produced by firm 1 and plays its BR $q_2(2)$, i.e., the quantity maximizing its profit, firm 1 readjusts its quantity to this reaction to $q_1(3)$ for its benefit to be maximal and so forth. Cournot proved that after "a while" this process converges to the so-called Cournot equilibrium, which can be shown to be the NE of the associated strategic-form game. This is what Figure 9 illustrates. A possible application of the dynamical procedure above can be found in [59] in which the authors consider a competitive spectrum sharing scheme based on GT for a CR network consisting of a primary user and multiple secondary users (SUs) sharing the same frequency spectrum. The spectrum-sharing problem is modeled as an oligopoly market and a static



[FIG9] An illustration of the Cournot tâtonnement. (a) This process, which is a special case of the sequential BRD algorithm, converges to the unique intersection point between the players' BRs (i.e., the unique pure NE of the game). (b) Also illustrated by the Cournot duopoly, convergence of sequential BRD is typically fast.

game has been used to obtain the NE for the optimal allocated spectrum size for the SUs. ■

The BRD can be formulated for a game with an arbitrary number of players. In its most used form, the BRD operates in a sequential manner (sequential BRD) such that players update their actions in a round-robin manner. Within round $t+1$ (with $t \geq 1$) the action chosen by player $k \in \mathcal{K}$ is computed as:

$$a_k(t+1) \in \mathrm{BR}_k[a_1(t+1), \ldots, a_{k-1}(t+1), a_{k+1}(t), \ldots, a_K(t)]. \tag{22}$$

If there is more than one best action, then one of them is chosen at random from the uniform probability distribution.

An alternative version of the BRD operates in a simultaneous way, meaning that all players update their actions simultaneously:

$$a_k(t+1) \in \mathrm{BR}_k[a_{-k}(t)]. \tag{23}$$

Algorithm 1: The BRD.

set $t = 0$
initialize $a_k(0) \in \mathcal{S}_k$ for all players $k \in \mathcal{K}$ (e.g., using a random initialization)
repeat
  for $k = 1$ to K do
    update $a_k(t+1)$ using (22) or (23)
  end for
  update $t = t+1$
until $|a_k(t) - a_k(t-1)| \leq \varepsilon$ for all $k \in \mathcal{K}$

The pseudocode of BRD for both instances is sketched in Algorithm 1. Observe that both can be applied to games in which the action sets are either continuous or discrete. If continuous, convergence means that the distance between two successive action profiles remains below a certain threshold $\varepsilon > 0$. If discrete, convergence means that the action profile does not change at all (i.e., $\varepsilon = 0$). When it converges, the convergence points are typically pure NE (e.g., see [22]). There are no convergence results for general games using BRD. Most of the existing results rely on application-specific proofs. For example, [5] considers an application example of the BRD in SP for which an ad hoc proof for convergence is provided. However, if some special classes of games are considered, then there exist sufficient conditions under which the convergence of the sequential BRD to a pure NE is always guaranteed. For example, it is ensured when exact potential games and supermodular games are considered (see "Special Classes of Strategic-Form Games" section and [22] for more details). In addition to this, the convergence of the sequential BRD is ensured when the BRs are standard functions [60]. These results are summarized next.

THEOREM 3 ([22])
In potential and supermodular games, the sequential BRD converges to a pure NE with probability one. ■

THEOREM 4 ([60])
If the BRs of a strategic-form game are standard functions, then the BRD converges to the unique pure NE with probability one. ■

Unlike the sequential BRD, there does not seem to exist general results that guarantee the convergence of the simultaneous BRD. As shown in [61], a possible way out to ensure convergence is to let player k update its action as $a_k(t+1) \in \overline{\mathrm{BR}}_k[a_{-k}(t)]$ where $\overline{\mathrm{BR}}_k[a_{-k}(t)]$ is defined as

$$\overline{\mathrm{BR}}_k[a_{-k}(t)] = \arg\max_{a_k \in \mathcal{A}_k} u_k(a_k, a_{-k}(t)) + \kappa \| a_k - a_k(t) \|^2 \quad (24)$$

with $\kappa \geq 0$. The term $\| a_k - a_k(t) \|^2$ acts as a stabilizing term, which has a conservative effect. If κ is large, this term is minimized by keeping the same action. By choosing κ in an appropriate manner, [61] shows that the simultaneous BRD associated with the modified utility converges.

Now we consider an application example that will be developed throughout this section to illustrate the different algorithms and notions under consideration. In particular, it allows us to extract sufficient conditions under which the sequential BRD converges.

EXAMPLE 5 (POWER-ALLOCATION GAMES IN MULTIBAND INTERFERENCE CHANNELS)

Consider a wireless communication system, which comprises K transmitter–receiver pairs. Each transmitter wants to communicate with its own receiver. More precisely, transmitter $k \in \{1, \ldots, K\}$ (player k) has to allocate its available power (denoted by P) among N orthogonal channels or frequency bands to maximize its own transmission rate $u_k = \sum_{n=1}^{N} \log_2(1 + \gamma_{k,n})$, where $\gamma_{k,n}$ is the signal-to-interference-plus-noise ratio (SINR) at receiver k over band n, which is defined as

$$\gamma_{k,n} = \frac{h_{kk,n} p_{k,n}}{\sigma^2 + \sum_{\ell \neq k} h_{\ell k,n} p_{\ell,n}}, \quad (25)$$

where $p_{k,n}$ is the power transmitter k allocates to band n, $h_{\ell k,n} \geq 0$ is the channel gain associated with the link from transmitter ℓ to receiver k over band n, and σ^2 accounts for the thermal noise. Denote by $p_k = (p_{k,1}, \ldots, p_{k,N})$ the power-allocation vector of transmitter k. Two scenarios in terms of action space are considered:

$$\mathcal{A}_k^{\mathrm{PA}} = \left\{ p_k \in \mathbb{R}_+^N : \sum_{n=1}^{N} p_{k,n} \leq P \right\} \text{ and } \mathcal{A}_k^{\mathrm{BS}} = \{ Pe_1, \ldots, Pe_N \}, \quad (26)$$

where PA stands for *power allocation* and BS for *band selection*, and $e_1, \ldots, e_N$ represents the canonical basis of $\mathbb{R}^N$ [i.e., $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$ and so on]. The two corresponding strategic-form game will be denoted by $\mathcal{G}^{\mathrm{PA}}$ and $\mathcal{G}^{\mathrm{BS}}$. ■

A sufficient condition for the sequential BRD to converge for the game $\mathcal{G}^{\mathrm{PA}}$ has been provided in [62]. The condition is that the spectral radius ρ of certain matrices $\mathbf{H}(n)$ are strictly less than one

$$\forall n, \rho(\mathbf{H}(n)) < 1 \text{ with } H_{k\ell}(n) = \begin{cases} 0 & \text{if } k = \ell \\ \dfrac{h_{\ell k,n}}{h_{k k,n}} & \text{if } k \neq \ell. \end{cases} \quad (27)$$

Condition (27) is useful for the general case of the multiband interference channel and roughly means that the interference level on every band should not be too high. However, as shown in [48], the sufficient condition holds with probability zero (randomness stems from the fact that the channel gains $h_{k\ell,n}$ are assumed to be realizations of a continuous random variable) in the special case of the multiband multiple access channel, which corresponds to have only one receiver of interest for all the transmitters. In the latter case, the SINR takes a more particular form, which is

$$\gamma_{k,n} = \frac{h_{k,n} p_{k,n}}{\sigma^2 + \sum_{\ell \neq k} h_{\ell,n} p_{\ell,n}}, \quad (28)$$

where $h_{k,n}$ is the channel gain of the link between transmitter k and the receiver for band n. Remarkably, in this particular setting, $\mathcal{G}^{\mathrm{PA}}$ and $\mathcal{G}^{\mathrm{BS}}$ can be shown to be exact potential games [48] with a potential function

$$\Phi = \sum_{n=1}^{N} \log_2 \left(\sigma^2 + \sum_{k=1}^{K} h_{k,n} p_{k,n} \right). \quad (29)$$

Exact potentiality of games guarantees the convergence of the sequential BRD to a pure NE. In game $\mathcal{G}^{\mathrm{PA}}$, the sequential BRD consists in updating the power level according to a water-filling formula

$$p_{k,n}(t+1) = \left[\frac{1}{\omega_k} - \frac{p_{k,n}(t)}{\gamma_{k,n}(t)} \right]^+, \quad (30)$$

where $[x]^+ = \max(0, x)$, ω_k is the Lagrangian multiplier associated with the inequality constraint $\sum_{n=1}^{N} p_{k,n} \leq P$, and $\gamma_{k,n}(t)$ is the SINR at receiver k over band n at time t. The solution is known as the *iterative water-filling algorithm* (*IWFA*) and was introduced for the multiband interference channel in [63]. In its most general form, the sequential BRD algorithm in (22) is quite demanding in terms of observation since each player has to observe the actions played by the others. In the case of the IWFA, it is seen that that only knowledge of the SINR $\gamma_{k,n}(t)$ is required to implement the BRD, which is basically an aggregate version of the played actions: this information can easily be estimated at the receiver and fed back to player k for updating its transmit power. When converging, the IWFA, and more generally the sequential BRD, does it quite fast: convergence is typically observed after a few iterations [48]. Intuitively, the feature of fast convergence stems from the fact that the BRD relies on a detailed knowledge of the problem at hand. Typically, the utility functions are assumed to be known. When this knowledge is not available, instead of considering highly structured distributed optimization algorithms such as the BRD, one may consider multiagent learning algorithms, which are typically much less demanding in terms of modeling the problem, as discussed in the next sections. However, before moving to such techniques, an alternative version of the BRD is considered, which operates on probability distributions over actions (instead of pure actions) and is referred to as the *fictitious play* (*FP*) algorithm. Considering the FP algorithm allows us to better understand the iterative structure of many learning algorithms, particularly the one considered in the "Reinforcement Learning" section.

The original version of the FP algorithm assumes discrete action sets, which is what is also assumed next. It should be stressed that the BRD is generally not well suited to the discrete case.

For example, when applied to $\mathcal{G}^{\mathrm{BS}}$, it converges in the scenario of multiband multiple access channels while it does not converge in the multiband interference channel case as cycles appear [64]. This is quite frequent in games with discrete actions. Therefore, learning algorithms such as the one described in the "Reinforcement Learning" section are not only useful to assume less structure on the problem but also to deal with the discrete case. From now on, we thus assume that

$$\mathcal{A}_k = \{a_{k,1}, \ldots, a_{k,N_k}\}, \tag{31}$$

where $|\mathcal{A}_k| < +\infty$. The FP algorithm, introduced by Brown in 1951 [65], is a BRD algorithm in which empirical frequencies are used. Working with probability distributions is very convenient mathematically. Although mixed strategies are exploited, this does not mean that mixed NE are sought. In fact, pure NE can be shown to be attracting points for all the dynamics, which are considered in this tutorial. This means that, under appropriate conditions, mixed strategies tend to pure strategies as the number of iterations grows large. The empirical frequency of use of action $a_k \in \mathcal{A}_k$ for player $k \in \mathcal{K}$ at time $t+1$ is defined by

$$\pi_{k,a_k}(t+1) = \frac{1}{t+1}\sum_{t'=1}^{t+1} \mathbb{1}_{\{a_{k,t'}=a_k\}}, \tag{32}$$

where $\mathbb{1}$ is the indicator function. If player k knows $\pi_{-k,a_{-k}}(t)$ (i.e., the empirical frequency of use of the action profile a_{-k} at time t), then it can compute its own expected utility and eventually choose the action maximizing it. Observe that the computation of $\pi_{-k,a_{-k}}(t)$ requires to observe the actions played by the others. As for BRD, this knowledge can be acquired only through an exchange of information among the players. For example, in the two-player CR's dilemma (Example 2), if CR1 has knowledge of the number of times that CR2 has picked narrowband or wideband up to time t, then CR1 can easily compute $\pi_{2,a_2}(t)$ through (32).

In its simultaneous form, the FP algorithm operates as follows:

$$a_k(t+1) \in \arg\max_{a_k \in \mathcal{A}_k} \sum_{k=1}^{K} \pi_{-k,a_{-k}}(t)\, u_k(a_k, a_{-k}). \tag{33}$$

The important point we want to make about the FP algorithm concerns the structure of the empirical frequencies. They can be computed in a recursive fashion as

$$\begin{aligned}\pi_{k,a_k}(t+1) &= \frac{1}{t+1}\sum_{t'=1}^{t+1} \mathbb{1}_{\{a_{k,t}=a_k\}} = \frac{1}{t+1} \\ &\quad \sum_{t'=1}^{t} \mathbb{1}_{\{a_{k,t'}=a_k\}} + \frac{1}{t+1}\mathbb{1}_{\{a_{k,t+1}=a_k\}} \\ &= \pi_{k,a_k}(t) + \lambda_k^{\mathrm{FP}}(t)\left[\mathbb{1}_{\{a_{k,t+1}=a_k\}} - \pi_{k,a_k}(t)\right],\end{aligned} \tag{34}$$

with $\lambda_k^{\mathrm{FP}}(t) = 1/(t+1)$. The last line translates the fact that the empirical frequency at time $t+1$ can be computed from its value at time t and the knowledge of the current action. More interestingly, it emphasizes a quite general structure, which is encountered with many iterative and reinforcement learning (RL) algorithms.

REINFORCEMENT LEARNING

Originally, RL was studied in the context of single-player (or single-automation) environments with a finite set of actions. A player receives a numerical utility signal and updates its strategy. The environment provides this signal as a feedback for the sequence of actions that has be taken by the player. Typically, the latter relates the utility signal to actions previously taken to learn a mixed strategy that performs well in terms of average utility. In a multiplayer setting, RL is inherently more complex since the learning process itself changes the thing to be learned. The main objective of this section is to show that feeding back to the players only the realizations of their utilities is enough to drive seemingly complex interactions to a steady state or, at least, to a predictable evolution of the state. In RL algorithms, players use their experience to choose or avoid certain actions based on their consequences. Actions that led to satisfactory outcomes will tend to be repeated in the future, whereas actions that led to unsatisfactory experiences will be avoided. One of the first RL algorithms was proposed by Bush and Mosteller in [66], wherein each player's strategy is defined by the probability of undertaking each of the available actions. After every player has selected an action according to its probability, every player receives the corresponding utility and revises the probability of undertaking that action according to a reinforcement policy. More formally, let $u_k(t)$ be the value of the utility function of player k at time t, and denote by $\pi_{k,a_{k,n}}(t)$ the probability player k assigns to action $a_{k,n}$ at time t. Then, the Bush and Mosteller RL algorithm operates as follows:

$$\pi_{k,a_{k,n}}(t+1) = \pi_{k,a_{k,n}}(t) + \lambda_k^{\mathrm{RL}}(t)\, u_k(t)\left[\mathbb{1}_{\{a_k(t)=a_{k,n}\}} - \pi_{k,a_{k,n}}(t)\right], \tag{35}$$

where $0 < \lambda_k^{\mathrm{RL}}(t) < 1$ is a known function that regulates the learning rate of player k (it plays the same role as the step-size in the gradient method). The updating rule given by (35) has the same form of (34), but one of the strengths of the algorithm corresponding to (35) is that each player only needs to observe the realization of its utility function and nothing else. It can, therefore, be applied to any finite game. Convergence is ensured for classes of games such as potential games and supermodular games, introduced in the "Special Classes of Strategic-Form Games" section. The convergence of RL algorithms is also ensured for dominance solvable games [22], which are not treated in this tutorial due to space limitations. As for the BRD, convergence points are either pure NE or boundary points. The price to be paid for the high flexibility regarding the environment and the absence of strong assumptions on its structure is that the RL algorithm in (35) usually requires a large number of iterations to converge compared to the BRD algorithm.

All the aforementioned distributed algorithms (i.e., the BRD algorithm, the FP algorithm, and the considered RL algorithm) are attractive since they only rely on partial knowledge of the problem. On the other hand, convergence points are typically pure NE, which, in most cases, are inefficient. Often, points that Pareto-dominate the NE points can be shown to exist. A nontrivial problem is how to reach one of them in a distributed manner.

We will not address this challenging task in this tutorial, rather, we will provide one learning algorithm that allows players to reach a CCE. This may be more efficient than a pure or mixed NE, since the latter is a special instance of it.

RM-LEARNING ALGORITHM

The key auxiliary notion that is exploited for RM-learning algorithms is the notion of regret [67], which is eventually exploited to assign a certain probability to a given action. The regret player k associates with action $a_{k,n}$ is the difference between the average utility the player would have obtained by always playing the same action $a_{k,n}$, and the average utility actually achieved with the current strategy. Mathematically, the regret at time t for player k is computed as

$$\forall n \in \{1, \ldots, N_k\}, r_{k,a_{k,n}}(t+1) = \frac{1}{t}\sum_{t'=1}^{t} u_k(a_{k,n}, a_{-k}(t')) - u_k(a_k(t'), a_{-k}(t')). \quad (36)$$

RM relies on the assumptions that, at every iteration t, player k is able to evaluate its own utility—i.e., to calculate $u_k(a_k(t), a_{-k}(t))$—and to compute the utility it would have obtained if it had played another action a'_k (i.e., $u_k(a'_k, a_{-k}(t))$. In [67], the rule for updating the probability player k assigns to action $a_{k,n}$ is as follows:

$$\pi_{k,a_{k,n}}(t+1) = \frac{[r_{k,a_{k,n}}(t+1)]^+}{\sum_{n'=1}^{N_k}[r_{k,a_{k,n}}(t+1)]^+}. \quad (37)$$

If, at time $t+1$, player k has a positive regret for every action, it implies that it would have obtained a higher utility by playing the same action during the whole game up to iteration $t+1$, instead of playing according to the distribution $\pi_k(t) = (\pi_{k,a_{k,1}}, \ldots, \pi_{k,a_{k,N_k}})$. The pseudocode for the RM algorithm is sketched in Algorithm 2. The updating rule (37) has a very attractive property: it is with no regret [67]. The consequence of this property is expressed through the following result.

THEOREM 5 (CONVERGENCE OF THE RM ALGORITHM)

In any finite game, when updated as (37), the empirical frequencies of the action profile always converge almost surely to the set of CCE. ■

Algorithm 2: The regret-matching-learning algorithm.

set $t = 0$
initialize $\pi_k(0)$ s.t. $\sum_{n=1}^{N_k}\pi_{k,n}(0) = 1$ for all players $k \in \mathcal{K}$ (e.g., using a random initialization)
repeat
  for $k = 1$ to K do
    for $n = 1$ to N_k do
      update $r_{k,n}(t+1)$ using (36)
    end for
    for $n = 1$ to N_k do
      update $\pi_{k,n}(t+1)$ using (37)
    end for
    choose $a_k(t+1)$ according to the distribution $\pi_k(t+1)$
  end for
  update $t = t+1$
until $|a_k(t) - a_k(t-1)| \leq \varepsilon$ for all $k \in \mathcal{K}$

Observe that in those games wherein CCE, CE, mixed NE, and pure NE coincide (e.g., in the simple CR's dilemma introduced in Example 2), then a unique CCE exists, which is a pure NE. In this particular setting, RM does not provide any performance gain over the BRD. However, in most cases, the RM algorithm has the potential to perform better than distributed algorithms such as the BRD. This is what is illustrated in the following section. In the CR context, an application example supporting this statement can be found in [68].

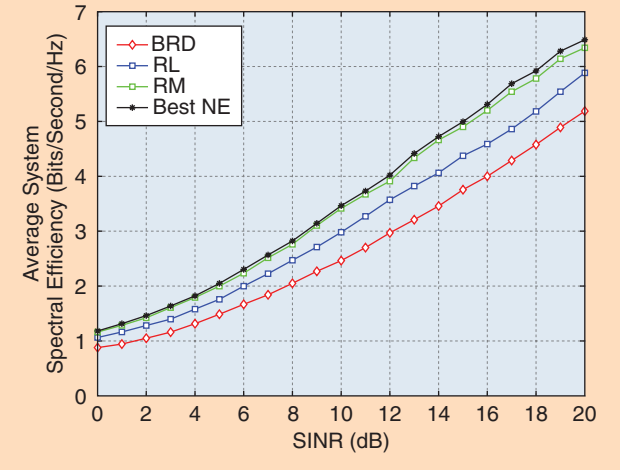
ILLUSTRATION AND COMPARISON ANALYSIS

Table 2 summarizes the different features of the three classes of distributed algorithms that have been discussed throughout this section. Here, we consider a special instance of game $\mathcal{G}^{\text{BS}}$ in which only two transmitters and two receivers are operating and two bands are assumed, and each transmitter has to select one single band [69].

Figure 10 depicts the performance in terms of sum-utility (i.e., the transmission sum-rate) as a function of the SINR for both BRD and RM algorithms. As shown in Figure 10, the RM-learning algorithm is more efficient in terms of sum-rate than the BRD algorithm. In fact, here, the performance of the CCE, which is obtained by implementing the RM-learning algorithm, is very close to the performance of the best pure NE of the

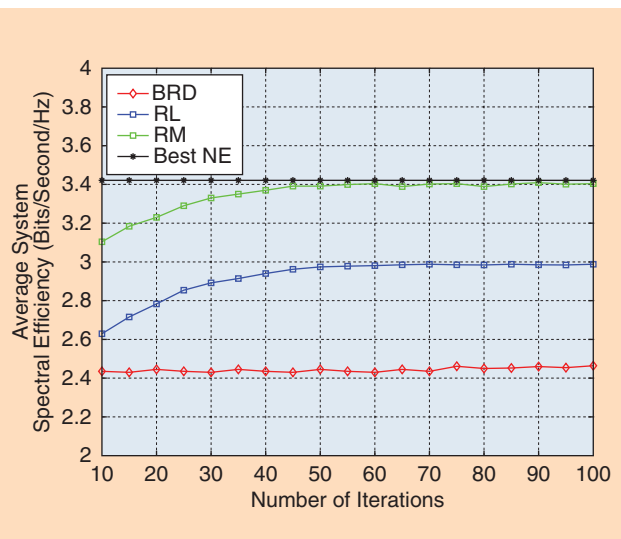
[TABLE 2] MAIN FEATURES FOR THE BRD, RL, AND RM ALGORITHMS.

	BRD	RL	RM
ACTION SETS	CONTINUOUS OR DISCRETE	DISCRETE	DISCRETE
CONVERGENCE	SUFFICIENT CONDITIONS	SUFFICIENT CONDITIONS	ALWAYS GUARANTEED
CONVERGENCE POINTS	PURE NE OR BOUNDARY POINTS	PURE NE OR BOUNDARY POINTS	CCE
CONVERGENCE SPEED	FAST	SLOW	MEDIUM
EFFICIENCY OF CONVERGENCE POINTS	TYPICALLY LOW	TYPICALLY LOW	TYPICALLY MEDIUM
OBSERVATION TYPICALLY REQUIRED	ACTIONS OF THE OTHERS	VALUE OF THE UTILITY FUNCTION	ACTIONS OF THE OTHERS
KNOWLEDGE TYPICALLY REQUIRED	UTILITY FUNCTIONS AND ACTION SETS	ACTION SETS	UTILITY FUNCTIONS AND ACTION SETS



[FIG10] RM always has the potential to perform better than BRD since a pure NE is a special case of a CCE. The figure shows that this is effectively the case for the sum-rate of the considered distributed power-allocation problem under the given simulation setup [64].

game. On the other hand, the BRD is observed to converge to a pure NE, which does not coincide with the best NE. Although this is not what is observed generically, there may exist some initial points for which the BRD performs better than the RM algorithm. This raises a problem, i.e., to characterize the relation between the initial and convergence points, which is a challenging and open issue. Note that if the RL algorithm is considered, the same issue would appear. The performance of the RL algorithm for the special case of interest would also strongly depend on the initial point. The main drawback of using the RL algorithm would be the number of iterations needed for convergence (when the algorithm effectively converges), as shown in Figure 11.



[FIG11] The average system spectral efficiency as a function of the number of iterations at a fixed SINR of 10 dB [64].

CONSENSUS ALGORITHMS

One last type of algorithm described in this section is the consensus algorithm. These algorithms rely on a strong coordination among the players. This is achieved at the price of a quite strong observation assumption: the corresponding updating rule requires explicit knowledge of the actions chosen by the other agents or players. As a result of this assumption, an efficient solution can be attained at convergence. For instance, assume that the players' actions are real numbers, $\forall k \in \mathcal{K}, a_k \in \mathbb{R}$, and assume that the network should be designed to operate at a given point $a^\star = (a_1^\star, \ldots, a_K^\star) \in \mathbb{R}^K$ referred to as *consensus*. This point must be attained by each player through a certain iterative and distributed procedure involving exchanges among the agents; of course reaching a point that is globally efficient may not be possible. A simple instance of a consensus algorithm (see, e.g., [70]) is as follows:

$$a_k(t+1) = a_k(t) + \sum_{j \in \mathcal{A}_k} \beta_{k,j}(a_j(t) - a_k(t)), \tag{38}$$

where t is the iteration index, $\mathcal{A}_k$ represents the neighbors of agent k, and $\beta_{k,j}$ is some weight that player k places on the action or state of player j. Simple sufficient conditions can be stated under which such an algorithm converges [70]. Indeed, the convergence analysis amounts to studying the dynamical system $a(t+1) = \mathbf{C}a(t)$, where the matrix $\mathbf{C}$ follows from (38). The convergence properties of consensus algorithms have been studied under several interaction models (delays in information exchange, connectivity issues, varying topologies, and noisy measurements) and can be usually ensured by construction of the algorithm itself. However, this requires a well-determined topology for the network and also a relatively large amount of information to be exchanged, especially in comparison with the other learning algorithms already described. Surprisingly, there has been relatively little research that explicitly links consensus problems or, more generally cooperative control problems, to the very relevant branches of learning in game literature or multiagent system literature that address coordination problems. Most of the attempts in this context aim at establishing a connection between coordination problems and potential games [51], [71]. A simple application of consensus is given next.

EXAMPLE 6 (DETECTION WITH SENSOR NETWORKS)

Consider a wireless sensor network in which each sensor can only communicate with the sensors within its transmitted signal range. Each sensor has to decide whether a tectonic plate is active (e.g., to detect earthquakes). The action to be taken by each sensor is assumed to be binary active or not active. To decide whether a plate is active or not by using all the measurements and associated decisions, a consensus algorithm such as (38) can be implemented [70]. ■

COALITION-FORM GAMES

As already discussed, strategic-form games mainly focus on the strategic choices of individual players and on what strategies each player would choose to reach its goal. More importantly, strategic-form representations often deal with noncooperative cases in which players are assumed to act selfishly, individually,

and without any side payment, cooperation, or exchange of communication. In contrast, many SP applications require some sort of cooperation among the players. For example, it is increasingly common to form virtual arrays of antennas, sensors, or telescopes to improve estimation or detection accuracy; this type of operations requires communication and partial-to-full cooperation between the players. Cooperative networking, in which devices can, e.g., cooperatively route their packets at the network layer, is also a typical application where cooperation is needed. In such cases, given the cooperative nature of the system, players may form groups among one another, in an effort to improve their state and position in the game. Thus, we now deal with groups of players or *coalitions* that act in a coordinated manner. Inside each such coalition, the players may still be choosing strategies, similar to a strategic-form game, but, overall, the goal here is to analyze the formation of the coalitions given the possibility of communication between the players.

THE COALITION FORM PROVIDES A SUITE OF TOOLS THAT ALLOW US TO EVALUATE FAIRNESS, STABILITY, AND EFFICIENCY WHEN PLAYERS IN A GAME ARE ABLE TO COORDINATE AND COMMUNICATE WITH EACH OTHER PRIOR TO MAKING DECISIONS.

Coalition-form games provide an appropriate representation for situations in which groups or coalitions (subsets) of players can work together in a game. In such games, we are typically concerned about the options available to coalitions, the possible coalitions that will form, and how the utility received by the coalition as a whole can be divided among its members in a way to sustain cooperation. This amounts to assuming the existence of a mechanism which imposes a particular action or, more in general, series of actions on each player. This mechanism can, e.g., result from a binding agreement among the players or from a rule imposed by a designer.

The coalition form is suitable to model a number of problems. On the one hand, it is the only game-theoretic tool available to predict and characterize how groups of players can weight and evaluate the mutual benefits and cost from cooperation and then decide on whether to work together and form binding agreements. On the other hand, when the coalition form is found to be suitable to model the problem at hand, one of its strengths is that it may lead to a solution that is more efficient than in the case in which no coordination occurs. Moreover, the coalition form provides a suite of tools that allow us to evaluate fairness, stability, and efficiency when players in a game are able to coordinate and communicate with each other prior to making decisions.

COALITION-FORM GAMES AND BARGAINING THEORY

One important distinction to make is that between NBS (more generally, the bargaining theory) and coalition-form games. In the game-theoretic literature [23], both Nash bargaining and coalition-form games are often grouped under the umbrella of cooperative games. This classification mainly stems from the fact that, in both cases, the players may coordinate their strategies and are, in general, cooperative. However, the NBS is restricted to the scenario in which two or more players want to share a resource, and they are, a priori, willing to cooperate in this resource sharing, provided that the "terms" of cooperation are fair. Then, the question becomes the following: given the players' initial positions (which is generally the max–min or NE solution using their individual utility functions), which have to be feasible, how should they split the rest of the resource being shared? Subsequently, as detailed in the "Solution Concepts" section, the NBS follows an axiomatic approach. In this regard, the NBS provides a unique allocation that answers this question.

Although the original solution proposed by Nash was restricted to two players, the idea of Nash bargaining has since been extended to the general multiplayer game. This extension has been particularly popular in the SP community, where the analogies between Nash bargaining and the famous proportional fair resource allocation mechanisms have been drawn and exploited. Important examples include consensus algorithms, resource allocation, and optimal beamforming [4], [9], [39]. Nonetheless, even with this extension, the overall Nash bargaining problem remains the same—how to share a resource among all players to 1) satisfy the Nash bargaining axiom and 2) improve the players' overall utility.

In contrast, coalition-form games address a different problem: how cooperative coalitions among different players can be formed given the mutual benefits and costs for cooperation. Therefore, coalition-form games are not restricted to a fair resource-sharing problem such as in the NBS. In contrast, they investigate a much more generic problem. Coalition-form games study how to stabilize and maintain cooperative coalitions between groups of players, in any situation, not just resource allocation. In contrast, for a bargaining problem, it is assumed that 1) all players are willing to cooperate, 2) there is no cost for cooperation, and 3) the cooperation is reduced to share a resource.

Therefore, in terms of objectives, the two approaches are different. However, the NBS can be used as an axiomatic solution for distributing the utility inside a formed coalition, in a fair manner (in the Nash bargaining sense). However, even though the bargaining solution will satisfy the NBS fairness axioms, it will not necessarily stabilize the coalition, in the sense that some players may still want to leave this coalition and form other coalitions, if the NBS is used to distribute the utilities. Thus, to study large-scale cooperation and coalition-formation processes, one must use solution concepts and algorithms that are much more general than the NBS. This motivates the need for coalition-form games.

COALITION-FORM GAME MODELS

In this section, we use the notation $\mathcal{C}$ to refer to a given subset of the set of players $\mathcal{K} = \{1, \ldots, K\}$. The notation $2^{\mathcal{K}}$ is used to denote the power set associated with $\mathcal{K}$. For example, if $\mathcal{K} = \{1,2\}$, then

$2^{\mathcal{K}} = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$. A coalition game is defined by the pair $(\mathcal{K}, v)$, where v, is the value of a coalition that is a function or mapping that provides a characterization of the utility (or utilities) achieved by the players that belong to a certain coalition.

In essence, for classical models of coalition games, depending on the definition of v, we can distinguish between nontransferable-utility (NTU) games and transferable utility (TU) games:

- *NTU games:* coalition actions result in utilities to individual coalition members.
- *TU games:* utilities are given to the coalition and then divided among its members.

In an NTU game, the formation of a coalition $C \subseteq \mathcal{K}$ leads to changes of the individual utilities of the players within C; however, there is no single value that can be used to describe the overall coalition utility. In contrast, in a TU setting, a single-valued function can be used to determine the overall utility of a coalition. Subsequently, the individual utilities can be viewed as a sharing of this single-valued gain.

NTU and TU games can be further categorized into characteristic-function (CF) games or partition-function (PF) games:

- *PF games:* the utility of a coalition $C \subseteq \mathcal{K}$ depends on the actions chosen by the other coalitions in $\mathcal{K} \backslash C$.
- *CF games:* the utility of a coalition C only depends on the action chosen by the members of C.

Both CF and PF games admit many applications in SP. The latter is particularly useful for cases in which externalities, such as interference or delay in communication networks, are present and depend on the coalition actions of the players. For simplicity, our focus will be on CF games. This classification of coalition-form games is shown in Figure 12, emphasizing the fact the TU and CF game are special classes of NTU and PF games, respectively.

NTU GAMES

The formal definition of a coalition-form NTU game with CF often follows the form introduced by Aumann and Peleg in [72], which states the following.

DEFINITION 13 (NTU COALITIONAL GAMES WITH CHARACTERISTIC FUNCTION)

An NTU game with CF is given by a pair $(\mathcal{K}, v)$: $\mathcal{K} = \{1, \ldots, K\}$ is called the *set of players,* and v is the CF The latter is a set-valued function

$$\begin{aligned} v\colon\ 2^{\mathcal{K}} &\rightarrow \mathbb{R}^{\mathcal{K}} \\ C &\mapsto v(C) \end{aligned} \tag{39}$$

such that, for every coalition $C \subseteq \mathcal{K}$, $v(C)$ is a closed convex subset of $\mathbb{R}^{\mathcal{K}}$ that contains the utility vectors that players in C can achieve. ■

In other words, in an NTU game, the value is a set of payoff vectors that can be achieved by the players in the game. A coalition game is therefore said to be NTU if the value or utility of a coalition cannot be arbitrarily apportioned among the coalition's members. For an NTU model, the players do not value a given coalition in the same way. Instead, for every coalition, one or more vectors of individual payoffs will be achieved. For example, when investigating a bargaining situation in which players cannot share their utilities, we can view the NBS vector as an example of an NTU allocation. In SP problems, casting a problem as an NTU coalition game strongly depends on the metrics being optimized. Some metrics such as energy are individual and thus NTU by design, while others (such as the sum-rate) are not necessarily NTU.

TU GAMES

A special case of NTU games is given by TU games. In TU games, $v(C)$ is a real value that represents the total utility obtained by the coalition C. This is what the following model translates.

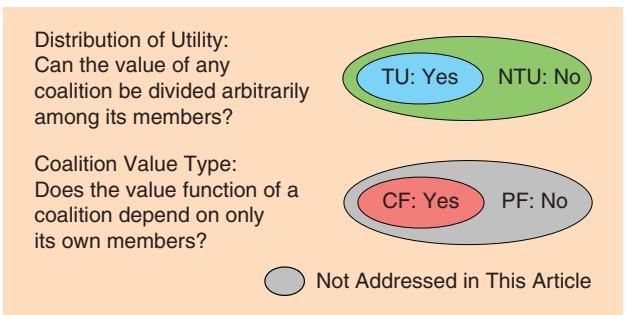
DEFINITION 14 (TU COALITIONAL GAMES WITH CF)

A TU game with CF is given by a pair $(\mathcal{K}, v)$: $\mathcal{K} = \{1, \ldots, K\}$ is called the *set of players,* and v is the CF. The latter is given by

$$\begin{aligned} v\colon\ 2^{\mathcal{K}} &\rightarrow \mathbb{R} \\ C &\mapsto v(C). \end{aligned} \tag{40}$$

■

The TU property means that this worth can be divided in any manner among the coalition members. The values in TU games are thought of as monetary values that the members in a coalition can distribute among themselves using an appropriate rule (one such rule being an equal distribution of the utility). In SP problems, one typical example in which the TU property is applicable is the case in which groups of devices aim to optimize a certain sum-rate. Given that a sum-rate can virtually be divided among the devices via a proper choice of transmit signal (or, more specifically, a power allocation), we can view the sum-rate as a TU metric.



[FIG12] **Classification of coalition-form games.**

Remark 3

In practice, we can convert an NTU game to a TU game for the purpose of analysis. One way to do so is to define the TU value function as being the sum of the individual payoffs of the players. Even though the actual division of this sum cannot be done in this case in an arbitrary manner, we can still use the TU model to understand how the system would behave under cooperation. In this case, we can consider this single-valued TU utility as being a total revenue achieved by the entire utility, with the individual divisions being the virtual monetary gain that is provided to each player, if those players are to act within a coalition. ■

CANONICAL GAME

For any type of coalition-form game, the primary goal is to develop strategic algorithms and mechanisms that allow to characterize and predict which coalitions will form, when, and how. Given this goal, we often refer to coalition games as *coalition-formation games*. However, one special case occurs when the value of a coalition is nondecreasing with respect to the size of the coalition. Here, cooperation is always beneficial and the costs of cooperation are negligible. In this specific case, the game is said to be *superadditive*, which is formally defined as follows for the TU case:

$$v(C_1 \cup C_2) \geq v(C_1) + v(C_2). \tag{41}$$

In this setting, it is trivial to see that the grand coalition of all players will yield the maximum utility. However, this does not mean that this grand coalition will always form. In fact, unless the total gains are properly distributed to the grand coalition's members, some of those members may deviate and form their own coalitions. In such scenarios, the coalition-formation game is simply reduced to the so-called canonical game model, in which the goal is no longer to form coalitions, but rather to study ways in which the grand coalition of all players can be sustained. This will lead to many solutions that look at fairness and stability, as detailed in the next section.

Remark 4

This basic definition of the various coalition-form game types can be used as a basis to develop more advanced model. For example, if a player may belong simultaneously to multiple coalitions, we can define the framework overlapping coalition-formation (OCF) games. In SP, this could be used to model applications such as sharing of sensor data between multiple cooperating groups. In OCF scenarios, one must redefine the way a coalition-form game is presented. One approach is to represent a coalition by a $1 \times K$ vector r whose element r_i represents the amount of resources that player i has shared with this coalition. For such OCF scenarios, notions of stability or fairness must now be extended to the new representation and definition of a coalition. ■

Given this overview on how to represent a coalition-form game, our next step is to discuss the solution concepts and main results.

SOLUTION CONCEPTS

For coalition-form games, we can distinguish two features for the solution: stability and fairness. On the one hand, the solution of coalition-formation game must ensure that the formed coalitions are not susceptible to deviations by individual members or even subgroups of members. On the other hand, given that coalition formation entails a division of utility, a suitable coalition-form solution must ensure fairness when dividing or allocating the various utilities. Balancing the two goals of fairness and stability is challenging and strongly dependent on factors such as the structure of the value function, the goals of the players, and the application being studied.

The solution of a coalition-form game can further be classified into two additional types: set-valued solutions and single-valued solutions. Set-valued solutions refer to solutions that can guarantee stability or fairness via more than one cooperative strategy. How to choose the most appropriate point within a set-valued solution becomes an important problem. This is reminiscent of the multiplicity of NE in strategic-form games. In contrast, single-valued solutions provide a unique strategy which achieves a given fairness or stability criteria. Practically, although both set-valued and single-valued solutions can be used for both fairness and stability, most existing set-valued solutions are focused on stability, while single-valued solutions are tailored toward fairness.

While both solutions can apply to any type of coalition-form game, for ease of exposition, in this section, we restrict our attention to CF games that are superadditive and TU. By doing so, the overall solution can be viewed as a distribution of utilities that can maintain the stability and fairness within the grand coalition. Nonetheless, throughout our discussions, we will point out the key aspects needed to extend the solutions to the more general coalition-formation cases. Moreover, in the "Algorithms for Coalition-Form Games" section, we will discuss algorithmic implementations that can provide more insights on solving coalition-formation games.

THE CORE

The most popular set-valued solution of a coalition-form game is the core [73]. The core is the set of payoff allocations which guarantees that no group of players has an incentive to leave the grand coalition $\mathcal{K}$ to form any other coalition $C \subset \mathcal{K}$. For a TU game, we let x be the $1 \times K$ vector of individual user utilities. Here, we must have group rationality, i.e., $\sum_{i \in \mathcal{K}} x_i = v(\mathcal{K})$. In other words, the total allocation must sum to the entire value of the grand coalition. In addition, we define a payoff vector x to be individually rational if every $x_i \geq v(\{i\}), \forall i$. This implies that an individually rational payoff vector ensures that no player will obtain a lower payoff by joining the grand coalition. Consequently, the core of a coalition game is defined as the set $\mathcal{S}$ of individually rational and group rational payoff vectors as follows:

$$\mathcal{S} = \left\{ x: \sum_{i \in \mathcal{K}} x_i = v(\mathcal{K}) \text{ and } \sum_{i \in C} x_i \geq v(C) \; \forall \, C \subseteq \mathcal{K} \right\}. \tag{42}$$

In simple terms, the core of a coalition game is the set of payoff allocations that ensure that no group of players would have an incentive to leave the grand coalition and form their own individual coalition. The core guarantees stability with respect to any deviation by any group of players. However, even though the core guarantees stability, and, for the superadditive case, we can easily see that the grand coalition is the most efficient, the core in this game may not be fair to the players. Examples of unfair allocations that lie in the core abound both in the GT and SP literature [74], [75]. Moreover, drawing yet another analogy with the NE, there is no guarantee that a coalition game will have a core solution. Indeed, the core, as a set-valued solution, may be empty.

Nonetheless, the core is one of the most popular set-valued solution concepts in a coalition-form game, which has led to many extensions. For instance, when dealing with a nonsuperadditive

coalition-formation game with TU, we can redefine the core, based on the partition of $\mathcal{K}$ that maximizes the total utility, as follows:

$$O = \left\{ x: \sum_{i\in\mathcal{K}} x_i = \max_{\pi\in\mathcal{P}} \sum_{C\in\pi} v(C) \text{ and } \sum_{i\in C} x_i \geq v(C)\ \forall C \subseteq \mathcal{K} \right\}, \tag{43}$$

where $\mathcal{P}$ is the set of all possible partitions of $\mathcal{K}$ and π is one such partition or coalition structure. Recall that the partition of the set $\mathcal{K}$ is a collection of disjoint subsets whose union would span the entire set $\mathcal{K}$. Thus, the partition constitutes the coalitions that are expected to form in the system. Essentially, the difference between (42) and (43) is that, in (42), the first core condition assumes that the sum of the individual payoffs is equal to the value of the grand coalition, which is guaranteed to form due to superadditivity. In contrast, in (43), because of the nonsuperadditive nature of the game, the grand coalition is not guaranteed to form. Consequently, the first condition of the core must now ensure that the sum of the individual payoffs must be equal to the sum of the values of all coalitions in the partition π that maximizes the total system value. Thus, this coalition-formation core notion implies that, instead of investigating a stable grand coalition, one would seek an allocation that will stabilize the partition π that maximizes the total social welfare of the system. This is particularly useful when coalition formation entails a cost, and, thus, the game is nonsuperadditive.

THE ε-CORE

One extension to the core is the ε-core. This notion bears an analogy with the notion of approximate or ε-equilibria in strategic-form games [64]. The basic idea is that the stability is not achieved exactly, but rather within an ε-approximation neighborhood as follows:

$$\mathcal{S}_1 = \left\{ x: \sum_{i\in\mathcal{K}} x_i = v(\mathcal{K}) \text{ and } \sum_{i\in C} x_i \geq v(C) - \varepsilon \forall C \subseteq \mathcal{K}, \varepsilon \geq 0 \right\}. \tag{44}$$

Interestingly enough, the value of ε can be viewed as a quantification of the "overhead" for deviating from the core. This overhead is incurred on the deviation of every possible coalition. This bears a very interesting analogy to SP—what is the overhead required by a group of devices to deviate from the stability concept, and will they be willing to incur this overhead? The above concept is also known as the weak ε-core, which is used to then define the so-called strong ε-core, where ε is divided between the members of a coalition, i.e., ε is substituted by $|C|\cdot\varepsilon$. In this case, the overhead ε is implicitly assumed to be equally divided between coalition members. The advantage of the ε-core is that it may be easier to establish its existence as well as to develop algorithms that can reach it. This simply mimics the advantages of any approximate solution concept in GT. In SP, there have been some recent works (e.g., [76]) that explored the ε-core as a suitable concept for investigating problems related to beamforming where the overhead of deviating from a certain beamforming strategy might be high enough to reach an ε-core, avoiding the need to reach the more stringent core definition.

THE SHAPLEY VALUE

The core and its variants constitute set-valued stability notions. In contrast, we can solve a coalition-form game using single-valued fairness notions. Single-valued solution concepts mainly associate with every coalition game $(\mathcal{K}, v)$ a unique payoff vector known as the *solution* or *value* of the game (which is different from the value of a coalition). One example of this notion is the NBS. In fact, most single-valued notions follow an axiomatic approach: a set of preset properties that are imposed on the sought after payoff allocation to find a desirable solution. One popular such solution is the Shapley value [23]. For a TU coalition-formation game, the Shapley value assigns to every player the payoff x_i, given by

$$x_i = \sum_{C\subseteq\mathcal{K}\setminus\{i\}} \frac{|C|!(|\mathcal{K}|-|C|-1)!}{|\mathcal{K}|!}[v(C\cup\{i\}) - v(C)]. \tag{45}$$

This allocation is interpreted as follows. In the event where the players join the grand coalition in an arbitrary order, the payoff allocated by the Shapley value to a player $i \in \mathcal{K}$ is the expected marginal contribution of player i when it joins the grand coalition. In other words, the contribution of a player is given by an expected value, assuming a random order of joining of the players to the grand coalition which, in a superadditive game, is known to be the most efficient solution. Shapley showed that this solution is unique and it satisfies the following four axioms:

1) Efficiency axiom: $\sum_{i\in\mathcal{K}} x_i = v(\mathcal{K})$.

2) Symmetry axiom: if player i and player j are such that $v(C\cup\{i\}) = v(C\cup\{j\})$ for every coalition C not containing player i and player j, then $x_i = x_j$.

3) Dummy axiom: if player i is such that $v(C) = v(C\cup\{i\})$ for every coalition C not containing i, then $x_i = 0$.

4) *Additivity axiom*: If u and v are CFs, then the Shapley value allotted to the game over $(u+v)$ is the sum of the values allotted to u and v, separately.

The Shapley value provides some form of fairness to allocate the payoffs of a grand coalition. Similarly to the core, the Shapley value has led to many extended notions such as the envy-free fairness [75], the Banzhaf index [23], and the Harsanyi index [23]. All of these notions follow the steps of the Shapley value in that they utilize certain axioms and attempt to find a coalition-form solution that satisfies these axioms. However, none of these solutions is guaranteed to be stable. For example, often, the Shapley value will not lie in the core, if that core exists. Therefore, one important challenge for coalition-form games is to balance fairness and stability by combining notions of core and Shapley value.

Remark 5

In summary, for solving coalition-form games, a great number of solution concepts exists. These are split into two categories: single-valued and set-valued. The focus is mainly on stability and fairness. The exact notion of stability or fairness depends largely on the type of the game and the scenario being considered. ■

Next, we will discuss some principle results from coalition-form games, and, then, we will delve more into algorithmic implementation and practical applications in the SP domain.

MAIN THEOREMS

Unlike strategic-form games, in which existence, efficiency, and uniqueness theorems are abundant, for coalition-form games, such results are sparse and often model dependent. However, when dealing with the core, we can discuss two seminal results that relate to the existence of the core and its fairness.

The first main result in this regard is given through the Bondareva–Shapley theorem [23]. This theorem is concerned with coalition-form games that are balanced.

DEFINITION 15 (BALANCED GAME)

A coalition TU game is said to be balanced if and only if we have

$$\sum_{C\subseteq\mathcal{K}} \mu(C)v(C) \le v(\mathcal{K}), \tag{46}$$

for all nonnegative weight collections $\mu = (\mu(C))_{C\subseteq\mathcal{K}}$. ■

Here, μ is simply a group of weights in $[0,1]$ that are assigned to each coalition $C\subseteq\mathcal{K}$ such that $\sum_{C\ni i}\mu(C) = 1, \forall i\in\mathcal{K}$. The main idea behind a balanced game can be explained as follows. Assuming that every player i has a unit of time that can be divided between all possible coalitions that i can form. Every coalition C is active for a time period $\mu(C)$ if all players in C are active during that time. The payoff of this active coalition would then be $\mu(C)v(C)$. Here, $\sum_{C\ni i}\mu(C) = 1, \forall i\in\mathcal{K}$, would then be a feasibility constraint on the players' time allocation. Consequently, a coalition-form game is balanced if there is no feasible allocation of time, which can yield an overall utility that exceeds the value $v(\mathcal{K})$ of the grand coalition. Thus, for a TU balanced game, the following result holds.

THEOREM 6 (BONDAREVA–SHAPLEY [73])

The core of a game is nonempty if and only if the game is balanced. ■

Although the Bondareva–Shapley theorem is a popular result for showing the nonexistence of the core, its applicability in SP may be very limited, as the required balancedness is quite restrictive on the coalition value. In this respect, yet another interesting result is given for convex coalition-form games. A coalition game with TU is said to be convex if its value function satisfies

$$v(C_1)+v(C_2) \le v(C_1\cup C_2)+v(C_1\cap C_2)\ \forall C_1, C_2\subseteq\mathcal{K}. \tag{47}$$

By observing (47), we can view directly its similarity with supermodular games, introduced in the "Special Classes of Strategic-Form Games" section. Now, supermodularity is defined with respect to subsets, rather than vectors in the Euclidean space. We note that the convexity conditions can also be written as follows:

$$v(C_1\cup\{i\}) - v(C_1) \le v(C_2\cup\{i\}) - v(C_2), \tag{48}$$

whenever $C_1\subseteq C_2\subseteq\mathcal{K}\backslash\{i\}$. This can be explained as follows. A game is convex if and only if the marginal contribution of each player to a coalition is nondecreasing with respect to set inclusion. For a convex game, we can state the following.

THEOREM 7 ([73])

For a convex coalition-form game, the core is nonempty and the Shapley value lies in the core. ■

This theorem provides a strong result that combines both stability and fairness. Indeed, for a convex game, the Shapley value is in the core and thus provides both stability and fairness. Although we stated the theorem here for TU games, it can also be extended to NTU games.

ALGORITHMS FOR COALITION-FORM GAMES

One key design challenge in coalition-form games is that of developing algorithms for characterizing and finding a suitable stable or fair solution. This is in general analogous with the algorithmic aspects of noncooperative games where learning is needed to reach a certain NE (see the "Learning Equilibria in Strategic-Form Games" section). In this respect, here, we discuss two algorithmic aspects: 1) finding a stable or a fair distribution for canonical games, and 2) characterizing stable partitions for coalition-formation games.

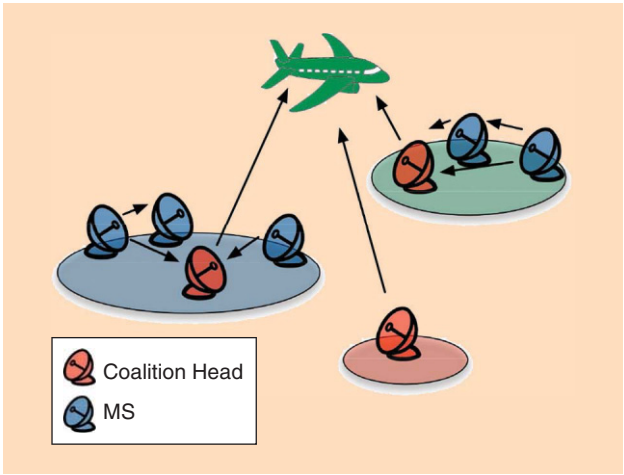
CANONICAL GAMES

For canonical games, the most important solution concept is the core and its variants. Despite being a strongly stable solution concept, computing the core can be relatively complex. In particular, to compute the core directly from the definition, we must solve the following linear program:

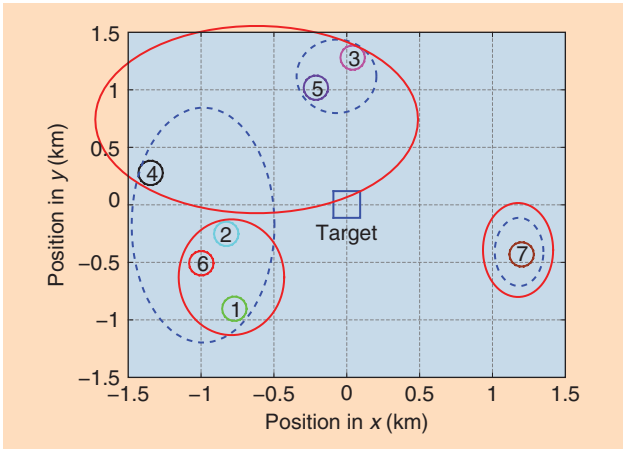
$$\text{minimize}_x \ \sum_{i\in\mathcal{K}} x_i \quad \text{s.t.} \quad \sum_{i\in C} x_i \ge v(C), \forall C\subseteq\mathcal{K}. \tag{49}$$

Solving (49) allows us to find all the solutions that lie in the core, as ensured by the constraint. Clearly, solving the linear program in (49) will require handling $2^{\mathcal{K}}$ constraints, which will grow exponentially as the number of players increase. While no generic rule exists for overcoming this complexity, we can exploit some properties of the game or application being sought. On one hand, we can use theorems such as the Bondareva–Shapley theorem or the convexity of the game to establishing the existence and nonemptiness of the core (see the "Main Theorems" section). On the other hand, for a given coalition-form game structure, we can evaluate the membership of known payoff division rules, such as the bargaining solution or a proportional fair division, in the core. Here, checking whether a certain allocation belongs in the core essentially becomes simpler than deriving all the solutions that are in the core.

Regarding the Shapley value, we can also observe a similar complexity limitation: computing the Shapley value via (45) calls for going again through all the possible coalitions. However, we note that, recently, some approximations for the Shapley value have been developed that allows us to compute it with reduced complexity. A popular approach in this context relies on the use of the multilinear extension method proposed by Owen [77] for a special class of games known as *voting games*. The basic idea is to observe that, in (45), the term inside the summation is the beta function, which can then be used to convert the Shapley value computation into a probability computation which is then approximated by exploiting some properties of voting games. Other approaches to approximate or to reduce the computational time of the Shapley value are surveyed in [78].



[FIG13] Distributed CTD as a coalition game.



[FIG14] The final coalition structure from both distributed (dashed line) and centralized (solid line) CTD for $K = 7$ MSs.

COALITION-FORMATION GAMES

Deriving suitable solutions for coalition-formation games is more challenging than the canonical case as it requires to jointly compute the payoff and the coalitional structure or network partition that will form. For example, computing coalitional structure that lies in the core, as per the definition in (43), can be highly complex, as it requires to look over all partitions of a set—which grow exponentially. However, some approaches using Markov chains or other related ideas have been proposed in [79] and [80], which were proven to work well for reasonably large games.

However, in practical SP applications, we must trade the strength of the core stability for the complexity of finding this solution. One baseline approach for a generic coalition-formation algorithm would consist of two key steps: 1) define a rule with which a player may decide to join or leave a coalition and 2) for the TU case, adopt a proper payoff allocation rule (e.g., the Shapley value, proportional fair, etc.) that is to be applied at the level of any formed coalition.

Regarding the coalition-formation rule, a number of approaches have been proposed within the SP community (e.g., [12], [74], [76], and [81]). Among them, the most popular ones are the merge and split rules, defined as follows (the symbol $\triangleright$ is a preference relation, discussed below):

- *Merge rule:* A group of coalitions $\{C_1, \ldots, C_p\}$ would merge into a single coalition $\cup_{k=1}^{p} C_k$ if and only if $\cup_{k=1}^{p} C_k \triangleright \{C_1, \ldots, C_p\}$.
- *Split rule:* A coalition $\cup_{k=1}^{p} C_k$ will split into a smaller group of coalitions $\{C_1, \ldots, C_p\}$ if and only if $\{C_1, \ldots, C_p\} \triangleright \cup_{k=1}^{p} C_k$.

Here, the preference relation $\triangleright$ can be defined based on the application being studied. A popular preference relation is the so-called Pareto order, whereby the merge or split rule would apply if at least one player improves its payoff via merge or split, without hurting the payoff of any other player. In other words, given the current payoff vector y of all players involved in a merge or split rule, the merge or split occurs when the vector x of the payoffs of all involved players is such that $x \geq y$ with at least one element x_i of x such that $x_i > y_i$. Essentially, this is reminiscent of the Pareto dominance rule used in noncooperative games (see the "Solution Concepts" section).

The advantages of using merge- and split-based algorithms include: 1) guaranteed convergence to a stable, merge- and split-proof coalition structure after a finite number of iterations, 2) convergence is ensured irrespective of the starting point of the network, and 3) the order of merge or split will not impact convergence. Another major advantage of using merge- and-split based algorithms includes the fact that, irrespective of implementation, such algorithms will reach the so-called $\mathbb{D}_c$-stable partition, when such a partition exists. The $\mathbb{D}_c$-stable partition is a partition that: 1) is strongly stable in the sense that no group of coalitions can do better by breaking away from this partition and 2) when using the Pareto order as a preference relation, is PO. Therefore, the merge and split can reach such an optimal and strongly stable partition if it exists. The existence of a $\mathbb{D}_c$-stable partition is highly application dependent and the condition for existence will depend on the domain being studied.

A CASE STUDY: COALITION FORMATION FOR COLLABORATIVE TARGET DETECTION

One SP application in which the coalition-form can be applied is that of collaborative target detection (CTD). For example, in radar systems, a number of monitoring stations (MSs) can collaborate to detect a certain target of interest at a given location. Such stations can be located at different points in the network, and, thus, their view on the target will be different. Here, we assume that the target is a wireless device that is transmitting a certain signal that must be detected. One major challenge in this scenario is the hidden terminal problem—because of fading and path loss, some MSs may receive a weaker signal from the target, hindering their detection performance.

To avoid this problem, CTD can be used. The basic idea is that MSs can share their individual detection results and, then, make a collective decision on the absence or presence of a target at a given location. By collaborating, the MSs can exploit the diversity of their observations to improve detection decisions. However, although CTD

[TABLE 3] THE STRATEGIC FORM VERSUS COALITION FORM.

	STRATEGIC FORM	COALITION FORM
COMPONENTS	PLAYERS, ACTIONS, PER-PLAYER UTILITY	PLAYERS, COALITION VALUE, PER-PLAYER UTILITY
PRIMARY PLAYER STRATEGY	CHOOSE A PARAMETER TO OPTIMIZE	CHOOSE A COALITION MEMBERSHIP
PRIMARY PLAYER OBJECTIVE	OPTIMIZE INDIVIDUAL UTILITY	OPTIMIZE INDIVIDUAL UTILITY (WHILE PART OF A COALITION)
GAME OBJECTIVES	FIND AN EQUILIBRIUM	FIND STABLE COALITIONS AND FAIR ALLOCATIONS
MAIN TYPES	STATIC, DYNAMIC	TU, NTU, CANONICAL, COALITION FORMATION, CF, PF
COMMUNICATION	NO COMMUNICATION BETWEEN PLAYERS	PLAYERS CAN FORM AGREEMENTS AND COMMUNICATE
MAIN SOLUTION CONCEPT	NE—NO PLAYER CAN UNILATERALLY DEVIATE	STABLE PARTITION—NO COALITION CAN DEVIATE
BASELINE ALGORITHMS	LEARNING BASED	MERGE-AND-SPLIT BASED
PRIMARY APPLICATION	DISTRIBUTED OPTIMIZATION	OPTIMIZED COOPERATION, RESOURCE DISTRIBUTION

can improve the probability of detecting the target as the number of collaborating MSs increases, collaboration can lead to an increasing probability of false alarms—the probability that a target is detected while it is not there. The tradeoff between probability of detection and probability of false alarms, as a function of the number of collaborating MSs, motivates the development of a coalition-form games in which the MSs can dynamically decide on how to collaborate while improving probability of detection and maintaining a tolerable false-alarm level.

As shown in Figure 13, we consider a coalition game between a set $\mathcal{K}$ of MSs that are seeking to cooperate to improve CTD performance. Since cooperation here entails a cost (in terms of increased false alarms), the game, in general, cannot be superadditive and, thus, it is classified as a coalition-formation game. In this game, each coalition C of SUs will be optimizing the following value function:

$$v(C) = Q_{d,C} - C(Q_{f,C}, \alpha_C), \tag{50}$$

where $Q_{d,C}$ is the collaborative probability of detection and $C(\cdot)$ is a cost function of the collaborative false-alarm level $Q_{f,C}$ and the target false-alarm constraint α_C. In this model, each coalition C will have a coalition head that will collect the detection results and fuse them to make a collective coalition decision. The fusion rule used will affect the way in which $Q_{d,C}$ and $Q_{f,C}$ are computed. However, it will not affect the way the game is formulated. Here, we notice that (50) is a probabilistic utility and, thus, it cannot be transferred between the members of C. As a result, the CTD coalition-formation game is an NTU game with a special property: the payoff x_i of each member i of C is simply equal to $v(C)$, since this value is a collective result, i.e., we assume that all players in a coalition abide by the entire coalition decision.

Given the utility and involved tradeoffs, a merge-and-split algorithm based on the Pareto order can be proposed, as shown in [81] to find and characterize stable partitions. In Figure 14, we show a snapshot of the network structure resulting from a merge-and-split collaborative spectrum sensing (CSS) algorithm (dashed line) as well as from a centralized approach (solid line) for seven randomly deployed MSs. We notice that the partitions resulting from both approaches are comparable, with neighboring MSs cooperating for improving spectrum sensing. However, this example allows us to highlight the difference between a distributed, coalition-formation game approach, in which each MS makes its own CTD decision, and a centralized optimization approach, in which the MSs have no say in the coalition-formation process. In particular, from Figure 14, we can see that, for the game solution, MS4 is part of coalition $\{1, 2, 4, 6\}$, while for the centralized approach, MS4 is member of $\{3, 4, 5\}$. This difference stems from the fact that, in the distributed case, MS4 acts selfishly while aiming at improving its own utility. In fact, by merging with $\{3, 5\}$, MS4 achieves a utility of 0.9859 with a probability of detection of 0.9976 whereas by merging with $\{1, 2, 6\}$ its utility will be 0.9957 with a probability of detection of 0.99901. Thus, in a coalition-based solution, MS4 prefers to merge with $\{1, 2, 6\}$ rather than with $\{3, 5\}$ regardless of the socially optimal partition.

In summary, the use of a coalition-formation game for CTD can also yield significant gains in terms of the probability of detection, while maintaining a required false-alarm level and without the need for a centralized optimization solution. Building on these results, we can develop a broad range of applications that adopt the coalition-form games for SP problems. For example, the aforementioned model for CSS is extended in [82] to the case in which an MS can belong simultaneously to multiple coalitions. In this regard, [82] shows that the merge-and-split algorithm can be extended to handle the cases of OCF games.

CONCLUSIONS

In this tutorial, we have provided a holistic view on the use of game-theoretic techniques in SP for networks. Particular emphasis has been given to games in strategic and coalitional forms. The key components of such games have been introduced and discussed while providing an SP-oriented view on the various types of games. Some of the primary differences and properties of strategy-form and coalition-form games were summarized in Table 3. Then, we have developed the main solution concepts and discussed the various advantages and drawbacks within SP domains. More importantly, this tutorial provides an in-depth discussion on the connections between GT and algorithmic aspects of SP techniques. The applications discussed range from traditional communication problems to modern-day SP problems such as cognitive radio and wireless sensor networks. Overall, this tutorial is expected to provide a comprehensive, self-contained reference on the challenges and opportunities for adopting GT in SP as well as to locate specific references either in applications or theory.

ACKNOWLEDGMENTS

Luca Sanguinetti is funded by the People Program (Marie Curie Actions) Seventh Framework Program PIEF-GA-2012-330731 Dense4Green and is also supported by the European Research Council starting grant 305123 Advanced Mathematical Tools for Complex Network Engineering. Walid Saad is supported by the U.S. National Science Foundation under grants CNS-1460316, CNS-1460333, CNS-1513697, and AST-1506297.

AUTHORS

Giacomo Bacci (gbacci@mbigroup.it) is a product manager for satellite and terrestrial communications at MBI srl, Pisa, Italy. After receiving his Ph.D. degree in information engineering from the University of Pisa in 2008, he was a postdoctoral research fellow there in 2008–2014 as well as a visiting postdoctoral research associate at Princeton University, New Jersey, in 2012–2014. He is the recipient of the Seventh Framework Program Marie Curie International Outgoing Fellowship 2011, the IEEE Wireless Communications and Networking Conference 2013 Best Paper Award, and the 2014 International Union of Radio Science Young Scientist Award.

Samson Lasaulce (samson.lasaulce@lss.supelec.fr) received his B.S. degree and agrégation degree in physics from École Normale Supérieure, Cachan, France, and his M.S. and Ph.D. degrees in signal processing from Telecom Paris, France. He has held positions with Motorola Labs (1999–2001) and Orange Labs (2002–2003). In 2004, he joined the French Centre National de la Recherche Scientifique and Supélec as a senior researcher. In 2004, he was a professor with École Polytechnique. He is the recipient of several paper awards. He is a coauthor of *Game Theory and Learning for Wireless Networks: Fundamentals and Applications* (Academic Press, 2011).

Walid Saad (walids@vt.edu) received his Ph.D. degree from the University of Oslo, Norway, in 2010. Currently, he is an assistant professor in the Bradley Department of Electrical and Computer Engineering at Virginia Tech, Blacksburg. He is also International Scholar at the Department of Computer Engineering, Kyung Hee University, South Korea. His research interests include wireless networks, game theory, cybersecurity, and cyberphysical systems. He was the recipient of the U.S. National Science Foundation CAREER Award in 2013, the Air Force Office of Scientific Research summer faculty fellowship in 2014, and the Young Investigator Award from the Office of Naval Research in 2015. He was the recipient of several conference best paper awards as well as the 2015 Fred W. Ellersick Prize.

Luca Sanguinetti (luca.sanguinetti@unipi.it) is an assistant professor at the University of Pisa, Italy. In 2007–2008, he was a postdoctoral associate at Princeton University, New Jersey. From 2013 to 2015, he was also with Centrale Supélec, France. His research topics are in the broad area of wireless communications and signal processing. He is currently serving as an associate editor of *IEEE Transactions on Wireless Communications*. He is the recipient of the Seventh Framework Program Marie Curie Intra-European Fellowship 2013 and the IEEE Wireless Communications and Networking Conference 2013 and 2014 Best Paper Awards.

REFERENCES

[1] S. Kassam and H. Poor, "Robust signal processing for communication systems," *IEEE Commun. Mag.*, vol. 21, no. 1, pp. 20–28, Jan. 1983.

[2] A. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Information Theory*, vol. 48, no. 6, pp. 1639–1667, June 2002.

[3] C. Saraydar, N. B. Mandayam, and D. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 291–303, Feb. 2002.

[4] E. Larsson and E. Jorswieck, "Competition versus cooperation on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sept. 2008.

[5] G. Scutari, D. Palomar, and S. Barbarossa, "The MIMO iterative waterfilling algorithm," *IEEE Trans. Signal Processing*, vol. 57, no. 5, pp. 1917–1935, May 2009.

[6] Y. Wu and K. Liu, "An information secrecy game in cognitive radio networks," *IEEE Trans. Information Forensics Security*, vol. 6, no. 3, pp. 831–842, Sept. 2011.

[7] B. Wang, K. Liu, and T. Clancy, "Evolutionary cooperative spectrum sensing game: how to collaborate?" *IEEE Trans. Commun.*, vol. 58, no. 3, pp. 890–900, Mar. 2010.

[8] F. Wang, M. Krunz, and S. Cui, "Price-based spectrum management in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 74–87, Feb. 2008.

[9] H. Park and M. van der Schaar, "Bargaining strategies for networked multimedia resource management," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3496–3511, July 2007.

[10] A. Chakraborty and J. Duncan, "Game-theoretic integration for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 1, pp. 12–30, Jan. 1999.

[11] B. Ibragimov, B. Likar, F. Pernus, and T. Vrtovec, "A game-theoretic framework for landmark-based image segmentation," *IEEE Trans. Med. Imaging*, vol. 31, no. 9, pp. 1761–1776, Sept. 2012.

[12] H. He, A. Subramanian, X. Shen, and P. Varshney, "A coalitional game for distributed estimation in wireless sensor networks," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 4574–4578.

[13] C. Jiang, Y. Chen, and K. Liu, "Distributed adaptive networks: A graphical evolutionary game-theoretic view," *IEEE Trans. Signal Processing*, vol. 61, no. 22, pp. 5675–5688, Nov. 2013.

[14] K. Han and A. Nehorai, "Joint frequency-hopping waveform design for MIMO radar estimation using game theory," in *Proc. IEEE Radar Conf.*, Ottawa, Canada, Apr. 2013, pp. 1–4.

[15] A. Mukherjee and A. Swindlehurst, "Jamming games in the MIMO wiretap channel with an active eavesdropper," *IEEE Trans. Signal Processing*, vol. 61, no. 1, pp. 82–91, Jan. 2013.

[16] X. Song, P. Willett, S. Zhou, and P. Luh, "The MIMO radar and jammer games," *IEEE Trans. Signal Processing*, vol. 60, no. 2, pp. 687–699, Feb. 2012.

[17] A. Moragrega, P. Closas, and C. Ibars, "Supermodular game for power control in TOA-based positioning," *IEEE Trans. Signal Processing*, vol. 61, no. 12, pp. 3246–3259, Dec. 2013.

[18] K. S. Narendra and M. A. L. Thathachar, *Learning Automata: An Introduction*. Harlow, UK: Prentice Hall, 1989.

[19] G. Scutari, D. Palomar, F. Facchinei, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 35–49, May 2010.

[20] J. C. Harsanyi and R. Selten, *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press, Mar. 2003.

[21] J. Y. Halpern, "A computer scientist looks at game theory," *Games Economic Behav.*, vol. 45, no. 1, pp. 114–131, Oct. 2003.

[22] S. Lasaulce and H. Tembine, *Game Theory and Learning for Wireless Networks: Fundamentals and Applications*. Waltham, MA: Academic Press, 2011.

[23] Z. Han, D. Niyato, W. Saad, T. Basar, and A. Hjørungnes, *Game Theory in Wireless and Communication Networks: Theory, Models and Applications*. Cambridge, UK: Cambridge Univ. Press, 2011.

[24] E. Jorswieck, E. Larsson, M. Luise, and H. Poor, "Game theory in signal processing and communications," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 17, 132, Sept. 2009.

[25] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, "A survey on networking games in telecommunications," *Comput. Oper. Res.*, vol. 33, no. 2, pp. 286–311, Feb. 2006.

[26] N. Mandayam, S. Wicker, J. Walrand, T. Basar, J. Huang, and D. Palomar, "Game theory in communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1042–1046, Sept. 2008.

[27] G. Bacci, L. Sanguinetti, and M. Luise, "Understanding game theory via wireless power control," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 132–137, July 2015.

[28] E. Björnson, E. Jorswieck, M. Debbah, and B. Ottersten, "Multiobjective signal processing optimization: The way to balance conflicting metrics in 5G systems," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 14–23, Nov. 2014.

[29] J. F. Nash, "Non-cooperative games," *Ann. Math.*, vol. 54, no. 2, pp. 286–295, 1951.

[30] R. J. Aumann, "Acceptable points in general cooperative n-person games," in *Contributions to the Theory of Games*, ser. Annals of Mathematics Studies 40, R. D. Luce and A. W. Tucker, Eds. Princeton, NJ: Princeton Univ. Press, 1959, vol. IV, pp. 287–324.

[31] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.

[32] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. Cambridge, MA: MIT Press, 1994.

[33] R. Mazumdar, L. Mason, and C. Douligeris, "Fairness in network optimal flow control: Optimality of product forms," *IEEE Trans. Commun.*, vol. 39, no. 5, pp. 775–782, May 1991.

[34] C. Touati, E. Altman, and J. Galtier, "Generalized Nash bargaining solution for bandwidth allocation," *Comput. Netw.*, vol. 50, no. 17, pp. 3242–3263, Dec. 2006.

[35] C. H. Papadimitriou, "Algorithms, games, and the Internet," in *Proc. ACM Annual Symp. Theory of Computing*, Heraklion, Greece, July 2001, pp. 749–753.

[36] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, "Selfish routing in capacitated networks," *Math. Operations Res.*, vol. 29, no. 4, pp. 961–976, Nov. 2004.

[37] R. B. Myerson, "Mechanism design," in *The New Palgrave Dictionary of Economics*, 2nd ed., S. N. Durlauf and L. E. Blume, Eds. London, U.K.: Palgrave Macmillan, 2008.

[38] J. F. Nash, "The bargaining problem," *Econometrica*, vol. 18, no. 2, pp. 155–162, Apr. 1950.

[39] H. Boche and M. Schubert, "Nash bargaining and proportional fairness for wireless systems," *IEEE/ACM Trans. Networking*, vol. 17, no. 5, pp. 1453–1466, Oct. 2009.

[40] Z. Liu, Z. Wu, P. Liu, H. Liu, and Y. Wang, "Layer bargaining: multicast layered video over wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 445–455, Apr. 2010.

[41] O. Compte and P. Jehiel, "The coalitional Nash bargaining solution," *Econometrica*, vol. 78, no. 5, pp. 1593–1623, Sept. 2010.

[42] E. Larsson and E. Jorswieck, "Competition versus cooperation on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sept. 2008.

[43] S. Gogineni and A. Nehorai, "Game-theoretic design for polarimetric MIMO radar target detection," *Signal Process.*, vol. 92, no. 5, pp. 1281–1289, May 2012.

[44] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.

[45] E. V. Belmega, S. Lasaulce, and M. Debbah, "Power allocation games for MIMO multiple access channels with coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3182–3192, June 2009.

[46] D. M. Topkis, *Supermodularity and Complementarity*. Princeton, NJ: Princeton Univ. Press, 1998.

[47] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2002.

[48] P. Mertikopoulos, E. V. Belmega, A. Moustakas, and S. Lasaulce, "Distributed learning policies for power allocation in multiple access channels," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 96–106, Jan. 2012.

[49] M. Piezzo, A. Aubry, S. Buzzi, A. D. Maio, and A. Farina, "Non-cooperative code design in radar networks: A game-theoretic approach," *EURASIP J. Adv. Signal Process.*, vol. 2013, p. 63, 2013.

[50] Y. Yang, F. D. Rubio, G. Scutari, and D. P. Palomar, "Multi-portfolio optimization: A potential game approach," *IEEE Trans. Signal Processing*, vol. 61, no. 22, pp. 5590–5602, Nov. 2013.

[51] J. Marden, G. Arslan, and J. Shamma, "Cooperative control and potential games," *IEEE Trans. Syst., Man, Cybernetics, Part B*, vol. 39, no. 6, pp. 1393–1407, Dec. 2009.

[52] J. C. Harsanyi, "Games with incomplete information played by Bayesian players," *Manage. Sci.*, vol. 14, pp. 320–334.

[53] G. Bacci and M. Luise, "A game-theoretic perspective on code synchronization for CDMA wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 107–118, Jan. 2012.

[54] V. Krishnamurthy and H. Poor, "Social learning and Bayesian games in multiagent signal processing: How do local and global decision makers interact?" *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 43–57, May 2013.

[55] D. G. Harper, "Competitive foraging in mallards: 'Ideal free' ducks," *Anim. Behav.*, vol. 30, no. 2, pp. 575–584, May 1982.

[56] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[57] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[58] B. Larrousse, O. Beaude, and S. Lasaulce, "Crawford-sobel meet Lloyd–Max on the grid," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 6127–6131.

[59] D. Niyato and E. Hossain, "Competitive spectrum sharing in cognitive radio networks: A dynamic game approach," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2651–2660, July 2008.

[60] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Select. Areas Commun.*, vol. 13, no. 9, pp. 1341–1347, Sept. 1995.

[61] L. Gan, U. Topcu, and S. H. Low, "Optimal decentralized protocol for electric vehicle charging," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 940–951, May 2013.

[62] G. Scutari, D. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory part I: Nash equilibria," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 1230–1249, Mar. 2008.

[63] W. Yu, G. Ginis, and J. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, June 2002.

[64] L. Rose, S. Lasaulce, S. Perlaza, and M. Debbah, "Learning equilibria with partial information in decentralized wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 136–142, Aug. 2011.

[65] G. W. Brown, "Iterative solutions of games by fictitious play," in *Activity Analysis of Production and Allocation*, T. C. Koopmans, Ed. New York: Wiley, 1951, pp. 374–376.

[66] R. R. Bush and F. Mosteller, *Stochastic Models of Learning*, 1st ed. New York: Wiley, 1955.

[67] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, Sept. 2000.

[68] M. Bennis, S. M. Perlaza, and M. Debbah, "Learning coarse correlated equilibrium in two-tier wireless networks," in *Proc. IEEE Int. Conf. Communications (ICC)*, Ottawa, Canada, June 2012, pp. 1592–1596.

[69] L. Rose, S. Perlaza, and M. Debbah, "On the Nash equilibria in decentralized parallel interference channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, June 2011, pp. 1–6.

[70] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[71] E. Semsar-Kazerooni and K. Khorasani, "Multi-agent team cooperation: A game theory approach," *Automatica*, vol. 45, no. 10, pp. 2205–2213, Oct. 2009.

[72] R. J. Aumann and B. Peleg, "Von Neumann-Morgenstern solutions to cooperative games without side payments," *Bull. Amer. Math. Soc.*, vol. 66, no. 3, pp. 173–179, 1960.

[73] G. Owen, *Game Theory*, 3rd ed. London, UK: Academic Press, 1995.

[74] S. Mathur, L. Sankaranarayanan, and N. Mandayam, "Coalitions in cooperative wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1104–1115, Sept. 2008.

[75] R. La and V. Anantharam, "A game-theoretic look at the Gaussian multiaccess channel," in *Proc. DIMACS Workshop on Network Information Theory*, Piscataway, NJ, Mar. 2003.

[76] R. Mochaourab and E. Jorswieck, "Coalitional games in MISO interference channels: Epsilon-core and coalition structure stable set," *IEEE Trans. Signal Processing*, vol. 62, no. 24, pp. 6507–6520, Dec. 2014.

[77] G. Owen, "Multilinear extensions of games," *Manage. Sci.*, vol. 18, no. 5, pp. 64–79, Jan. 1972.

[78] S. S. Fatima, M. Woodridge, and N. R. Jennings, "A linear approximation method for the Shapley value," *Artificial Intell.*, vol. 172, no. 14, pp. 1673–1699, Sept. 2008.

[79] T. Arnold and U. Schwalbe, "Dynamic coalition formation and the core," *J. Economic Behav. Organization*, vol. 49, no. 3, pp. 363–380, Nov. 2002.

[80] D. Niyato, P. Wang, W. Saad, and A. Hjørungnes, "Controlled coalitional games for cooperative mobile social networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1812–1824, May 2011.

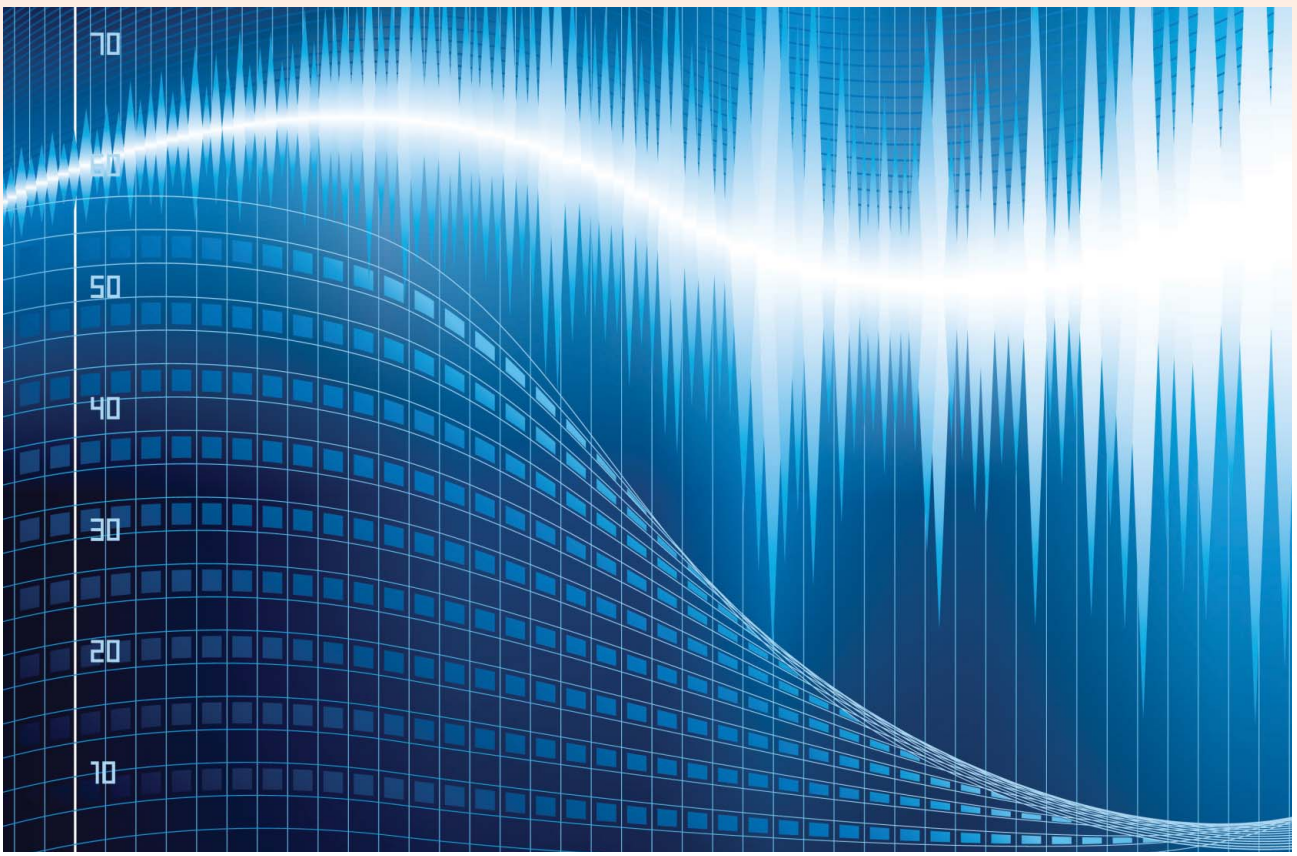
[81] W. Saad, Z. Han, T. Basar, M. Debbah, and A. Hjørungnes, "Coalition formation games for collaborative spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 276–297, Jan. 2011.

[82] T. Wang, L. Song, Z. Han, and W. Saad, "Distributed cooperative sensing in cognitive radio networks: An overlapping coalition formation approach," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3144–3160, Sept. 2014.

[SP]

Combinations of Adaptive Filters

Performance and convergence properties



©ISTOCKPHOTO.COM/POLYGRAPHUS

Jerónimo Arenas-García, Luis A. Azpicueta-Ruiz, Magno T.M. Silva, Vítor H. Nascimento, and Ali H. Sayed

Adaptive filters are at the core of many signal processing applications, ranging from acoustic noise supression to echo cancelation [1], array beamforming [2], channel equalization [3], to more recent sensor network applications in surveillance, target localization, and tracking. A trending approach in this direction is to recur to in-network distributed processing in which individual nodes implement adaptation rules and diffuse their estimation to the network [4], [5].

Digital Object Identifier 10.1109/MSP.2015.2481746
Date of publication: 29 December 2015

1053-5888/16©2016IEEE

Ranging from the simple least mean squares (LMS) to sophisticated state-space algorithms, significant research has been carried out over the last 50 years to develop effective adaptive algorithms, in an attempt to improve their general properties in terms of convergence, tracking ability, steady-state misadjustment, robustness, or computational cost [6]. Many design procedures and theoretical models have been developed, and many novel adaptive structures are continually proposed with the objective of improving filter behavior with respect to well-known performance tradeoffs (such as convergence rate versus steady-state performance), or incorporating available a priori knowledge into the learning mechanisms of the filters (e.g., to enforce sparsity).

When the a priori knowledge about the filtering scenario is limited or imprecise, selecting the most adequate filter structure and adjusting its parameters becomes a challenging task, and erroneous choices can lead to inadequate performance. To address this difficulty, one useful approach is to rely on combinations of adaptive structures. Combinations of adaptive schemes have been studied in several works [7]–[21] and have been exploited in a variety of applications such as the characterization of signal modality [22], acoustic echo cancelation [23]–[26], adaptive line enhancement [27], array beamforming [28], [29], and active noise control [30], [31].

The combination of adaptive filters exploits to some extent the same divide and conquer principle that has also been successfully exploited by the machine-learning community (e.g., in bagging or boosting [32]). In particular, the problem of combining the outputs of several learning algorithms (mixture of experts) has been studied in the computational learning field under a different perspective: rather than studying the expected performance of the mixture, deterministic bounds are derived that apply to individual sequences and, therefore, reflect worst-case scenarios [33]–[35]. These bounds require assumptions different from the ones typically used in adaptive filtering, which is the emphasis of this overview article. We review the key ideas and principles behind these combination schemes, with emphasis on design rules. We also illustrate their performance with a variety of examples.

PROBLEM FORMULATION AND NOTATION

We generally assume the availability of a reference signal and an input regressor vector, $d(n)$ and $\mathbf{u}(n)$, respectively, which satisfy a linear regression model of the form

$$d(n) = \mathbf{u}^{\top}(n)\mathbf{w}_o(n) + v(n), \tag{1}$$

where $\mathbf{w}_o(n)$ represents the (possibly) time-varying optimal solution, and $v(n)$ is a noise sequence, which is considered independent and identically distributed, and independent of $\mathbf{u}(m)$, for all n, m. In this article, we will mostly restrict ourselves to the case in which all involved variables are real, although the extension to the complex case is straightforward, and has been analyzed in other works [6]. To estimate the optimal solution at time n, adaptive filters typically implement a recursion of the form

$$\mathbf{w}(n+1) = \mathbf{f}[\mathbf{w}(n), d(n), \mathbf{u}(n), \mathbf{s}(n)],$$

where different adaptive schemes are characterized by their update functions $\mathbf{f}[\cdot]$, and $\mathbf{s}(n)$ represents any other state information that is needed for the update of the filter.

We define the following error variables that are commonly used to characterize the performance of adaptive filters [6]:

- Weight error: $\tilde{\mathbf{w}}(n) = \mathbf{w}_o(n) - \mathbf{w}(n)$
- A priori filter error: $e_a(n) = \mathbf{u}^{\top}(n)\,\tilde{\mathbf{w}}(n)$
- Filter error: $e(n) = d(n) - \mathbf{u}^{\top}(n)\mathbf{w}(n) = e_a(n) + v(n)$
- Mean square error (MSE): $\mathrm{MSE}(n) = \mathbb{E}\{e^2(n)\}$
- Excess MSE (EMSE): $\zeta(n) = \mathbb{E}\{e_a^2(n)\} = \mathrm{MSE}(n) - \mathbb{E}\{v^2(n)\}$
- Mean square deviation: $\mathrm{MSD}(n) = \mathbb{E}\{\|\tilde{\mathbf{w}}(n)\|^2\}$.

During their operation, adaptive filters normally go from a convergence phase, where the MSE decreases, to a steady-state regime in which the MSE tends toward some asymptotic value. Thus, for steady-state performance, we also define the steady-state MSE, EMSE, and MSD as their limiting values as n increases. For instance, the steady-state EMSE is defined as

$$\zeta(\infty) = \lim_{n\to\infty} \mathbb{E}\{e_a^2(n)\}.$$

A BASIC COMBINATION OF TWO ADAPTIVE FILTERS

The most simple combination scheme incorporates two adaptive filters. Figure 1 illustrates the configuration, which shows that combination schemes have two concurrent adaptive layers: adaptation of individual filters and adaptation of the combination layer. As illustrated in the figure, both adaptive filters have access to the same input and reference signals and produce their individual estimates of the optimum weight vector, $\mathbf{w}_o(n)$. The goal of the combination layer is to learn which filter component is performing better dynamically at any particular time, assigning them weights to optimize the overall performance.

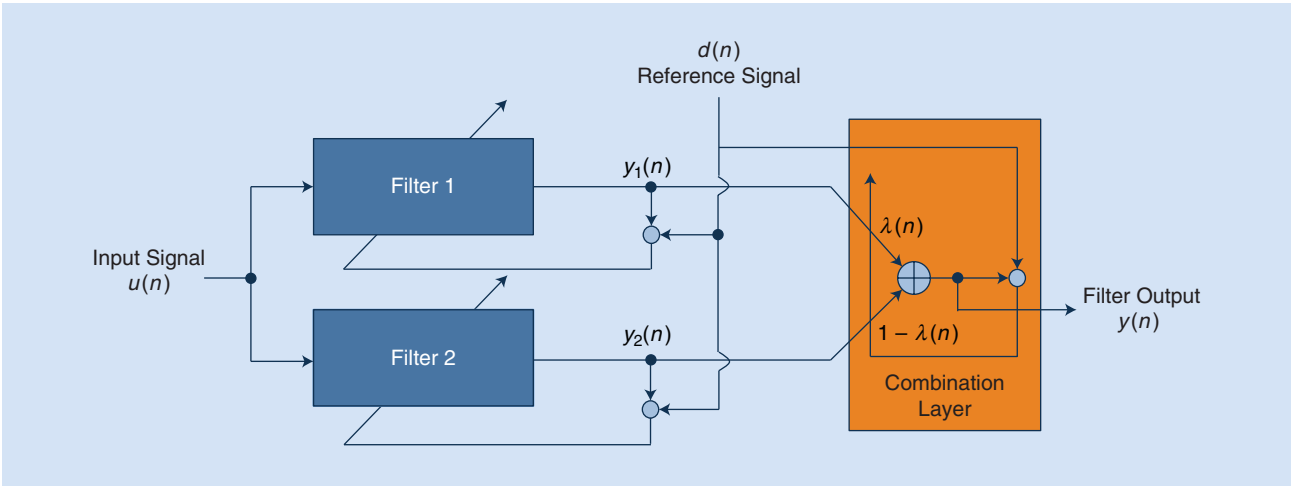
In this article, we focus on affine and convex combinations, where the overall filter output is given by

$$y(n) = \lambda(n)y_1(n) + [1-\lambda(n)]y_2(n), \tag{2}$$

where $y_i(n) = \mathbf{u}^{\top}(n)\mathbf{w}_i(n)$, $i = 1,2$, are the outputs of the two adaptive filters characterized by weights $\mathbf{w}_i(n)$, and $\lambda(n)$ is a mixing parameter. Similarly, the estimated weight vector, the error, and the a priori error of the combination scheme are given by

$$\mathbf{w}(n) = \lambda(n)\mathbf{w}_1(n) + [1-\lambda(n)]\mathbf{w}_2(n), \tag{3}$$
$$e(n) = \lambda(n)e_1(n) + [1-\lambda(n)]e_2(n), \tag{4}$$
$$e_a(n) = \lambda(n)e_{a,1}(n) + [1-\lambda(n)]e_{a,2}(n), \tag{5}$$

where $e_i(n) = d(n) - y_i(n)$, $i = 1, 2$, are the errors of the two adaptive filter components, and $e_{a,i}(n)$ their a priori errors. The combination parameter $\lambda(n)$ is constrained to lie in the interval $[0, 1]$ for convex combinations, while it can take any real value for



[FIG1] **The basic scheme for the combination of two adaptive filters.**

affine combinations. There are also combinations where the sum of weights assigned to the component filters is not equal to one [10]. These unconstrained combinations will not be addressed in this article. To obtain an enhanced behavior from the combination, it is crucial to design mechanisms to learn the mixing parameter $\lambda(n)$ and to adapt it to track drifts in the model. Different algorithms for the adaptation of the combination have been proposed in the literature; some of them will be reviewed in the section "Estimating the Combination Parameter."

In (3), we are implicitly assuming component filters with the same length. When this is not the case, we can still use (3) by extending the shortest filter and filling with zeros the nonexistent coefficients. Using this observation, most of the algorithms discussed in this article continue to be valid if the component filters have different lengths. Moreover, in practice, construction (3) does not need to be evaluated since only the output combination (2) is necessary. Furthermore, although not required by the analysis, in the experiments we will assume that the component filter weight vectors match the length of the optimum solution, $\mathbf{w}_o(n)$. Nevertheless, there have been useful studies in the literature where the length of the filters is treated as a design variable and the adaptation rules are used to learn the length of the optimum solution as well to avoid degradation due to under- or overmodeling [9], [13], [19].

The selection of the individual filters is often influenced by the application itself. For example, combination structures can be used to:

- Facilitate the selection of filter parameters (such as the step size or forgetting factor, regularization constants, filter length, projection order, etc). Existing schemes fix the values of these parameters or recur to complex schemes that learn them over time. Alternatively, we can consider a combination structure where the two filters belong to the same family of algorithms but differ in the values of the filter parameters. The combination layer would then select and give more weight to the best filter at every time instant, making parameter selection less critical.
- Increase robustness against unknown environmental conditions. A common objective in the design of adaptive filters consists in the incorporation of a priori knowledge about the filtering scenario to improve the performance of the filter. For instance, sparse-aware adaptive filters try to benefit from the knowledge that many coefficients of the optimal solution are (close to) zero, and Volterra filters (VFs) try to model nonlinearities with polynomial kernels. However, when the filtering conditions are not known accurately, or change over time, these schemes can show very suboptimal performance. Combination schemes can be exploited to increase the robustness of these techniques against lack of or imprecise knowledge of the adaptive filtering scenario.
- Provide diversity that can be exploited to enhance performance beyond the capabilities of an individual adaptive filter. An example will be given of the tracking abilities of a combination of one LMS and one recursive least-squares (RLS) filter in the section "Improving the Tracking Performance of LMS and RLS Filters."

The main disadvantage of this approach is the increased computational cost required for the adaptation of the two adaptive filters. However, it should be noted that the filters can be adapted in parallel and some strategies allow for the joint adaptation of the two filters with just a slight increment with respect to that of a single adaptive filter. Regarding the adaptation of the mixing parameter, the computational requirement of most available schemes is not significant.

In the next section, we review the theoretical limits of the combination scheme consisting of just two filters. We then review different combination rules that have already appeared in the literature, and compare them in a variety of simulated conditions.

OPTIMUM MIXING PARAMETER AND COMBINATION PERFORMANCE

Here we derive the expression for the optimal mixing parameter in the affine and convex cases, in the sense of minimizing the MSE of the combination. Since for $\lambda(n) = 0$ and $\lambda(n) = 1$

convex and affine combinations are equivalent to each of the component filters, we know that an optimal selection of the mixing parameter would guarantee that these combinations perform at least as the best component filter. Here we examine the behavior of the combination scheme for other nontrivial choices of the mixing parameter.

The EMSE of the combination (2) is given by

$$\zeta(n) = \mathbb{E}\{e_a^2(n)\} = \lambda^2(n)\zeta_1(n) + [1-\lambda(n)]^2\zeta_2(n) + 2\lambda(n)[1-\lambda(n)]\zeta_{12}(n), \quad (6)$$

where $\zeta_i(n)$, $i = 1, 2$, are the EMSEs of the two individual filters, and $\zeta_{12}(n)$ denotes the cross-EMSE defined as [12]

$$\zeta_{12}(n) = \mathbb{E}\{e_{a,1}(n)e_{a,2}(n)\}.$$

This cross-EMSE variable is an important measure and is closely related to the ability of the combination to improve the performance of both filter components, as we will see shortly.

An important property of the cross-EMSE that will be useful when discussing the properties of the optimal combination is derived from the Cauchy-Schwartz inequality

$$|\zeta_{12}(n)|^2 \leq \zeta_1(n)\zeta_2(n), \quad (7)$$

which implies that the cross-EMSE can never be simultaneously larger (in absolute terms) than both individual EMSEs.

Equation (6) reveals a quadratic dependence on the mixing factor, $\lambda(n)$. It follows that, in the affine case, the optimal choice for $\lambda(n)$ is given by

$$\lambda_{\text{aff}}^o(n) = \frac{\zeta_2(n) - \zeta_{12}(n)}{\zeta_1(n) + \zeta_2(n) - 2\zeta_{12}(n)} = \frac{\Delta\zeta_2(n)}{\Delta\zeta_1(n) + \Delta\zeta_2(n)}, \quad (8)$$

where we have defined $\Delta\zeta_i(n) = \zeta_i(n) - \zeta_{12}(n)$, $i = 1, 2$. In the case of a convex combination, the minimization of (6) needs to be carried out by constraining $\lambda(n)$ to lie in the interval $[0, 1]$. Given that (6) is nonnegative and quadratic in $\lambda(n)$, the optimum mixing parameter is either given by (8) or lies in the limits of the considered interval, i.e.,

$$\lambda_{\text{cvx}}^o(n) = \frac{\Delta\zeta_2(n)}{\Delta\zeta_1(n) + \Delta\zeta_2(n)}\Bigg|_0^1, \quad (9)$$

where the vertical line denotes truncation to the indicated values.

Substituting either (8) or (9) into (6), we find that the optimum EMSE of the filter is given by

$$\zeta^o(n) = \zeta_1(n) - [1-\lambda^o(n)]\Delta\zeta_1(n) \quad (10a)$$
$$= \zeta_2(n) - \lambda^o(n)\Delta\zeta_2(n), \quad (10b)$$

where $\lambda^o(n)$ stands for either $\lambda_{\text{aff}}^o(n)$ or $\lambda_{\text{cvx}}^o(n)$. Note that these expressions are valid for any time instant n, and also for the steady-state performance of the filter (i.e., for $n \to \infty$).

The analysis of the expressions for the optimum mixing parameter and for the combination EMSE allows us to conclude that, depending on the cross-EMSE value, the combination can be performing in one of four possible conditions at every time instant. Taking into account that the denominator of (8) is nonnegative, $\Delta\zeta_1(n) + \Delta\zeta_2(n) = \mathbb{E}\{[e_{a,1}(n) - e_{a,2}(n)]^2\}$, Table 1 summarizes the main properties of these four possible cases (the time index is omitted in the first column for the sake of compactness).

From the information included in the table we can deduce the following properties:

- In cases 1 and 2, the convexity constraints are active. As a consequence, the optimum mixing parameter of the convex combination is either one or zero, respectively, making this scheme behave just like the best filter component. However, an affine combination can outperform both components in this case. To see this, consider (10a) and (10b) for cases 1 and 2, respectively, and note that a positive value is subtracted from the smallest filter EMSE (ζ_1 and ζ_2) to obtain the EMSE of the combination.
- In case 3, affine and convex combinations share the same optimum mixing parameter. As in the previous case, using (10a) and (10b) we can easily conclude that, in this case, combinations outperform both component filters simultaneously. This performance can be explained from the small cross-EMSE in this case: since the correlation between the a priori errors of both component filters is small, their weighted combination provides an estimation error of reduced variance.
- For completeness, we have included a fourth case where $\zeta_1 = \zeta_2 = \zeta_{12}$. Simplifying all terms that involve $\lambda(n)$ in (6), we can conclude that, in this case, we have $\zeta^o(n) = \zeta_1(n) = \zeta_2(n)$, irrespectively of the value of the mixing parameter. It is important to remark that for case 4 to hold, condition $\zeta_1 = \zeta_2 = \zeta_{12}$ should be satisfied exactly, so that this case will be rarely encountered in practice.

We should emphasize that the results in this section refer to the theoretical performance of optimally adjusted affine and convex combinations. Practical algorithms to adjust the mixing parameter will be presented in the section "Estimating the Combination Parameter" and will imply the addition of a certain amount of noise that, in many cases, justify the adoption of convexity constraints.

CROSS-EMSE OF LMS AND RLS FILTERS

The cross-EMSE plays an important role in clarifying the performance regime of a combination of two adaptive filters. Several

[TABLE 1] POSSIBLE OPERATION REGIMES FOR THE AFFINE AND CONVEX COMBINATION OF TWO ADAPTIVE FILTERS.

	$\zeta_{12}(n)$	$\Delta\zeta_1(n)$	$\Delta\zeta_2(n)$	$\lambda_{\text{aff}}^o(n)$	$\lambda_{\text{cvx}}^o(n)$
CASE 1:	$\zeta_1 \leq \zeta_{12}$ $(< \sqrt{\zeta_1\zeta_2} < \zeta_2)$	≤ 0	> 0	≥ 1	$= 1$
CASE 2:	$\zeta_2 \leq \zeta_{12}$ $(< \sqrt{\zeta_1\zeta_2} < \zeta_1)$	> 0	≤ 0	≤ 0	$= 0$
CASE 3:	$(-\sqrt{\zeta_1\zeta_2} <)\ \zeta_{12} < \min\{\zeta_1, \zeta_2\}$	> 0	> 0	$\in (0, 1)$	$\in (0, 1)$
CASE 4:	$\zeta_1 = \zeta_2 = \zeta_{12}$	$= 0$	$= 0$	—	—

[TABLE 2] STEADY-STATE CROSS-EMSE FOR THE LMS AND RLS ADAPTIVE FILTER FAMILIES.

COMBINATION TYPE	STEADY-STATE CROSS-EMSE, $\zeta_{12}(\infty)$
μ_1-LMS AND μ_2-LMS	$\frac{\mu_1\mu_2\sigma_v^2\mathrm{Tr}\{\mathbf{R}\}+\mathrm{Tr}\{\mathbf{Q}\}}{\mu_1+\mu_2}$
β_1-RLS AND β_2-RLS	$\frac{\beta_1\beta_2\sigma_v^2 M+\mathrm{Tr}\{\mathbf{QR}\}}{\beta_1+\beta_2}$
μ-LMS AND β-RLS	$\mu\beta\sigma_v^2\mathrm{Tr}\{(\mu\mathbf{R}+\beta\mathbf{I})^{-1}\mathbf{R}\}+\mathrm{Tr}\{\mathbf{Q}(\mu\mathbf{R}+\beta\mathbf{I})^{-1}\mathbf{R}\}$

works have extended the available theoretical models for the EMSE of adaptive filters to the derivation of expressions for the cross-EMSE [6], [12], [15], [16]. In this section, we review some of the main results for LMS and RLS filters.

Assume that the optimal solution satisfies the following random-walk model

$$\mathbf{w}_o(n) = \mathbf{w}_o(n-1) + \mathbf{q}(n), \tag{11}$$

where $\mathbf{q}(n)$ is the change of the solution at every step that is assumed to have zero-mean and covariance $\mathbf{Q} = \mathbb{E}\{\mathbf{q}(n)\mathbf{q}^\top(n)\}$. Although this simple model implies divergence of the optimum solution, it gives qualitatively the same results as more accurate models; see [6] for further explanations on this issue. It is known [6] that in this case there is a tradeoff between filter performance and tracking ability. Table 2 shows an approximation of the steady-state EMSE of LMS and RLS adaptive filters derived using the energy conservation approach described in [6]. From these expressions, it is straightforward to obtain the optimum step size for the LMS adaptation rule (μ_o) and the optimal forgetting factor for the RLS adaptive filter with exponentially decaying memory ($1-\beta_o$). These expressions are also indicated in the table, together with the associated optimum EMSE for each filter family, $\zeta_o(\infty)$.

To study the behavior of combinations of this kind of filters, [14] and [15] derived expressions for the steady-state cross-EMSE, for cases in which the two component filters belong to the same family (LMS or RLS) and also for the combination of one LMS and one RLS filter. These results are summarized in Table 2, and constitute the basis for the study that we carry out in the next two sections.

COMBINATION OF TWO LMS FILTERS WITH DIFFERENT STEP SIZES

A first example to illustrate the possibilities of combination schemes consists in putting together two filters with different adaptation speeds to obtain a new filter with improved tracking capabilities. Here, we illustrate the idea with a combination of two LMS filters with different step sizes, as was done in [12], but the idea can be straightforwardly generalized to other adaptive filter families (see, among others, [15], [19], [21], [23], [25], and [30]).

Using the tracking model given by (11), in this section we analyze the theoretical steady-state EMSE for varying $\mathrm{Tr}\{\mathbf{Q}\}$. In the studied scenario, the product $\sigma_v^2\mathrm{Tr}\{\mathbf{R}\}$ remains fixed and is equal to 10^{-2}. To facilitate the comparison among filters, we will recur to the normalized squared deviation (NSD), which we define as the EMSE excess with respect to the optimum LMS filter for a particular $\mathrm{Tr}\{\mathbf{Q}\}$, i.e.,

$$\mathrm{NSD}_i(\infty) = \frac{\zeta_i(\infty)}{\zeta_o(\infty)},\ i = 1, 2,$$

and

$$\mathrm{NSD}(\infty) = \frac{\zeta(\infty)}{\zeta_o(\infty)},$$

for the component LMS filters and their combination, respectively.

Figure 2 illustrates the theoretical NSD of individual LMS filters and combined schemes in two cases. Figure 2(a) considers LMS components with step sizes $\mu_1 = 0.1$ and $\mu_2 = \mu_1/20$. We can see that both individual filters become optimal (NSD = 0 dB) for a certain value of $\mathrm{Tr}\{\mathbf{Q}\}$, and very rapidly increase their NSD level both for larger and smaller $\mathrm{Tr}\{\mathbf{Q}\}$, i.e., for solutions that change relatively faster or slower than the speed of changes for which the filter is optimum. We can see that the NSD of each individual filter remains smaller than 2 dB for a range of about two orders of magnitude of $\mathrm{Tr}\{\mathbf{Q}\}$. In contrast to this, we can check that the NSD of the combination of the two filters remains smaller than 2 dB when $\mathrm{Tr}\{\mathbf{Q}\}$ changes by more than four orders of magnitude. It should be mentioned that, in this case, both the affine and convex combinations provide almost identical behavior.

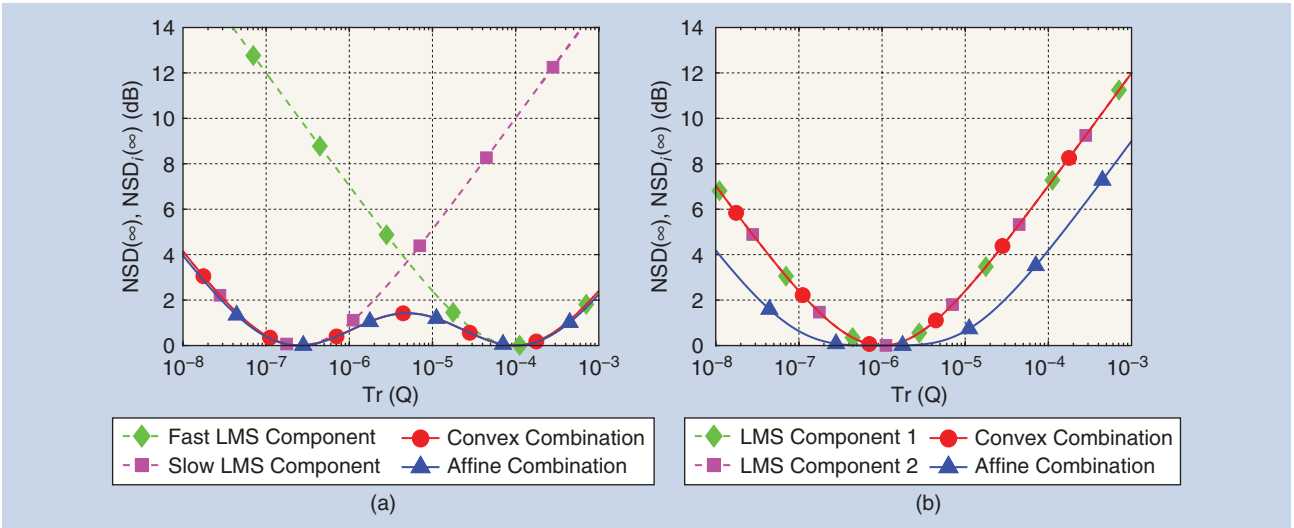
A second possibility is illustrated in Figure 2(b), for which the step sizes of the two LMS filters are set at almost identical values ($\mu_1 = 0.01$ and $\mu_2 = 0.010001$). In this case, both LMS components obtain approximately the same NSD for all values of $\mathrm{Tr}\{\mathbf{Q}\}$. According to the theoretical results, it can be shown that, in this example, cases 1 and 2 enumerated in Table 1 are observed when $\mathrm{Tr}(\mathbf{Q})$ is, respectively, smaller or larger than approximately 10^{-6}. The limitation in the range of $\lambda(n)$ makes the convex combination perform exactly as any of the components, whereas the affine combination achieves an almost systematic gain of 3 dB with respect to the single filters [18], with the NSD remaining at values of less than 2 dB when $\mathrm{Tr}\{\mathbf{Q}\}$ is changed by three orders of magnitude.

The examples in this section illustrate the following points:

- By employing a combination structure, we can improve the tracking ability of the adaptive implementation. This is important because the actual speed of change of the solution is rarely known a priori, and can even change over time. We compare the combination framework with variable step-size (VSS) schemes in the section "Combination Schemes Versus VSS LMS," and show that combinations are more robust than VSS.
- When combining two LMS filters with different step sizes, the resulting combination cannot outperform the optimum filter for a given $\mathrm{Tr}\{\mathbf{Q}\}$ [15]. Therefore, in this situation, if enough statistical information is a priori available, it would be preferable to use a single LMS filter with optimum step size. A different situation will be described next.

IMPROVING THE TRACKING PERFORMANCE OF LMS AND RLS FILTERS

When studying a combination of LMS and RLS filters following the same approach, an interesting result is obtained. Assume the



[FIG2] The steady-state NSD of a combination of two adaptive filters with different step sizes as a function of the speed of changes of the optimum solution. The figure illustrates the NSD incurred by the individual LMS filters and by their optimum combination. LMS step sizes are: (a) $\mu_1 = 0.1$ and $\mu_2 = 0.005$; (b) $\mu_1 = 0.01$ and $\mu_2 = 0.010001$.

tracking model (11) and consider the optimum LMS and RLS filters with adaptation parameters given by the expressions in Table 3. As explained in [6], LMS will outperform RLS if $\mathbf{Q}$ is proportional to the autocorrelation matrix of the input signal, $\mathbf{R}$, and the opposite will occur when $\mathbf{Q} \propto \mathbf{R}^{-1}$. To illustrate this behavior, let us consider a synthetic example where $\mathbf{Q}$ is a mixture of $\mathbf{R}$ and $\mathbf{R}^{-1}$:

$$\mathbf{Q} = 10^{-5}\left[\alpha \frac{\mathbf{R}}{\mathrm{Tr}\{\mathbf{R}\}} + (1-\alpha)\frac{\mathbf{R}^{-1}}{\mathrm{Tr}\{\mathbf{R}^{-1}\}}\right], \tag{12}$$

with $\alpha \in (0, 1)$.

Figure 3 plots the steady-state optimum EMSE that can be achieved by both types of filters with optimum settings (μ_o-LMS and β_o-RLS), when $\mathbf{Q}$ smoothly changes between $\mathbf{R}^{-1}$ (for $\alpha = 0$) and $\mathbf{R}$ (for $\alpha = 1$). Other settings for this scenario are: optimal solution and filters length $M = 7$, noise variance $\sigma_v^2 = 10^{-2}$, and $\mathbf{R}$ is a Toeplitz matrix whose first row is given by $(1/7)[1, 0.8, 0.8^2, \cdots, 0.8^6]$. According to (12), the value of the trace of $\mathbf{Q}$ remains constant when changing the value of α, meaning that the EMSE of the optimum LMS is approximately equal to -35 dB for any value of α. If a nonoptimum step size is selected, the LMS filter will incur a larger EMSE. Thus, the region of feasible EMSE values for this family of filters is given by the area above the -35 dB level in Figure 3. Similarly, the region above the curve for the optimum RLS performance represents all feasible EMSE values for the family of RLS filters. We can see that, depending on the value of α, the optimal EMSE for RLS filters can be larger or smaller than the optimal EMSE for LMS filters. Finally, the dark green area in Figure 3 is a feasible region for both LMS and RLS filters.

The performance that can be achieved using a combination of one LMS and one RLS filters is also illustrated. The red curve represents the steady-state tracking EMSE of the combination when using μ_o and β_o for the LMS and RLS filters in the combination. As expected, the optimal EMSE for the combination is never larger than the smallest of the filter component EMSEs. More importantly, in this case we can check the existence of an area of EMSE values which is exclusive to the combination scheme. In other words, the red area in Figure 3 is a feasible region for the operation of the combined scheme, but not for single LMS or RLS filters. It can be shown that, in this case, the optimum mixing parameter of the combination lies in interval (0, 1), so this conclusion is valid for convex combination schemes [36]. This is a useful conclusion that can be used to improve the tracking performance beyond the limits of individual filters.

ESTIMATING THE COMBINATION PARAMETER

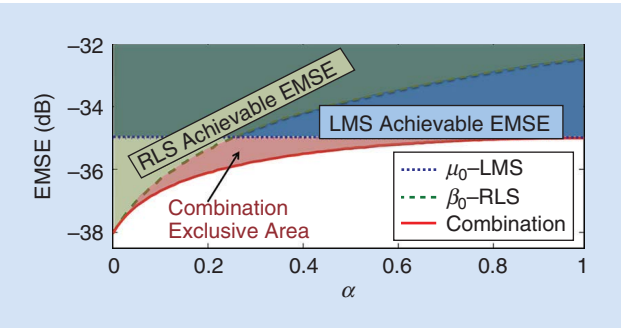
Since the optimum linear combiner (8) is unrealizable, many practical algorithms have been proposed in the literature to adjust the mixing parameter in convex [12], [37], [38] and affine combinations [16]–[18], [39], [40]. Rewriting (2) as

$$y(n) = y_2(n) + \lambda(n)[y_1(n) - y_2(n)], \tag{13}$$

[TABLE 3] TRACKING THE STEADY-STATE PERFORMANCE OF LMS AND RLS ADAPTIVE FILTERS. THE TABLE SHOWS STEADY-STATE EMSE AS A FUNCTION OF THE LMS STEP SIZE (μ) AND RLS FORGETTING FACTOR ($1-\beta$), THE OPTIMAL VALUES OF THESE PARAMETERS, AND THE MINIMUM EMSE THAT CAN BE ACHIEVED BY BOTH ADAPTIVE FILTERS [$\zeta_o(\infty)$].

ALGORITHM	$\zeta(\infty)$	μ_o OR β_o	$\zeta_o(\infty)$
LMS	$\frac{1}{2}[\mu\sigma_v^2 \mathrm{Tr}\{\mathbf{R}\} + \mu^{-1}\mathrm{Tr}\{\mathbf{Q}\}]$	$\sqrt{\frac{\mathrm{Tr}\{\mathbf{Q}\}}{\sigma_v^2 \mathrm{Tr}\{\mathbf{R}\}}}$	$\sqrt{\sigma_v^2 \mathrm{Tr}\{\mathbf{R}\}\mathrm{Tr}\{\mathbf{Q}\}}$
RLS	$\frac{1}{2}[\beta\sigma_v^2 M + \beta^{-1}\mathrm{Tr}\{\mathbf{QR}\}]$	$\sqrt{\frac{\mathrm{Tr}\{\mathbf{QR}\}}{\sigma_v^2 M}}$	$\sqrt{\sigma_v^2 M \mathrm{Tr}\{\mathbf{QR}\}}$

$\sigma_v^2 = \mathbb{E}\{v^2(n)\}$; $\mathbf{Q} = \mathbb{E}\{\mathbf{q}(n)\mathbf{q}^T(n)\}$; $\mathbf{R} = \mathbb{E}\{\mathbf{u}(n)\mathbf{u}^T(n)\}$; M IS THE FILTER LENGTH



[FIG3] The tracking performance of a combination of LMS and RLS filters when Q smoothly changes between $\mathbf{R}^{-1}$ (for $\alpha = 0$) and R (for $\alpha = 1$). The performance of optimally adjusted filters is depicted with curves. The blue and light green regions represent EMSEs that can be obtained with LMS and RLS filters, respectively, whereas the dark green area comprises feasible EMSE values for both LMS and RLS filters. Finally, the red region contains EMSE values that can be obtained with combinations of LMS and RLS (but not with these filters individually).

we can reinterpret the adaptation of $\lambda(n)$ as a "second layer" adaptive filter of length one, so that in principle any adaptive rule can be used for adjusting the mixing parameter. However, this filtering problem has some particularities, specifically, the strong and time-varying correlation between $y_1(n)$ and $y_2(n)$. This implies that the power of the difference signal, $[y_1(n) - y_2(n)]$, is also time-varying depending, e.g., on the signal-to-noise conditions and on whether the individual filters are operating in convergence phase or steady state, etc.

Using a stochastic gradient search to minimize the quadratic error $e^2(n)$ of the overall filter, defined in (4), [16] proposed the following adaptation for the mixing parameter in an affine combination (i.e., without imposing restrictions on $\lambda(n)$),

$$\begin{aligned}\lambda(n+1) &= \lambda(n) - \frac{\mu_\lambda}{2}\frac{\partial e^2(n)}{\partial \lambda(n)} \\ &= \lambda(n) + \mu_\lambda e(n)[y_1(n) - y_2(n)]. \end{aligned} \tag{14}$$

We will refer to this case as *aff-LMS* adaptation, with μ_λ being a step-size parameter. As discussed in [16], a large step size should be used in this rule to ensure an adequate behavior of the affine combination. However, this can cause instability during the initial convergence of the algorithm, which was circumvented in [16] by constraining $\lambda(n)$ to be less than or equal to one. Even with this constraint, the combination scheme may stay away from the optimum EMSE (see, e.g., [18]). This is a direct consequence of the time-varying power of $[y_1(n) - y_2(n)]$, which makes the selection of μ_λ a difficult task.

To obtain a more robust scheme, it is possible to recur to normalized adaptation schemes. Using a rough (low-pass filtered) estimation of the power of $[y_1(n) - y_2(n)]$, a power normalized version of aff-LMS was proposed in [18], and is summarized in Table 4. This aff-PN-LMS algorithm does not impose any constraints on $\lambda(n)$, is less sensitive to the filtering scenario, and converges in the mean-square sense if the step size is selected in the interval $0 < \mu_\lambda < 2$.

Rather than directly adjust $\lambda(n)$ as in the affine case, convex combination schemes recur to activation functions to keep the mixing parameter in the range of interest. For example, [12] proposed an adaptation scheme for an auxiliary parameter $a(n)$ that is related to $\lambda(n)$ via the sigmoid function

$$\lambda(n) = \mathrm{sgm}[a(n)] = \frac{1}{1 + e^{-a(n)}}. \tag{15}$$

Recurring to this activation function (or similar ones), $a(n)$ can be adapted without constraints, and $\lambda(n)$ will be automatically kept inside the interval $(0, 1)$ at all times. Using a gradient descent method to minimize the quadratic error of the overall filter, $e^2(n)$, two algorithms were proposed to update $a(n)$: the cvx-LMS algorithm [12] and its power normalized version, cvx-PN-LMS [37], whose update equations are given by

$$\begin{aligned} a(n+1) &= a(n) + \mu_a \lambda(n) \\ &\quad [1 - \lambda(n)]e(n)[y_1(n) - y_2(n)], \qquad \text{(cvx-LMS)} \end{aligned} \tag{16}$$

$$\begin{aligned} a(n+1) &= a(n) + \frac{\mu_a}{p(n)}\lambda(n) \\ &\quad [1 - \lambda(n)]e(n)[y_1(n) - y_2(n)]. \qquad \text{(cvx-PN-LMS)} \end{aligned} \tag{17}$$

Here, $p(n)$ is a low-pass filtered estimation of the power of $[y_1(n) - y_2(n)]$. These algorithms are also shown in Table 4 for further reference. Compared to the cvx-LMS scheme, cvx-PN-LMS is more robust to signal-to-noise ratio (SNR) changes, and simplifies the selection of step size μ_a [37]. [According to the linear regression model (1), the SNR is defined as $\mathrm{SNR} = [\mathbf{w}_o^\top(n)\mathbf{R}\mathbf{w}_o(n)]/\sigma_v^2$].

The factor $\lambda(n)[1 - \lambda(n)]$, that appears in (17), reduces the gradient noise when $\lambda(n)$ gets too close to the values of zero or one. In this situation, adaptation would virtually stop if $a(n)$ were allowed to grow unchecked. To avoid this problem, the auxiliary parameter $a(n)$ is restricted by saturation to a symmetric interval $[-a^+, a^+]$, which ensures a minimum level of adaptation [12], [39]. A common choice in the literature is $a^+ = 4$. This truncation procedure constrains $\lambda(n)$ to the interval $\mathrm{sgm}[-a^+] \le \lambda(n) \le \mathrm{sgm}[a^+]$. To allow $\lambda(n)$ to take a value equal to zero or one, a scaled and shifted version of the sigmoid was proposed in [38],

$$\lambda(n) = \varphi[a(n)] = \frac{\mathrm{sgm}[a(n)] - \mathrm{sgm}[-a^+]}{\mathrm{sgm}[a^+] - \mathrm{sgm}[-a^+]}, \tag{18}$$

[TABLE 4] THE POWER NORMALIZED ADAPTATION OF $\lambda(n)$ IN AFFINE COMBINATIONS AND OF AUXILIARY PARAMETER $a(n)$ IN CONVEX COMBINATIONS. UNNORMALIZED RULES aff-LMS AND cvx-LMS ARE OBTAINED SETTING $p(n) = 1$.

ALGORITHM	UPDATE EQUATIONS
aff-PN-LMS [18]	$\lambda(n+1) = \lambda(n) + \frac{\mu_\lambda}{\varepsilon + p(n)}e(n)[y_1(n) - y_2(n)]$
	$p(n) = \eta\, p(n-1) + (1-\eta)[y_1(n) - y_2(n)]^2$
	$0 \ll \eta < 1;\ \varepsilon > 0$ IS A SMALL CONSTANT
cvx-PN-LMS [37]	$a(n+1) = a(n) + \frac{\mu_a}{p(n)}\lambda(n)[1 - \lambda(n)]e(n)[y_1(n) - y_2(n)]$
	$p(n) = \eta\, p(n-1) + (1-\eta)[y_1(n) - y_2(n)]^2$
	$0 \ll \eta < 1$; CONSTRAINT: $-a^+ \le a(n+1) \le a^+$

which ensures that $\lambda(n)$ attains values 1 and 0 for $a(n) = a^+$ and $a(n) = -a^+$, respectively. Using (18), the cvx-PN-LMS rule becomes

$$a(n+1) = a(n) + \frac{\mu_a}{p(n)} e(n)[y_1(n) - y_2(n)] \operatorname{sgm}[a(n)]\{1 - \operatorname{sgm}[a(n)]\}\Big|_{-a^+}^{a^+}. \quad (19)$$

As we can see, $a(n)$ is truncated at every iteration to keep it inside the interval $[-a^+, a^+]$.

When deciding between affine or convex combinations with the above gradient-based rules, one should be aware of the following facts:

- Optimally adjusted affine combinations attain smaller EMSE than convex ones in some situations (cases 1 and 2 of Table 1).
- Due to the constraints on the value of $\lambda(n)$, cvx-LMS and cvx-PN-LMS do not diverge, and large values of the step size μ_a can be used.
- Gradient adaptation of the combination parameter implies that combinations introduce some additional gradient noise, which should be minimized with an adequate selection of μ_a or μ_λ. In this sense, the factor $\lambda(n)[1 - \lambda(n)]$ that appears in the adaptation of $a(n)$ with cvx-PN-LMS reduces the gradient noise when the mixing parameter becomes close to 0 or 1.

In our experience, the last two issues have an important impact on the performance of the combination, to the extent that convex combinations may actually be preferred over affine ones, unless in situations where significant EMSE gains can be anticipated for the affine combination from the theoretical analysis. The next section will compare the performance of these rules through simulations.

We should point out that other rules can be found in the literature for the adaptation of the mixing parameter, such as the least squares rule from [39], the sign algorithm proposed in [41], or a method relying on the ratio between the estimated errors of the filter components [16]. Nevertheless, in the following we will restrict our discussion to the methods that have been presented in this section, which are the most frequently used in the literature.

CONVERGENCE PROPERTIES OF COMBINATION FILTERS

To examine the convergence properties of aff-PN-LMS and cvx-PN-LMS, we consider the combination of two normalized LMS (NLMS) filters with step sizes $\mu_1 = 0.5$ and $\mu_2 = 0.01$. The optimum solution is a stationary vector of length seven, the covariance matrix of the input signal is $\mathbf{R} = (1/7)\mathbf{I}$, and the variance of the observation noise is adjusted to get an SNR of 20 dB. Different step sizes have been explored for the combination: $\mu_a = [0.25, 0.5, 1]$ for cvx-PN-LMS, while $\mu_\lambda = \mu_a/800$ is used for aff-PN-LMS to get comparable steady-state error. Regarding these step-size values, we can see that the range of practical values for μ_a is within the usual range of steps sizes used with normalized schemes, whereas for the affine combination much smaller values are required for comparable performance. This fact simplifies the selection of the step size in the convex case.

Figure 4 illustrates the performance of affine and convex combinations averaged over 1,000 experiments. Figure 4(b)–(d) compares the convergence of both schemes with respect to the optimum selection of the mixing parameter given by (8). In all cases, the combination schemes converge first to the EMSE level of the fast filter (−30 dB), and after a while follow the slow component to get a final EMSE of around −50 dB. It is interesting to see that cvx-PN-LMS shows near optimum selection of the mixing parameter for all three values of μ_a, while the affine combination may incur a significant delay, especially for the smallest μ_λ. Figure 4(a) plots the excess steady-state error of both schemes with respect to $\zeta_2(\infty)$ as a function of the step size, and shows that this faster convergence of cvx-PN-LMS with respect to aff-PN-LMS is not in exchange of larger residual error.

The fact that cvx-PN-LMS has the ability to switch rapidly between the fast and slow filter components while at the same time minimizing the residual error in steady state is due to the incorporation of the activation function, whose derivative propagates to the update rule. In other words, we can view the effective step size of cvx-PN-LMS as being $\mu_a\lambda(n)[1 - \lambda(n)]$ (see Table 4), and the evolution of the multiplicative factor, represented in Figure 4(e), shows that this effective step size becomes large when the combination needs to switch between filter components, while becoming small in steady state, thus minimizing the residual error after convergence is complete.

BENEFITS OF POWER NORMALIZED UPDATING RULES

To illustrate the benefits of power normalized updating rules, we compare the behavior of a convex combination of two NLMS filters with step sizes $\mu_1 = 0.5$ and $\mu_2 = 0.01$, employing both the cvx-LMS rule and its power normalized version (cvx-PN-LMS) for the combination layer. The optimum solution is a length-30 nonstationary vector, which varies according to the random-walk model given by (11). The covariance matrices of the change of the optimum solution and of the input signal are given respectively by $\mathbf{Q} = \sigma_q^2\mathbf{I}$ and $\mathbf{R} = (1/30)\mathbf{I}$. The variance of the observation noise is adjusted to get different SNR levels. The step size for the cvx-LMS rule has been set to $\mu_a = 1{,}000$ and for its power normalized version (cvx-PN-LMS), we have set $\mu_a = 1$.

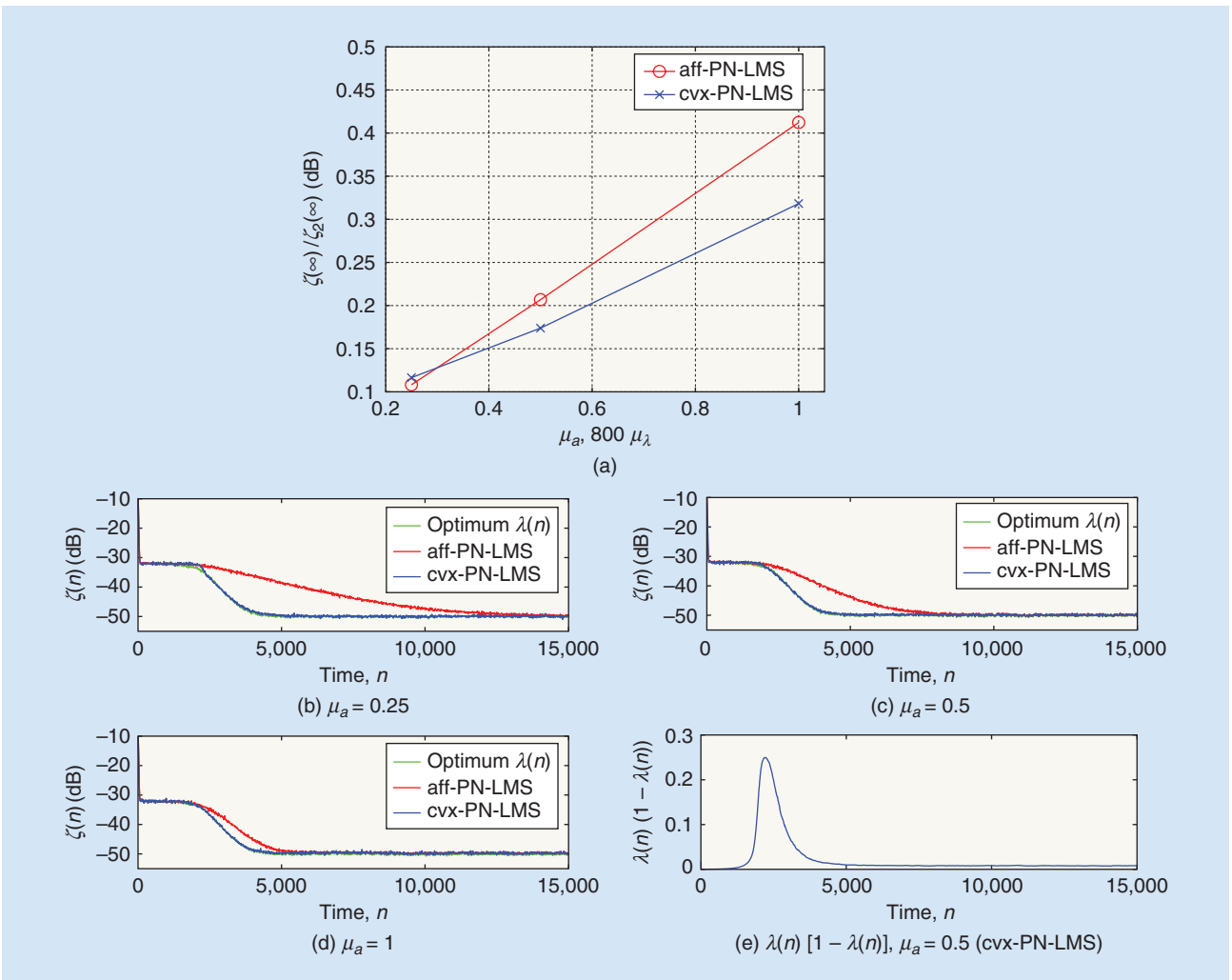
Figure 5 shows the steady-state NSD of the individual filters and of their convex combinations obtained with the cvx-LMS rule and with the cvx-PN-LMS scheme, as a function of the speed of changes of the optimum solution [$\operatorname{Tr}(\mathbf{Q}) = 30\sigma_q^2$]. The left panel considers an SNR of 5 dB and the right panel, an SNR of 30 dB. We can observe that the combination scheme with the cvx-LMS rule results in a suboptimal performance when the optimum solution changes very fast [for large $\operatorname{Tr}(\mathbf{Q})$]. This is due to the fact that both component filters are incurring a very significant error, resulting in a nonnegligible gradient noise when updating the auxiliary parameter $a(n)$ with cvx-LMS [37]. The performance of cvx-LMS also degrades, whatever the value of $\operatorname{Tr}(\mathbf{Q})$, as the SNR decreases. On the other hand, when the cvx-PN-LMS rule is employed, the combination shows a very stable operation, and behaves as well as the best component filter not only for any SNR, but also for all values of $\operatorname{Tr}(\mathbf{Q})$. A similar behavior is also observed when we compare the aff-LMS rule to its power normalized version (aff-PN-LMS) to update $\lambda(n)$ in affine combinations [18].

COMBINATION SCHEMES VERSUS VSS LMS

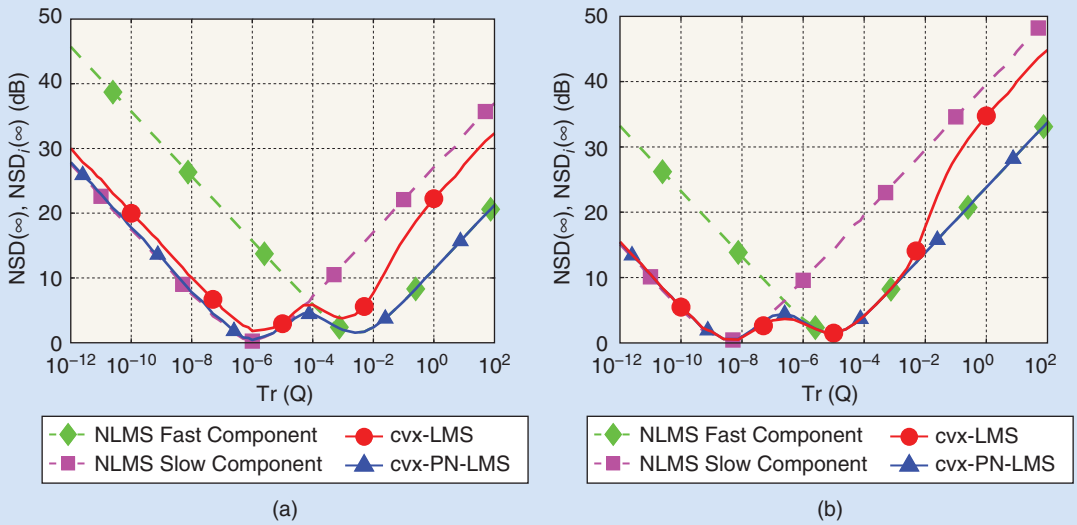
So far we have illustrated how combination schemes can be exploited to provide filters with improved performance, relying on adaptive rules to learn the most adequate value for the mixing parameter. A reasonable question is whether the same improvement could be achieved by an individual but more complex filter structure. Although the answer to this question is generally positive, these more complex solutions usually require further statistical knowledge about the filtering scenario that is normally unavailable, while combination approaches offer a more versatile and robust approach. To illustrate this point, we will examine again the identification of a time-varying solution.

Adaptive filters usually have two conflicting goals: on one hand, to track variations of the parameter vector they are trying to estimate and, on the other hand, to suppress measurement noise and errors due to undermodeling. The first goal requires a fast filter, one that quickly corrects any mismatch between the input and the estimates, while the second goal would benefit from a slower filter that averages out measurement noise. We can emphasize one goal over the other by the choice of a design variable such as the step size in LMS or the forgetting factor in RLS. However, in a nonstationary environment, the optimum choice of the step size or the forgetting factor will continuously change, which has motivated the proposal of different methods to update the step size (or forgetting factor); see, e.g., [42]–[46]. Most existing VSS algorithms implement procedures that rely on the available input and error signals. However, in the absence of additional information (e.g., background noise level, whether the filter is still converging or has reached steady state), these algorithms can fail to identify the most convenient step size, especially if the filtering conditions change over time.

As an alternative to the VSS algorithms, combinations of adaptive filters with different step sizes (or forgetting factors) can be particularly useful. One advantage of this approach is that it performs reasonably well for large variations in input SNR and speed of change of the true parameter vector. Figure 6 compares the



[FIG4] The performance of aff-PN-LMS and cvx-PN-LMS when combining two NLMS adaptive filters with step sizes $\mu_1 = 0.5$ and $\mu_2 = 0.01$. (a) The steady-state excess error with respect to the slow filter component for three different values of the step sizes. (b)–(d) compare the convergence of the affine and convex rules (for three values of μ_a and $\mu_\lambda = \mu_a/800$) with the combination scheme using the optimum mixing parameter. (e) Evolution of factor $\lambda(n)[1-\lambda(n)]$ for cvx-PN-LMS.

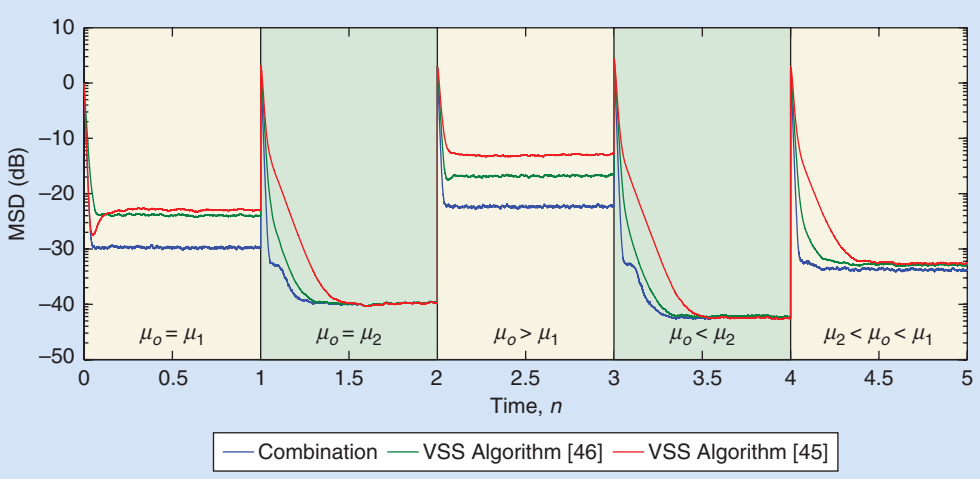


[FIG5] The steady-state NSD of a convex combination of two NLMS filters with different step sizes as a function of the speed of change of the optimum solution. The figure illustrates the NSD incurred by the individual NLMS filters and by their convex combinations. NLMS step sizes are $\mu_1 = 0.5$ and $\mu_2 = 0.01$; combination rules: cvx-LMS ($\mu_a = 1{,}000$) and cvx-PN-LMS ($\mu_a = 1$, $\eta = 0.9$); SNRs: (a) SNR = 5 dB and (b) SNR = 30 dB. All results were averaged over 50,000 iterations once the algorithms reached the steady state and more than 50 independent runs.

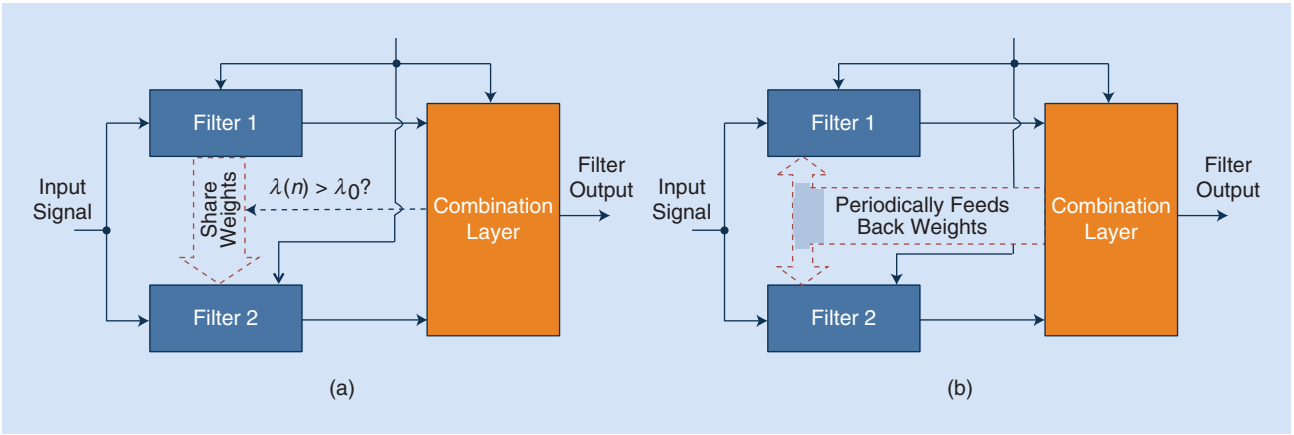
performance, in an identification problem, of the robust VSS methods proposed in [45] and [46] against the performance of convex combination of two LMS adaptive filters with different step sizes. The fast filter uses $\mu_1 = 0.01$ and the slow filter, $\mu_2 = 0.001$. The figure shows the mean-square deviation (MSD) of a ten-tap filter in a tracking situation like the one described by (11), with $\mathbf{Q} = \sigma_q^2\mathbf{I}$, for different values of σ_q^2 (and, correspondingly, of μ_o). To provide a fair comparison, the parameters for the VSS methods were optimized for step sizes in the range $[\mu_2, \mu_1]$, and Figure 6 shows the performance of the three methods for five different values of the optimum step size μ_o for a single LMS filter: from less than μ_2 to more than μ_1. It is seen that the convex combination scheme is able to deliver the best possible performance in all conditions. In other words, it shows a more robust performance given the lack of knowledge about the true value of σ_q^2, and attains an overall performance that could not be achieved by a standard LMS filter or by the VSS schemes.

COMMUNICATION BETWEEN COMPONENT FILTERS

One simple way to improve the performance of combination schemes is to allow some communication between the component filters. Since the component filters are usually designed to



[FIG6] MSD obtained by VSS algorithms and by the convex combination of two LMS filters with $\mu_1 = 0.01$ and $\mu_2 = 0.001$. The value of σ_q^2 changes with time, and the corresponding optimum LMS step sizes (μ_o) are indicated in the figure.

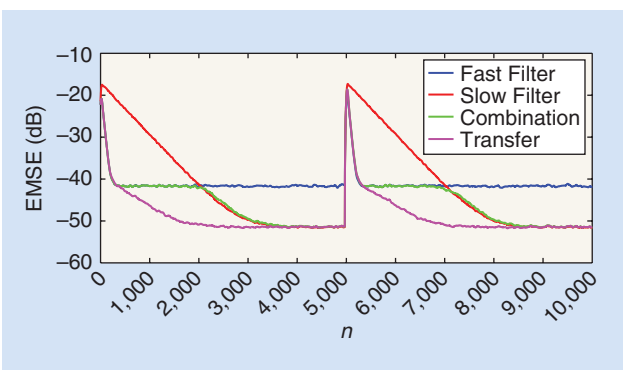


[FIG7] **Communication between subfilters. (a) Weight transfer and (b) weight feedback.**

perform better in different environmental conditions, information from the best filter in a given set of conditions can be used to boost the performance of the other filters. In some situations, the gain for the overall output can be significant.

Here we describe two approaches for interfilter communication: the weight transfer scheme of [47] and [48], and the weight feedback of [49]. Both are applicable when the component filters have the same length. The difference between these methods is that the first allows only communication from one of the filters to the other, whereas the second approach feeds back the overall combined weights to all component filters, as depicted in Figure 7.

Consider for example the combination of two component filters, one designed for a fast-changing environment (fast filter), and the other for a slow-changing environment (slow filter). In this situation, during the initial convergence, or after an abrupt change in the optimum weight vector $\mathbf{w}_o(n)$, the slow filter may lag behind the fast filter (see Figure 8). After the fast filter reaches the steady state, the combination would need to wait until the slow filter catches up. This can be avoided by either leaking [47] or copying [48] the fast filter coefficients to the slow filter at the appropriate times.



[FIG8] **The performance of weight transfer schemes. The plots show the EMSE obtained with the component filters and standard convex combination without weight transfer, and the second method of Table 5, using $N_0 = 2$ and $\lambda_0 = 0.982$. The components are two NLMS algorithms, one using $\mu_1 = 0.1$, the other using $\mu_2 = 0.01$.**

The good news is that the mixing parameter is already available to help decide when to transfer the coefficients—we only need to check if $\lambda(n)$ is close to one (so that the overall output is essentially the fast filter). Therefore, the idea is to choose a threshold λ_0, and modify the slow filter update whenever $\lambda(n) \geq \lambda_0$, as shown in Table 5. The two methods are similar: in the gradual transfer scheme of [47], whenever $\lambda(n)$ is larger than the threshold, we allow the fast filter coefficients to gradually "leak" into the slow filter. The other method simply copies the fast filter coefficients to the slow filter, and thus has a smaller computational cost; however, the condition for transfer has to be modified. If we allowed the fast filter coefficients to be copied to the slow filter whenever $\lambda(n) \geq \lambda_0$, the combination would be stuck forever at the fast filter (the slow filter adaptation would be irrelevant). For this reason the transfer is only allowed at periodic intervals of length $N_0 \geq 2$.

The weight feedback scheme, on the other hand, copies the combined weight vector $\mathbf{w}(n) = \lambda(n)\mathbf{w}_1(n) + (1-\lambda)\mathbf{w}_2(n)$ periodically to $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$, as shown in Figure 7 and Table 5. This method is simpler to set up, since only one parameter needs to be tuned, but requires the explicit computation of $\mathbf{w}(n)$ at intervals of N_0 samples. Since N_0 can be chosen reasonably large, the additional cost per sample is modest. Even though these methods introduce additional parameters that need to be fixed, their selection is not very problematic and the transfer methods show a behavior rather robust with respect to them (a detailed discussion is not included here; see [47]–[49]).

Figure 8 provides an example of the operation of the simple-copy method for length $M = 7$ filters (the other methods have similar performance). The value of $\mathbf{w}_o(n)$ changes abruptly at the middle of the simulation $(n = 5 \times 10^4)$.

COMBINATION OF SEVERAL ADAPTIVE FILTERS

Until now, we have considered combinations of just two adaptive filters. However, combining K adaptive filters (with $K > 2$) makes it possible to further increase the robustness and versatility

of combination schemes. The combination of an arbitrary number of filters can be mainly used to:

- Simplify the selection of a parameter. For instance, paying attention to the tracking scenario depicted in Figure 2(a), the robustness of the scheme would be increased for $\mathrm{Tr}\{\mathbf{Q}\} > 10^{-3}$ if a third filter with step size equal to one were incorporated to the combination [50].
- Alleviate several compromises simultaneously. For instance, regarding the selection of the step size, μ, and the length of an adaptive filter, M, we can combine four adaptive filters with settings $\{\mu_1, M_1\}$, $\{\mu_2, M_1\}$, $\{\mu_1, M_2\}$, and $\{\mu_2, M_2\}$. Another example was proposed in [25] and it is included in the section "Acoustic Echo Cancelation," where a combination of several filters (linear and nonlinear) is designed to alleviate simultaneously the compromise related with the step size selection, and with the presence or absence of nonlinearities in the filtering scenario.

In the literature, two different approaches for the combination of several adaptive filters have been proposed, both for affine and convex approaches. These schemes differ in the topology employed to perform the combination as described next.

[TABLE 5] INTERFILTER COMMUNICATION SCHEMES (EXEMPLIFIED FOR COMBINATIONS OF TWO FILTERS).

	GRADUAL TRANSFER [47]	SIMPLE COPY [48]	FEEDBACK [49]
FILTER COMMUNICATION	$\mathbf{w}_2(n) \leftarrow \ell\mathbf{w}_2(n) + (1-\ell)\mathbf{w}_1(n)$	$\mathbf{w}_2(n) \leftarrow \mathbf{w}_1(n)$	$\mathbf{w}_1(n), \mathbf{w}_2(n) \leftarrow \mathbf{w}(n)$
CONDITION	$\lambda(n) \geq \lambda_0$	$\lambda(n) \geq \lambda_0$ AND $n = kN_0$	$n = kN_0$
PARAMETERS	$0 < \ell < 1;\ 0 \ll \lambda_0 < 1$	$N_0 \geq 2;\ 0 \ll \lambda_0 < 1$	$N_0 \geq 2$

THE HIERARCHICAL SCHEME

This approach combines K adaptive filters employing different layers, where only combinations of two elements are considered at a time. For instance, the output of a hierarchical combination of four filters depicted in Figure 9(a) reads:

$$y(n) = \lambda_{21}(n)\{\lambda_{11}(n)y_1(n) + [1-\lambda_{11}(n)]y_2(n)\} + [1-\lambda_{21}(n)]\{\lambda_{12}(n)y_3(n) + [1-\lambda_{12}(n)]y_4(n)\}, \quad (20)$$

where $\lambda_{ij}(n)$ refers to the mixing parameter of the jth combination in the ith layer. All mixing parameters are adapted to minimize the power of the local combined error. For their update, we can follow the same adaptive rules (convex or affine) reviewed previously for the case of a combination of two filters.

THE ONE-LAYER SCHEME

We can combine an arbitrary number of filters employing an alternative approach based on one-layer combination as depicted in Figure 9(b). Focusing on affine combinations of M adaptive filters with outputs $y_k(n)$, with $k = 1,\ldots,K$, [19] and [50] proposed two one-layer combination schemes whose output is given by

$$y(n) = \sum_{k=1}^{K-1} \lambda_k(n)y_k(n) + \left[1 - \sum_{k=1}^{K-1} \lambda_k(n)\right] y_K(n). \quad (21)$$

Different adaptive rules were proposed in the literature to update $\lambda_k(n)$, with $k = 1,\ldots,K-1$, following, for instance, LMS or RLS approaches [19], or estimating the $K-1$ affine mixing parameters as the solution of a least-squares problem [50]. The incorporation of convex combination constraints forces the inclusion of an additional mechanism (similar to the sigmoidal activation) to make all mixing parameters remain positive and add up to one. This scheme was proposed by [51] and [19] as an extension to the standard convex combination of two adaptive filters, and obtains the combined output as

$$y(n) = \sum_{k=1}^{K} \lambda_k(n)y_k(n). \quad (22)$$

As in the case of combining two filters, instead of adapting directly the K mixing parameters, K auxiliary parameters $a_k(n)$ are updated following a gradient descent algorithm. The relation between $\lambda_k(n)$ and $a_k(n)$ is based on the softmax activation function

$$\lambda_k(n) = \frac{\exp[a_k(n)]}{\sum_{j=1}^{K} \exp[a_j(n)]}, \quad \text{for } k = 1,\ldots,K. \quad (23)$$

This activation function is a natural extension of the sigmoid used in the binary case to map several real parameters to a probability distribution [52], as required by a convex combination, where all parameters must remain positive and sum up to one.

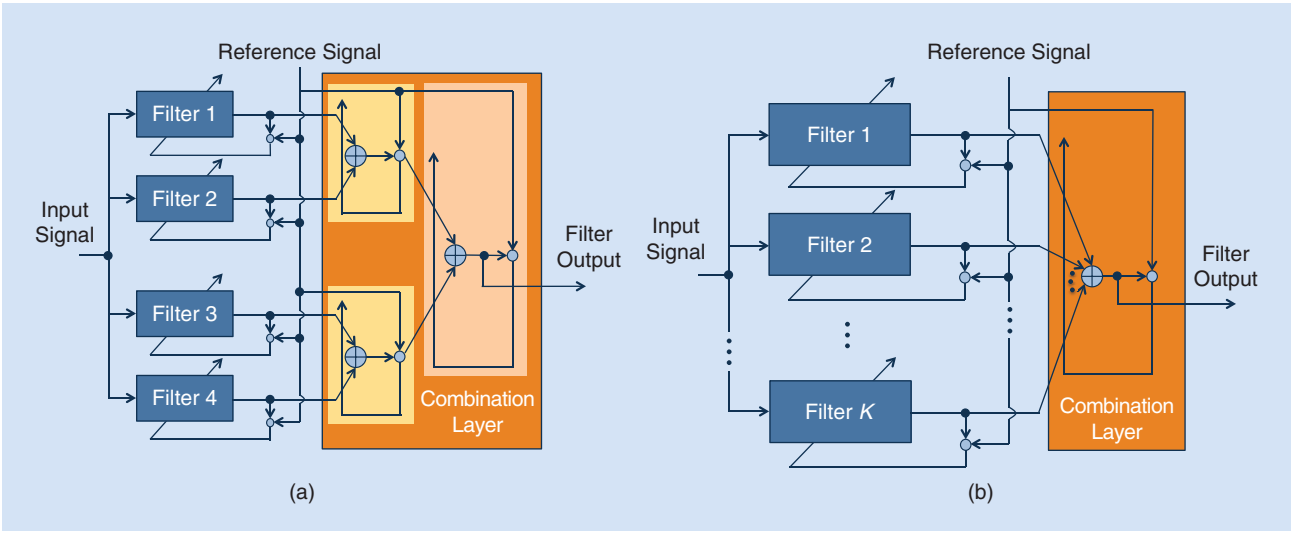
Although both multifilter structures can improve the performance beyond the combination of just two filters, one useful characteristic of hierarchical schemes is its ability to extract more information about the filtering scenario from the evolution of the mixing parameters, since each combination usually combines two adaptive filters that only differ in the value of a setting (step size, length, etc.), and the combination parameter selects the best of both competing models.

REDUCED-COST COMBINATIONS

Combination schemes require running two or more filters in parallel, which may be a concern in applications in which computational cost is at a premium. In many situations, however, the additional computational cost of adding one or more filters can be made just slightly higher than the cost of running a single filter. In the following sections, we describe a few methods to reduce the cost of combination schemes.

USE A LOW-COST FILTER AS COMPANION TO A HIGH-COST ONE

Although straightforward, several useful results fall in this class. For example, consider the previously mentioned case of a combination of an RLS and an LMS filter. This structure enhances the tracking performance of a single RLS filter, and only requires a modest increase in computational complexity. A lattice implementation of RLS has a computational cost of about $16M$



[FIG9] Hierarchical and one-layer combination schemes for an arbitrary number of filters K. (a) The hierarchical scheme: In the first layer, we carry out two combinations of two adaptive filters with complementary settings, and their outputs are combined in the upper layer. (b) The combination of K adaptive filters following the one-layer approach. The affine schemes of [50] and [19] adapt $K-1$ mixing parameters, whereas the convex combination of [19] and [51] requires the adaptation of K parameters.

multiplications, $8M$ additions, and $8M$ divisions [6]. Compared with the approximately $2M$ additions and $2M$ multiplications required by LMS, the combination scheme will increase the cost by only about 10% over that of a single RLS. For other stable implementations of RLS, such as QR methods, the computational cost is even higher, so the burden of adding an LMS component would be almost negligible. Note that, for their operation, combination schemes require just the outputs of component filters (see the section "Estimating the Combination Parameter"), so lattice or QR implementations of RLS can readily be incorporated into combination schemes without modifications.

Other situations in which it is useful to combine a low-cost and a higher-cost filter are: the combination of a simple linear filter with a Volterra nonlinear filter, as described in the section "Acoustic Echo Cancelation"; and the use of a short filter combined with a longer filter, as described, for example, in [53]. The short filter is responsible for tracking fast variations in the weight vector, whereas the long filter is responsible for guaranteeing a small bias in times of slower variation of the optimum weight vector.

PARALLELIZATION

Since the component filters are running independently (apart from occasional information exchanges as in the section "Communication Between Component Filters"), combination schemes are easy to implement in parallel form. In a field programmable gate array implementation, for example, this means that implementing a combination of filters will not require a faster clock rate, compared to running a single filter. Depending on the hardware in which the filters are to be implemented, this is an important advantage.

TAKE ADVANTAGE OF REDUNDANCIES IN THE COMPONENT FILTERS

This approach can provide substantial gains in some situations, but depends on the kind of filters that are being combined. For example, transform-domain or subband filters can share the transform or filter-bank blocks that process the input signal $\mathbf{u}(n)$.

A second example is based on the following observation [54]: if filters using the same input regressor $\mathbf{u}(n)$ are combined, the weight estimates $\mathbf{w}_i(n)$ will be, most of the time, similar. Therefore, one could update only one of the filters [say, $\mathbf{w}_1(n)$], and the difference between this filter and the others. The important observation is that, if the entries of $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$ are real values ranging from, say, -1 to $+1$, the range of the entries of $\delta\mathbf{w}_2(n) = \mathbf{w}_2(n) - \mathbf{w}_1(n)$ will most of the time be smaller, from -2^{-B_c} to $+2^{-B_c}$, for some number of bits $B_c > 0$. If the filters are implemented in custom or semicustom hardware, this observation can be used to reduce the number of bits allocated to $\delta\mathbf{w}_2(n)$, reducing the complexity of all multiplications related to the second filter, both for computing the output and for updating the filter coefficients. In summary, this idea amounts to making the filters share the most significant bits in all coefficients, and adapting them with the step size of the fast filter. During convergence, when the difference between the filters is expected to be more significant, disregarding the most significant bits of the difference filter has the effect of working just as a weight transfer mechanism from the fast to the slow filter.

To take full advantage of the reduced word length used to store $\delta\mathbf{w}_2(n)$, the output of the second filter and the difference filter update should be computed as shown in Table 6. Steps 3 and 6 of the algorithm can be computed with a reduced word length. As can be seen, all the costly operations related to the second filter are computed with a reduced word length. Using this technique, it is possible to use for the difference filter a word length of about half that of the first filter, with little to no degradation in performance, compared to using full word length for all variables.

SIGNAL MODALITY CHARACTERIZATION

As we have already discussed, one of the main advantages of the combination approach to adaptive filters is its versatility. In the

remainder of the article, we illustrate the applicability of this strategy in different fields where adaptive filters are generally used.

As a first application example, a combination of adaptive filters can be employed to characterize signal modality, for instance, whether a particular signal is generated through a linear or nonlinear process. This observation has been successfully applied to track changes in the modality of different kinds of signals, including the electroencephalogram [55], complex representations of wind and maritime radar [56], or speech [22]. To illustrate the idea, we consider the characterization of the linear/nonlinear nature of a signal along the lines of [22], which uses a convex combination of an NLMS filter and a normalized nonlinear gradient descent (NNGD) algorithm to predict future values of the input signal

$$\hat{u}(n+1) = \lambda(n)\hat{u}_{\mathrm{NLMS}}(n+1) + [1-\lambda(n)]\hat{u}_{\mathrm{NNGD}}(n+1).$$

The mixing parameter of the combination, $\lambda(n)$, is adapted to minimize the square error of the prediction. A value $\lambda(n) \approx 1$ means that an NLMS filter suffices to achieve a good prediction allowing us to conclude that the input signal $u(n)$ is intrinsically linear. In the other extreme, when $\lambda(n)$ approaches zero, the predominance of the NNGD prediction suggests a nonlinear input signal. Figure 10 shows the time evolution of the mixing parameter for a single realization of the algorithm, when we evaluate a signal that alternates between linear autoregressive processes (of orders 1 and 4) and two benchmark nonlinear signals described in [22, eqs. (9)–(12)]. Information about the linear/nonlinear nature of the signal can be easily extracted from the evolution of the mixing parameter $\lambda(n)$.

[TABLE 6] THE REDUCED-COST COMBINATION THROUGH DIFFERENCE FILTER (USING TWO LMS FILTERS AS AN EXAMPLE).

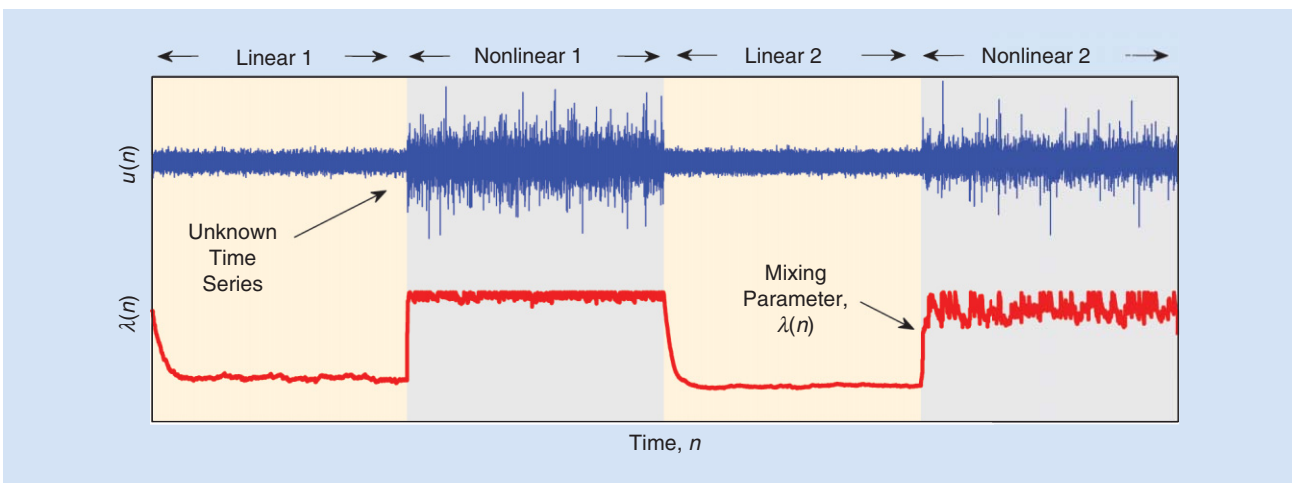
ALGORITHM STEP	EQUATION	WORD LENGTH
1—FIRST FILTER ERROR:	$e_1(n) = d(n) - \mathbf{u}^{\mathsf{T}}(n)\mathbf{w}_1(n)$	FULL
2—UPDATE FIRST FILTER:	$\mathbf{w}_1(n+1) = \mathbf{w}_1(n) + \mu_1 e_1(n)\mathbf{u}(n)$	FULL
3—OUTPUT OF DIFFERENCE FILTER:	$\delta y_2(n) = \mathbf{u}^{\mathsf{T}}(n)\delta\mathbf{w}_2(n)$	REDUCED
4—SECOND FILTER ERROR:	$e_2(n) = e_1(n) + \delta y_2(n)$	FULL
5—OUTPUT ERROR:	$\lambda(n)e_1(n) + (1-\lambda(n))e_2(n)$	FULL
6—UPDATE DIFFERENCE FILTER:	$e_\delta(n) = \mu_2 e_2(n) - \mu_1 e_1(n)$ $\delta\mathbf{w}_2(n+1) = \delta\mathbf{w}_2(n) + e_\delta(n)\mathbf{u}(n)$	REDUCED

ADAPTIVE BLIND EQUALIZATION

Adaptive equalizers are widely used in digital communications to remove the intersymbol interference introduced by dispersive channels. To avoid the transmission of pilot sequences and use the channel bandwidth in an efficient manner, these equalizers can be initially adapted using a blind algorithm and switched to a decision-directed (DD) mode after the blind equalization achieves a sufficiently low steady-state MSE. The selection of an appropriate switching threshold is crucial, and this selection depends on many factors such as the signal constellation, the communication channel, or the SNR.

An adaptive combination of blind and DD equalization modes can be used, where the combination layer is itself adapted using a blind criterion. This combination scheme provides an automatic mechanism for smoothly switching between the blind and DD modes. The blind algorithm mitigates intersymbol interference during the initial convergence or when abrupt changes in the channel occur. Then, when the steady-state MSE is sufficiently low, the overall equalizer automatically switches to the DD mode to get an even lower error. The main advantage of this scheme is that it avoids the need to set a priori the transition MSE level [57], and automatically changes to DD mode as soon as a sufficient equalization level is attained.

Figure 11 illustrates the performance of the combination scheme as well as of other well-known schemes for blind equalization [58]–[61]. In all cases, the multimodulus algorithm (MMA) and the LMS algorithm were used for the blind and DD



[FIG10] **The signal modeling characterization using a combination of linear and nonlinear filters. The mixing parameter is an effective indicator of the linear/nonlinear nature of the input signal generation process.**

equalization phases, respectively. The communication scenario considers the transmission of a 256-quadrature amplitude modulation signal through a telephonic channel with an SNR of 40 dB, including an abrupt change at $n = 2.5 \times 10^5$. Figure 11(a) shows the MSE evolution for all methods averaged over 1,000 independent runs, whereas Figure 11(b) illustrates the distribution of the switching time between blind and DD modes, considering that the switching to DD mode is complete when the MSE reaches -16 dB for a particular run. We can conclude that the combination scheme does a better job at detecting the right moment for commuting to DD mode. Furthermore, the smaller variance observed in the switching iteration suggests that the combination method is more robust than other methods to the particularities of the simulation scenario.

This example shows that the combination approach can be used beyond MSE adaptive schemes. Here, we considered blind cost functions that exploit properties of high-order moments of the involved signals, both at the individual filter and at the combination layer levels. Similar uses can be conceived in other applications where high-order moments are frequently used, such as blind source separation.

SPARSE SYSTEM IDENTIFICATION

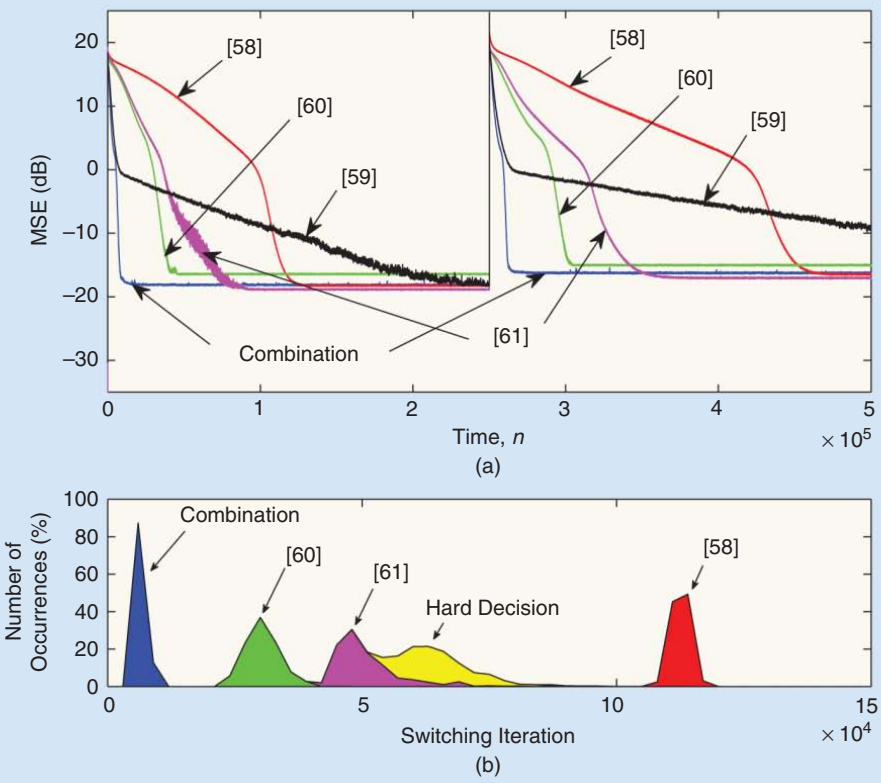
Sparse systems have gathered a lot of attention. These systems are characterized by long impulse responses with only a few nonzero coefficients, and are frequently encountered in applications such as network and acoustic echo cancelation [62], compressed sensing [63], high-definition television [64] and wireless communication and multiple-input, multiple-output channel estimation [65], [66], among many others.

In the literature, several adaptive schemes have exploited such prior information to accelerate the convergence toward the sparse optimum solution. For instance, proportionate adaptive filters assign to each coefficient a different step size proportional to its amplitude [23], [67]. A recent approach [68] promotes sparsity based on adaptive projections where the sparsity constraints are included considering ℓ_1 balls and giving rise to a convex set whose shape is a hyperslab. An important family of sparse adaptive algorithms [69]–[71] incorporates sparsity enforcing norms into the cost function minimized by the filter.

The zero-attracting LMS (ZA-LMS) filter [69] uses a stochastic gradient descent rule to minimize a cost function that mixes the square power error and the ℓ_1-norm penalty of the weight vector. The update equation for this adaptive scheme reads:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\,\mathbf{u}(n) - \rho sgn\,[\mathbf{w}(n)], \qquad (24)$$

where function sgn $[\cdot]$ extracts the sign of each element of the vector, and parameter ρ controls how the sparsity is favored in the coefficient adaptation. In fact, $\rho = 0$ turns (24) into the standard



[FIG11] The peformance of blind equalization methods. (a) MSE time evolution averaged over 1,000 runs. (b) The distribution of switching times from blind to DD mode (–16 dB).

LMS update, removing any ability to promote sparsity. This scheme has shown improved performance with respect to the standard LMS algorithm when the unknown system is highly sparse; however, standard LMS outperforms ZA-LMS scheme when the system is dispersive. For this reason, there exists a trade-off related to the degree of sparsity of the system, which unfortunately is usually unknown a priori, or can even vary over time.

Here we cover two approaches specially designed to alleviate this compromise:

- Scheme A: An adaptive convex combination of the ZA-NLMS filter, i.e., a sparsity-norm regularized version of the NLMS scheme and a standard NLMS algorithm [72]. This approach constitutes the straightforward application of combination schemes to alleviate the compromise regarding the selection of parameter ρ in the ZA-NLMS filter.
- Scheme B: A block-wise biased adaptive filter, depicted in Figure 12, where the outputs of Q nonoverlapping blocks of coefficients of an adaptive filter are weighted by different scaling factors to obtain the overall output [73]

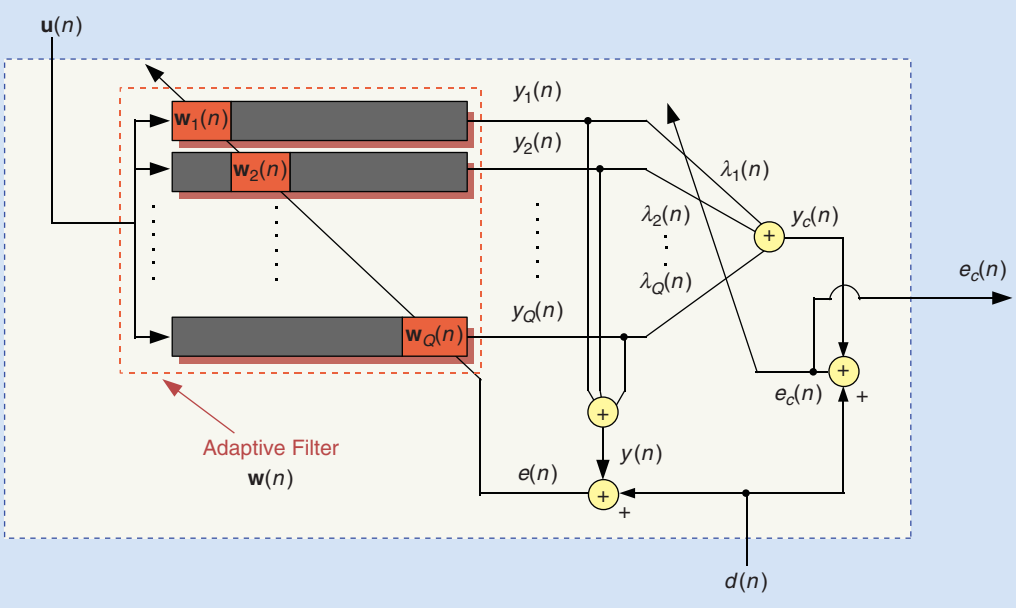
$$y_c(n) = \sum_{q=1}^{Q} \lambda_q(n) y_q(n), \tag{25}$$

where $y_q(n)$ is the partial output of each block, and $\lambda_q(n) \in [0,1]$ with $q = 1,...,Q$ are the shrinkage factors that adapt to minimize the power of $e_c(n) = d(n) - y_c(n)$. This scheme manages the well-known bias versus variance compromise in a block-wise manner [73]: if the MSD in the qth block is much higher than the energy of the unknown system in this block, $\lambda_q(n)$ will tend to be zero, biasing the output of the block $y_q(n)$ toward zero but reducing the output error. The shrinkage factors $\lambda_q(n)$ therefore act as estimators for the support (set of nonzero coefficients) of the filter.

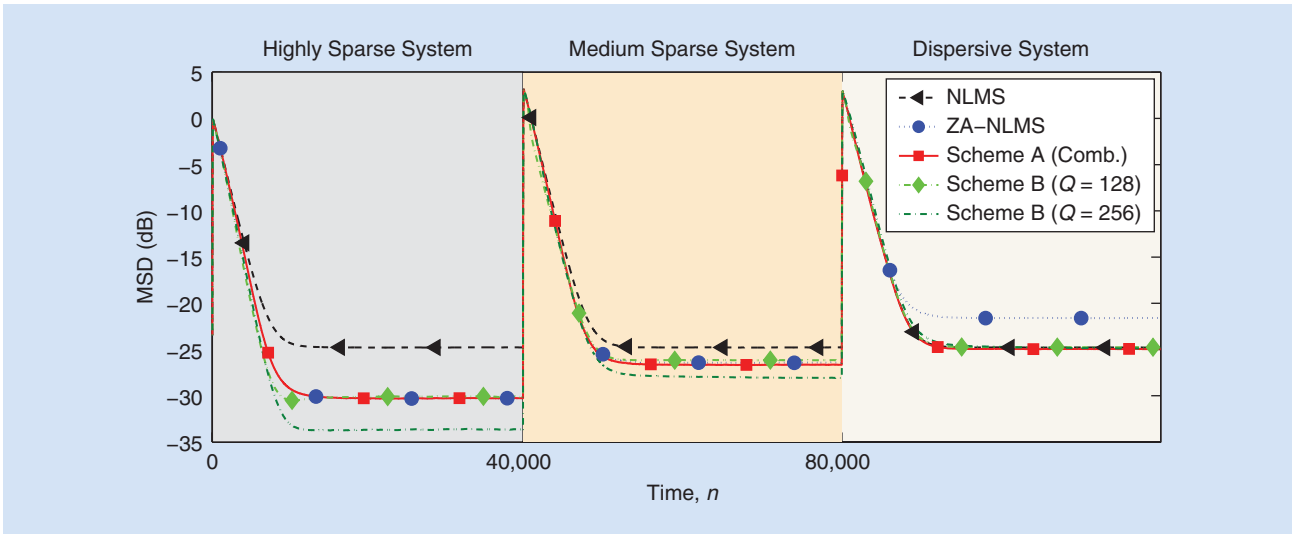
Regarding the estimation of these scaling factors $\lambda_q(n)$, with $q = 1,...,Q$, it can be shown that their optimal values lie on the interval $(0,1)$ [73]. Therefore, proceeding as in [38], we can reinterpret each element in the sum of (25) as a convex combination between $y_q(n)$ and a virtual filter whose output remains constant and equal to zero. This is useful, because it implies that we can employ rules similar to cvx-LMS and cvx-PN-LMS reviewed in the section "Estimating the Combination Parameter" for adjusting parameters $\lambda_q(n)$. It should be noted that, for this kind of scheme, the ability to model sparse systems increases with the number of blocks Q, since the length of each block decreases. However, the computational cost associated with the adaptation of the mixing parameters also increases with Q.

To illustrate both approaches, we have carried out a system identification experiment with an unknown plant with 1,024 taps, whose sparsity degree changes over time considering white noise as input signal and SNR = 20 dB. We start with a very sparse system (only 16 active taps), that abruptly changes at $n = 40{,}000$ to a plant with 128 active coefficients, and finally, at $n = 80{,}000$, we employ a more dispersive unknown system with 512 active taps. Figure 13 shows the MSD reached by Scheme A and Scheme B, considering two different possibilities for the number of blocks Q.

As expected, Scheme A shows a robust behavior with respect to the sparsity degree of the unknown plant, behaving as well as its best component filter. Scheme B provides an attractive alternative, whose performance improves when the length of each block is reduced. Comparing both schemes, the computational



[FIG12] A block diagram of Scheme B. Each block of the adaptive filter $\mathbf{w}(n)$ is represented as a complete filter computing its output using just the indicated coefficients.



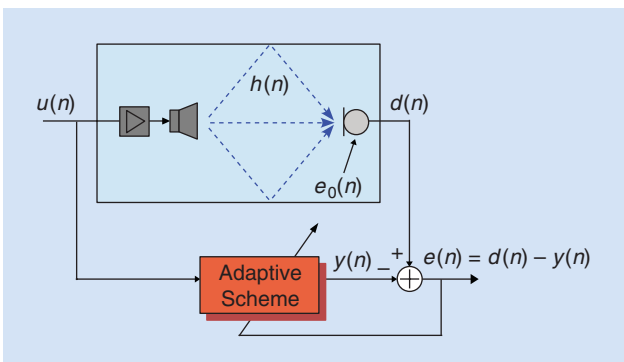
[FIG13] MSD performance of both proposed algorithms: Scheme A (and its constituent filters: an NLMS filter with $\mu = 0.5$ and a ZA-NLMS with $\mu = 0.5$ and $\rho = 10^{-6}$) and Scheme B (based on an NLMS filter with $\mu = 0.5$) for two different configurations: $Q = 128$ and $Q = 256$ blocks. The sparseness of the plant varies over time.

cost of Scheme B is smaller than that of Scheme A, even when $Q = 256$. In this situation, we also observe that Scheme B outperforms Scheme A in terms of MSD for highly and medium sparse systems.

ACOUSTIC ECHO CANCELATION

Combination filters have been successfully employed in acoustic signal processing applications, such as active noise cancelation [30], [31], adaptive beamforming [29], and acoustic echo cancelation, where combination schemes have been used in linear cancelation (considering both time-domain [23] and frequency-domain [24] schemes), and in nonlinear acoustic echo cancelation [25], [26], [74].

The abundance of portable devices has been motivating the inclusion of nonlinear models in the adaptive scheme of acoustic echo cancelers (see Figure 14), to compensate for the nonlinear



[FIG14] The acoustic echo cancelation scenario. Signal $\mathbf{u}(n)$ stands for the input signal, $e_0(n)$ is the background noise at microphone location, $h(n)$ represents the room impulse response that describes the acoustic propagation (linear in nature). Depending on the behavior of the loudspeaker and the power amplifier, nonlinear distortion can be found in the desired signal $d(n)$. The adaptive scheme tries to minimize the power of the cancelation error $e(n)$, incorporating nonlinear models if necessary.

distortion caused by low-cost loudspeakers fed by power amplifiers driven at high levels. One of the main solutions used for this purpose is the VF, which is able to represent a large class of nonlinearities with memory [75].

The output of a VF can be calculated as the addition of partial outputs of different kernels, each one representing an order of nonlinearity, i.e.,

$$y(n) = \sum_{p=1}^{P} y_p(n) = \sum_{p=1}^{P} \sum_{i_1=0}^{N_p-1} \cdots \sum_{i_p=i_p-1}^{N_p-1} h_{p,i_1,\ldots,i_p} \prod_{q=1}^{p} u(n-i_q), \quad (26)$$

where $y_p(n)$ corresponds with the output of the pth kernel with length N_p, P is the number of kernels that compose the VF and $h_{p,i_1,\ldots,i_p}$ represents the adaptive coefficients of the pth kernel. For instance, the output of a VF of order $P = 2$ can be obtained as $y(n) = y_1(n) + y_2(n)$, where $y_1(n)$ is the output of the linear kernel, and $y_2(n)$ represents the output of the quadratic kernel, responsible for the modeling of second-order nonlinearities. Despite the large computational cost associated with the operation of a VF, the success of this nonlinear filter lies in the simplicity of the adaptation of the coefficients in each kernel $h_{p,i_1,\ldots,i_p}$, which permits the employment of update rules similar to those of linear adaptive filters. For this reason, VFs present similar tradeoffs to those of linear filters, in particular the selection of the step size, and the memory of the kernels.

In [25], two different schemes were proposed to alleviate such compromises, called the *combination of VFs* and *combination of kernels* (*CKs*); see Figure 15. The first approach constitutes a straightforward application of the combination idea, i.e., the output of the combined scheme is calculated mixing different VF outputs, each one obtained by means of (26), and following any of the schemes included in the section "Combination of Several Adaptive Filters." However, the second scheme proposes a special kind of VF, where each kernel is replaced by a combination of

kernels with complementary settings. For the case of a combination of two kernels per order, the output of the CK scheme reads

$$y(n) = \sum_{p=1}^{P} y_p(n) = \sum_{p=1}^{P} \lambda_p(n) y_{p,1}(n) + [1 - \lambda_p(n)] y_{p,2}(n), \quad (27)$$

where the outputs of two kernels of order p, $y_{p,1}(n)$, and $y_{p,2}(n)$, are combined by means of the mixing parameter $\lambda_p(n)$. Both algorithms reach similar performance but CK is a more attractive scheme regarding the computational cost [25].

We should note that the performance of VFs is also subject to a tradeoff related to the number of kernels, since the degree of nonlinearity present in the filtering scenario is rarely known a priori. This compromise directly affects the nonlinear acoustic echo cancelers: if only linear distortion is present in the echo path, the adaptation of nonlinear kernels degrades the performance of the whole VF, which behaves worse than a simple linear adaptive filter.

To deal with this compromise, a nonlinear acoustic echo canceler, represented in Figure 15 using both the combination of VFs and the CK approaches, was proposed in [25] considering two requirements: the robustness with respect to the presence or absence of nonlinear distortion, and the fact that the acoustic room impulse response $h(n)$ is strongly variant. The first requirement is dealt with considering a combination of a quadratic kernel and a virtual kernel named *all-zeros kernel*, with all coefficients equal to zero and no adaptation, whose partial output is always zero. Regarding the second requisite, a combination of two linear kernels with different step sizes serves to provide a fast reconvergence and a low steady-state error.

The output of this algorithm, considering the CK scheme, is

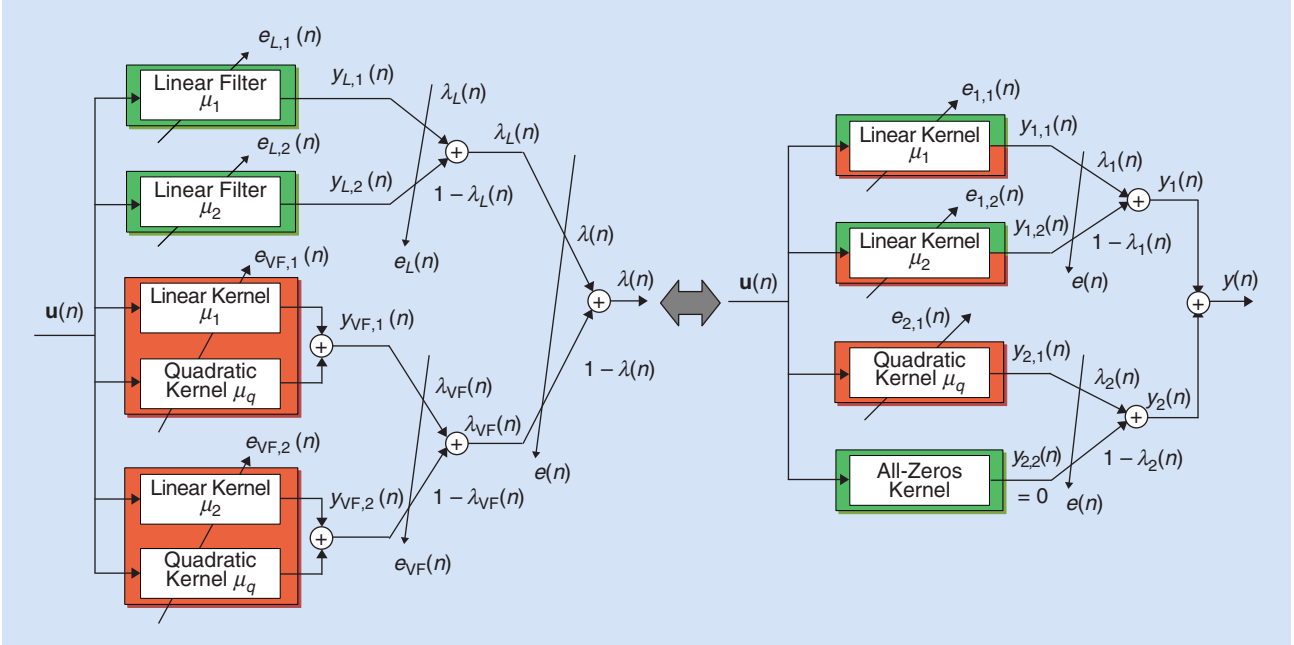
$$y(n) = \underbrace{\lambda_1(n) y_{1,1}(n) + [1 - \lambda_1(n)] y_{1,2}(n)}_{\text{linear part}} + \underbrace{\lambda_2(n) y_{2,1}(n) + [1 - \lambda_2(n)] \cdot 0}_{\text{nonlinear part}}, \quad (28)$$

where $y_{1,1}(n)$ and $y_{1,2}(n)$ are the outputs of two linear kernels with different step sizes, $\lambda_1(n)$ is the mixing parameter to combine them, $y_{2,1}(n)$ is the output of the quadratic kernel, and $\lambda_2(n)$ is the mixing parameter to weight the estimation of the nonlinear echo.

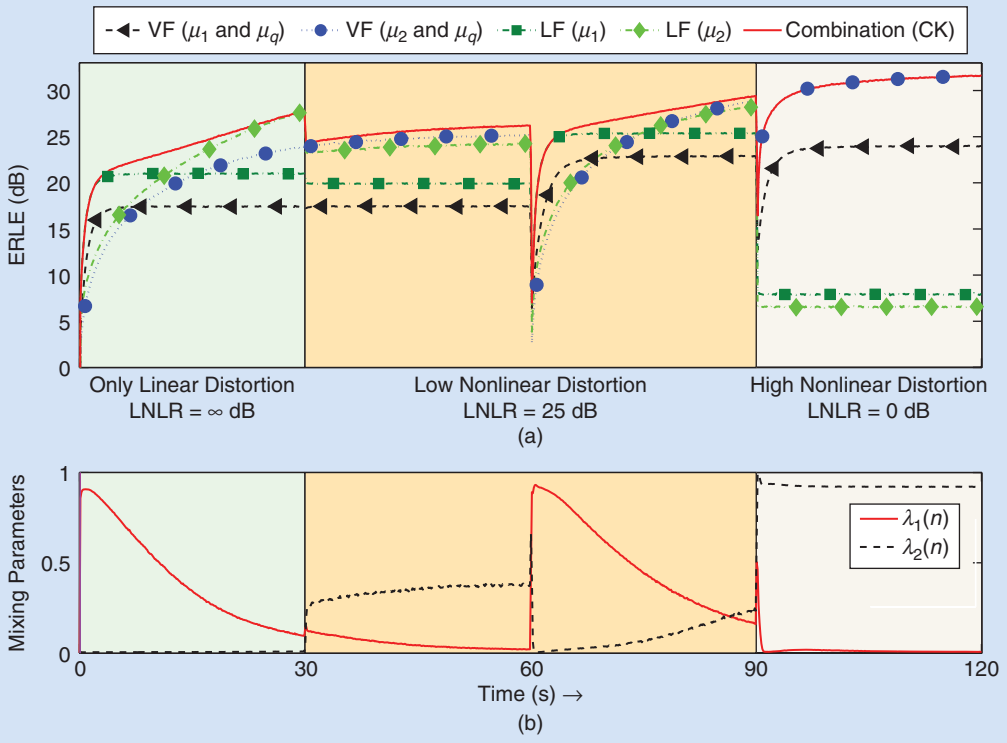
Figure 16 represents the performance of the proposed nonlinear acoustic echo canceler, compared with that of all the adaptive filters (linear and nonlinear) that could be designed with its kernels, in terms of the ERLE computed as

$$\text{ERLE}(n) \triangleq 10 \log \frac{\mathbb{E}\{[d(n) - e_0(n)]^2\}}{\mathbb{E}\{[e(n) - e_0(n)]^2\}}. \quad (29)$$

The proposed scheme reaches the best performance, behaving as a combination of linear filters when only linear distortion is present ($t \leq 30$ s). In this case, $\lambda_2(n) \approx 0$ and the quadratic kernel is not considered in the output of the filter, avoiding the performance degradation of the proposed scheme with respect to that of linear filters. At the end of the experiment ($t > 90$ s), when strong nonlinear distortion is considered, $\lambda_2(n) \approx 1$, and the output of the quadratic kernel is fully incorporated, providing the ability to model nonlinearities. In an intermediate case, during $30 \text{ s} < t \leq 90$ s, the proposed scheme slightly outperforms both



[FIG15] A nonlinear acoustic echo canceler based on combination of filters. (a) A hierarchical combination of filters (linear and VFs) where each filter adapts employing its own error signal, for instance $e_{L,1}(n) = d(n) - y_{L,1}(n)$. (b) A special VF based on the combination of kernels, where each kernel adapts using an error signal employing its own output and the combined outputs of the kernels of different orders (see [25]). The mixing parameters $\lambda_1(n)$ and $\lambda_2(n)$ update to minimize the power of the global error signal, i.e., $e(n) = d(n) - y(n)$.



[FIG16] The performance of the proposed scheme as a function of the linear-to-nonlinear ratio (LNLR) of echo powers, including an abrupt change in the room impulse response at $t = 60$ s. (a) The echo return loss enhancement (ERLE) achieved by the CKs scheme and of all possible filters (linear and VFs) using its constituent kernels. (b) Time evolution of the mixing parameters. In this experiment, we employ Laplacian speech-like colored noise as input signal; experiments with real speech can be found in [25].

linear and nonlinear filters. In addition, the algorithm shows a suitable reconvergence when the room impulse response abruptly changes, thanks to the combination of two linear kernels with different step sizes, as it can be seen after $t = 60$ s.

CONCLUSIONS AND OPEN PROBLEMS

Combinations of adaptive filters constitute a powerful approach to improve the performance of adaptive filters. In this article, we have reviewed some of the most popular combination schemes, highlighting their theoretical properties and limits. Practical algorithms to combine adaptive filters need to implement estimation methods to adjust the combination layer, taking into account the possibly time-varying conditions in which the filter operates.

We have reviewed several of the methods that have been proposed in the literature, paying special attention to gradient methods. Power-normalized algorithms are particularly interesting, since they simplify the selection of parameters and result in a more robust behavior when the statistics of the filtering scenario are (partly) unknown, which is frequently the case. The versatility of the approach has been demonstrated through several examples in a variety of applications. We have seen that, in all studied scenarios, combination schemes offer competitive performance when compared to state-of-the-art methods for each application. This fact, together with the inherent simplicity of the approach, make combination structures attractive for demanding applications requiring enhanced performance, as illustrated by examples and references given in the article.

In our opinion, some of the most interesting open problems to be addressed are:

- the selection and design of component filters with reduced cross-EMSE, with the goal to minimize the overall EMSE
- providing and exploiting strategies to reduce the cost of combined schemes, trying to develop new structures whose complexity is close to that of an individual filter
- extensions to other application domains where different kinds of compromises and performance tradeoffs may be present.

ACKNOWLEDGMENTS

The work of Jerónimo Arenas-García and Luis Azpicueta-Ruiz was partially supported by the Spanish Ministry of Economy and Competitiveness (under projects TEC2011-22480 and PRI-PIBIN-2011-1266. The work of Magno M.T. Silva was partially supported by CNPq under Grant 304275/2014-0 and by FAPESP under Grant 2012/24835-1. The work of Vítor H. Nascimento was partially supported by CNPq under grant 306268/2014-0 and FAPESP under grant 2014/04256-2. The work of Ali Sayed was supported in part by NSF grants CCF-1011918 and ECCS-1407712. We are grateful to the colleagues with whom we have shared discussions and coauthorship of papers along this research line, especially Prof. Aníbal R. Figueiras-Vidal.

AUTHORS

Jerónimo Arenas-García (jarenas@tsc.uc3m.es) received a Ph.D. degree in telecommunication technologies (honors) from Universidad Carlos III de Madrid, Spain, in 2004. After a postdoctoral stay at the Technical University of Denmark, he returned to Universidad Carlos III, where he is now a lecturer of digital signal and information processing. His research interests are focused on statistical learning theory, particularly in adaptive algorithms and advanced machine learning techniques. He has coauthored more than 75 papers on these topics. He has served as a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (SPS), associate editor of *IEEE Signal Processing Letters*, and president of the IEEE SPS Spain Chapter. He is a Senior Member of the IEEE.

Luis A. Azpicueta-Ruiz (azpicueta@tsc.uc3m.es) received the telecommunication engineer degree in 2004 from Universidad Politécnica de Madrid, Spain, and the Ph.D. degree in telecommunication technologies from Universidad Carlos III de Madrid, Leganés, Spain, in 2011. He is a lecturer of electroacoustics and acoustic engineering with the Department of Signal Theory and Communications at Universidad Carlos III, Madrid. His research interests are focused in the fields of adaptive signal processing and their applications, mainly in audio and acoustic processing. He is a Member of the IEEE.

Magno T.M. Silva (magno@lps.usp.br) received the B.S. degree in 1999, the M.S. degree in 2001, and the Ph.D. degree in 2005, all in electrical engineering from Escola Politécnica, University of São Paulo, Brazil. Since August 2006, he has been with the Department of Electronic Systems Engineering at Escola Politécnica, University of São Paulo, where he is currently an associate professor. In 2012, he worked as a postdoctoral researcher at Universidad Carlos III de Madrid, Leganés, Spain. He is currently an associate editor of *IEEE Signal Processing Letters*. His research interests include linear and nonlinear adaptive filtering. He is a Member of the IEEE.

Vítor H. Nascimento (vitor@lps.usp.br) is an associate professor at the University of São Paulo, Brazil. He was an associate editor of *IEEE Signal Processing Letters*, *IEEE Transactions on Signal Processing,* and *EURASIP Journal on Advances in Signal Processing*. He was a member of the IEEE Signal Processing Society (SPS) Signal Processing Theory and Methods Technical Committee and founded and chaired the São Paulo IEEE-SPS Chapter from 2010 to 2014. He received a 2002 IEEE SPS Best Paper Award. His research interests include signal processing theory and applications, microphone arrays, and adaptive and sparse estimation. He is a Senior Member of the IEEE.

Ali H. Sayed (sayed@ee.ucla.edu) is a distinguished professor of electrical engineering at the University of California, Los Angeles (UCLA), where he directs the UCLA Adaptive Systems Laboratory. An author of more than 460 scholarly publications and six books, his research involves several areas including adaptation and learning, network science, information processing theories, and biologically inspired designs. His work has been recognized with several awards including the 2014 Papoulis Award from the European Association for Signal Processing, the 2013 Meritorious Service Award and the 2012 Technical Achievement Award from the IEEE Signal Processing Society, the 2005 Terman Award from the American Society for Engineering Education, the 2003 Kuwait Prize, and the 1996 IEEE Donald G. Fink Prize. He was awarded several Best Paper Awards from the IEEE (2002, 2005, 2012, and 2014) and is a Fellow of the IEEE and the American Association for the Advancement of Science. He is recognized as a highly cited researcher by Thomson Reuters.

REFERENCES

[1] E. Hänsler and G. Schmidt (Eds.), *Topics in Acoustic Echo and Noise Control*. Berlin, Germany: Springer, 2006.

[2] B. Widrow, "A microphone array for hearing aids," *IEEE Circuits Syst. Mag.*, vol. 1, no. 2, pp. 26–32, 2001.

[3] C. R. Johnson, Jr., et al., "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.

[4] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[5] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Found. Trends Mach. Learn.*, vol. 7, nos. 4–5, pp. 311–801, 2014.

[6] A. H. Sayed, *Adaptive Filters*. Hoboken, NJ: Wiley, 2008.

[7] P. Andersson, "Adaptive forgetting factor in recursive identification through multiple models," *Int. J. Control*, vol. 42, no. 5, pp. 1175–1193, May 1985.

[8] M. Niedźwiecki, "Multiple-model approach to finite memory adaptive filtering," *IEEE Trans. Signal Processing*, vol. 40, no. 2, pp. 470–473, Feb. 1990.

[9] A. C. Singer and M. Feder, "Universal linear prediction by model order weighting," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2685–2699, Oct. 1999.

[10] S. S. Kozat and A. C. Singer "Multi-stage adaptive signal processing algorithms," in *Proc. SAM Signal Process. Workshop*, Cambridge, MA, Mar. 16–17, 2000, pp. 380–384.

[11] M. Martínez-Ramón, J. Arenas-García, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "An adaptive combination of adaptive filters for plant identification," in *Proc. 14th IEEE Int. Conf. Digital Signal Processing*, Santorini, Greece, July 2002, pp. 1195–1198.

[12] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 1078–1090, Mar. 2006.

[13] Y. Zhang and J. Chambers, "Convex combination of adaptive filters for a variable tap-length LMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 628–631, Oct. 2006.

[14] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Tracking properties of convex combinations of adaptive filters," *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, Bordeaux, France, July 17–20, 2005, pp. 109–114.

[15] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 3137–3149, July 2008.

[16] N. J. Bershad, J. C. M. Bermudez, and J.-Y. Tourneret, "An affine combination of two LMS adaptive filters – Transient mean-square analysis," *IEEE Trans. Signal Processing*, vol. 56, no. 5, pp. 1853–1864, May 2008.

[17] J. C. M. Bermudez, N. J. Bershad, and J.-Y. Tourneret, "Stochastic analysis of an error power ratio scheme applied to the affine combination of two LMS adaptive filters," *Signal Process.*, vol. 91, no. 11, pp. 2615–2622, Nov. 2011.

[18] R. Candido, M. T. M. Silva, and V. H. Nascimento, "Transient and steady-state analysis of the affine combination of two adaptive filters," *IEEE Trans. Signal Processing*, vol. 58, no. 7, pp. 4064–4078, July 2010.

[19] S. S. Kozat, A. T. Erdogan, A. C. Singer, and A. H. Sayed, "Steady-state MSE performance analysis of mixture approaches to adaptive filtering," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4050–4063, Aug. 2010.

[20] S. S. Kozat, A. T. Erdogan, A. C. Singer, and A. H. Sayed, "Transient analysis of adaptive affine combinations," *IEEE Trans. Signal Processing*, vol. 59, no. 12, pp. 6227–6232, Dec. 2011.

[21] T. Trump, "Transient analysis of an output signal based combination of two NLMS adaptive filters," in *Proc. Intl. Workshop on Mach. Learning for Signal Processing*, Kittilä, Finland, Aug. 29–Sept. 1, 2010, pp. 244–249.

[22] B. Jelfs, S. Javidi, P. Vayanos, and D. P. Mandic, "Characterisation of signal modality: Exploiting signal nonlinearity in machine learning and signal processing," *J. Signal Process. Syst.*, vol. 61, no. 1, pp. 105–115, Oct. 2010.

[23] J. Arenas-García, and A. R. Figueiras-Vidal, "Adaptive combination of proportionate filters for sparse echo cancelation," *IEEE Trans. Audio, Speech Lang. Processing*, vol. 17, no. 6, pp. 1087–1098, Aug. 2009.

[24] J. Ni and F. Li, "Adaptive combination of subband adaptive filters for acoustic echo cancelation," *IEEE Trans. Consumer Electron.*, vol. 56, no. 3, pp. 1549–1555, Aug. 2010.

[25] L. A. Azpicueta-Ruiz, M. Zeller, A. R. Figueiras-Vidal, J. Arenas-García, and W. Kellermann, "Adaptive combination of Volterra kernels and its application to nonlinear acoustic echo cancelation," *IEEE Trans. Audio, Speech Lang. Processing*, vol. 19, no. 1, pp. 97–110, Jan. 2011.

[26] L. A. Azpicueta-Ruiz, M. Zeller, A. R. Figueiras-Vidal, W. Kellermann, and J. Arenas-García, "Enhanced adaptive Volterra filtering by automatic attenuation of memory regions and its application to acoustic echo cancelation," *IEEE Trans. Signal Processing*, vol. 61, no. 11, pp. 2745–2750, June 2013.

[27] T. Trump, "A combination of two NLMS filters in an adaptive line enhancer," in *Proc. Intl. Conf. Digital Signal Processing*, Corfu, Greece, July 6–8, 2011, pp. 1–6.

[28] S. Lu, J. Sun, G. Wang, and J. Tian, "A novel GSC beamformer using a combination of two adaptive filters for smart antenna array," *IEEE Antennas Wireless Propagat. Lett.*, vol. 11, pp. 377–380, Apr. 2012.

[29] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini "Combined adaptive beamforming schemes for nonstationary interfering noise reduction," *Signal Process.*, vol. 93, no. 12, pp. 3306–3318, Dec. 2013.

[30] M. Ferrer, A. Gonzalez, M. de Diego, and G. Piñero, "Convex combination filtered-X algorithms for active noise control systems," *IEEE Trans. Audio, Speech Lang. Processing*, vol. 21, no. 1, pp. 156–167, Jan. 2013.

[31] N. V. George, and A. Gonzalez, "Convex combination of nonlinear adaptive filters for active noise control," *Appl. Acoust.*, vol. 76, pp. 157–161, Feb. 2014.

[32] L. Kuncheva, *Combining Pattern Classifiers*. Hoboken, NJ: Wiley, 2004.

[33] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth, "How to use expert advice," *J. ACM*, vol. 44, no. 3, pp. 427–485, 1997.

[34] V. Vovk, "Competitive on-line statistics," *Int. Statist. Rev.*, vol. 69, no. 2, pp. 213–248, 2001.

[35] H. Ozkan, M. A. Donmez, S. Tunc and S. S. Kozat, "A deterministic analysis of an online convex mixture of experts algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1575–1580, July 2015.

[36] V. H. Nascimento, M. T. M. Silva, L. A. Azpicueta-Ruiz, and J. Arenas-García, "On the tracking performance of combinations of least mean squares and recursive least squares adaptive filters," in *Proc. ICASSP*, Dallas, TX, Mar. 14–19, 2010, pp. 3710–3713.

[37] L. A. Azpicueta-Ruiz, A.R. Figueiras-Vidal, and J. Arenas-García, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proc. ICASSP*, Dallas, TX, Mar. 31–Apr. 4, 2008, pp. 3301–3304.

[38] M. Lázaro-Gredilla, L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "Adaptively biasing the weights of adaptive filters," *IEEE Trans. Signal Processing*, vol. 58, no. 7, pp. 3890–3895, July 2010.

[39] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "A new least squares adaptation scheme for the affine combination of two adaptive filters," in *Proc. Int. Workshop on Machine Learning for Signal Processing*, Cancun, Mexico, Oct. 16–19, 2008, pp. 327–332.

[40] T. Trump, "Output statistics of a line enhancer based on a combination of two adaptive filters," *Central Eur. J. Eng.*, vol. 1, no. 3, pp. 244–252, Sept. 2011.

[41] L. Shi, Y. Lin, and X. Xie, "Combination of affine projection sign algorithms for robust adaptive filtering in non-Gaussian impulsive interference," *Electron. Lett.*, vol. 50, no. 6, pp. 466–467, Mar. 2014.

[42] R. Harris, D. Chabries, and F. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 2, pp. 309–316, Apr. 1986.

[43] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Processing.*, vol. 40, no. 7, pp. 1633–1642, July 1992.

[44] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Processing*, vol. 41, no. 6, pp. 2075–2087, June 1993.

[45] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 631–639, Mar. 1997.

[46] M. H. Costa and J. C. M. Bermudez, "A robust variable step size algorithm for LMS adaptive filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)* Toulouse, France, May 14–19, 2006, vol. III, pp. 93–96.

[47] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "Plant identification via adaptive combination of transversal filters," *Signal Process.*, vol. 86, no. 9, pp. 2430–2438, Sept. 2006.

[48] V. H. Nascimento and R. C. de Lamare, "A low-complexity strategy for speeding up the convergence of convex combinations of adaptive filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 25–30, 2012, pp. 3553–3556.

[49] L. F. O. Chamon, W. B. Lopes, and C. G. Lopes, "Combination of adaptive filters with coefficients feedback," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 25–30, 2012, pp. 3785–3788.

[50] L. A. Azpicueta-Ruiz, M. Zeller, A.R. Figueiras-Vidal, and J. Arenas-García, "Least-squares adaptation of affine combinations of multiple adaptive filters," in *Proc. IEEE Int. Conf. Circuits and Systems*, Paris, France, 2010, pp. 2976–2979.

[51] J. Arenas-García, V. Gómez-Verdejo, and A. R. Figueiras-Vidal, "New algorithms for improved adaptive convex combination of LMS transversal filters," *IEEE Trans. Instrum. Measure.*, vol. 54, no. 6, pp. 2239–2249, Dec. 2005.

[52] C. M. Bishop, *Neural Networks for Pattern Recognition*. London: Oxford Univ. Press, 1995.

[53] L. A. Azpicueta-Ruiz, J. Arenas-Garcia, V. H. Nascimento, and M. T. M. Silva, "Reduced-cost combination of adaptive filters for acoustic echo cancelation," in *Proc. Int. Telecommunications Symp. (ITS)*, Sao Paulo, Brazil, Aug. 17–20, 2014, pp. 1–5.

[54] V. H. Nascimento, M. T. M. Silva, and J. Arenas-García, "A low-cost implementation strategy for combinations of adaptive filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, May 26–31, 2013, pp. 5671–5675.

[55] L. Li, Y. Xia, B. Jelfs, and D. P. Mandic, "Modelling of brain consciousness based on collaborative filters," *Neurocomputing*, vol. 76, no. 1, pp. 36–43, Jan. 2012.

[56] B. Jelfs, D. P. Mandic, and S. C. Douglas "An adaptive approach for the identification of improper signals," *Signal Process.*, vol. 92, no. 2, pp. 335–344, Feb. 2012.

[57] M. T. M. Silva and J. Arenas-García, "A soft-switching blind equalization scheme via convex combination of adaptive filters," *IEEE Trans. Signal Processing*, vol. 61, no. 5, pp. 1171–1182, Mar. 2013.

[58] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a "stop-and-go" decision-directed algorithm," *IEEE Trans. Commun.*, vol. 35, no. 9, pp. 877–887, Sept. 1987.

[59] V. Weerackody and S.A. Kassam, "Dual-mode type algorithms for blind equalization," *IEEE Trans. Commun.*, vol. 42, no. 1, pp. 22–28, Jan. 1994.

[60] F. C. C. De Castro, M. C. F. De Castro, and D. S. Arantes, "Concurrent blind deconvolution for channel equalization," in *Proc. ICC*, vol. 2, Helsinki, Finland, June 11–14, 2001, pp. 366–371.

[61] J. Mendes-Filho, M. D. Miranda, and M. T. M. Silva, "A regional multimodulus algorithm for blind equalization of QAM signals: Introduction and steady-state analysis," *Signal Process.*, vol. 92, no. 11, pp. 2643–2656, Nov. 2012.

[62] C. Paleologu, J. Benesty and S. Ciochina, *Sparse Adaptive Filters for Echo Cancelation*. San Raphael, CA: Morgan & Claypool, 2011.

[63] N. Kalouptsidis, G. Mileounis, B. Babadi and V. Tarokh, "Adaptive algorithms for sparse system identification," *Signal Process.*, vol. 91, no. 8, pp. 1910–1919, Aug. 2011.

[64] W. Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," *Proc. IEEE*, vol. 83, no. 6, pp. 958–981, June 1995.

[65] Guan Gui and Li Xu, "Affine combination of two adaptive sparse filters for estimating large scale MIMO channels," in *Proc. Annu. Summit and Conf. Asia-Pacific Signal and Information Processing Association (APSIPA)*, Chiang Mai, Thailand, Dec. 9–12, 2014, pp. 1–7.

[66] Guan Gui, S. Kumagai, A. Mehbodniya, F. Adachi, and Li Xu, "Variable is good: Adaptive sparse channel estimation using VSS-ZA-NLMS algorithm," in *Proc. Int. Conf. Wireless Communications and Signal Processing (WCSP)*, Hangzhou, China, Oct. 24–26, 2013, pp. 1–5.

[67] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Audio, Speech Lang. Processing*, vol. 8, no. 5, pp. 508–518, Sept. 2000.

[68] Y. Kopsinis, K. Slavakis and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted ℓ_1 balls," *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 936–952, Mar. 2011.

[69] Y. Chen, Y. Gu and A.O. Hero III, "Sparse LMS for system identification," in *Proc. ICASSP*, Taipei, Taiwan, Apr. 19–24, 2009, pp. 3125–3128.

[70] S. Zhang and J. Zhang, "Transient analysis of zero-attracting NLMS algorithm without Gaussian inputs assumption," *Signal Process.*, vol. 97, pp. 100–109, Apr. 2014.

[71] P. Di Lorenzo and A. H. Sayed, "Sparse distributed learning based on diffusion adaptation," *IEEE Trans. Signal Processing*, vol. 61, no. 6, pp. 1419–1433, Mar. 2013.

[72] B. K. Das, and M. Chakraborty, "Sparse adaptive filtering by an adaptive convex combination of the LMS and the ZA-LMS algorithms," *IEEE Trans. Circuits Syst. I*, vol. 61, no. 5, pp. 1499–1507, May 2014.

[73] L. A. Azpicueta-Ruiz, M. Lázaro-Gredilla, A.R. Figueiras-Vidal, and J. Arenas-García, "A block-based approach to adaptively bias the weights of adaptive filters," in *Proc. Int. Workshop on Machine Learning for Signal Processing*, Beijing, China, Sept. 18–21, 2011, pp. 1–6.

[74] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-García, and A. Uncini, "Functional link adaptive filters for nonlinear acoustic echo cancelation," *IEEE Trans. Audio, Speech Lang. Processing*, vol. 21, no. 7, pp. 1502–1511, July 2013.

[75] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Process. Mag.*, vol. 8, no. 3, pp. 10–26, July 1991.

[SP]

IEEE SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®

[REFLECTIONS]

Lawrence Rabiner

James L. Flanagan: A Scholar and a True Gentleman

James (Jim) L. Flanagan was born 26 August 1925 and grew up on a cotton farm in Greenwood, Mississippi. He graduated from high school in 1943 and completed his freshman year at Mississippi State University before joining the U.S. Army at age 18. He returned home about three years later, picked up his studies with the help of the G.I. Bill, and graduated with a B.S. degree in electrical engineering. He continued his education at the Massachusetts Institute of Technology (MIT), where he received the M.S. and doctor of science degrees in electrical engineering in 1950 and 1955, respectively. He was assistant professor of electrical engineering at Mississippi State University (1950–1952) and an electronic scientist at the Air Force Cambridge Research Lab (1955–1957). In 1957, he joined AT&T Bell Laboratories. He was with Bell Laboratories for 33 years, retiring in 1990 as director, Information Principles Research Laboratory. Subsequently, Jim served 15 years as a Board of Governors professor and university vice president for research at Rutgers University, New Jersey. Jim retired from Rutgers in 2005 at the age of 80.

James L. Flanagan

I first met Jim Flanagan in 1963 when I was given the opportunity to join his department as a fresh new MIT cooperative (co-op) education student. Jim quickly became my mentor, boss, colleague, and, most of all, friend. I knew him for more than 50 years, and it has been an honor and a privilege to know the man as well and as long as I did.

TECHNOLOGICAL CONTRIBUTIONS

It goes without saying that first and foremost, Jim Flanagan was an outstanding technologist. His research career spanned digital communications, speech processing, and acoustics.

His individual research included comprehensive modeling of basilar membrane motion in the inner ear, leading to useful engineering models of auditory signal processing. It also provided the theoretical basis and experimental development of a physiologically based model of vocal excitation for speech production, which provided a basis for advanced types of vocoders. Jim's early research included theoretical and practical studies of formant and phase vocoders along with perceptual experiments that quantified the relationship between hearing and speech models and led to an understanding of fundamental discrimination limits of the ear.

JAMES L. FLANAGAN (1925–2015)

Birthplace
Greenwood, Mississippi, United States

Education
B.S.E.E. degree, 1948, Mississippi State University; S.M. and Sc.D. degrees, 1950 and 1955, respectively, Massachusetts Institute of Technology

First Job
Assembling Christmas toys for J.C. Penney Co. in Greenwood

Established in 2002, the IEEE James L. Flanagan Speech and Audio Processing Award was named after Flanagan.

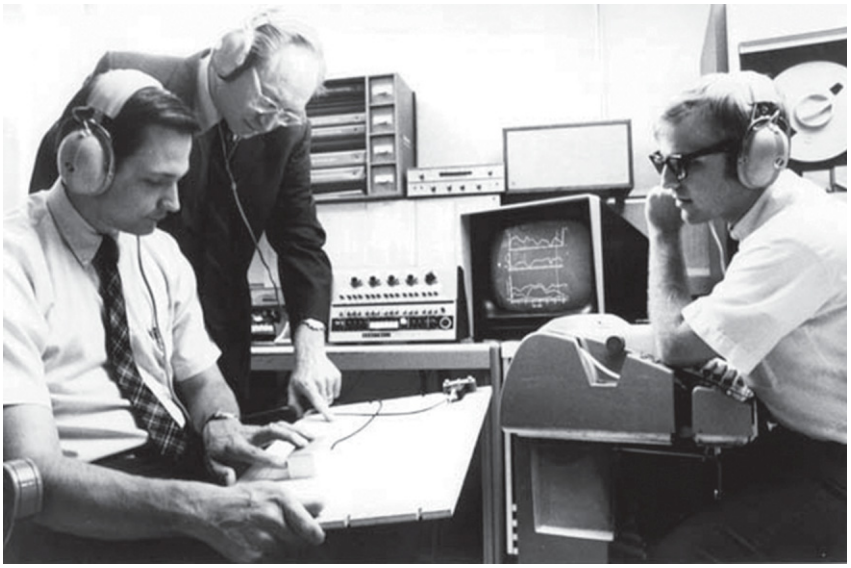
Major Awards
IEEE Medal of Honor, L.M. Ericsson International Prize in Telecommunications, National Medal of Science, Marconi International Fellowship, Gold Medal of the Acoustical Society of America; Member of National Academy of Sciences and National Academy of Engineering

Connections to IEEE Signal Processing Society (SPS)
President, IEEE SPS; Society Award, SPS, 1975; IEEE James L. Flanagan Speech and Audio Processing Award, sponsored by SPS

In His Own Words
- Oral History Interview with J.L. Flanagan, IEEE Global History Network, April 1997. [Online]. Available: http://ethw.org/Oral-History:James_L._Flanagan
- "Curious Science," *IEEE Signal Processing Mag.*, vol. 26, no. 3, pp. 10–36, May 2009. DOI: 10.1109/MSP.2009.93221

Read More About Him
- T. S. Perry, "Sultan of Sound," *IEEE Spectrum,* vol. 42, no. 5, pp. 44–48, May 2005
- Obituary, IEEE Inside Signal Processing eNewsletter, Sept. 2015.

Digital Object Identifier 10.1109/MSP.2015.2490758
Date of publication: 29 December 2015



James L. Flanagan, center, with fellow Bell Laboratories scientists, Ron Schafer (left) and Larry Rabiner (right), in 1970.

He led research efforts that resulted in the development of what are known as adaptive waveform coders. These coders automatically adjust to characteristics of the speech that they encode. His work on waveform coders provided the basis for many of the low bit rate coders presently used.

He was a pioneer in the field of speech and audio processing, and he had outstanding insights that changed the way we communicate both people to people and people to machines. Jim always had an eye on long-term goals, at the same time working on current technologies that greased the wheels for the many technical contributions made during his long career. By way of example, Jim set as his early research goal a tentimes reduction in bandwidth of the channel for human-to-human voice communication. When queried as to why such an aggressive goal, he would invariably reply that AT&T would derive the most benefit from attaining the goal.

Another example of Jim's ability to see into the future was his long-range goal of inventing ways to give a computer a mouth to speak and an ear to listen and learn. Much of the research that led to today's working synthesis and recognition systems originated in Jim's lab, thereby realizing his vision of customer care by machine-generated voice commands.

Perhaps the best validation of Jim's vision in this area was his paper "Computers that Talk and Listen; Man-Machine Communication by Voice," which was published in 1976 in *Proceedings of the IEEE*. This paper predicted user agents such as Siri and Cortana—39 years before their appearance in today's smartphones.

THE BEST VALIDATION OF JIM'S VISION WAS HIS 1976 PAPER "COMPUTERS THAT TALK AND LISTEN; MAN-MACHINE COMMUNICATION BY VOICE"—40 YEARS BEFORE SIRI AND CORTANA IN SMARTPHONES.

Finally, Jim had a clear vision of how a range of disparate multimedia technologies could work in unison to create something bigger and more useful as a whole. He called this idea the HuMaNet (Human-to-Machine Network) system integrating voice and image processing technologies along with advanced networking capability, leading to the concept of agent-based visual systems.

No set of reflections on the technical achievements of Jim Flanagan could be considered authoritative without the mention of Jim's classic and pioneering textbook, *Speech Analysis, Synthesis, and Perception*. This book has correctly been referred to as the "bible of speech processing" and has been used to resolve issues and provide insight in speech processing for the past 50 years and will undoubtedly continue in this role for many years to come.

Finally, Jim was the author or coauthor of more than 200 publications and more than 50 patents, including the design patent on the artificial larynx (providing speaking capability to people who had tracheotomies) and one patent on handling voice in a data network, which was a forerunner to voice over IP services.

In addition to the numerous technological contributions, Jim was widely recognized as an insightful technical speaker and writer. He had a real knack for getting to the essence of complex concepts and making them clear to an audience with a wide range of experience and technical expertise.

MANAGERIAL SKILLS

Jim spent most of his technical career managing other individuals as a department head and then as a lab director. He guided the careers of more than two generations of individuals who grew to positions of prominence in their own right. Starting with MIT co-op students, Jim was always available to discuss the options people had in their technical careers and give them well thought out advice as to how best to proceed with virtually any aspect of their technical careers. I was one of the early MIT co-ops, as were Joe Hall and Aaron Rosenberg, and all of us continued to work with Jim until his retirement as manager at Bell Labs.

Jim Flanagan was an outstanding judge of technical talent and thereby was able to attract and hire the best and the brightest individuals. Jim also realized that a lot of good talent was outside Bell Labs, and he continually thought of ways to bring such talent into Bell Labs to work alongside members of his department. Through such programs, Jim attracted Kenzo Ishizaka to work with him on vocal cord models for the human vocal tract, Fumitada

Itakura to work on speech recognition, and Sadaoki Furui to work on speaker verification, among dozens of other such collaborations.

A hallmark of Jim's managerial skills was the general feeling of the broad research community that every time one research challenge was met and solved by members of Jim's team, Jim was already prepared with a new set of challenges for the team, thereby illustrating his out-of-the-box thinking skills.

As a manager, Jim considered it his responsibility to call attention to the accomplishments of the researchers under his supervision. He did this for 33 years at Bell Labs and for 15 years at Rutgers. Although he contributed many ideas as part of his management style, he was scrupulous in giving credit where it was due to those who picked up on his ideas and carried them forward.

Jim inspired individuals to be the best that they could be, and he took interest in all aspects of their technical growth and technical maturity. He guided individuals by such basic principles such as "you never get a second chance to make a great first impression," generally followed by the sage advice to "do it right the first time."

Flanagan demonstrating an artificial larynx to high school students visiting Bell Labs in 1965.

OUTSTANDING SERVICE TO THE TECHNICAL COMMUNITY AND TO THE NATION

Jim Flanagan was a model in his role of providing outstanding service to the technical community and to the nation. While at Bell Labs, Jim served the nation at a critical time in history by being part of a Blue Ribbon committee that analyzed the infamous 18-min gap in the Watergate tapes, and by his service in the analysis of the final spoken words in the Challenger explosion.

Jim believed strongly in the role of service as a way of paying back the debt you accumulated by taking advantage of all that the various technical societies offered. The way to pay off that debt was to volunteer and assume leadership positions in both the IEEE (where Jim served as president of the Group on Audio and Electroacoustics) and the Acoustical Society of America, where he served as president. Jim also had a way of making sure that all the people that he mentored also assumed positions of leadership at the appropriate times in their technical careers.

RECOGNITION OF TECHNICAL ACHIEVEMENTS

Jim received numerous awards throughout his career, including the 1996 National Medal of Science, presented at the White House by U.S. President Bill Clinton, the Gold Medal of the Acoustical Society of America in 1986, and the IEEE Medal of Honor in 2006. Jim was elected to the National Academy of Engineering in 1978 and to the National Academy of Sciences in 1983.

For his groundbreaking contributions, the IEEE established the IEEE James L. Flanagan Speech and Audio Processing Award in 2002. This award is sponsored by the IEEE Signal Processing Society and is awarded to individuals for an outstanding contribution to the advancement of speech and/or audio signal processing.

Jim is survived by his wife Mildred Bell (Flanagan); sons Stephen, James, and Aubrey; and grandchildren Aubrey, James, Bryant, Antonia, and Hanks.

Jim Flanagan was a very special individual, and he will be missed both by his family and by his many friends.

ACKNOWLEDMENT

This article was written with input from Ron Schafer and Aaron Rosenberg.

AUTHOR

Lawrence Rabiner (lrr@jove.rutgers.edu) was with Bell Laboratories Research (1962–1995) and then AT&T Labs Research (1995–2002), where he served as vice president of research. Since 2002, he has been a distinguished professor of Rutgers University. He is a member of the U.S. National Academy of Engineering and National Academy of Sciences. [SP]

[lecture NOTES]

P. Murali Krishna
and T.P. Sameer Babu

Polyphase Channelizer Demystified

Since the late 1970s, numerous instances of design and applications of channelizers or channelization receivers have been widely reported. With the advent of high-speed field programmable gate arrays, new-generation architectures for implementation with reduced hardware have emerged in the past decade. However, these discrete processing techniques have evolved independently of the classical analog techniques, which were used until a generation earlier. These digital processing methods are not the digitized forms of their analog counterparts and are far superior in performance and flexibility in comparison to the legacy systems. Most of the reported literature in this area requires a level of analytical comprehension that may be too demanding for a student at an entry level in the subject. Keeping this in mind, we are, in this lecture note, presenting an alternative graphical analysis of a polyphase channelizer with a four-channel case study. This example can be easily generalized for an M-channels receiver case.

RELEVANCE

This lecture note provides a detailed pictorial representation of the channelization receiver operation in a manner that even an undergraduate student with an elementary knowledge of multirate digital signal processor (DSP) will find easy to grasp. In most traditionally oriented designs, it is observed that a digital receiver tries to emulate an existing analog receiver implementation that can be constrained in performance, and it usually results in the suboptimal utilization of the features in the currently available hardware. Here, we illustrate a typical case wherein, the adoption of improved processing techniques like multirate DSP will lead to better system design. It is imperative to explain alternate views of receiver architecture designs and novel applications of a DSP at an undergraduate stage itself. This lecture note emphasizes visualization and does not resort to a thorough analytical treatment to provide an intuitive understanding of the operations of a channelizer in line with popular titles such as [1]. Such an approach can complement material already available on this subject, which describes the underlying mathematical principles [2]–[5].

Digital Object Identifier 10.1109/MSP.2015.2477423
Date of publication: 29 December 2015

PREREQUISITE

The reader only requires a basic understanding of discrete-time Fourier signal processing and associate spectra due to the effect of sampling. An appreciation of the polyphase decomposition is also useful but not necessary.

DOWNSAMPLING AND POLYPHASE REPRESENTATION

One of the basic operations in multirate DSP is decimation, or downsampling, in which samples from a higher-sampled input sequence are dropped to yield a lower-sampled one. The downsampling factor M (M is a positive integer) is the ratio of the input to output sampling rates between the input sequence $x[n]$ and the downsampled sequence $x_d[n] = x[Mn]$. The downsampling process, while easy to visualize in the time domain, presents a different view in the frequency domain [6]. The discrete-time Fourier transform (DTFT) of the input sequence $x[n]$, which is a sampled version of a continuous time signal $x_c(t)$ with a sampling frequency $f_s = 1/T$, is

$$X(e^{j\omega}) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_c\left[j\left(\frac{\omega}{T} - \frac{2\pi k}{T}\right)\right]. \quad (1)$$

Similarly the DTFT of the output sequence $x_d[n]$, which is at a lower sampling rate of f_s/M, is

$$X_d(e^{j\omega}) = \frac{1}{\mathrm{MT}}\sum_{r=-\infty}^{\infty} X_c\left[j\left(\frac{\omega}{\mathrm{MT}} - \frac{2\pi r}{\mathrm{MT}}\right)\right]. \quad (2)$$

Substituting $r = m + kM$, where $0 \le m \le M-1$ and $-\infty < k < \infty$, (2) is simplified to obtain

$$X_d(e^{j\omega}) = \frac{1}{M}\sum_{m=0}^{M-1} \left\{\frac{1}{T}\sum_{k=-\infty}^{\infty} X_c\left[j\left(\frac{\omega - 2\pi m}{\mathrm{MT}} - \frac{2\pi k}{T}\right)\right]\right\}. \quad (3)$$

Comparing (3) with (1), the expression for the down sampled spectrum is given by

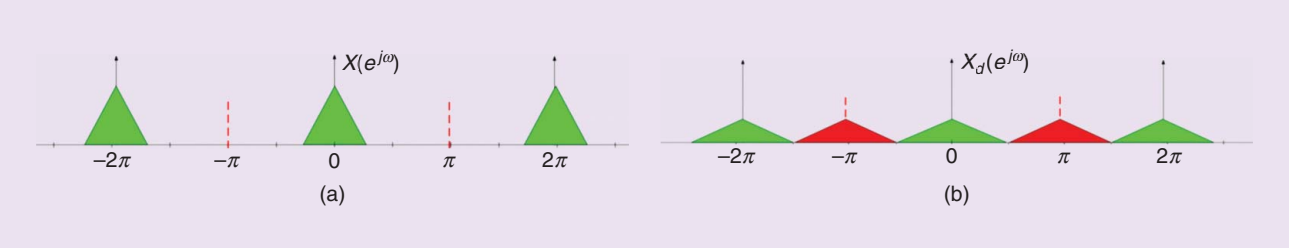
$$X_d(e^{j\omega}) = \frac{1}{M}\sum_{m=0}^{M-1} X\left(e^{j\frac{\omega - 2\pi m}{M}}\right). \quad (4)$$

The effect is seen to be twofold: the original spectrum is scaled by a factor of M and linearly combined with $(M-1)$ shifted versions of the same. For example, substituting $M = 2$ in (4), we get:

$$X_d(e^{j\omega}) = \frac{1}{2}\left[X(e^{j\frac{\omega}{2}}) + X\left(e^{j\left(\frac{\omega}{2} - \pi\right)}\right)\right]. \quad (5)$$

For a simple band-limited spectrum, the effect can be easily illustrated in a graphical form for $M = 2$ as shown in Figure 1(a) and (b). Note that the frequency axis in both cases are the same even though for the downsampled spectra, the new normalized Nyquist fold-over frequency will be

1053-5888/16©2016IEEE



[FIG1] **(a) The input spectrum to the decimator. (b) The output spectrum of decimator ($M = 2$).**

$(\pi/2M)$. The green shaded portion in Figure 1(b) is the original spectrum while the red shaded portion is the aliased, or shifted, spectrum. The effective computations involved in filtering the downsampled sequence can be reduced by the use of the Noble identity for downsampling, whose general form is given in Figure 2(a) [2], [3]. In the case of a unit delay, the application of the Noble identity for downsampling results in a phase shift in the frequency domain, which is illustrated in Figure 2(b).

For a sequence which is downsampled with a delay of p samples, the resulting spectrum is given by [3]:

$$X_d(e^{j\omega}) = \frac{1}{M}\sum_{m=0}^{M-1} X\left(e^{j\frac{\omega-2\pi m}{M}}\right)e^{-j\frac{2\pi pm}{M}}. \tag{6}$$

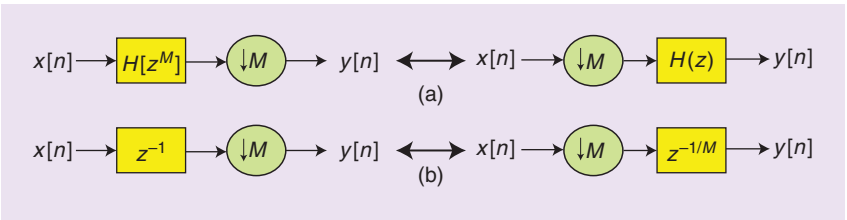
For $M = 2$ along with a unit delay $(p = 1)$, (6) gives

$$X_d(e^{j\omega}) = \frac{1}{2}\left[X\left(e^{j\frac{\omega}{2}}\right) - X\left(e^{j\left(\frac{\omega}{2}-\pi\right)}\right)\right]. \tag{7}$$

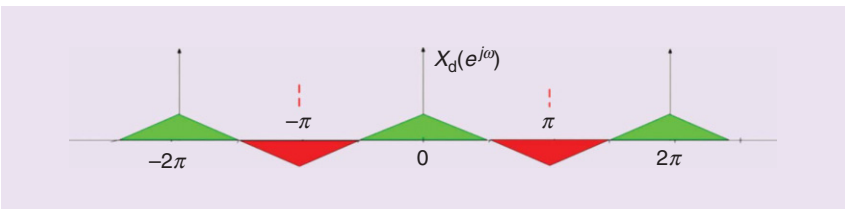
In the case of the spectrum given in Figure 1(a), the effect of an additional unit delay on the downsampling operation is indicated in Figure 3. As in the previous case, the green shaded portion is the original spectrum while the red shaded portion is the aliased or shifted spectrum. As a result of the additional unit delay, the shifted spectra is also flipped in amplitude. Similar results apply for upsampling operations as well (which we skip for brevity).

The very fact that downsampling introduces aliasing implies that the input signal must be further band limited prior to downsampling. This filtering operation performed prior to downsampling is not an appealing choice since it involves discarding some of the convolution results once they have been computed. An alternative is to only compute the necessary samples required and forwarding the same. This approach yields a very useful polyphase structure, which can be used to implement downsampling efficiently. Given an input sequence $u[n]$ and a filter impulse response $h[n]$, e.g., consider a finite impulse response (FIR) with length N, the output $x[n]$ after filtering can be written as the linear convolution given by

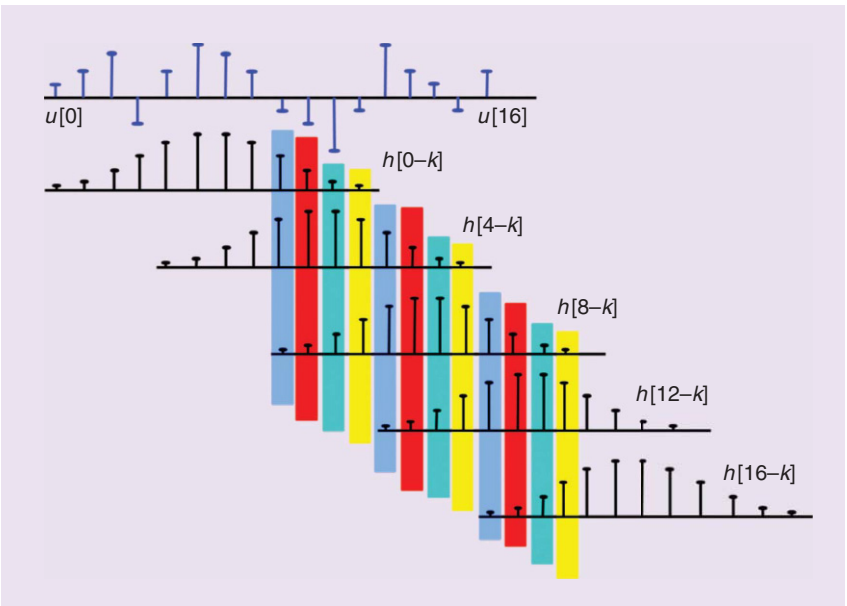
$$x[n] = u[n] * h[n] = \sum_{k=0}^{N-1} u[k]h[n-k]. \tag{8}$$



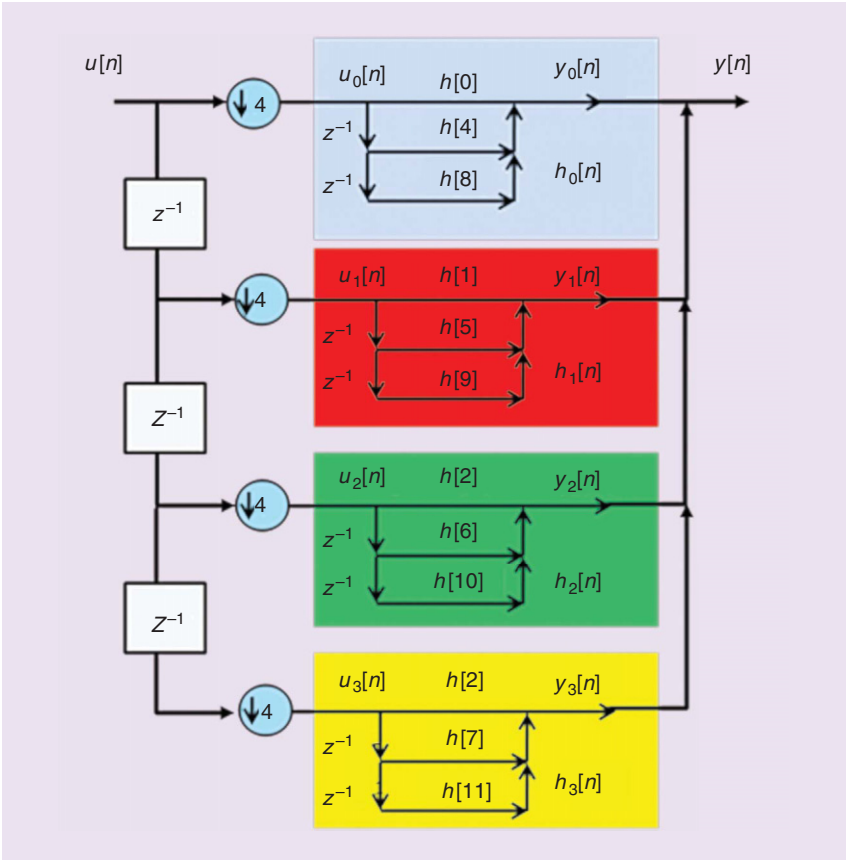
[FIG2] **(a) The Noble identity for downsampling operation. (b) The phase shift as a result of the Noble identity in downsampling.**



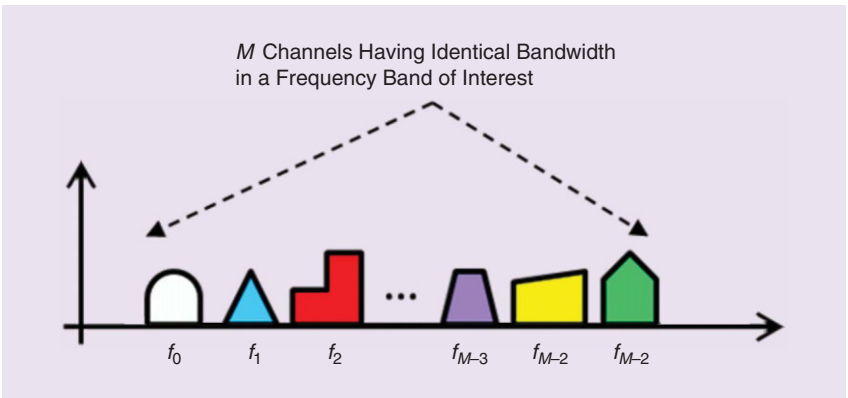
[FIG3] **The output spectrum after unit delay and decimation by two.**



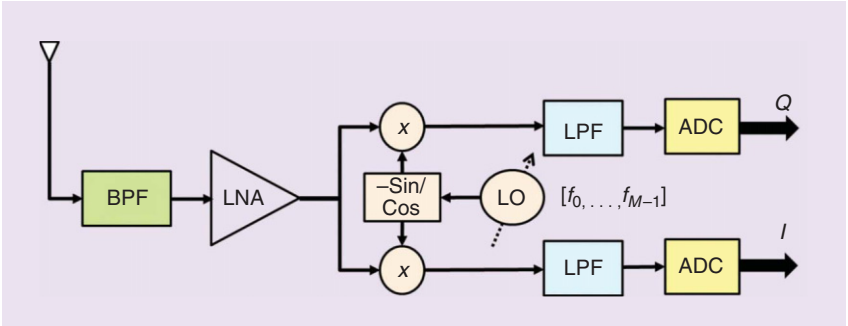
[FIG4] **The polyphase subsequences for downsampling.**



[FIG5] The polyphase downsampling implementation structure ($M = 4, N = 12$).



[FIG6] A multichannel reception scenario.



[FIG7] The receiver architecture with digital baseband processing.

The downsampled (downsampling factor M) output y is then given by $y[n] = x[Mn]$, and (8) can be modified as

$$y[n] = x[Mn] = \sum_{k=0}^{N-1} u[k]h[Mn-k]. \quad (9)$$

Figure 4 provides a graphical illustration of (9) for an FIR filter with impulse length $N = 12$ and a downsampling factor $M = 4$. The shifts for which the output is discarded by downsampling are not shown. It is apparent that certain groups of the filter coefficients (shaded by appropriate colors) are operating on specific sample indices of the input sequence $u[n]$.

By introducing an index mapping $k = m + rM$ where $(m = 0, 1, \ldots, M-1)$, (9) is modified as

$$y[n] = \sum_{m=0}^{M-1} \sum_{r=-\infty}^{\infty} u[rM+m]h[(n-r)M-m]. \quad (10)$$

Substituting, $u_m[r] = u[rM+m]$ and $h_m[r] = h[rM+m]$ in (10), we get

$$y[n] = \sum_{m=0}^{M-1} \sum_{r=-\infty}^{\infty} u_m[r]h_m[n-r] = \sum_{m=0}^{M-1} u_m[n] * h_m[n]. \quad (11)$$

The convolution followed by downsampling can be simplified and represented as the algebraic sum of several lower-order convolutions between subsequences of the input $u[n]$ and the filter coefficients $h[n]$. These subsequences are termed the *polyphase components* of $h[n]$ and $u[n]$. Given a sequence $h[n]$, the mth polyphase component of $h[n]$ is given by the relation

$$h_m[Mn] = h[Mn+m]. \quad (12)$$

The system transfer function $H(z)$ can be expressed in terms of the polyphase components. The sample index is mapped according to $n = Mn + k$ to obtain the expression

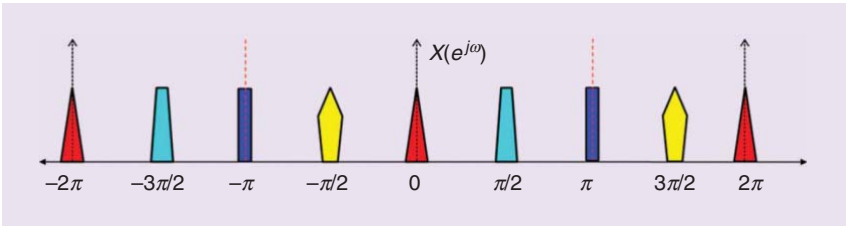
$$H(z) = \sum_{n=-\infty}^{\infty} \sum_{m=0}^{M-1} h[Mn+m]z^{-(Mn+m)} = \sum_{m=0}^{M-1} H_m(z^M)z^{-m}, \quad (13)$$

where $H_m(z^M)$ are the z-transforms of the polyphase components of $h[n]$. The polyphase downsampling implementation structure for $M = 4$ in the case of an $N = 12$ point FIR filter with impulse response $h[n]$ is illustrated in Figure 5. The input sequence is downsampled and fed into each of the sub filters. At any given point of time, only one of the subfilter sections will provide an output sample. This structure can also be interpreted as an input switch which is dispatching the samples from the input sequence to the appropriate subfilter for implementing the convolution.

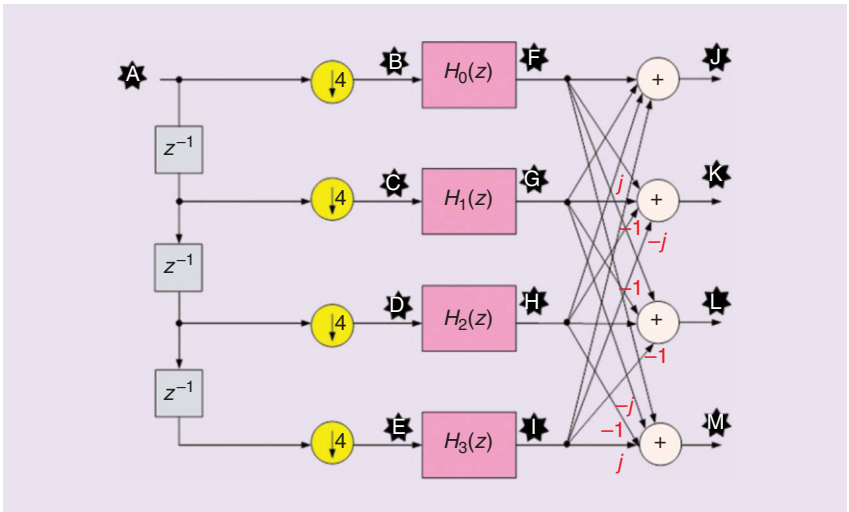
PROBLEM STATEMENT

The polyphase downsampling structure discussed in the previous section can be used to simplify the design and implementation of a multichannel receiver. Figure 6 illustrates the case of a frequency division multiplexing (FDM) system where M identical channels are used for communication in a frequency band of interest. In practice, several popular communication services (AM/FM broadcast) and wireless standards (Global System for Mobile Communications, code division multiple access) conform to such a situation. There can also be applications, such as cognitive radio, where spectrum monitoring and reception needs to be done for all the available channels in the given band of interest. If a conventional design strategy is adopted, M identical receivers are needed to receive and process the entire band simultaneously. Such an approach will result in replication of the hardware, which will result in increased cost, power, and space requirements. The architecture of a typical receiver front end for one of the receivers in such an implementation is shown in Figure 7.

The radio-frequency signal is bandpass filtered, amplified, and then downconverted to baseband where it is digitized and further demodulated using DSP techniques. To receive and process multiple channels simultaneously, the signal flow illustrated in Figure 8 is usually duplicated with different local oscillator parameters corresponding to the channels of interest.



[FIG8] **Complex $(I + jQ)$ spectrum of a four-channel FDM input signal.**



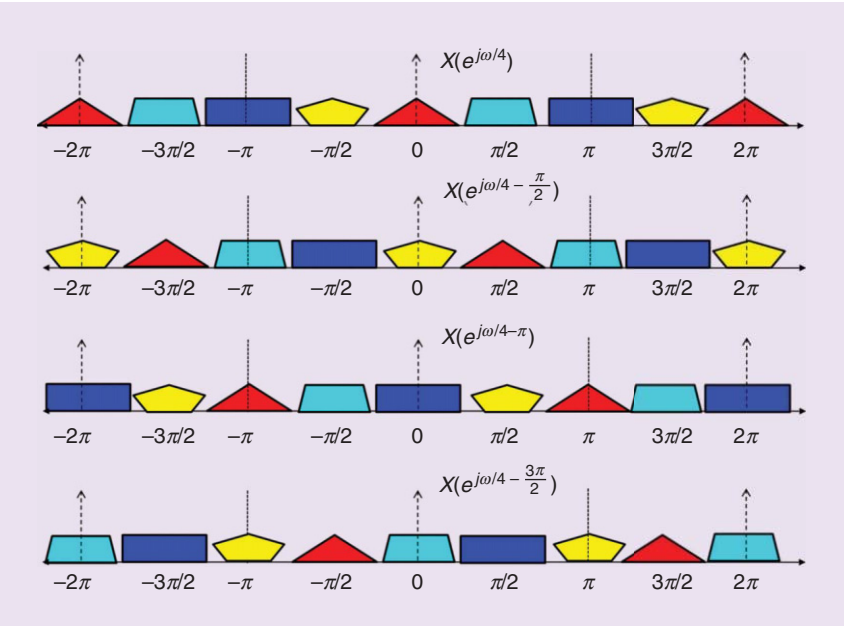
[FIG9] **The polyphase structure with signal flow indicated.**

POLYPHASE CHANNELIZER DESIGN

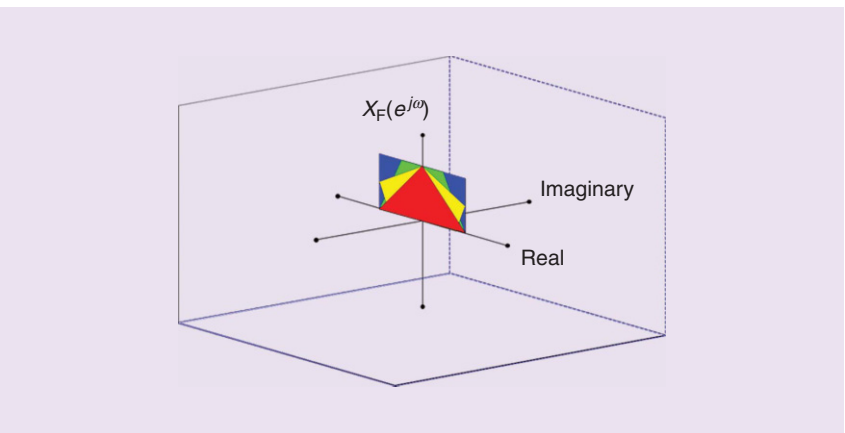
An elegant implementation for the multichannel receiver is possible by incorporating a polyphase downsampling structure followed by a suitable linear combiner. Each individual polyphase branch introduces a specific phase shift to the downsampled spectrum. This phase shift, together with the spectral shift introduced by the downsampling operation, is useful to achieve channel separation in the case of a multichannel input signal, e.g., FDM input signals [7]. Using appropriate phase shifters (complex multipliers), the outputs of the polyphase branches can be combined linearly to separate the individual channels in a manner similar to that of solving a set of consistent linear equations [8]. The sequence of these operations has already been analyzed in detail, and the results are available in published literature [2]–[9].

AN ELEGANT IMPLEMENTATION FOR THE MULTICHANNEL RECEIVER IS POSSIBLE BY INCORPORATING A POLYPHASE DOWNSAMPLING STRUCTURE FOLLOWED BY A SUITABLE LINEAR COMBINER.

In this lecture note, we aim to describe the solution graphically to provide a better intuitive insight to the whole process. Let us consider a four-channel FDM receiver. Figure 8 gives the sampled baseband $(I + jQ)$ complex spectrum obtained after the initial downconversion and digitization. There are four channels of equal bandwidth $(= \pi/16)$: 1) Channel 0 represented by the red triangle at $\omega = 0$, 2)



[FIG10] **Individual spectral components in (14).**



[FIG11] **Spectra at (*F*) obtained by filtering $X_B(e^{j\omega})$.**

Channel 1 by the green parallelogram at $\omega = \pi/2$, 3) Channel 2 by the blue rectangle at $\omega = \pi$, and 4) Channel 3 by the yellow pentagon at $\omega = -\pi/2$. All frequency values are normalized with respect to the sampling frequency such that $f_s/2 = \pi$.

This baseband signal is passed through a polyphase structure with the signal flow given by Figure 9. There are four polyphase paths starting with time delays and a downsampling with $M = 4$. The outputs of the polyphase paths are combined linearly using complex weighing factors to obtain the final outputs. For reference purposes, labels have been inserted at selected points in the signal flow from (A) to (M), with the input spectrum (A) depicted in Figure 8.

The polyphase filters $H_0(z), \ldots, H_3(z)$ are derived from a common prototype finite length low-pass filter with a normalized cutoff frequency equal to $\pi/4$. Since the downsampling factor is equal to $M = 4$, the new folding frequency is $\pi/4$. Denoting the input spectrum as $X(e^{j\omega})$, the spectrum at (B) is given by substituting $M = 4$ in (4)

$$X_B(e^{j\omega}) = \frac{1}{4}\Big[X(e^{j\frac{\omega}{4}}) + X\Big(e^{j\left(\frac{\omega}{4}-\frac{\pi}{2}\right)}\Big) + X(e^{j\left(\frac{\omega}{4}-\pi\right)}) + X\Big(e^{j\left(\frac{\omega}{4}-\frac{3\pi}{2}\right)}\Big)\Big]. \tag{14}$$

Ignoring the magnitude scaling factor, Figure 10 illustrates each of the shifted spectral components that contribute toward the composite spectrum defined in (14). The spectra from the individual channels get linearly combined as a result of the downsampling operation. Since there is no time delay preceding the first polyphase branch, all the spectral components are aligned with the real axis.

The subsequent low-pass filtering operation limits the spectrum to within the pass band of $\pm\pi/4$, thereby eliminating the aliased shifted spectral components. The resulting spectrum at location (F) is thus illustrated in Figure 11.

The expression for the spectra at points (C), (D), and (E) are obtained by substituting $M = 4$ in (6) with $p = 1, 2$, and 3 (time delays) for the additional phase shift term

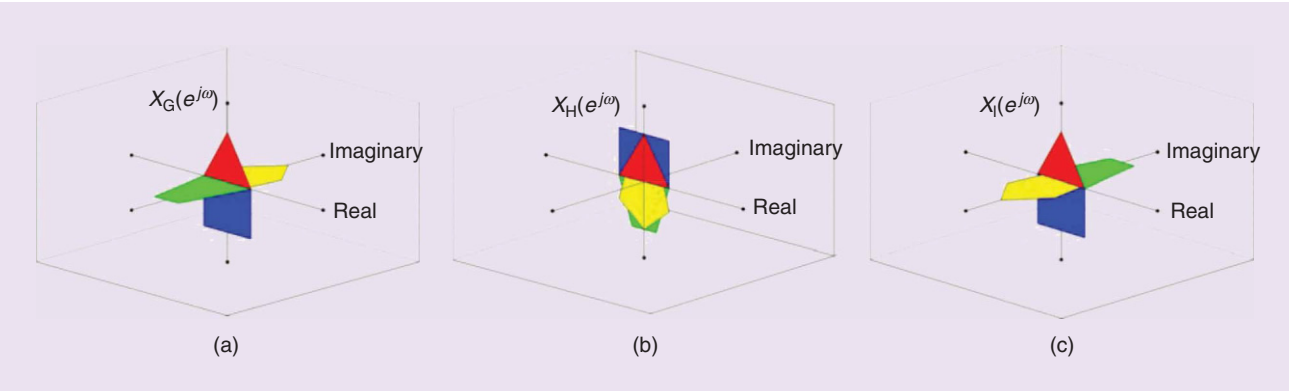
$$X_C(e^{j\omega}) = \frac{1}{4}\Big[X(e^{j\frac{\omega}{4}}) + jX(e^{j\left(\frac{\omega}{4}-\frac{\pi}{2}\right)}) - X(e^{j\left(\frac{\omega}{4}-\pi\right)}) - jX\Big(e^{j\left(\frac{\omega}{4}-\frac{3\pi}{2}\right)}\Big)\Big] \tag{15}$$

$$X_D(e^{j\omega}) = \frac{1}{4}\Big[X(e^{j\frac{\omega}{4}}) - X(e^{j\left(\frac{\omega}{4}-\frac{\pi}{2}\right)}) + X(e^{j\left(\frac{\omega}{4}-\pi\right)}) - X\Big(e^{j\left(\frac{\omega}{4}-\frac{3\pi}{2}\right)}\Big)\Big] \tag{16}$$

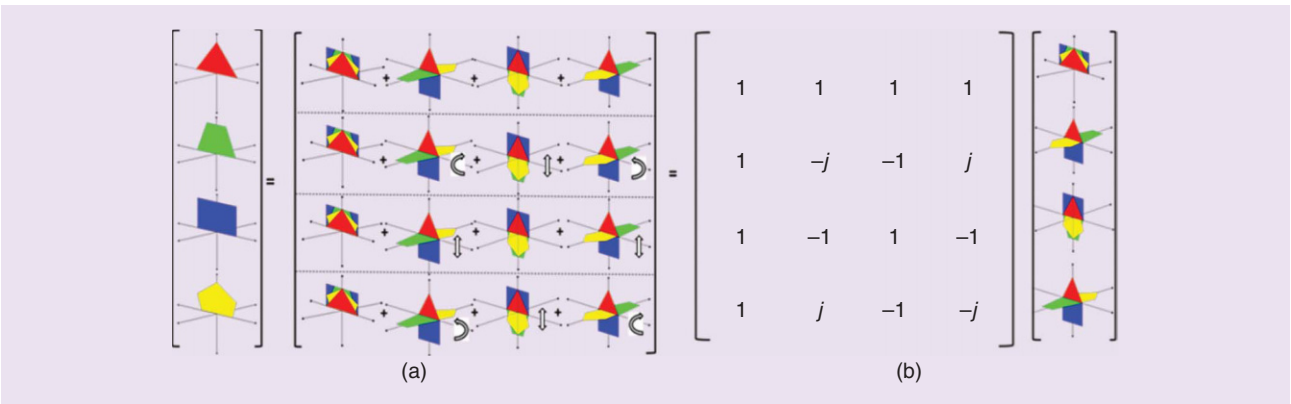
$$X_E(e^{j\omega}) = \frac{1}{4}\Big[X(e^{j\frac{\omega}{4}}) - jX(e^{j\left(\frac{\omega}{4}-\frac{\pi}{2}\right)}) - X(e^{j\left(\frac{\omega}{4}-\pi\right)}) + jX\Big(e^{j\left(\frac{\omega}{4}-\frac{3\pi}{2}\right)}\Big)\Big]. \tag{17}$$

The individual channels are combined linearly as described by (15)–(17). In two cases, there are components that are aligned with the imaginary axis. The spectra at locations (G), (H), and (I) are obtained after low-pass filtering and are illustrated in Figure 12. All channels are present in the spectrum after the polyphase filter. The relative orientation of each input channel is different due to the phase shift encountered in each individual branch.

Substituting mathematical rigor with geometrical intuition, Figure 13 exemplifies how it is possible to recover the individual channels. The individual spectra are rotated by $\pm\pi/2$ or π as indicated by the different arrows and are added along each row. It is clear that the resulting summations yield the desired individual



[FIG12] **Spectra after the low-pass filter in various polyphase branches.**



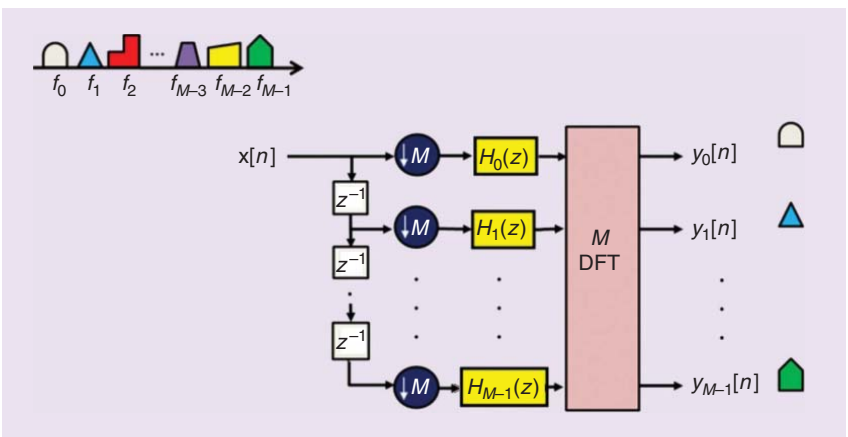
[FIG13] **(a) and (b) show the folding and summation used to recover the channels.**

channels. The same operation can be represented as a matrix operation as depicted in Figure 13(b). In this case, a rotation by $\pi/2$ in the counterclockwise (respectively, clockwise) direction is converted as j (respectively, $-j$). The rotation by π amounts to a sign change.

Equations (14)–(17) can be summarized in matrix notation

$$\begin{bmatrix} X_B(e^{j\omega}) \\ X_C(e^{j\omega}) \\ X_D(e^{j\omega}) \\ X_E(e^{j\omega}) \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix}\begin{bmatrix} X(e^{j\frac{\omega}{4}}) \\ X(e^{j(\frac{\omega}{4}-\frac{\pi}{2})}) \\ X(e^{j(\frac{\omega}{4}-\pi)}) \\ X(e^{j(\frac{\omega}{4}-\frac{3\pi}{2})}) \end{bmatrix}. \tag{18}$$

On inverting the coefficient matrix, the individual channels can be separated. This matrix is unitary and its inverse is given by the conjugate transpose, which yields the reconstruction matrix depicted in Figure 13 and which happens to be the four-point discrete Fourier transform (DFT) matrix. In other words, the DFT operation acts as a phase shifter to coherently combine the respective polyphase outputs and yield the respective channels. The beautiful relationship between the polyphase decomposition, the downsampling operation, and the DFT has made such an operation realizable. Interestingly, the dual of these operations, the inverse DFT followed by a polyphase upconverter, can be used to design efficient multichannel transmitter architectures. Case studies of efficient hardware realizations of multichannel receivers and transmitters that exploit this structure are available in the literature [9], [10].



[FIG14] **Generalized polyphase channelizer for *M* channels.**

Figure 14 summarizes the channelizer receiver architecture described in

this lecture note for a generic M-channel case [4], [8]. The connection between multichannel transceivers and multicarrier communications involving orthogonal frequency-division multiplexing also becomes evident when one investigates the topic further [11], [12]. A variant of the same filter-bank multicarrier modulation scheme has been proposed for applications in upcoming wireless standards [13], [14].

CONCLUSIONS

The link between a processing technique meant to simplify the downsampling implementation and an efficient realization of an FDM receiver through a DFT operation is not an easy concept to appreciate. The usual analytical approach is to start with the derivation of the polyphase decomposition of a modulated filter and later on generalize this relationship for center frequencies that are evenly spaced [15]. In this lecture note, a channelizer case study is visualized for a typical configuration to provide a more intuitive insight. Such graphical representations can be considered to complement the analytical approach without compromising the mathematical rigor. We hope that such illustrative examples will motivate signal processing students and practitioners enrich their courses and research results with similar graphical interpretations. To conclude, we quote the verses from T.S. Eliot: "We shall not cease from exploration and the end of all our exploring will be to arrive where we started and know the place for the first time."

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to Prof. Fred Harris of San Diego State University for teaching us polyphase channelizers during a course held in India in January 2010. It was these lectures that provided the crucial insight needed for understanding the beautiful theory linking polyphase decomposition, filter banks, and the DFT. We would like to thank the director of the Naval Physical and Oceanographic Laboratory for his support and encouragement.

AUTHORS

P. Murali Krishna (muralikrishna@npol.drdo.in) works with the Naval Physical and Oceanographic Laboratory at Cochin under the Defence Research and Development Organisation as a scientist. His areas of interest include digital signal processing and wireless communications.

T.P. Sameer Babu (sameerbabu@npol.drdo.in) works with the Naval Physical and Oceanographic Laboratory at Cochin under the Defence Research and Development Organisation as a scientist. His areas of interest include wireless communications and signal processing.

REFERENCES

[1] R. B. Nelson, *Proof without Words: Exercises in Visual Thinking*. Washington, D.C.: Mathematical Assoc. of America, 2000.

[2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[3] N. J. Fliege, *Multirate Digital Signal Processing*. New York: Wiley, 1993.

[4] F. J. Harris, *Multirate Signal Processing for Communication Systems*. Upper Saddle River, NJ: Prentice Hall, 2004.

[5] A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*. New York: Academic, 1992.

[6] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice Hall, 2009.

[7] M. Bellanger and J. Daguet, "TDM-FDM transmultiplexer: Digital polyphase and FFT," *IEEE Trans. Commun.*, vol. 22, pp. 1199–1205, Sept. 1974.

[8] F. J. Harris, C. Dick, and M. Rice, "Digital receivers and transmitters using polyphase filter banks for wireless communications," *IEEE Trans. Microwave Theory Tech.*, vol. 51, no. 4, pp. 1395–1412, Apr. 2003.

[9] K. C. Zangi and R. D. Koilpillai, "Efficient filter bank channelizers for software radio receivers", in *Proc. IEEE ICASSP*, 1998, pp. 3497–3500.

[10] F. J. Harris, D. Vuletic, and W. Lowdermilk "How to pack a room of analog FM modulators into a Xilinx FPGA," *DSP Mag.*, pp 16–21, Apr. 2007.

[11] Y. P. Lin, S. M. Phoong, and P. P. Vaidyanathan, *Filter Bank Transceivers for OFDM and DMT Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[12] P. P. Vaidyanathan, S.-M. Phoong, and Y.-P. Lin, *Signal Processing and Optimization for Transceiver Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[13] B. F. Boroujeny, *Signal Processing Techniques for Software Radios*. Raleigh, NC: Lulu Publishing House, 2011.

[14] B. F. Boroujeny, "OFDM versus filter bank multicarrier," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 92–112, May 2011.

[15] F. J. Harris, "On the relationship between multiratepolyphase FIR filters and windowed, overlapped, FFT processing," in *Proc. Asilomar Conf. Signals, Systems and Computers*, 1989, pp. 485–488.

[SP]

Society's flagship conferences, and 3) financial support for each SPS Chapter worldwide to send two practicing engineers to attend SPS conferences, specifically the IEEE Global Conference on Signal and Information Processing; the IEEE International Conference on Acoustics, Speech, and Signal Processing; and ICIP.

During the cocktail reception, attendees discussed how the SPS could assist young professionals in their careers. Examples included hosting tutorial sessions on topics of interest, organizing networking events and social functions that encourage young professionals to join industry and academia together, and facilitating the evolution of students to become world-class professionals. This first meeting was hopefully the first edition of a long series. Be sure to check out the SPS eNewsletter for the latest news and don't hesitate to contact Dr. Mahsa T. Pourazad at mahsa.pourazad@gmail.com for more information.

REFERENCE

[1] Online. Available: http://www.ieee.org/membership_services/membership/young_professionals/index.html

[SP]

[sp TIPS&TRICKS]

David Shiung, Ya-Yin Yang, and Chu-Sing Yang

Cascading Tricks for Designing Composite Filters with Sharp Transition Bands

This article presents novel tricks regarding cascading two digital filters to produce composite filters with very sharp transition bands for high-performance applications. The key point of the proposed tricks is to shape the magnitude frequency response of a prototype infinite impulse response (IIR) filter by a two-tap finite impulse response (FIR) filter using its nulls. In particular, we choose either a comb filter or a complementary comb filter of coefficients +1/−1, also called a *shaping* filter, to sharpen the transition bands of a prototype filter. The magnitude frequency response of the shaping filter compensates the Gibbs phenomenon commonly appearing in the passband edge and produces sharp transition bands for the cascaded filter. As compared to an equivalent IIR filter, the price paid is an additional comb/complementary comb filter of low complexity.

In contrast to conventional cascaded structures that combine two IIR filters in either cascade or parallel form, our new approach is relatively simple since we avoid using any multiplier. The tricks can be applied to design digital lowpass filters (LPFs), highpass filters (HPFs), and bandpass filters (BPFs).

CASCADE FILTER PROPERTIES

In general, an IIR filter has much sharper transition bands than an FIR filter of the same order. In this article, we focus on designing composite filters with sharp transition bands by cascading an FIR filter and an IIR filter as indicated in Figure 1. We improve the frequency response of a prototype filter using a shaping filter that can be either a comb filter or a complementary comb filter. The resultant transfer function of two cascaded filters is the product of those functions; or, we have [1]

$$H_{\text{cas}}(z) = H_{\text{sha}}(z)H_{\text{pro}}(z). \quad (1)$$

$H_{\text{sha}}(z)$ and $H_{\text{pro}}(z)$ denote the transfer function of the shaping filter and the prototype filter, respectively. $X(z)$ and $Y(z)$ denote the input sequence z-transform and the output sequence z-transform of the system, respectively. Owing to stability issues, an IIR filter is usually realized using an order of only two [1]. However, there are many types of IIR filters that happen to have an overshoot, also known as the *Gibbs* phenomenon, at the passband edge [1]. This overshoot is usually regarded as a drawback and is hard to avoid in filter design. One of our tricks is to use a comb/complementary comb filter for compensating the overshoot and to produce a composite filter with a flat passband response.

Since both the shaping filter and the prototype filter are linear, the frequency response of the cascade filter is unchanged if we reorder the cascade of filters. If one stage of shaping filter is not enough to meet the design specifications, we can cascade the prototype filter and two shaping filters of appropriate lengths.

SHAPING FILTER PROPERTIES

The transfer function of the shaping filter can be either

$$H_{\text{sha}}(z) = H_1(z) \triangleq 1 - z^{-K} \quad (2)$$

or

$$H_{\text{sha}}(z) = H_2(z) \triangleq 1 + z^{-K}, \quad (3)$$

where $K = 1, 2, 3, \ldots$. Clearly, both choices of $H_{\text{sha}}(z)$ are either symmetric or antisymmetric and thus have a linear phase response [2]. $H_1(z)$ is a comb filter. We call $H_2(z)$ its complementary filter. For both filters, their frequency responses can be easily derived if we substitute $z = e^{j\omega}$ into (2) and (3). The magnitude frequency responses of the two shaping filters for $K = 5$ and $K = 6$ are shown in Figure 2(a) and (b), respectively. Obviously, the number of nulls increases when K increases.

$H_1(\omega)$ and $H_2(\omega)$ can be used to shape some kinds of LPF, BPF, and HPF depending on the choice of filter order K. For $K = 5$, we can see in Figure 2(a) that $H_1(\omega)$ cannot be used to shape a LPF because it has a null at $\omega = 0$. On the other hand, $H_2(\omega)$ cannot be used to shape a HPF because it has a null at $\omega = \pi$. Other than that mentioned above, $H_1(\omega)$ can be used to shape BPFs and HPFs; and the filter order K can be used to determine the passband width. For $K = 6$, $H_1(\omega)$ can only be used to shape a BPF since there are nulls at both $\omega = 0$ and $\omega = \pi$. On the other hand, $H_2(\omega)$ can be used to shape all the lowpass, bandpass, and highpass filters since all the nulls are not located at $\omega = 0$ and $\omega = \pi$. We may tabulate the filter types that can be shaped by $H_1(\omega)$ and $H_2(\omega)$ in Table 1. We summarize Table 1 as the following two rules:

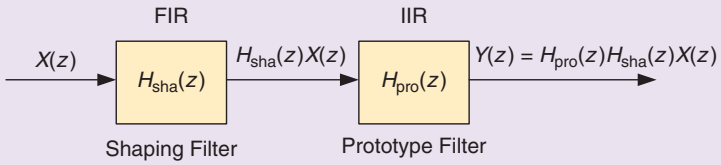
$$H_1(\omega) \text{ is for } \begin{cases} \text{HPF shaping, when } K \text{ is odd and } K \geq 1, \\ \text{LPF shaping, when } K \geq 2, \end{cases} \quad (4)$$

and

$$H_2(\omega) \text{ is for } \begin{cases} \text{LPF shaping, when } K \geq 1, \\ \text{HPF shaping, when } K \text{ is even and } K \geq 2, \\ \text{BPF shaping, when } K \geq 3. \end{cases} \quad (5)$$

Digital Object Identifier 10.1109/MSP.2015.2477420
Date of publication: 29 December 2015

IEEE SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®



[FIG1] **Combining two filters in a cascaded form.**

There are K zeros for every shaping filter, and we use its sharp transition band to reshape $H_{\text{pro}}(\omega)$. The zeros of $H_1(z)$ can be found by setting (2) to zero. Applying De Moivre's formula [3] to $z^K = 1$, we have

$$z_k = \sqrt[k]{1}\left[\cos\left(\frac{2\pi k}{K}\right) + j\sin\left(\frac{2\pi k}{K}\right)\right], \quad k = 0, 1, \ldots, K-1. \tag{6}$$

Comparing (6) with

$$z = e^{j\omega} = \cos(\omega) + j\sin(\omega),$$

we can locate zeros of $H_1(\omega)$ at

$$\omega_{1k} = \frac{2\pi k}{K},\ k = 0, 1, \cdots, K-1. \tag{7}$$

The null-to-null width of $H_1(\omega)$ defined as the width between consecutive nulls becomes

$$\Delta\omega_1 = \frac{2\pi}{K},\ K \geq 1. \tag{8}$$

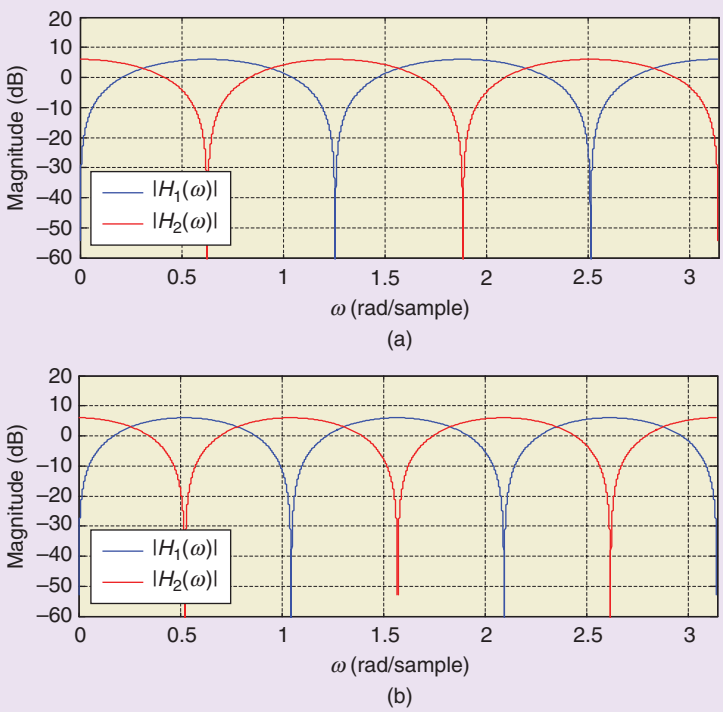
The null-to-null width places an upper bound on the passband width of the prototype filter. Similarly, the zeros of $H_2(z)$ can be found by setting (3) to zero.

The zeros of $H_2(\omega)$ are

$$\omega_{2k} = \frac{\pi + 2\pi k}{K},\ k = 0, 1 \ldots, K-1, \tag{9}$$

and the null-to-null width of $H_2(\omega)$ is

$$\Delta\omega_2 = \frac{2\pi}{K},\ K \geq 1, \tag{10}$$



[FIG2] **(a) The magnitude frequency responses of comb/complementary comb filter for $K = 5$. (b) The magnitude frequency responses of comb/complementary comb filter for $K = 6$.**

which is identical to that of $H_1(\omega)$.

From (7)–(10), we can formulate the passband centers of $H_1(\omega)$ and $H_2(\omega)$ as

$$\text{passband centers of } H_1(\omega) = \begin{cases} \frac{\pi}{K}, \frac{3\pi}{K}, \ldots, \frac{(K-1)\pi}{K}, & K \text{ is even}, \\ \frac{\pi}{K}, \frac{3\pi}{K}, \ldots, \frac{K\pi}{K}, & K \text{ is odd}, \end{cases} \tag{11}$$

and

$$\text{passband centers of } H_2(\omega) = \begin{cases} \frac{0}{K}, \frac{2\pi}{K}, \ldots, \frac{K\pi}{K}, & K \text{ is even}, \\ \frac{0}{K}, \frac{2\pi}{K}, \ldots, \frac{(K-1)\pi}{K}, & K \text{ is odd}. \end{cases} \tag{12}$$

When shaping $H_{\text{pro}}(\omega)$, the passband width of the prototype filter must be less than either $\Delta\omega_1$ or $\Delta\omega_2$, depending on which shaping filter is used. The shaping filter passband center should be made to coincide with that of the prototype filter. Thus, (8) and (10)–(12) are essential to our filter design. When we design the shaping filter, we first choose either the comb filter or the complementary filter according to Table 1. The filter length K is then determined such that (8) and (10)–(12) are satisfied.

DESIGN PROCEDURE

The primary advantage of IIR filters over FIR filters is that they generally achieve a set of design specifications with a much lower filter order [1]. The classic IIR filters such as Butterworth, type-1 and type-2 Chebyshev, and elliptic all approximate the ideal "brick-wall" filter in various ways [4]. A Butterworth filter can be designed to have a frequency response with a maximally flat passband and stopband; however, its transition band is the widest among the four types of filters previously mentioned. A Chebyshev filter can be designed to have a frequency response with ripples in the passband and maximally flat in the stopband (type-1), or to have a frequency response with ripples in the stopband and maximally flat in passband (type-2) [1]. It is worth noting that there is one ripple in the passband edge of an order-two Chebyshev type-1 LPF/HPF. Thus, a

Chebyshev type-1 filter of filter order two is ideal to be chosen as the prototype filter. An order-four Chebyshev type-1 filter has three ripples in the passband and is also an ideal candidate to be cascaded with the shaping filters. One trick for designing a LPF/HPF/BPF with a flat passband and a sharp transition band is to utilize the ripples in the passband to be cascaded with the shaping filter. In other words, the existing passband ripples of the prototype filter are flattened by the shaping filter; the transition band of the prototype filter is sharpened by the nulls of the shaping filter. Thus, a filter closer to an ideal "brick-wall" is realized. An elliptic filter has a frequency response with the sharpest transition band among all the candidate filters mentioned above but at the price that its frequency response has ripples in both the passband and the stopband. An order-two elliptic filter is also an ideal prototype filter to be combined with our shaping filter to produce a cascaded filter with a flat passband and a sharp transition band. We choose Chebyshev type-1 LPF as an example to illustrate the design tricks.

Figure 3 shows the magnitude frequency response of the comb (or complementary comb) filter that shapes the frequency response of a prototype filter. One peak magnitude of the filter is located at frequency ω_c. Since the null-to-null width of both $H_1(\omega)$ and $H_2(\omega)$ is $2\pi/K$, a zero next to ω_c is at $\omega_c + (\pi/K)$ (the other one is at $\omega_c - [\pi/K]$). From (2) and (3), we have $|H_i(\omega_c)| = 2$, $i = 1, 2$. This implies $|H_i(\omega_c)| = |1 \mp e^{-jK\omega_c}| = 2$ and we conclude that $e^{-jK\omega_c} = \mp 1$ for $i = 1, 2$, respectively. The magnitude frequency response of $H_i(\omega)$, at frequency deviation $\Delta\omega$, is

$$\begin{aligned} |H_i(\omega_c + \Delta\omega)| &= |1 \mp e^{-jK(\omega_c + \Delta\omega)}| \\ &= |1 \mp e^{-jK\omega_c} e^{-jK\Delta\omega}| \\ &= |1 + e^{-jK\Delta\omega}|. \end{aligned} \tag{13}$$

The attenuation in decibel with respect to $H(\omega_c)$ is

$$\Delta H_i(\Delta\omega) = -20\log_{10}\left(\frac{|1 + e^{-jK\Delta\omega}|}{2}\right). \tag{14}$$

[TABLE 1] FILTER TYPES THAT CAN BE SHAPED BY $H_1(\omega)$ AND $H_2(\omega)$ FOR $K = 1, \cdots, 9$.

K	1	2	3	4	5	6	7	8	9
$H_1(\omega)$	HPF	BPF	HPF, BPF	BPF	HPF, BPF	BPF	HPF, BPF	BPF	HPF, BPF
$H_2(\omega)$	LPF	LPF, HPF	LPF, BPF	LPF, BPF, HPF	LPF, BPF	LPF, BPF, HPF	LPF, BPF	LPF, BPF, HPF	LPF, BPF

We assume the design specifications that should be met by the cascaded filter of frequency response $H_{\text{cas}}(\omega)$ are 1) R decibels of peak-to-peak ripples in the passband and 2) cut-off frequency located at ω_p. When designing a BPF, ω_p becomes a two-element vector, denoted as $\omega_p = [\omega_{p1}\ \omega_{p2}]$, with passband width defined as $\omega_{p2} - \omega_{p1}$. Figure 4 shows two responses for $H_{\text{cas}}(\omega)$ to meet the design specifications. Also shown in Figure 4 is the response of a digital prototype filter; its peak-to-peak ripple is R' and its cut-off frequency is ω'_p. Obviously, its specifications are slightly over our design target. The peak magnitude response of the prototype filter occurs at $\omega'_{\max}$, while that of the cascaded filter is at $\omega_{\max}$. We have $R' > R$ and $\omega'_p > \omega_p$. The null-to-null width of the shaping filter defined in (8) and (10) should be larger than the prototype filter passband width so that we can utilize their sharp transition bands to reshape the frequency response of the prototype filter. In other words, we have the criterions

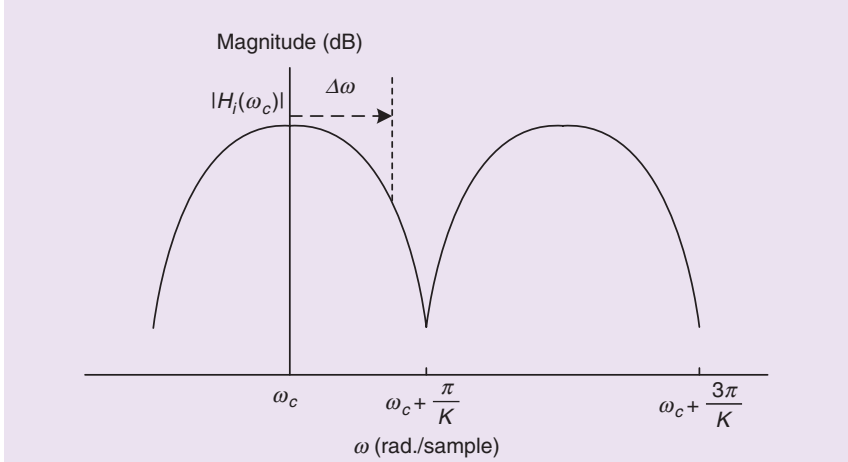
$$\text{prototype filter passband width} \le \begin{cases} \dfrac{\pi}{K}, & \text{if designing a LPF or HPF,} \\ \dfrac{2\pi}{K}, & \text{if designing a BPF.} \end{cases} \tag{15}$$

We use curve (I) in Figure 4 to present the theory behind our filter design procedures and assume that a total of two shaping filters are used to shape the prototype filter. The frequency responses of the two shaping filters are denoted as $H_{\text{sha1}}(\omega)$ and $H_{\text{sha2}}(\omega)$, respectively. The filter orders of the two shaping filters are K_1 and K_2, respectively. Designing a cascaded filter is in fact a problem of fitting the objective function

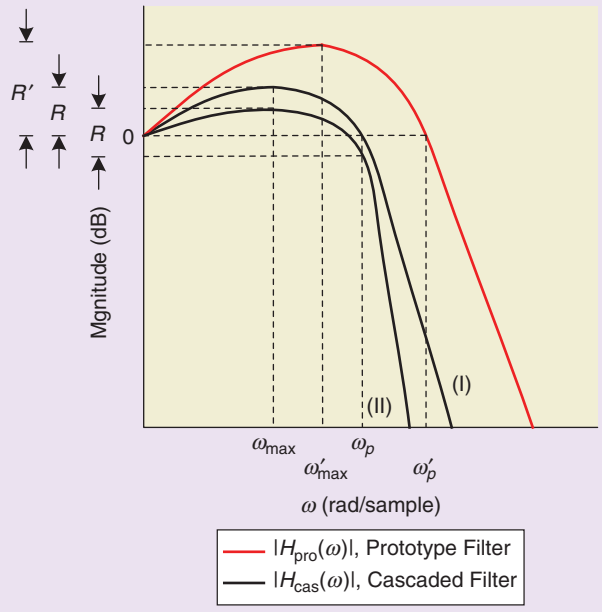
$$\begin{aligned} |H_{\text{cas}}(\omega_p)| &= |H_{\text{sha1}}(\omega_p)| \\ &\quad .|H_{\text{sha2}}(\omega_p)|.|H_{\text{pro}}(\omega_p)| \\ &= |H_{\text{pro}}(0)| \end{aligned} \tag{16}$$

using the following four parameters

1) filter order K_1 for $H_{\text{sha1}}(\omega)$
2) filter order K_2 for $H_{\text{sha2}}(\omega)$



[FIG3] The magnitude difference of $H_i(\omega)$ at frequency deviation $\Delta\omega$ from $\omega = \omega_c$ for $i = 1, 2$.



[FIG4] The magnitude frequency responses of a lowpass prototype filter and two cascaded filters.

3) magnitude of the peak-to-peak ripple of the prototype filter, R'

4) cut-off frequency of the prototype filter, ω_p'.

Among the four parameters, items 3) and 4) are the specifications of the prototype filer. Clearly, the four design parameters are mutually correlated with each other. It seems hard to find the complete solution set that satisfies (16). We thus turn to finding one feasible solution from an engineering-heuristics perspective. Observing Figure 2(a) and (b), we find that the magnitude frequency response at the passband edge of a comb/complementary-comb filter has an attenuation of around 1–4 dB for filter orders $K = 5, 6$. The passband peak-to-peak ripple of a practical filter is typically targeted at 1–4 dB. In addition, cascading the shaping filters with the prototype filter does narrow down its passband width to some degree. It is thus reasonable to choose $R' = 2R$ and $\omega_p' = 1.2\omega_p$ as the initial guess of our filter design problem for a maximally flat passband response. However, if only one shaping filter is cascaded with the prototype filter, we choose $R' = R$ and $\omega_p' = 1.1\omega_p$ as the initial guess. It is also highly probable there are multiple different designs that meet the design target. We use a digital Chebyshev type-1 filter as the prototype filter and we use two shaping filters for our design. The design procedures are summarized as follows: Initial settings: $R' = 2R$, $\omega_p' = 1.2\omega_p$.

STEP 1

Obtain the frequency response of prototype filter, $H_{\text{pro}}(\omega)$. The filter order for lowpass/highpass prototype filters is set to two and that for a bandpass prototype filter is set to four. $H_{\text{pro}}(\omega)$ can be easily obtained by transforming an analog Chebyshev type-1 filter using bilinear IIR filter design method [1], [4].

STEP 2

Determine the shaping filter type (comb filter or complementary comb filter) and its filter order K using Table 1 and (15). For example, the design specifications are designing a LPF with $R = 2$ dB and $\omega_p = \pi/10$ rad/sample. According to Table 1 and (15), the filter order K for the shaping filter, in fact, a complementary comb filter, should be lower than 10. However, when $K = 10$, there exists a notch at $\omega = \pi/10$ rad/sample, which prevents the cascaded filter from meeting the design specifications. Thus, we can test K from $K = 9$ to $K = 1$.

STEP 3

Check if the frequency response $H_{\text{cas1}}(\omega) = H_{\text{sha1}}(\omega)H_{\text{pro}}(\omega)$, or $H_{\text{cas2}}(\omega) = H_{\text{cas2}}(\omega) = H_{\text{sha2}}(\omega)H_{\text{sha1}}(\omega)H_{\text{pro}}(\omega)$ when determining K_2 for $H_{\text{sha2}}(\omega)$, is close to the design specification. Let the magnitude of the passband ripple and the cut-off frequency of $H_{\text{cas1}}(\omega)$ be $\hat{R}_1$ and $\hat{\omega}_{p1}$, respectively. If $\hat{R}_1 > 2R$, return to step 2 and try a smaller K. This is because a large $\hat{R}_1$ induces stability issues with the prototype filter (will be redesigned in step 4). Clearly, the resulting cut-off frequency is $\hat{\omega}_{p1} < 1.2\omega_p$ since it is reduced by the shaping filter. Table 2 shows the relation between K_1 and the corresponding $\hat{R}_1$ and $\hat{\omega}_{p1}$. Note that both cases $K_1 = 6$ and $K_1 = 5$ meet $\hat{R}_1 < 2R$. We just arbitrarily choose $K_1 = 6$. On the other hand, $K_1 = 7$ will result $\hat{R}_1 > 2R$. When determining K_2 for $H_{\text{sha2}}(\omega)$, $H_{\text{sha1}}(\omega)$ and $H_{\text{pro}}(\omega)$ are lumped together as a single filter and the procedure to determine K_1 is repeated again. To meet the design specifications, K_2 is selected only from $K = K_1 - 1, \ldots, 1$. Let the magnitude of the passband ripple and the cut-off frequency of $H_{\text{cas2}}(\omega)$ be $\hat{R}_2$ and $\hat{\omega}_{p2}$, respectively. If $\hat{R}_2 > 2R$, return to step 2 and try a relatively smaller K_2. Table 3 shows the relation between K_2 and the corresponding $\hat{R}_2$ and $\hat{\omega}_{p2}$. $K_2 = 4, 3, 2, 1$ can be arbitrarily selected among K_2—we just choose $K_2 = 3$. We then further trim the parameters R' and ω_p' for

[TABLE 2] THE RELATION BETWEEN K_1 AND ASSOCIATED $\hat{R}_1$ AND $\hat{\omega}_{p1}$.

K_1	R_1 (dB)	$\hat{\omega}_{p1}$ (rad/sample)
9	13.0	0.076 π
8	7.1	0.083 π
7	4.2	0.089 π
6	2.7	0.094 π
5	2.0	0.105 π

the prototype filter to meet our design specifications in step 4.

STEP 4

Check if the frequency response $H_{cas2}(\omega) = H_{sha2}(\omega)H_{sha1}(\omega)H_{pro}(\omega)$ meets the design specifications. If either $\hat{R}_2$ or $\hat{\omega}_{p2}$ does not meet the design target, we first make a one-step adjustment for ω_p' and then manually adjust R' so that $H_{cas2}(\omega)$ meets the design specifications. From Table 3, we already know that $\hat{R}_2 = 3.2$ dB and $\hat{\omega}_{p2} = 0.095\pi$ rad/sample for $K_1 = 6$ and $K_2 = 3$. In other words, the cut-off frequency $\hat{\omega}_{p2}$ is 0.005π rad/sample below the design target, while $\hat{R}_2$ is 1.2 dB larger than the design target. We first extend the resulting cut-off frequency by increasing ω_p' to 0.13π rad/sample and then manually adjust R' to around 4 dB. The resulting $\hat{R}_2$ and $\hat{\omega}_{p2}$ for various choices of $R' = 5$ are shown in Table 4. Because the design specifications are just met when $R' = 5$ dB, we finally set the parameters of the prototype filter to $R' = 5$ dB and $\omega_p' = 0.13\pi$ rad/sample.

STEP 5

Substitute $e^{jw} = z$ into $H_{pro}(\omega)$, $H_{sha1}(\omega)$, and $H_{sha2}(\omega)$, to obtain

$$H_{pro}(z) = \frac{\sum_{k=0}^{N} b(k)z^{-k}}{1 - \sum_{k=1}^{M} a(k)z^{-k}} \quad (17)$$

and

$$H_{sha}(z) = 1 \pm z^{-K}, \quad (18)$$

respectively.

STEP 6

By inspection, we can express the time-domain equations for the prototype filter and the shaping filter as

$$\begin{aligned} y(n) &= b(0)\,x(n) + b(1)\,x(n-1) \\ &\quad + \cdots + b(N)\,x(n-N) \\ &\quad + a(1)\,y(n-1) + a(2)\,y(n-2) \\ &\quad + \cdots + a(M)\,y(n-M) \end{aligned} \quad (19)$$

and

$$y(n) = x(n) \pm x(n-K), \quad (20)$$

respectively. The input and output sequence of each filter are denoted by $x(n)$ and $y(n)$. We can swap the order of filtering because both filters are linear and time invariant.

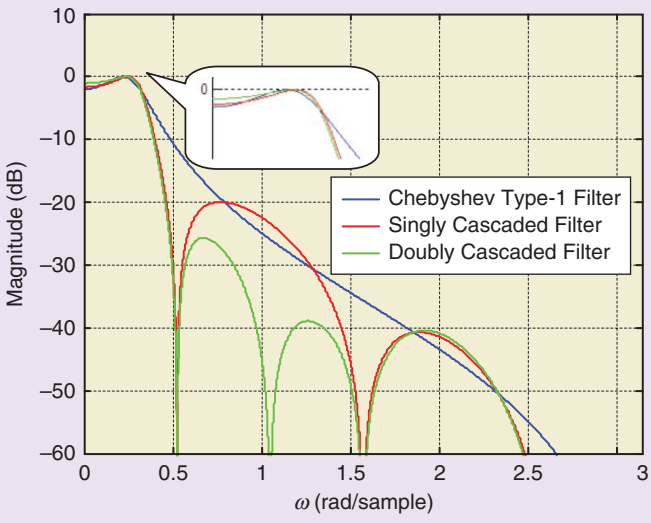
The trick shown in steps 1–6 lies in first determining the filter orders and then fitting the design specifications through trimming the prototype filter frequency response. This is a simple yet efficient way to solve (16). Steps 1–3 try to find a suitable shaping filters through an initial guess for the prototype filter. After the two shaping filters order are set, we proceed to trim the prototype filter so that the design specifications are met. If we want to achieve a sharper transition band, we can cascade more shaping filters. We present two examples regarding the design of a LPF and a BPF. Note that

[TABLE 3] THE RELATION BETWEEN K_2 AND ASSOCIATED $\hat{R}_2$ AND $\hat{\omega}_{p2}$ WHEN K_1 IS SET TO 6.

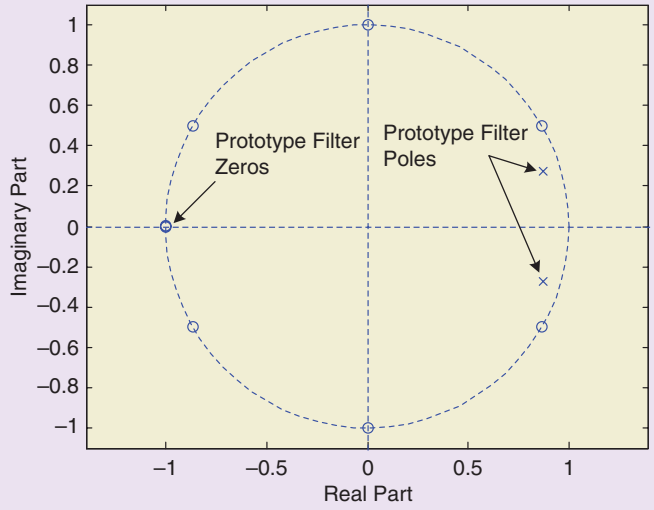
K_2	$\hat{R}_2$ (dB)	$\hat{\omega}_{p2}$ (rad/sample)
5	4.7	0.090 π
4	3.9	0.092 π
3	3.2	0.095 π
2	3.0	0.096 π
1	2.9	0.097 π

[TABLE 4] THE RELATION BETWEEN R' AND ASSOCIATED $\hat{R}_2$ AND $\hat{\omega}_{p2}$. K_1 IS SET TO 6 AND K_2 IS SET TO 3.

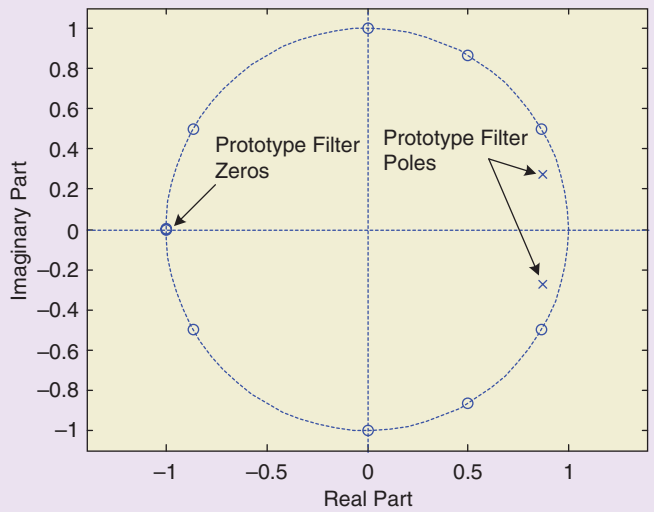
R' (dB)	$\hat{R}_2$ (dB)	$\hat{\omega}_{p2}$ (rad/sample)
3.8	2.3	0.098 π
4.0	2.2	0.098 π
4.2	2.1	0.099 π
4.4	2.1	0.099 π
4.6	2.1	0.099 π
4.8	2.0	0.099 π
5.0	2.0	0.100 π



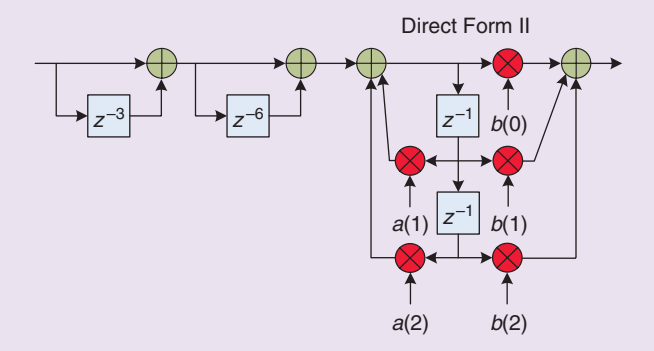
[FIG5] The magnitude frequency responses of $H_{cas1}(\omega)$ and $H_{cas2}(\omega)$ for the design specifications $R = 2$ dB and $\omega_p = \pi/10$ rad/sample. Also shown in the figure is a Chebyshev type-1 digital filter of the same specifications.



[FIG6] Poles and zeros of the singly cascaded LPF on the *z*-plane.



[FIG7] The poles and zeros of the doubly cascaded LPF on the *z*-plane.



[FIG8] The doubly cascaded filter structure. The prototype filter is in direct form II.

in the following examples, we normalize the peak magnitude-response to 0 dB for easy comparison.

DESIGN EXAMPLE: LPF

In this first example, we introduce the design of LPFs. The design specifications are $R = 2$ dB and $\omega_p = \pi/10$ rad/sample, which are the same as that illustrated in steps 1–6. Figure 5 shows the magnitude responses of two examples for LPFs (singly and doubly cascaded filter). Also shown in Figure 5 for comparison is a Chebyshev type-1 filter that meets the design specifications. The prototype filter peak-to-peak ripple is set to $R' = 5$ dB and the cut-off frequency is set to $\omega_p' = 0.13\pi$ rad/sample, which are obtained in step 4. The frequency response of the prototype filter is

$$H_{\text{pro}}(\omega) = \frac{0.013 + 0.026e^{-j\omega} + 0.013e^{-j2\omega}}{1 - 1.7384e^{-j\omega} + 0.8309e^{-j2\omega}}. \quad (21)$$

According to step 3 and Table 2, a complementary comb filter with $K_1 = 6$ is selected as the first shaping filter. The design guidelines in step 3 prohibit the choice of $K_1 = 7$ since it will result in $\hat{R}_1 > 2R$ and may raise the stability issues of the prototype filter redesigned in step 4. Thus, the singly cascaded filter frequency response is

$$H_{\text{cas1}}(\omega) = (1 + e^{-j6\omega}) \left(\frac{0.013 + 0.026e^{-j\omega} + 0.013e^{-j2\omega}}{1 - 1.7384e^{-j\omega} + 0.8309e^{-j2\omega}} \right). \quad (22)$$

A doubly cascaded filter is obtained by cascading in step 3 an additional complementary comb filter with order $K_2 = 3$. We obtain $\hat{R}_1 = 2$ dB and $\hat{\omega}_{p2} = 0.100\pi$ rad/sample as shown in Table 4. The doubly cascaded filter frequency response is

$$H_{\text{cas2}}(\omega) = (1 + e^{-j3\omega})(1 + e^{-j6\omega}) \left(\frac{0.013 + 0.026e^{-j\omega} + 0.013e^{-j2\omega}}{1 - 1.7384e^{-j\omega} + 0.8309e^{-j2\omega}} \right). \quad (23)$$

By substituting $e^{j\omega} = z$ into (22) and (23), we have

$$H_{\text{cas1}}(z) = (1 + z^{-6}) \left(\frac{0.013 + 0.026z^{-1} + 0.013z^{-2}}{1 - 1.7384z^{-1} + 0.8309z^{-2}} \right) \quad (24)$$

and

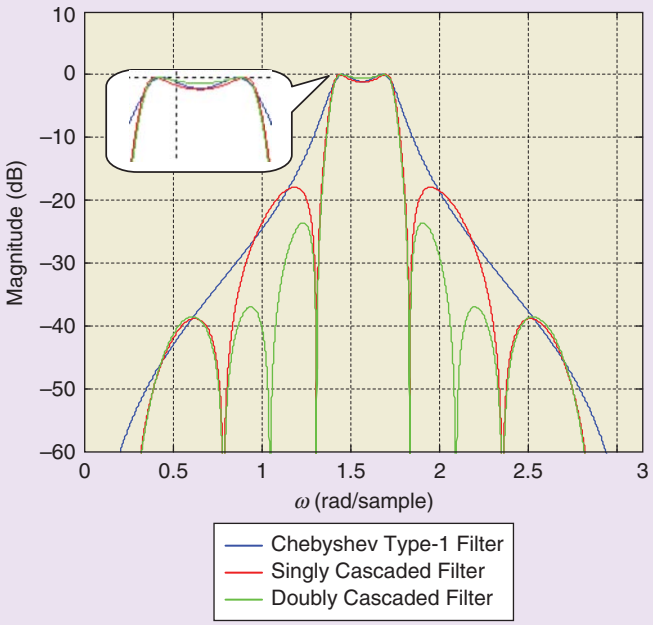
$$H_{\text{cas2}}(z) = (1+z^{-3})(1+z^{-6})\left(\frac{0.013+0.026z^{-1}+0.013z^{-2}}{1-1.7384z^{-1}+0.8309z^{-2}}\right). \quad (25)$$

Further observing Figure 5, we find that both the singly cascaded filter and the doubly cascaded filter achieve much sharper transition bands than that of the conventional Chebyshev type-1 filter. The difference in response can reach up to 20 dB around $\omega = 0.5$ rad/sample. This characteristic is especially useful when the intended signal happens to have a nearby interferer in the frequency domain. In addition, the shaping filter nulls can also be used to suppress narrowband noise such as in [1] and [5]. For the case of singly cascaded filter, the price paid for the sharp transition band is less attenuation around $\omega = 1.2$ rad/sample. This is because the prototype filter has a larger cut-off frequency than the competing Chebyshev type-1 filter and the shaping filter happens to have passband centers at the aforementioned frequencies. For the doubly cascaded filter case, lower attenuation occurred only around $\omega = 2$ rad/sample. This is because the second shaping filter has a notch around $\omega = 1$ rad/sample. The attenuation at $\omega = 2$ rad/sample is less than –40 dB and clearly has a minor impact on signal filtering. We can still cascade a third shaping filter if we intend to further shape the magnitude frequency response around $\omega = 2$ rad/sample.

Figure 6 shows on the z-plane the singly cascaded LPF pole-zero diagram. We can see that the prototype filter poles are inside the unit circle; this guarantees the stability of the filter. In addition, all the zeros are on the unit circle and two of them are placed near the prototype filter poles. This indicates that the transition band is sharpened by the shaping filter.

Figure 7 shows on the z-plane the doubly cascaded LPF pole-zero diagram. In contrast to Figure 6, two additional zeros are placed near the two poles in the z-plane to further sharpen the transition band.

The structure of the doubly cascaded filter is shown in Figure 8. The prototype filter is in direct form II. The shaping filters do not contain any multiplier and are simple for implementation.



[FIG9] **The magnitude frequency responses of the two BPFs in the example Chebyshev type-1 digital filter.**

DESIGN EXAMPLE: BPF

In this example, we present the design of BPFs. The design specifications are 1.1-dB peak-to-peak ripples and cut-off frequencies $\omega_p = [0.45\pi \;\; 0.55\pi]$ rad/sample. The singly cascaded filter and the doubly cascaded filter magnitude responses are shown in Figure 9. Also shown in Figure 9 is the response of a Chebyshev type-1 digital filter that meets the same specifications. We choose the cut-off frequencies $\omega_p{}' = [0.43\pi \;\; 0.57\pi]$ rad/sample and the peak-to-peak ripples $R' = 5$ dB for the prototype filter. Using bilinear transform filter design method with the help of [4], and choosing the complementary comb filter for frequency shaping, we get the singly and doubly cascaded BPF

$$H_{\text{cas1}}(z) = \left(1+z^{-12}\right) \times \left(\frac{0.015-0.03z^{-2}+0.015z^{-4}}{1+1.7127z^{-2}+0.8194z^{-4}}\right) \quad (26)$$

and

$$H_{\text{cas2}}(z) = \left(1+z^{-6}\right)\left(1+z^{-12}\right) \times \left(\frac{0.015-0.03z^{-2}+0.015z^{-4}}{1+1.7127z^{-2}+0.8194z^{-4}}\right), \quad (27)$$

respectively. Notice that the prototype filter order is four. We see that the singly and doubly cascaded filters both have sharper transition bands outside the cut-off frequencies than that of the Chebyshev type-1 digital filter. The singly cascaded filter stopbands achieve a larger attenuation than that of the Chebyshev type-1 digital filter except for the frequencies around $\omega = 0.5,\ 1.1,\ 2.1,\ 2.6$ rad/sample; meanwhile, the doubly cascaded filter achieves a larger attenuation than that of the Chebyshev type-1 digital filter except for the frequencies around $\omega = 0.5$ and 2.6 rad/sample. However, at these frequencies the magnitude frequency responses reach –40 dB and clearly have little impact on most applications.

CONCLUSIONS

In this article, we presented tricks for designing composite filters with very sharp transition bands. We first design a prototype IIR filter whose specifications in terms of passband ripples and passband width are slightly over the design objective. After cascading it with comb/complementary comb filter(s), we achieve a cascaded filter with even sharper transition bands than a comparable Chebyshev type-1 digital filter. The price paid is small since the comb/complementary comb filters are of low complexity

(continued on page 162)

[standards IN A NUTSHELL]

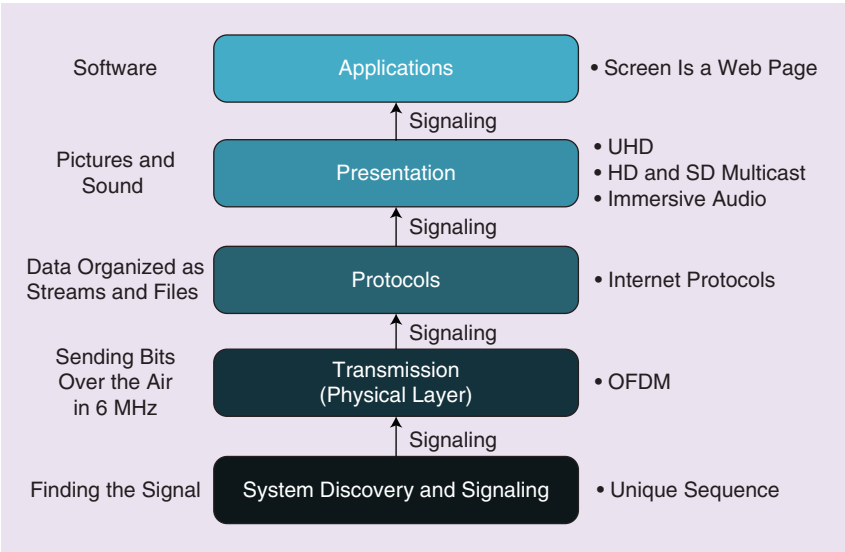
Richard Chernock and Jerry C. Whitaker

Next-Generation Broadcast Television: ATSC 3.0

Although the current digital terrestrial broadcast system (now referred to as ATSC 1.0) is widespread and quite successful, the basic component technologies have been in use for 20 years. Technology has evolved and viewer expectations have changed. Television broadcasters are under increasing pressure, due to regulatory and spectrum issues, as well as increasing competition for the viewer's attention. For these reasons, the Advanced Television Committee (ATSC) has been working on the next-generation broadcast television system, known as ATSC 3.0.

The goals of ATSC 3.0 are to improve the television viewing experience with higher audio and video quality; improved and more flexible reception on both fixed and mobile devices; and more accessibility, personalization, and interactivity. The ATSC is also addressing changing consumer behavior and preferences, providing TV content on a wide variety of devices. Furthermore, the ATSC is working to add value to the broadcasting service platform, extending its reach and adding new business models—all without the restriction of backward compatibility with the legacy system.

ATSC 3.0 is being built to last. Technology advances rapidly and so methods to gracefully evolve must be built into the core of the system. An essential capability is to signal when a layer or components of a layer evolve—including the ability to signal minor version changes and updates, and major version changes and updates. The goal of this flexibility is to avoid disruptive technology transitions, and to enable graceful transitions from one technology to another.



[FIG1] **Layers of the ATSC 3.0 system.**

The ATSC 3.0 standard enables system options for the broadcast industry (not just system parameter options) that include high robustness, spectrally efficient operating points, flexible configuration of operating modes, a robust bootstrap process that includes signaling basic system parameters, channel bonding options for spectrum sharing, and many other capabilities. It should be noted that these standards provide toolboxes for the broadcasters and specify how the tools are used. The selection of which tools are utilized and how depends upon the broadcasters' needs and business models.

Work on ATSC 3.0 has been divided into functional layers as illustrated in Figure 1 and described in the following sections. More information about the working groups for the different layers is available at http://atsc.org/subcommittees/technology-group-3/.

PHYSICAL LAYER

The physical layer for ATSC 3.0 embodies the theme of flexibility and future evolution by starting every physical layer frame with a bootstrap signal. This bootstrap is extremely robust, able to be received in very challenging radio-frequency conditions. Besides providing synchronization, the bootstrap also signals basic information about the technology used in the physical layer itself (major and minor version, which enables graceful evolution of the physical layer itself in the future) as well as an Emergency Alert Service wake-up bit, system bandwidth, time to the next frame of a similar service, and the sampling rate of the current frame.

The bootstrap is followed by a preamble, which carries the information needed to define the payload framing, including the information needed for the receiver to acquire the data frame. The remainder of the physical layer

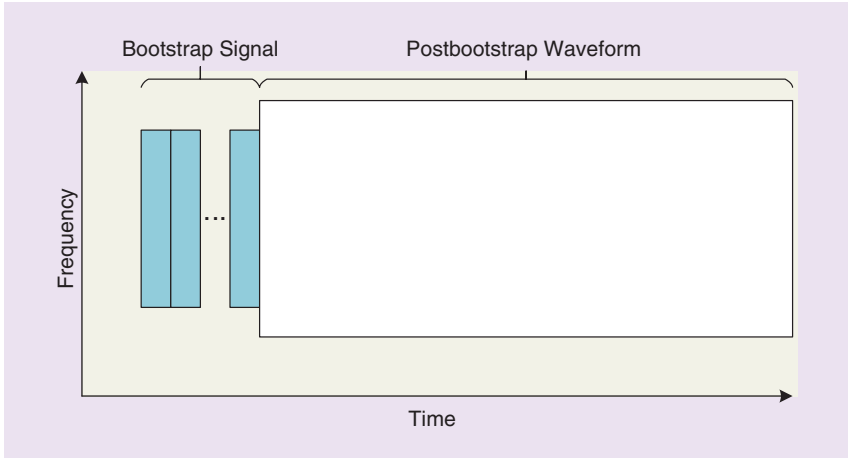
Digital Object Identifier 10.1109/MSP.2015.2479109
Date of publication: 29 December 2015

1053-5888/16©2016IEEE

structure is the data payload itself. This structure is illustrated in Figure 2.

By taking advantage of recent advances in modulation, coding, error correction, constellations, and multiplexing, the ATSC 3.0 physical layer offers performance in the bit interleaver, coding, and modulation (BICM) chain that is very close to the Shannon limit (the theoretical limit for the amount of information that can be carried in a noisy channel) as shown in Figure 3. Of particular note is the wide range of operating points available in the new system, especially considering that the current ATSC DTV system (described in ATSC Standard A/53 [1]) offers a single operating point of 19.39 megabytes/second, 15 dB C/N.

The ATSC 3.0 physical layer offers broadcasters the capability to operate in a robust/lower bit rate fashion for mobile services (lower left portion of the curve in Figure 3) and/or a less robust/higher bitrate fashion for services to large screens in the home (upper right portion of the curve). If desired, the broadcaster can also operate with a simultaneous mixture of types of services using either time division multiplexing or layer division multiplexing, or both. This allows broadcasters to construct their broadcast emission to support a variety of different business models and to experiment with new ones.

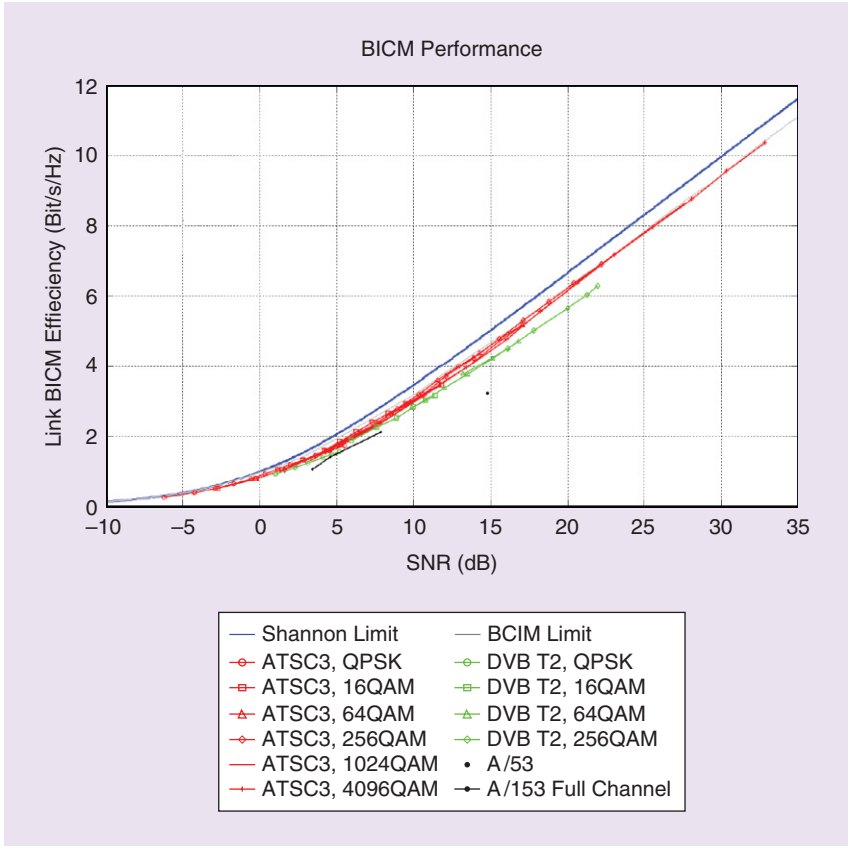


[FIG2] **Structure of the bootstrap/preamble/payload.**

TRANSPORT LAYER

The ATSC 3.0 transport layer is based on using Internet Protocol (IP) encapsulation for both streaming and file delivery, rather than the MPEG-2 Transport Stream (TS) encapsulation as is currently used in today's systems. When the original digital television (DTV) system (ATSC 1.0, A/53 [1]) was developed, the Internet was in its infancy. Entertainment content was not (yet) delivered over the Internet to homes. Broadcast television and digital cable were the primary entertainment delivery mechanisms, both based on MPEG-2 TS [2]. The Internet evolved very quickly (due to numerous factors, including the life cycle of computers and the ease of updating codecs and software over a two-way delivery system) and is now used for a large portion of entertainment content delivered to homes (for example, streaming services such as Netflix and Hulu). Rather than remaining an independent silo, ATSC 3.0 allows broadcasting to become part of the Internet. This allows the creation of new services and business models for broadcasters and enables evolution nearer the pace of how the Internet evolves. A conceptual view of the ATSC 3.0 protocol stack is given in Figure 4.

Besides similarities to the Internet, the use of IP transport enables incorporation of hybrid services, where components of services can be delivered by broadcast and broadband in a way that



[FIG3] **The ATSC 3.0 BICM performance.**

IEEE SignalProcessing | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | Qmags THE WORLD'S NEWSSTAND®

they can be synchronized. Woven throughout the structure of ATSC 3.0 is the basic notion that components of services can be delivered by different mechanisms (broadcast, broadband, or even predelivered by push in advance) and combined as needed to create services. This gives the broadcaster a large degree of flexibility and control over the services they offer.

An example use of this hybrid delivery might be a linear television services with multiple audio languages. Imagine a situation where the broadcaster's audience is composed of diverse ethnicities. There might two languages in widespread use, but enough people speaking other languages to want to include them as alternate audio tracks. With hybrid delivery, it will be possible to include the two more common language audio tracks in the broadcast delivery and make the others available over broadband. The viewer would have the choice of which language to listen to and it would be synchronized with the video, no matter what the delivery path might be.

The use of IP transport, coupled with another decision—to deliver "streaming" content as chunks of ISO Base Media File Format (ISOBMFF) files (similar to streaming delivery over the Internet) rather than continuous streams of bits—enables another business model that has been difficult to support in the past: the ability to do advertising (or content) insertion at the receiver in a personalized fashion. With elementary streams delivered over MPEG-2 TS, this was a very difficult task, requiring sophisticated buffer manipulation that was beyond the capabilities of an inexpensive receiver. By delivering file segments (along with an associated playlist), insertion in the receiver becomes relatively simple—essentially amounting to changing the pointer to the next segment to be played.

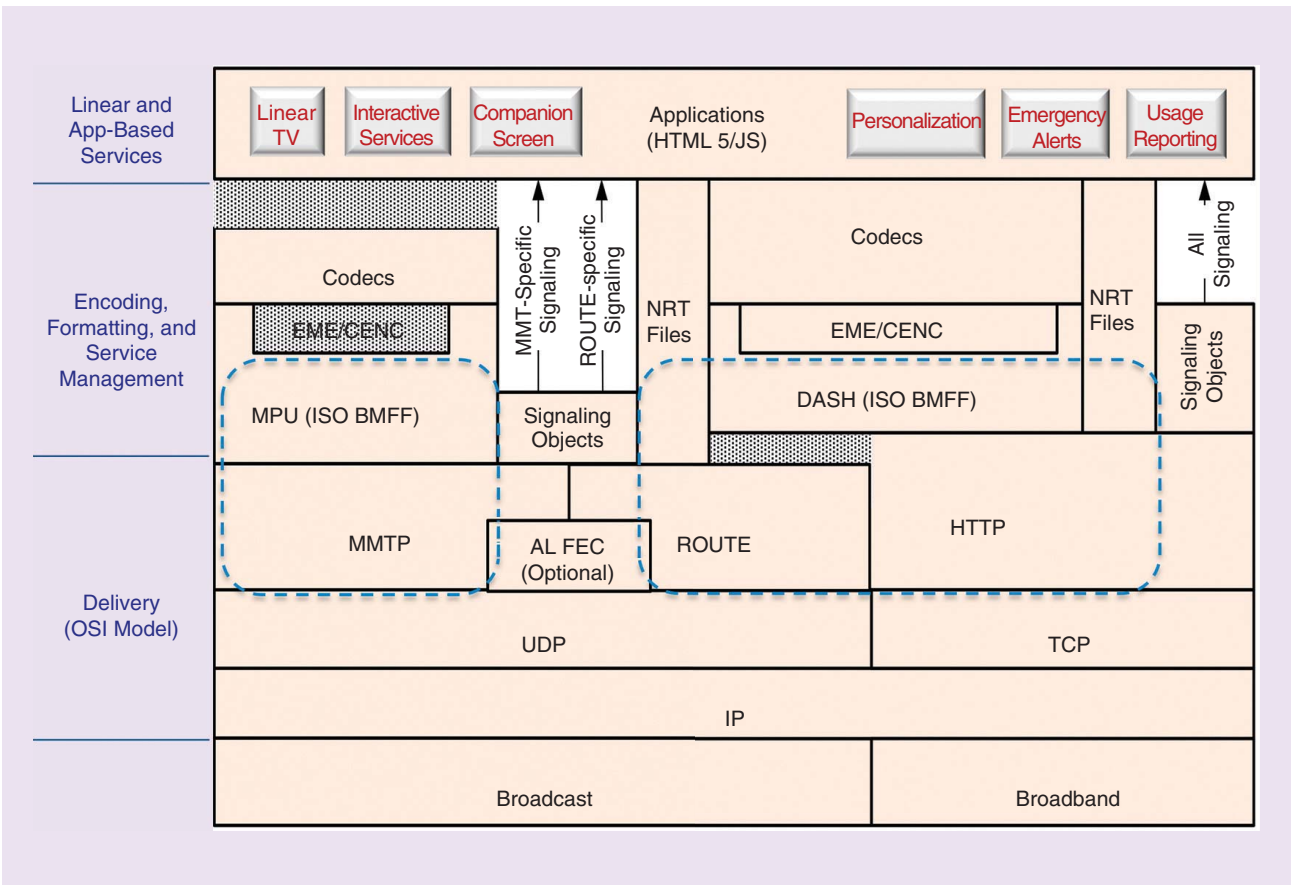
APPLICATIONS AND PRESENTATION LAYER

The applications layer encompasses video and audio coding, interactive features, accessibility, and other services.

VIDEO

The ATSC 1.0 system was a breakthrough when it emerged, offering high-definition television (HDTV) video (along with the option of multiple standard definition video simulcast). ATSC 3.0 offers the capability to move to ultrahigh definition (UHD) video. With current advances in video coding technology (HEVC/H.265 [3]), 4K video can fit within the available broadcast bandwidth. Similarly, multiple high-definition (HD) videos can also be delivered. While not currently in the design, the move to 8K video may be supported in the future (should it be desired).

However, besides more pixels, UHD also supports "better" pixels, specifically:



[FIG4] **A conceptual protocol stack for ATSC 3.0.**

IEEE SignalProcessing | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | Qmags THE WORLD'S NEWSSTAND®

- high dynamic range: 1,000-nit (or more) color grading rather than 100-nit
- wider color gamut: approaching Rec. 2020 rather than Rec. 709
- 10 bits/pixel (rather than 8)
- higher frame rate: 120 Hz.

All of these features provide significant improvements in picture quality, even at HD resolution.

Given the current screen sizes used in the home, as well as the typical viewing distance, many suggest that the other improvements listed above might offer a larger impact to the viewer than simply increasing the pixel count. Counterintuitively, 4K may be more impactful for smaller devices (such as tablets), which are often held at the optimal viewing distance.

Scalable video coding is being considered for inclusion in the video toolbox. Scalability allows additional efficiency and flexibility. For example, spatial scalability can be used to provide a base HD video layer sent over a robust physical layer pipe (PLP) aimed at mobile devices and an enhancement layer sent over a higher capacity (lower robustness) PLP, which provides UHD video at 4K. This approach offers a significant savings in overall bandwidth compared to transmitting both an HD video and a 4K video.

AUDIO

While the selection of audio technology for ATSC 3.0 is still under discussion at this writing, the new audio system will provide a number of new capabilities. Interactivity will be an important feature, providing the capability to personalize audio rendering. A broadcaster can choose to give the viewer control over a number of aspects of the audio, including the ability to adjust the dialog level (which can help with certain hearing impairments) or the ability to select which tracks of audio are heard (for example, being able to choose which announcer to listen to during a sporting event—or even to turn the announcer off altogether and simply listen to the crowd).

Television audio has evolved from the original monaural to stereo (in the analog days), to 5.1 surround (for ATSC 1.0), and now to full immersion in ATSC 3.0, which includes height information (typically 7.1+4). The additional level of immersion adds a significant degree of immediacy to the experience.

A major departure from the past is that in the ATSC 3.0 system a common audio stream is broadcast and rendered at the receiver appropriately. Besides decreasing the required bitrate, this local rendering allows dynamic range control (DRC) to be applied accurately for the type of receiver (for example home theater as opposed to a portable device with earbuds). Additionally, the receiving device can render correctly for the actual speaker configuration, including incorrectly positioned speakers.

INTERACTIVITY

Work is underway to develop a robust application runtime environment for ATSC 3.0. Supporting HTML5, the system will be based on HbbTV 2.0 [4] (with restrictions and extensions). Hybrid broadcast broadband TV (HbbTV) 2.0 was published in mid-2015 and more than 20 extensions are being documented, several based on documents written for ATSC 1.0 (ATSC standard A/105 [5]). Interactivity capabilities are expected to include:

- targeted ad insertion
- on-demand content launcher
- flexible use of display real estate (e.g., "screen mash-up")
- picture-in-picture, tickers, graphics, etc.
- T-commerce
- voting and polling
- games
- news- and sportsfeeds
- notifications and reminders.

ATSC is working with the Consumer Electronics Association (CEA) to bundle features into "profiles." One or more "universal profile(s)" are desired. The goal is to promote adoption and use by application developers and to simplify receiver implementations.

ACCESSIBILITY AND PUBLIC SERVICE

Accessibility functions are obviously a key element of any new broadcast system. The current work is initially focused on support for closed captioning in the new system. One point of discussion is whether closed captioning can be its own essence, rather than always inserted in the video stream. This approach provides for increased flexibility in carriage and presentation.

In an emergency, broadcasting is the most reliable means of disseminating information to the public. ATSC 3.0 will have support for advance emergency alerting, including a flag in the most robust part of the physical layer that can wake up sleeping receivers in case of an emergency. In a similar manner to what was developed for the Mobile Emergency Alert System (M-EAS) in ATSC Mobile DTV (A/153 Part 10 [6]), broadcasters will have the ability to send not only the basic alert information (text and audio), but also associated rich media (such as evacuation and weather maps, video clips, information on shelters, and recovery after the event), which can be geographically targeted.

MOVING FORWARD

The transition to a new technology can be difficult for any industry. For ATSC 3.0, the transition challenges can be broadly divided into two groups:

- The spectrum challenge: In most countries, additional spectrum will not be available for a second service. One possible solution might be for stations within a region to channel-share to create a dual service during transition.
- The receiver challenge: New devices are needed for consumers. Options include a "stick tuner" (dongle, thumb drive, etc.) for current flat panel TVs, and/or set-top converters. An industry-driven campaign to include a tuner in portable devices (tablets, laptops, smartphones, etc.) will help move the transition forward. In addition, adoption by other countries will foster product development.

In summary, ATSC 3.0 represents a significant step forward in capabilities for a broadcast television system. It provides a set of flexible capabilities for broadcasters that enable new services and new business cases. The concepts of flexibility, extensibility, and scalability

are in the core of the system and will allow graceful evolution over a long period of time. For more information, visit http://www.atsc.org.

AUTHORS

Richard Chernock (rchernock@trivenidigital.com) is the chief science officer, Triveni Digital, Inc.

Jerry Whitaker (jwhitaker@atsc.org) is the vice president for standards development, Advanced Television Systems Committee.

REFERENCES

[1] *ATSC Digital Television Standard*, Parts 1–6, Doc. A/53, Advanced Television Systems Committee (ATSC), 2009–2014.

[2] *Information Technology—Generic Coding of Moving Pictures and Associated Audio Information: Systems*, ISO/IEC 13818-1: 2007 + Amendment #3.

[3] *Information Technology—High Efficiency Coding and Media Delivery in Heterogeneous Environments—Part 2: High Efficiency Video Coding*, ISO/IEC 23008-2: 2015, ISO/IEC.

[4] *HbbTV 2.0 Specification*, HbbTV, 1 Aug. 2015.

[5] *ATSC Candidate Standard-Interactive Services Standard*, Doc. A/105, Advanced Television Systems Committee (ATSC), 8 April 2015.

[6] *Mobile Emergency Alert System*, Doc. A/153 Part 10:2013, Advanced Television Systems Committee (ATSC), 11 Mar. 2013.

[SP]

sp TIPS&TRICKS (continued from page 157)

and contain no multiplier. We present two design examples, which contain LPFs and BPFs, to validate the usefulness of the tricks. In the transition bands, the improvement in magnitude can easily reach 20 dB as compared to a conventional Chebyshev type-1 digital filter.

AUTHORS

David Shiung (davids@cc.ncue.edu.tw) is an associate professor at the National Changhua University of Education, Changhua, Taiwan.

Ya-Yin Yang (ivyyang64@gmail.com) is a postdoctoral fellow at the National Cheng Kung University, Tainan, Taiwan.

Chu-Sing Yang (csyang@mail.ee.ncku.edu.tw) is a professor at the National Cheng Kung University, Tainan, Taiwan.

REFERENCES

[1] R. G. Lyons, *Understanding Digital Signal Processing*. Boston: Pearson Education, 2011.

[2] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1999.

[3] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1965.

[4] MathWorks. (2014, Mar.) IIR filter design. [Online]. Available: http://www.mathworks.com/help/signal/ug/iir-filter-design.html

[5] D. Shiung, H. Ferng, and R. Lyons, "Filtering tricks for FSK demodulation," *IEEE Signal Processing Mag.*, vol. 22, no. 3, pp. 80–82, May 2005.

[SP]

special REPORTS (continued from page 13)

reducing device visibility and providing new control and customization capabilities. The UT Dallas researchers are taking an approach that leverages a mobile device's greater processing power to support an automatic environmental noise detection, identification and classification process, leading to the almost instant selection of the best speech enhancement algorithm for a particular situation.

Panahi believes that precise, automatic classification of hundreds of various noise environments is essential for obtaining optimal hearing devices performances. "When we started looking at the problem, we saw that there are possibilities in automatically identifying the environmental noise and then finding the best algorithm that can suppress the noise and enhance the speech," he says.

In a world pulsing with all sorts of annoying sounds, pinpointing and addressing each type of noise with a specific algorithm would be an overwhelming, highly impractical task, Panahi believes. "There are many different types of background noises, yet we cannot have 100 or 1,000 speech enhancement algorithms for all of them," he explains. "What we can do is classify all of the noises into a finite number of classes—say three, five, or ten—and then, for each one of them, we can have a set of parameters for speech enhancement—different algorithms—that will lead to the best possible enhancement of speech for that particular noise environment."

To the hearing-device user, the sound enhancement process would be both transparent and seamless. As the user moves from one environment to another—from a busy street into a bustling restaurant, for example—a statistical algorithm would evaluate each new background noise as soon as it emerges. The algorithm would then automatically search for the speech enhancement classification providing the closest match. "If it is closest to that particular background noise, then it will be considered that type of noise and immediately associated with the best speech enhancement algorithm," Panahi says.

Panahi predicts that the project will open the door to the development of a wide collection of mobile device apps designed for use with hearing-assist devices. "What we are going to come up with is a system that is adjustable and adaptable to different people," he says. "While we may end up with a finite number of algorithms, the platform will allow end users to make their own personal adjustments," he says. "For example, suppose I am a hearing aid user, and I have this app sitting on my smartphone. I can adjust that app by either touching the screen or speaking a command, allowing me to adapt the system to meet my needs and situation."

Panahi is particularly enthusiastic about the technology's potential to help children with mild to severe hearing impairment. He notes that smartphones are becoming progressively easier to use and more widely available. "This technology would help the student to better understand what the teacher is saying, even if the teacher walks around the room, so the child will not miss anything," Panahi says. "That is very important, and I would very much like to get to that point."

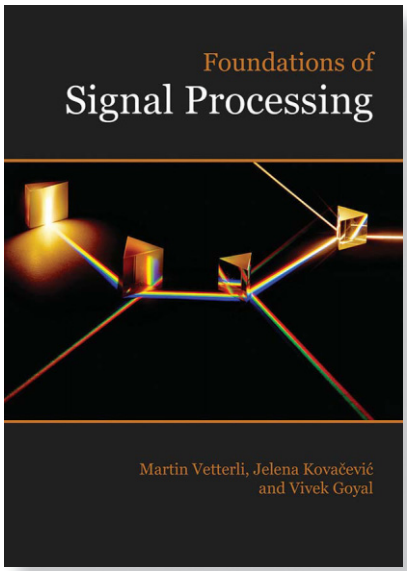
AUTHOR

John Edwards (jedwards@johnedwards-media.com) is a technology writer based in the Phoenix, Arizona, area. [SP]

[book REVIEWS]

Andres Kwasinski

Foundations of Signal Processing



Martin Vetterli, Jelena Kovačević, and Vivek Goyal
Publisher: Cambridge University Press
Year: 2014, first edition
ISBN: 978-1-107-03860-8.

The landscape of signal processing books is quite populated, and it is a tall order to produce a new book that can constructively add to this landscape. Yet, *Foundations of Signal Processing* not only accomplishes this but also stands out for its distinct, consistent, and strong personality. The book's identity starts from a presentation that is faithful to its title since the topics covered in this book are thorough in explaining the foundations of signal processing. However, foundations can be built in many ways and with different materials. The distinct element in this book is that these foundations are built based on the use of Hilbert space geometry, which allows extending Euclidean geometric insights to signals. As such, the geometry of Hilbert spaces forms the common thread across the multiple topics explained in the book. This allows for a presentation where the topics, such as Fourier representations, sampling, interpolation, approximation, compression, and filter design, can be seamlessly unified across finite dimensions, discrete time, and continuous time.

The use of Hilbert space geometry, while establishing a common thread throughout the book, serves as a pathway to understand the fundamental concepts in signal processing and how they relate to each other. The approach entails a larger weight for mathematical abstraction than what is usually found in a typical book written about signal processing. In much the same way as with interpersonal relations, the identity and personality of this book is distinctive for mathematical abstraction. No personality is widely accepted; in that sense, this book will become a must-have and absolute favorite for a certain type of reader, while other readers may not be willing to appreciate the depth and style of its presentation.

Digital Object Identifier 10.1109/MSP.2015.2470455
Date of publication: 29 December 2015

INTENDED AUDIENCE

Foundations of Signal Processing will be greatly enjoyed by the reader who revels in investing effort to find, through the language of mathematical abstraction, a deep understanding of signal processing concepts. This book requires a reader who is filled with curiosity for discovering the deep concepts that form the foundations for signal processing. For these readers, the rewards will be plenty, as this book presents an almost continuum of interrelated deep concepts. At the same time, this book is not for the reader wishing to find a quick explanation or a reference material of a certain signal processing technique. These readers may find an area of improvement in that, for example, the book explains that the Fourier series is the projection onto subspaces associated with frequencies and, from this point, the reader will need to work to the insight that the Fourier series coefficients represent the component of a signal for a particular frequency.

As a textbook, *Foundations of Signal Processing* is so complete in the range of covered topics that it would be applicable to more than one type of course. The book assumes the typical background knowledge on linear algebra that engineering students learn in their first year of studies and some basic knowledge of probability. More advanced topics in vector spaces are given in the second chapter. Because of the reliance on the mathematical abstraction of Hilbert space geometry, at the undergraduate level the book will find a better fit with courses with a clear applied mathematics character.

For the typical signal processing course in the undergraduate curriculum of departments such as electrical engineering and computer engineering, where students would arrive after taking courses in linear systems, circuits, and electronics (with laboratories), this book would work for a minority of students (as an arbitrary guess, maybe a third of the students). Unfortunately, it may be that the majority of students, pressed by their packed class schedules, will be reticent to invest the effort needed to reach the whole depth of the concepts in the book, or they will lack the introspection time to fully realize the implications of the abstract concept presented.

However, *Foundations of Signal Processing* would be the book that bold faculty, who have a love for signal processing, may be waiting for to redesign a signal

IEEE SignalProcessing | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | Qmags THE WORLD'S NEWSSTAND®

processing course into one that challenges the students to achieve that deeper level of understanding where mathematical abstraction and technology come together. Since the book links the concepts of Fourier representation with multirate system, generalized sampling theory, signal compression, computational complexity, and wavelets representations, it would serve as a good source for an advanced signal processing course at the upper undergraduate or at the graduate level. It is clear that this book has been written with great care and dedication. It includes more than 160 homework problems (about half of them solved) and 220 examples. Also, the book is supported by companion *Mathematica* software, lecture slides, and an instructor's solution manual, accessible from the website that accompanies the book.

CONCLUSIONS

Foundations of Signal Processing is written by highly accomplished researchers in the field. The book shows not only the authors' expertise but also their passion for signal processing and a genuine interest in communicating and teaching the ins and outs of this area. Beyond its classroom use, *Foundations of Signal Processing* is a must-have for scientists and engineers who have sufficient interest in owning more than a couple books on signal processing. For scientists and engineers working in signal processing, it will be a pleasure to read, even when already familiar with the presented concepts. Over time, because of its distinct, consistent, and strong personality, *Foundations of Signal Processing* may be considered a "classic" book on signal processing.

REVIEWER

Andres Kwasinski (axkeec@rit.edu) is an associate professor in the Department of Computer Engineering, Rochester Institute of Technology, New York.

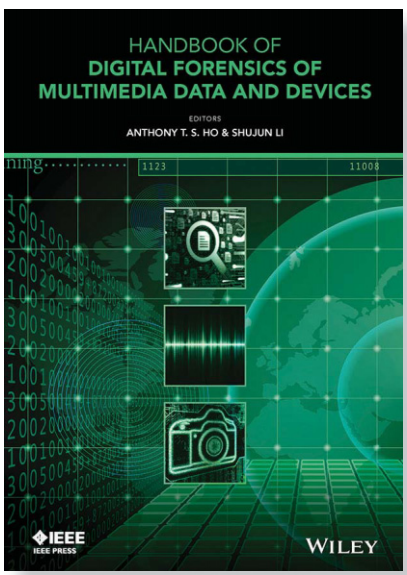
Luisa Verdoliva

Handbook of Digital Forensics of Multimedia Data and Devices

The recovery and analysis of digital information has become a major component of many criminal investigations today. Given the ever-increasing number of personal digital devices, such as notebooks, tablets, and smartphones, as well as the development of communication infrastructures, we all gather, store, and generate huge amounts of data. Some of this information may be precious evidence for investigation and may be used in courts. During the last several decades, increasing research efforts have therefore been dedicated toward defining tools and protocols for the analysis of evidence coming from digital sources.

BRIDGING THE GAP BETWEEN COMMUNITIES

Despite this fast-growing momentum, digital forensics has much suffered from the cultural gap between its major core disciplines: computer science and signal processing. For years, the focus on computer science methods and tools has been dominant. In comparison, signal processing research entered the digital forensics arena only recently, in large part to analyze the underlying structure of visual and audio evidence.



Anthony T.S. Ho and Shujun Li, Editors
Publisher: Wiley-IEEE Press
Year: 2015
ISBN: 978-1-118-64050-0
704 pages.

Handbook of Digital Forensics of Multimedia Data and Devices, edited by Anthony T.S. Ho and Shujun Li, attempts to link research in these two communities by providing a wide-ranging and up-to-date reference for both researchers and practitioners. The digital forensics ecosystem is surveyed with the necessary breadth in the first half of the book, by exploring all phases of the forensics workflow and detailing several tools of interest. Gaining insight into these aspects is of paramount importance for practitioners, but also for academic researchers who are often not aware of the standard practices and processes required to preserve digital evidence, e.g., for legal purposes. Similarly, practitioners have the opportunity to discover the state of the art in forensics research in the second half of the book, which is written from a signal

Digital Object Identifier 10.1109/MSP.2015.2488018
Date of publication: 29 December 2015

IEEE SignalProcessing | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | Qmags THE WORLD'S NEWSSTAND®

processing perspective. This balanced mix is a major asset of this book, making it suitable for readers of diverse background.

Handbook of Digital Forensics of Multimedia Data and Devices is organized in four parts, comprising four chapters each, starting from practical forensic labs problems, through the most advanced underlying technologies used for their solution.

The first part deals with operational aspects of digital forensic services for multimedia data and digital devices. The forensic procedures of three laboratories managed by law enforcement agencies are presented. Then, current standards and best practices are reviewed, followed by digital triage forensics and audio/video file authentication.

The second part focuses on the recovery of relevant pieces of evidence from multimedia data. It considers photogrammetry, file-carving approaches, the interplay between biometrics and forensics, and the analysis of large collections of multimedia data by means of visual analytics.

The challenges of tracing the source of multimedia data and verifying their authenticity, which is a central issue in criminal investigations, is addressed in the third part. Several approaches are presented for different applications (camera, printer/scanner, microphone), spanning different levels of depth. Then, a hands-on chapter is dedicated to the authentication of printed documents.

The fourth and final part considers the main approaches for passive multimedia content analysis. It starts by describing techniques for region duplication and splicing detection in images and videos, moving on to review camera-based approaches for image forgery detection, with a special focus on chromatic aberration methods. The last two chapters are devoted to emerging research challenges: 1) processing history recovery, which aims at identifying the correct sequence of processing steps that have been applied to an image/video, and 2) counterforensics, which establishes the techniques that could be developed by a malicious adversary to deceive current forensic methods.

WHEN PRACTICE FEEDS RESEARCH

Another virtue of this book is that, although every chapter is written by a different set of authors, thus ensuring the highest degree of expertise, the editors successfully managed to develop and organize the book in a uniform and well-thought-out way. Fundamental concepts are explained beforehand and in simple terms, followed by specific techniques, examples, and case studies. This helps to gain insight about practical applications and yields a good overall balance between theory and practice. The various parts, and even individual chapters, are largely self-contained, allowing selective readings (and giving rise to some repetitions).

I personally found that reading *Handbook of Digital Forensics of Multimedia Data and Devices* has been a particularly rewarding experience. As an academic researcher, I was initially more interested in the later parts of the book that deal with the technical details of the most recent forensics techniques. Nevertheless, the first few chapters were actually enlightening for me, as they opened a window to a world of which I knew little. Understanding how real-world investigations are conducted in laboratories shed some light on the actual relevance and applicability of forensic methods, and even suggested some promising areas for future research. Moreover, it was interesting to learn more about the numerous case studies that deal with some pressing societal issues (e.g., child pornography). It was also instructive to gain insight into the human–machine interaction and the importance of machine learning techniques, both badly needed when a large bulk of data is available. All these practical considerations contributed to strengthening the connection between the different points of views discussed throughout the book.

In line with the comprehensive approach undertaken, one could regret the absence of a chapter on social network forensics, which is a rapidly growing field of study. More generally, a weak point of the book is that it misses some of the most recent developments. This shortcoming is not unusual and is due to a combination of the lengthy editorial process for any book and the fast progress of research in emerging areas.

To mitigate this issue, the editors set up a companion website (www.wiley.com/go/digitalforensics) to complement the book. Besides granting access to multimedia material, the objective is to maintain an up-to-date source of information covering available datasets, reference source code and software, standards and best practices, related books and journals, and upcoming conferences and events.

CONCLUSIONS

In summary, *Handbook of Digital Forensics of Multimedia Data and Devices* is a well-conceived and timely book. Besides providing a reference for this field, it also offers a fresh and open perspective, promoting a long-needed cross-fertilization of diverse worlds and academic fields. The editorial choices make the book accessible to a broad audience—from newcomers, who will find a solid and smooth guide to this vast field of study—to established researchers, who may enjoy the aspects of systematic analysis of specific topics.

REVIEWER

Luisa Verdoliva (verdoliv@unina.it) is an assistant professor with the Department of Electrical Engineering and Information Technology, University Federico II of Naples, Italy.

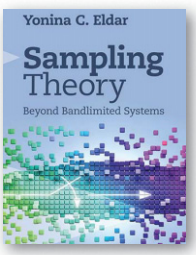
[SP]

[BOOK digest]

EDITORS' INTRODUCTION

The IEEE Signal Processing Society (SPS) and our *IEEE Signal Processing Magazine (SPM)* receive a constant stream of requests from academic publishers and researchers to provide visibility to their newly published books. SPS offers several platforms to publicize books, such as the online Book Beat on the SPS website and the "New Book" section of the SPS monthly eNewsletter. In *SPM*, the traditional platform for books is to invite volunteers who have strong expertise in related areas and have enough bandwidth to read through a whole book and write a comprehensive review, which can be a relatively slow process. Given the magazine's large readership, who have diverse interests, the editorial team would like to introduce a complementary "lightweight" column to the "Book Review" column to inform our readers of more recently published books in a timely fashion. This new column, "Book Digest," provides a visually appealing summary of books that have been selected by a pool of senior editors based on criteria such as timeliness of the topic, track record of the authors, training material for students, signal processing focus, and other considerations beneficial to readers. Should you have comments or should you wish to have your book considered for publication in this column, do not hesitate to contact Gwenaël Doërr (gwenael.doerr@technicolor.com) and Kenneth Lam (enkmlam@polyu.edu.hk), *SPM*'s area editors for columns and forum.

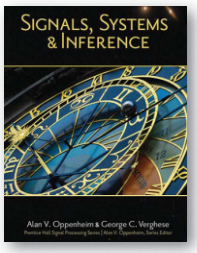
Yonina C. Eldar, *Sampling Theory Beyond Bandlimited Systems*, Cambridge University Press, 2015, ISBN: 9781107003392.



Covering the fundamental mathematical underpinnings together with key principles and applications, this book provides a comprehensive guide to the theory and practice of sampling from an engineering perspective. Beginning with traditional ideas such as uniform sampling in shift-invariant spaces and working through to the more recent fields of compressed sensing and sub-Nyquist sampling, the key concepts are addressed in a unified and coherent way. Emphasis is given to applications in signal processing and communications, as well as hardware considerations, throughout. With 200 worked examples and over 200 end-of-chapter problems, this is an ideal course textbook for senior undergraduate and graduate students. It is also an invaluable reference or self-study guide for engineers and students across industry and academia.

Digital Object Identifier 10.1109/MSP.2015.2478122
Date of publication: 29 December 2015

Alan V. Oppenheim and George C. Verghese, *Signals, Systems & Inference*, Prentice Hall, 2016, ISBN: 9780133943283.



The concepts and mathematics of signals, systems, and probability are usefully combined in studying fields such as communication, control, and signal processing, or other domains that involve data sequences, time series, or waveforms. This text assumes two prerequisites: an introductory subject in time- and frequency-domain analysis of deterministic signals and systems, and an introductory subject in probability. The book is divided into four parts.

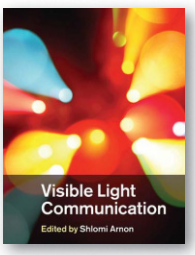
The first part begins with a brief review of the desired prerequisites in signals and linear time-invariant (LTI) systems, though parts of the material (e.g., group delay) may be unfamiliar to some readers with the assumed background. This is followed by the application of some of this material in the setting of digital communication by pulse amplitude modulation.

The second part is devoted to state-space models, concentrating on the single-input, single-output LTI case. The development is largely built around the eigenmodes of such systems, under the simplifying assumption of distinct natural frequencies. This part of the book introduces the idea of model-based inference in the context of state observers for LTI systems and examines associated feedback control strategies.

A short review of the desired probability prerequisites, including estimation and hypothesis testing for random variables, is discussed in the third part of the text. Again, parts of this material (e.g., normal equations, Neyman–Pearson testing and receiver operating characteristic curves) may be unfamiliar to some readers with the assumed background.

The final part of the book characterizes wide-sense stationary random signals, and the outputs that result from LTI filtering of such signals, in both the time- and frequency-domains. Correlation functions and power spectral densities are then used to study canonical signal estimation and detection problems, specifically linear minimum mean-square-error signal estimation (i.e., Wiener filtering) and signal detection problems whose optimum solutions involve matched filtering.

Shlomi Arnon, *Visible Light Communication*, Cambridge University Press, 2015, ISBN: 9781107061552.

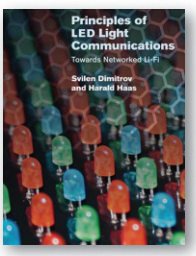


Visible light communications (VLC) is an evolving communication technology for short-range applications. Exploiting recent advances in the development of

high-power visible light light-emitting diodes (LEDs), VLC offers an clean, energy-efficient alternative to RF technology, enabling the development of optical wireless communication systems that make use of existing lighting infrastructure.

Drawing on the expertise of leading researchers from across the world, this concise book sets out the theoretical principles of VLC and outlines key applications of this cutting-edge technology. Providing insight into modulation techniques, positioning and communication, synchronization, and industry standards, as well as techniques for improving network performance, this is an invaluable resource for graduate students and researchers in the fields of VLC and optical wireless communication, and for industrial practitioners in the field of telecommunications.

Svilen Dimitrov and Harald Haas, *Principles of LED Light Communications Towards Networked Li-Fi*, Cambridge University Press, 2015, ISBN: 9781107049420.

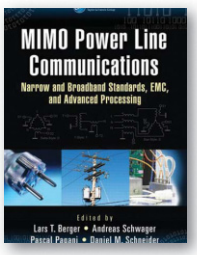


Balancing theoretical analysis and practical advice, this book describes all the underlying principles required to build high-performance indoor optical wireless communication systems based on visible and infrared light, alongside essential techniques for optimizing systems by maximizing throughput, reducing hardware complexity, and measuring performance effectively.

It provides a comprehensive analysis of information rate-, spectral-, and power-efficiencies for single- and multicarrier transmission schemes, and novel analysis of nonlinear signal distortion, enabling the use of off-the-shelf LED technology. Other topics covered include cellular network throughout and coverage, static resource partitioning via dynamic interference-aware scheduling, realistic light propagation modeling, orthogonal frequency-division multiplexing (OFDM), optical multiple-input, multiple-output (MIMO) transmission, and nonlinearity modeling.

Covering practical techniques for building indoor optical wireless cellular networks supporting multiple users, and guidelines for 5G cellular system studies, in addition to physical layer issues, this is an indispensable resource for academic researchers, professional engineers, and graduate students working in optical communications.

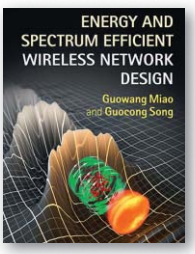
Lars T. Berger, Andreas Schwager, Pascal Pagani, and Daniel M. Schneider, *MIMO Power Line Communications: Narrow and Broadband Standards, EMC, and Advanced Processing*, CRC Press, 2014, ISBN: 9781466557529.



One of the first publications of its kind in the exciting field MIMO power line communications (PLC), *MIMO Power Line Communications: Narrow and Broadband Standards, EMC, and Advanced Processing* contains contributions from experts in industry and academia, making it practical enough to provide a solid understanding of how PLC technologies work, yet scientific enough to form a base for ongoing R&D activities.

This book is subdivided into five thematic parts. Part I looks at narrow- and broadband channel characterization based on measurements from around the globe. Taking into account current regulations and electromagnetic compatibility (EMC), part II describes MIMO signal processing strategies and related capacity and throughput estimates. Current narrow- and broadband PLC standards and specifications are described in the various chapters of part III. Advanced PLC processing options are treated in part IV, drawing from a wide variety of research areas such as beamforming/precoding, time reversal, multiuser processing, and relaying. Finally, part V contains case studies and field trials, where the advanced technologies of tomorrow are put into practice today.

Guowang Miao and Guocong Song, *Energy and Spectrum Efficient Wireless Network Design*, Cambridge University Press, 2014, ISBN: 9781107039889.



Covering the fundamental principles and state-of-the art cross-layer techniques, this practical guide provides the tools needed to design MIMO- and OFDM-based wireless networks that are both energy and spectrum efficient. Technologies are introduced in parallel for both centralized and distributed wireless networks to give you a clear understanding of the similarities and differences between their energy- and spectrum-efficient designs, which is essential for achieving the highest network energy saving without losing performance. Cutting-edge green cellular network design technologies, enabling you to master resource management for next-generation wireless networks based on MIMO and OFDM, and detailed real-world implementation examples are provided to guide your engineering design in both theory and practice.

[SP]

dates AHEAD

Please send calendar submissions to:
Dates Ahead, c/o Jessica Barragué
IEEE Signal Processing Magazine
e-mail: j.barrague@ieee.org

2016

MARCH

41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
21–25 March, Shanghai, China.
General Chairs: Zhi Ding, Zhi-Quan Luo, and Wenjun Zhang
URL: http://icassp2016.org

Data Compression Conference (DCC)
29 March–1 April, Snowbird, Utah, USA.
URL: http://www.cs.brandeis.edu/~dcc/

APRIL

15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)
11–14 April, Vienna, Austria.
General Chair: Guoliang Xing
URL: http://ipsn.acm.org/2016/

IEEE International Symposium on Biomedical Imaging (ISBI)
13–16 April, Prague, Czech Republic.
General Chairs: Jan Kybic and Milan Sonka
URL: http://biomedicalimaging.org/2016/

JUNE

IEEE Workshop on Statistial Signal Processing (SSP)
26–29 June, Palma de Mallorca, Spain.
General Chairs: Antonio Artés-Rodríguez and Joaquín Miguez
URL: http://ssp2016.tsc.uc3m.es/

JULY

IEEE Ninth IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)
10–13 July, Rio de Janeiro, Brazil.
General Chairs: Rodrigo C. de Lamare and Martin Haardt
URL: http://delamare.cetuc.puc-rio.br/sam2016/index.html

IEEE International Conference on Multimedia and Expo (ICME)
11–15 July, Seattle, Washington, USA.
General Chairs: Tsuhan Chen, Ming-Ting Sun, and Cha Zhang
URL: http://www.icme2016.org/

SEPTEMBER

IEEE International Conference on Image Processing (ICIP)
25–28 September, Phoenix, Arizona, USA.
General Chair: Lina Karam
URL: http://www.ieeeicip2016.org

DECEMBER

IEEE Global Conference on Signal and Information Processing (GlobalSIP)
7–9 December, Greater Washington, D.C., USA.
General Chairs: Zhi Tian and Brian Sadler
URL: http://2016.ieeeglobalsip.org

[SP]

Digital Object Identifier 10.1109/MSP.2015.2479478
Date of publication: 29 December 2015

SigPort.org Do you know? Your colleagues archived the slides of their signal processing work on IEEE SigPort.

The slides and posters you spent hours to make are highlights of your work. Aren't they "forgotten" soon after conference presentations or thesis defense?

IEEE Signal Processing Society's new SigPort repository helps extend the life of your slides and posters, and raise the visibility of your work by sharing these appealing summaries. Checkout the material from recent GlobalSIP, WIFS, ICIP and MLSP. Upload yours today!

- **Raise visibility:** Recent conference slides/posters posted on SigPort had an average of 25 downloads in a month, more than IEEE Xplore!
- **Easy to use:** Watch a short video tutorial and upload your work in less than 15 minutes; link slides to full papers on IEEE Xplore
- **Well indexed:** Each has a unique URL picked up by major search engines; plus additional highlights through the editor's picks

As an IEEE SPS community repository on signal and information processing, SigPort welcomes research drafts, white papers, theses, slides, posters, lecture notes, dataset descriptions, product brief, and more. Try out SigPort with promotion code **you14100** (expires 31 May 2016).

Digital Object Identifier 10.1109/MSP.2015.2507501

General Chairs
Zhi Tian
George Mason Univ.
Brian M. Sadler
Army Research Lab.

Technical Program Chairs
Phillip Regalia
NSF
Trac Tran
John Hopkins Univ.
Brian Mark
George Mason Univ.

Finance Chair
Jill Nelson
George Mason Univ.

Local Arrangements Chair
Nathalia Peixoto
George Mason Univ.

Publications Chair
Kathleen Wage
George Mason Univ.

Publicity Chairs
Piya Pal
Univ. MD, College Park
Seung-Jun Kim
Univ. MD, Baltimore Cty

Technical Workshop Liaison Chair
Min Wu
Univ. MD, College Park

Government Panel Chairs
Joel Goodman
Amy Bell
Naval Research Lab

Industrial Liaison Chairs
Kristine Bell
Metron Inc.
Hang Liu
Catholic Univ.

International Liaison Chairs
Chengyang Yang
BUAA, China
Mounir Ghogho
Univ. of Leeds, UK

Advisory Committee
Monson Hayes
George Mason Univ.

Fourth IEEE Global Conference on Signal and Information Processing

December 7–9, 2016, Greater Washington D.C., USA

http://2016.ieeeglobalsip.org/

The fourth IEEE Global Conference on Signal and Information Processing (GlobalSIP) will be held in Washington DC, USA on Dec. 7–9, 2016. GlobalSIP is a flagship conference of the IEEE Signal Processing Society, comprising of co-located symposia and workshops. GlobalSIP 2016 will feature government panels, lectures, tutorials, exhibits, oral and poster sessions on signal and information processing, with an emphasis on up-and-coming themes. Such topics include but are not limited to :

- Signal processing in communications, networks, and massive MIMO
- Big-data signal processing
- Signal processing for information secrecy, privacy, and security
- Information forensics
- Image and video processing
- Selected topics in speech and language processing
- Signal processing in finance
- Signal processing in energy and power systems
- Signal processing in genomics and bioengineering
- Selected topics in statistical signal processing
- Graph theoretic signal processing
- Machine learning
- Compressed sensing and applications
- Music and multimedia transmission, indexing and retrieval and playback challenges
- Real time DSP
- Other novel and significant areas

Timeline:

- **February 5, 2016 : Symposium proposals due**
- February 19, 2016 : Symposium selection decision made
- February 29, 2016 : Call for Papers for accepted symposia will be publicized
- June 5, 2016 : Paper Submission Due
- August 5, 2016 : Final Acceptance decisions notifications sent to all authors
- September 5, 2016: Camera-ready papers due.

About Washington, D.C.:

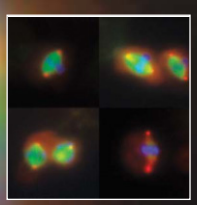
The nation's capital has a rich, vibrant cultural, educational and scientific identity. It houses numerous museums, monuments, art centers, universities, federal agencies and research labs, attracting a global population. Washington D.C. cordially invites you to share this global spirit of scientific collaboration by participating in GlobalSIP 2016!

Digital Object Identifier 10.1109/MSP.2015.2484499

MathWorks®
Accelerating the pace of engineering and science

あなたは MATLAB を話しますか?

Over one million people around the world speak MATLAB. Engineers and scientists in every field from aerospace and semiconductors to biotech, financial services, and earth and ocean sciences use it to express their ideas. Do you speak MATLAB?



Cells in mitosis: high-throughput microscopy for image-based screens. Provided by Roy Wollman, Univ. California, Davis.

Article available at mathworks.com/ltc

©2012 The MathWorks, Inc.

MATLAB®
The language of technical computing

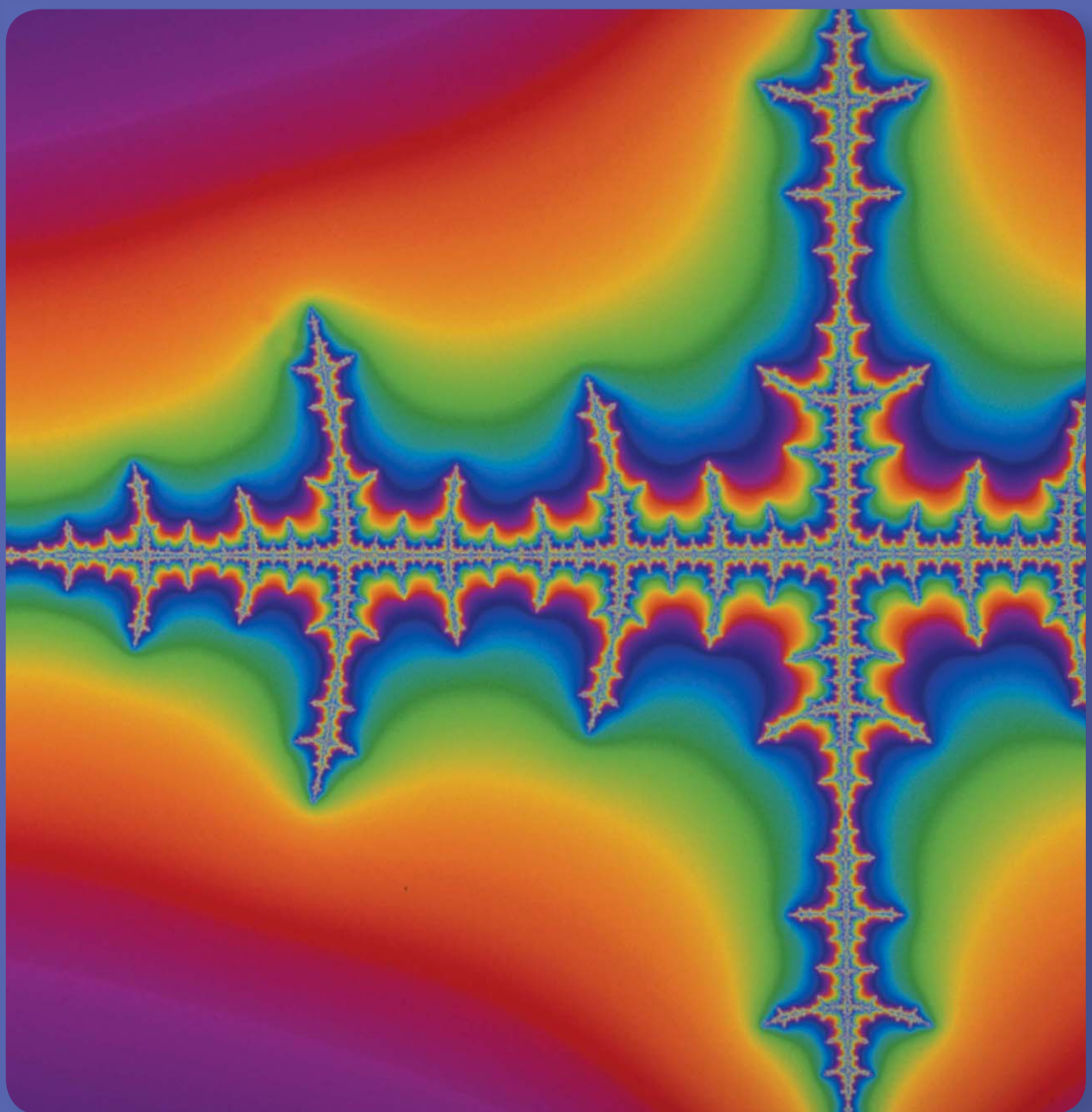
SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®

IEEE SIGNAL PROCESSING SOCIETY

ContentGazette

JANUARY 2016

ISSN 2167-5023



IEEE Signal Processing Society

IEEE

SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®

GlobalSIP 2016 – Washington DC

Call for Symposium Proposals

The fourth IEEE Global Conference on Signal and Information Processing (GlobalSIP) will be held in Greater Washington, DC, USA on December 7–9, 2016. GlobalSIP has rapidly assumed flagship status within the IEEE Signal Processing Society. It focuses on signal and information processing and up-and-coming signal processing themes. The conference aims to feature world-class speakers, tutorials, exhibits, and oral and poster sessions. GlobalSIP is comprised of co-located symposia selected based on responses to the Call for Symposium Proposals. Topics include but are not limited to:

* Signal processing in communications and networks, including green communications, and signal processing in massive MIMO and millimeter-wave technology
* Big data signal processing
* Signal processing in information secrecy, privacy and security
* Information forensics
* Image and video processing
* Selected topics in speech and language processing
* Signal processing in finance
* Signal processing in energy and power systems
* Signal processing in genomics and bioengineering
* Selected topics in statistical signal processing
* Graph-theoretic signal processing
* Machine learning
* Compressed sensing, sparsity analysis and applications
* Music and multimedia transmission, indexing and retrieval, and playback challenges
* Real-time DSP implementations
* Other novel and significant applications of selected areas of signal processing

Symposium proposals should have the following information: title of the symposium; length of the symposium; paper length requirement; names, addresses, and a short CV (up to 250 words) of the organizers including the general organizers and the technical chairs of the symposium; a two-page description of the technical issues that the symposium will address, including timeliness and relevance to the signal processing community; names of (potential) participants on the technical program committee; names of the invited speakers; and a draft Call for Papers. Please pack everything together in pdf format, and email the proposals to conference TPC Chairs at tpc-chairs@2016.ieeeglobalsip.org. More detailed information can be found in the GlobalSIP Symposium Proposal Preparation Guide, available at http://2016.ieeeglobalsip.org/.

Conference Timeline:

* **February 5, 2016: Symposium proposals due**
* February 19, 2016: Symposium selection decision made
* February 29, 2016: Call for Papers for accepted symposia will be publicized
* June 5, 2016: Paper submission due
* August 5, 2016: Final Acceptance decisions notifications sent to all authors
* September 5, 2016: Camera-ready papers due.

IEEE TRANSACTIONS ON
SIGNAL PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed MEDLINE U.S. National Library of Medicine

DECEMBER 1, 2015 VOLUME 63 NUMBER 23 ITPRED (ISSN 1053-587X)

REGULAR PAPERS

Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO http://dx.doi.org/10.1109/TSP.2015.2463260 ..... *Z. Gao, L. Dai, Z. Wang, and S. Chen* 6169

Robust Matched Filtering in ℓ_p-Space http://dx.doi.org/10.1109/TSP.2015.2464179 ..... *X. Jiang, W.-J. Zeng, H. C. So, S. Rajan, and T. Kirubarajan* 6184

Power Versus Spectrum 2-D Sensing in Energy Harvesting Cognitive Radio Networks http://dx.doi.org/10.1109/TSP.2015.2464191 ..... *Y. Zhang, W. Han, D. Li, P. Zhang, and S. Cui* 6200

Fundamental Initial Frequency and Frequency Rate Estimation of Random-Amplitude Harmonic Chirps http://dx.doi.org/10.1109/TSP.2015.2463251 ..... *Y. Doweck, A. Amar, and I. Cohen* 6213

Learning the Nonlinear Geometry of High-Dimensional Data: Models and Algorithms http://dx.doi.org/10.1109/TSP.2015.2469637 ..... *T. Wu and W. U. Bajwa* 6229

Robust Scheduling Filter Design for a Class of Nonlinear Stochastic Poisson Signal Systems http://dx.doi.org/10.1109/TSP.2015.2465305 ..... *B.-S. Chen and C.-F. Wu* 6245

IEEE

SignalProcessing | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | Qmags

Distributed Estimation of Variance in Gaussian Graphical Model via Belief Propagation: Accuracy Analysis and Improvement http://dx.doi.org/10.1109/TSP.2015.2465303 ..... *Q. Su and Y.-C. Wu* 6258

Uplink Downlink Rate Balancing in Cooperating Cellular Networks http://dx.doi.org/10.1109/TSP.2015.2464184 ..... *I. Bergel, Y. Perets, and S. Shamai* 6272

Artificial Noise Assisted Secure Transmission Under Training and Feedback http://dx.doi.org/10.1109/TSP.2015.2465301 ..... *H.-M. Wang, C. Wang, and D. W. K. Ng* 6285

Importance-Weighted Adaptive Search for Multi-Class Targets http://dx.doi.org/10.1109/TSP.2015.2472370 ..... *G. E. Newstadt, B. Mu, D. Wei, J. P. How, and A. O. Hero* 6299

Joint Tensor Factorization and Outlying Slab Suppression With Applications http://dx.doi.org/10.1109/TSP.2015.2469642 ..... *X. Fu, K. Huang, W.-K. Ma, N. D. Sidiropoulos, and R. Bro* 6315

Diversity Effects in Kalman Filtering Over Rayleigh Fading Channels http://dx.doi.org/10.1109/TSP.2015.2464196 ..... *R. Parseh and K. Kansanen* 6329

Analyzing Sparse Dictionaries for Online Learning With Kernels http://dx.doi.org/10.1109/TSP.2015.2457396 ..... *P. Honeine* 6343

Optimal Resource Sharing and Network Beamforming in Multi-Carrier Bidirectional Relay Networks http://dx.doi.org/10.1109/TSP.2015.2454487 ..... *A. Gavili and S. Shahbazpanahi* 6354

Robust Kronecker Product PCA for Spatio-Temporal Covariance Estimation http://dx.doi.org/10.1109/TSP.2015.2472364 ..... *K. Greenewald and A. O. Hero* 6368

A Distributed Framework for Dimensionality Reduction and Denoising http://dx.doi.org/10.1109/TSP.2015.2465300 ..... *I. D. Schizas and A. Aduroja* 6379

MUSIC for Multidimensional Spectral Estimation: Stability and Super-Resolution http://dx.doi.org/0.1109/TSP.2015.2463255 ..... *W. Liao* 6395

Performance Analysis of an Improved MUSIC DoA Estimator http://dx.doi.org/0.1109/TSP.2015.2465302 ..... *P. Vallet, X. Mestre, and P. Loubaton* 6407

Analysis of Fisher Information and the Cramér–Rao Bound for Nonlinear Parameter Estimation After Random Compression http://dx.doi.org/0.1109/TSP.2015.2464183 ..... *P. Pakrooh, A. Pezeshki, L. L. Scharf, D. Cochran, and S. D. Howard* 6423

Iterative Demodulation and Channel Estimation for Asynchronous Joint Multiple Access Reception http://dx.doi.org/0.1109/TSP.2015.2468681 ..... *C. Schlegel and M. Jar* 6429

IEEE TRANSACTIONS ON

SIGNAL PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed MEDLINE U.S. National Library of Medicine

DECEMBER 15, 2015 VOLUME 63 NUMBER 24 ITPRED (ISSN 1053-587X)

REGULAR PAPERS

A Convergence and Asymptotic Analysis of the Generalized Symmetric Fast ICA Algorithm http://dx.doi.org/10.1109/TSP.2015.2468686 ..... *T. Wei* 6445

Accurate Localization of a Rigid Body Using Multiple Sensors and Landmarks http://dx.doi.org/10.1109/TSP.2015.2465356 ..... *S. Chen and K. C. Ho* 6459

Optimal Parameter Estimation Under Controlled Communication Over Sensor Networks http://dx.doi.org/10.1109/TSP.2015.2469639 ..... *D. Han, K. You, L. Xie, J. Wu, and L. Shi* 6473

Wideband Sparse Signal Acquisition With Dual-rate Time-Interleaved Undersampling Hardware and Multicoset Signal Reconstruction Algorithms http://dx.doi.org/10.1109/TSP.2015.2469648 ..... *T. Moon, H. W. Choi, N. Tzou, and A. Chatterjee* 6486

Mathematical Framework for Pseudo-Spectra of Linear Stochastic Difference Equations http://dx.doi.org/10.1109/TSP.2015.2469640 ..... *M. Bujosa, A. Bujosa, and A. Garca-Ferrer* 6498

Discrete Signal Processing on Graphs: Sampling Theory http://dx.doi.org/10.1109/TSP.2015.2469645 ..... *S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević* 6510

$\mathbf{L}_1$-Constrained Normalized LMS Algorithms for Adaptive Beamforming http://dx.doi.org/10.1109/TSP.2015.2474302 ..... *J. F. de Andrade, M. L. R. de Campos, and J. A. Apolinário* 6524

IEEE

Binary Compressive Sensing Via Analog Fountain Coding http://dx.doi.org/10.1109/TSP.2015.2472362 ..... *M. Shirvanimoghaddam, Y. Li, B. Vucetic, J. Yuan, and P. Zhang* 6540

Efficient Timing Mismatch Correction for Low-Cost Digital-Mixing Transmitter http://dx.doi.org/10.1109/TSP.2015.2477045 ..... *C. Li, M. Li, M. Verhelst, A. Bourdoux, M. Ingels, L. Van der Perre, and S. Pollin* 6553

Hardware Impairments Aware Transceiver for Full-Duplex Massive MIMO Relaying http://dx.doi.org/10.1109/TSP.2015.2469635 ..... *X. Xia, D. Zhang, K. Xu, W. Ma, and Y. Xu* 6565

A Factor Analysis Framework for Power Spectra Separation and Multiple Emitter Localization http://dx.doi.org/10.1109/TSP.2015.2464194 ..... *X. Fu, N. D. Sidiropoulos, J. H. Tranter, and W.-K. Ma* 6581

Improving M-SBL for Joint Sparse Recovery Using a Subspace Penalty http://dx.doi.org/10.1109/TSP.2015.2477049 ..... *J. C. Ye, J. M. Kim, and Y. Bresler* 6595

Coherence Optimization and Best Complex Antipodal Spherical Codes http://dx.doi.org/10.1109/TSP.2015.2477052 ..... *H. Zörlein and M. Bossert* 6606

To Harvest and Jam: A Paradigm of Self-Sustaining Friendly Jammers for Secure AF Relaying http://dx.doi.org/10.1109/TSP.2015.2477800 ..... *H. Xing, K.-K. Wong, Z. Chu, and A. Nallanathan* 6616

Bayesian Estimation in the Presence of Deterministic Nuisance Parameters—Part I: Performance Bounds http://dx.doi.org/10.1109/TSP.2015.2468684 ..... *S. Bar and J. Tabrikian* 6632

Bayesian Estimation in the Presence of Deterministic Nuisance Parameters—Part II: Estimation Methods http://dx.doi.org/10.1109/TSP.2015.2468680 ..... *S. Bar and J. Tabrikian* 6647

Attacks on Sensor Network Parameter Estimation With Quantization: Performance and Asymptotically Optimum Processing http://dx.doi.org/10.1109/TSP.2015.2477804 ..... *B. Alnajjab, J. Zhang, and R. S. Blum* 6659

Joint-Bin Monopulse Processing of Rayleigh Targets http://dx.doi.org/10.1109/TSP.2015.2478749 ..... *J. D. Glass, W. D. Blair, and A. D. Lanterman* 6673

A Robust Statistics Approach to Minimum Variance Portfolio Optimization http://dx.doi.org/10.1109/TSP.2015.2474298 ..... *L. Yang, R. Couillet, and M. R. McKay* 6684

Reduced Interference Sparse Time-Frequency Distributions for Compressed Observations http://dx.doi.org/10.1109/TSP.2015.2477056 .... *B. Jokanovic and M. Amin* 6698

List of Reviewers http://dx.doi.org/10.1109/TSP.2015.2488118 6710

EDICS—Editors' Information Classification System http://dx.doi.org/10.1109/TSP.2015.2502638 6720

Information for Authors http://dx.doi.org/10.1109/TSP.2015.2502639 6721

2015 INDEX http://dx.doi.org/10.1109/TSP.2015.2509838 .....Available online at http://ieeexplore.ieee.org

new

Now with no page charges!

IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS

Now accepting paper submissions

The new *IEEE Transactions on Signal and Information Processing over Networks* publishes high-quality papers that extend the classical notions of processing of signals defined over vector spaces (e.g. time and space) to processing of signals and information (data) defined over networks, potentially dynamically varying. In signal processing over networks, the topology of the network may define structural relationships in the data, or may constrain processing of the data. Topics of interest include, but are not limited to the following:

Adaptation, Detection, Estimation, and Learning
- Distributed detection and estimation
- Distributed adaptation over networks
- Distributed learning over networks
- Distributed target tracking
- Bayesian learning; Bayesian signal processing
- Sequential learning over networks
- Decision making over networks
- Distributed dictionary learning
- Distributed game theoretic strategies
- Distributed information processing
- Graphical and kernel methods
- Consensus over network systems
- Optimization over network systems

Communications, Networking, and Sensing
- Distributed monitoring and sensing
- Signal processing for distributed communications and networking
- Signal processing for cooperative networking
- Signal processing for network security
- Optimal network signal processing and resource allocation

Modeling and Analysis
- Performance and bounds of methods
- Robustness and vulnerability
- Network modeling and identification

Modeling and Analysis (cont.)
- Simulations of networked information processing systems
- Social learning
- Bio-inspired network signal processing
- Epidemics and diffusion in populations

Imaging and Media Applications
- Image and video processing over networks
- Media cloud computing and communication
- Multimedia streaming and transport
- Social media computing and networking
- Signal processing for cyber-physical systems
- Wireless/mobile multimedia

Data Analysis
- Processing, analysis, and visualization of big data
- Signal and information processing for crowd computing
- Signal and information processing for the Internet of Things
- Emergence of behavior

Emerging topics and applications
- Emerging topics
- Applications in life sciences, ecology, energy, social networks, economic networks, finance, social sciences, smart grids, wireless health, robotics, transportation, and other areas of science and engineering

Editor-in-Chief: Petar M. Djurić, Stony Brook University (USA)

To submit a paper, go to: https://mc.manuscriptcentral.com/tsipn-ieee

IEEE COMMUNICATIONS SOCIETY

IEEE Signal Processing Society

IEEE COMPUTER SOCIETY

http://www.signalprocessingsociety.org/

Call for Papers
IEEE Signal Processing Society
IEEE Transactions on Signal and Information Processing over Networks

SPECIAL ISSUE ON INFERENCE AND LEARNING OVER NETWORKS

Networks are everywhere. They surround us at different levels and scales, whether we are dealing with communications networks, power grids, biological colonies, social networks, sensor networks, or distributed Big Data depositories. Therefore, it is not hard to appreciate the ongoing and steady progression of network science, a prolific research field spreading across many theoretical as well as applicative domains. Regardless of the particular context, the very essence of a network resides in the interaction among its individual constituents, and Nature itself offers beautiful paradigms thereof. Many biological networks and animal groups owe their sophistication to fairly structured patterns of cooperation, which are vital to their successful operation. While each individual agent is not capable of sophisticated behavior on its own, the *combined interplay* among simpler units and the *distributed processing* of dispersed pieces of information, enable the agents to solve complex tasks and enhance dramatically their performance. Self-organization, cooperation and adaptation emerge as the essential, combined attributes of a network tasked with distributed information processing, optimization, and inference. Such a network is conveniently described as an ensemble of spatially dispersed (possibly moving) agents, linked together through a (possibly time-varying) connection topology. The agents are allowed to interact locally and to perform in-network processing, in order to accomplish the assigned inferential task. Correspondingly, several problems such as, e.g., network intrusion, community detection, and disease outbreak inference, can be conveniently described by signals on graphs, where the graph typically accounts for the topology of the underlying space and we obtain multivariate observations associated with nodes/edges of the graph. The goal in these problems is to identify/infer/learn patterns of interest, including anomalies, outliers, and existence of latent communities. Unveiling the fundamental principles that govern distributed inference and learning over networks has been the common scope across a variety of disciplines, such as signal processing, machine learning, optimization, control, statistics, physics, economics, biology, computer, and social sciences. In the realm of signal processing, many new challenges have emerged, which stimulate research efforts toward delivering the theories and algorithms necessary to (a) designing networks with sophisticated inferential and learning abilities; (b) promoting truly distributed implementations, endowed with real-time adaptation abilities, needed to face the dynamical scenarios wherein real-world networks operate; and (c) discovering and disclosing significant relationships possibly hidden in the data collected from across networked systems and entities. This call for papers therefore encourages submissions from a broad range of experts that study such fundamental questions, including but not limited to:

- Adaptation and learning over networks.
- Consensus strategies; diffusion strategies.
- Distributed detection, estimation and filtering over networks.
- Distributed dictionary learning.
- Distributed game-theoretic learning.
- Distributed machine learning; online learning.
- Distributed optimization; stochastic approximation.
- Distributed proximal techniques, sub-gradient techniques.
- Learning over graphs; network tomography.
- Multi-agent coordination and processing over networks.
- Signal processing for biological, economic, and social networks.
- Signal processing over graphs.

Prospective authors should visit http://www.signalprocessingsociety.org/publications/periodicals/tsipn/ for information on paper submission. Manuscripts should be submitted via Manuscript Central at http://mc.manuscriptcentral.com/tsipn-ieee.

Important Dates:
- Manuscript submission: February 1, 2016
- First review completed: April 1, 2016
- Revised manuscript due: May 15, 2016
- Second review completed: July 15, 2016
- Final manuscript due: September 15, 2016
- Publication: December 1, 2016

Guest Editors:

Vincenzo **Matta**, University of Salerno, Italy, vmatta@unisa.it
Cédric **Richard**, University of Nice Sophia-Antipolis, France, cedric.richard@unice.fr
Venkatesh **Saligrama**, Boston University, USA, srv@bu.edu
Ali H. **Sayed**, University of California, Los Angeles, USA, sayed@ucla.edu

Call for Papers
IEEE Signal Processing Society
IEEE Transactions on Signal and Information Processing over Networks

Special Issue on Distributed Information Processing in Social Networks

Over the past few decades, online social networks such as *Facebook* and *Twitter* have significantly changed the way people communicate and share information with each other. The opinion and behavior of each individual are heavily influenced through interacting with others. These local interactions lead to many interesting collective phenomena such as herding, consensus, and rumor spreading. At the same time, there is always the danger of mob mentality of following crowds, celebrities, or gurus who might provide misleading or even malicious information. Many efforts have been devoted to investigating the collective behavior in the context of various network topologies and the robustness of social networks in the presence of malicious threats. On the other hand, activities in social networks (clicks, searches, transactions, posts, and tweets) generate a massive amount of decentralized data, which is not only big in size but also complex in terms of its structure. Processing these data requires significant advances in accurate mathematical modeling and computationally efficient algorithm design.

Many modern technological systems such as wireless sensor and robot networks are virtually the same as social networks in the sense that the nodes in both networks carry disparate information and communicate with constraints. Thus, investigating social networks will bring insightful principles on the system and algorithmic designs of many engineering networks. An example of such is the implementation of consensus algorithms for coordination and control in robot networks. Additionally, more and more research projects nowadays are data-driven. Social networks are natural sources of massive and diverse big data, which present unique opportunities and challenges to further develop theoretical data processing toolsets and investigate novel applications. This special issue aims to focus on addressing distributed information (signal, data, etc.) processing problems in social networks and also invites submissions from all other related disciplines to present comprehensive and diverse perspectives.

Topics of interest include, but are not limited to:

- Dynamic social networks: time varying network topology, edge weights, etc.
- Social learning, distributed decision-making, estimation, and filtering
- Consensus and coordination in multi-agent networks
- Modeling and inference for information diffusion and rumor spreading
- Multi-layered social networks where social interactions take place at different scales or modalities
- Resource allocation, optimization, and control in multi-agent networks
- Modeling and strategic considerations for malicious behavior in networks
- Social media computing and networking
- Data mining, machine learning, and statistical inference frameworks and algorithms for handling big data from social networks
- Data-driven applications: attribution models for marketing and advertising, trend prediction, recommendation systems, crowdsourcing, etc.
- Other topics associated with social networks: graphical modeling, trust, privacy, engineering applications, etc.

Important Dates:

- Manuscript submission due: September 15, 2016
- First review completed: November 1, 2016
- Revised manuscript due: December 15, 2016
- Second review completed: February 1, 2017
- Final manuscript due: March 15, 2017
- Publication: June 1, 2017

Guest Editors:

- Zhenliang Zhang, Qualcomm Corporate R&D (zhenlian@qti.qualcomm.com)
- Wee Peng Tay, Nanyang Technological University (wptay@ntu.edu.sg)
- Moez Draief, Imperial College London (m.draief@imperial.ac.uk)
- Xiaodong Wang, Columbia University (xw2008@columbia.edu)
- Edwin K. P. Chong, Colorado State University (edwin.chong@colostate.edu)
- Alfred O. Hero III, University of Michigan (hero@eecs.umich.edu)

IEEE/ACM TRANSACTIONS ON

AUDIO, SPEECH, AND LANGUAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

acm Association for Computing Machinery

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed MEDLINE U.S. National Library of Medicine

DECEMBER 2015 VOLUME 23 NUMBER 12 ITASFA (ISSN 2329-9290)

REGULAR PAPERS

Sentence Compression for Aspect-Based Sentiment Analysis http://dx.doi.org/10.1109/TASLP.2015.2443982 ..... *W. Che, Y. Zhao, H. Guo, Z. Su, and T. Liu* 2111

Binaural Reproduction of Finite Difference Simulations Using Spherical Array Processing http://dx.doi.org/10.1109/TASLP.2015.2468066 ..... *J. Sheaffer, M. van Walstijn, B. Rafaely, and K. Kowalczyk* 2125

Joint Optimization of Masks and Deep Recurrent Neural Networks for Monaural Source Separation http://dx.doi.org/10.1109/TASLP.2015.2468583 ..... *P.-S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis* 2136

Finding Complex Features for Guest Language Fragment Recovery in Resource-Limited Code-Mixed Speech Recognition http://dx.doi.org/10.1109/TASLP.2015.2469634 ..... *A. Heidel, H.-H. Lu, and L.-S. Lee* 2148

Interaural Coherence Preservation in Multi-Channel Wiener Filtering-Based Noise Reduction for Binaural Hearing Aids http://dx.doi.org/10.1109/TASLP.2015.2471096 ..... *D. Marquardt, V. Hohmann, and S. Doclo* 2162

Constrained Markov Bayesian Polynomial for Efficient Dialogue State Tracking http://dx.doi.org/10.1109/TASLP.2015.2470597 ..... *K. Yu, K. Sun, L. Chen, and S. Zhu* 2177

Spatially Robust Far-field Beamforming Using the von Mises(-Fisher) Distribution http://dx.doi.org/10.1109/TASLP.2015.2473684 ..... *C. A. Anderson, P. D. Teal, and M. A. Poletti* 2189

Spectro-Temporal Gabor Filterbank Features for Acoustic Event Detection http://dx.doi.org/10.1109/TASLP.2015.2467964 ..... *J. Schröder, S. Goetze, and J. Anemüller* 2198

Classification Based on Speech Rhythm via a Temporal Alignment of Spoken Sentences http://dx.doi.org/10.1109/TASLP.2015.2475155 ..... *I. Heo and W. A. Sethares* 2209

IEEE

An Efficient Parameterization of the Room Transfer Function http://dx.doi.org/10.1109/TASLP.2015.2475173 ..... P. Samarasinghe, T. Abhayapala, M. Poletti, and T. Betlehem 2217
Spread Spectrum-Based High Embedding Capacity Watermarking Method for Audio Signals http://dx.doi.org/10.1109/TASLP.2015.2476755 ..... Y. Xiang, I. Natgunanathan, Y. Rong, and S. Guo 2228
Formant-Based Robust Voice Activity Detection http://dx.doi.org/10.1109/TASLP.2015.2476762 ..... I.-C. Yoo, H. Lim, and D. Yook 2238
Speaker-Adaptive Acoustic-Articulatory Inversion Using Cascaded Gaussian Mixture Regression http://dx.doi.org/10.1109/TASLP.2015.2464702 ..... T. Hueber, L. Girin, X. Alameda-Pineda, and G. Bailly 2246
Late Reverberation Synthesis: From Radiance Transfer to Feedback Delay Networks http://dx.doi.org/10.1109/TASLP.2015.2478116 ..... H. Bai, G. Richard, and L. Daudet 2260
A Multichannel Audio Denoising Formulation Based on Spectral Sparsity http://dx.doi.org/10.1109/TASLP.2015.2479042 ..... I. Bayram 2272
Fast Single- and Cross-Show Speaker Diarization Using Binary Key Speaker Modeling http://dx.doi.org/10.1109/TASLP.2015.2479043 ..... H. Delgado, X. Anguera, C. Fredouille, and J. Serrano 2286
A New HMM-Based Voice Conversion Methodology Evaluated on Monolingual and Cross-Lingual Conversion Tasks http://dx.doi.org/10.1109/TASLP.2015.2479040 ..... W. S. Percybrooks and E. Moore 2298
Statistical Framework with Knowledge Base Integration for Robust Speech Understanding of the Tunisian Dialect http://dx.doi.org/10.1109/TASLP.2015.2464687 ..... M. Graja, M. Jaoua, and L. H. Belguith 2311
Adaptive Feedback Cancellation for Realistic Hearing Aid Applications http://dx.doi.org/10.1109/TASLP.2015.2479038 ..... F. Strasser and H. Puder 2322
Supervised Single-Microphone Multi-Talker Speech Separation with Conditional Random Fields http://dx.doi.org/10.1109/TASLP.2015.2479039 ..... Y. T. Yeung, T. Lee, and C.-C. Leung 2334
Theory and Design of Multizone Soundfield Reproduction Using Sparse Methods http://dx.doi.org/10.1109/TASLP.2015.2479037 ..... W. Jin and W. B. Kleijn 2343
A Time–Frequency Masking Based Random Finite Set Particle Filtering Method for Multiple Acoustic Source Detection and Tracking http://dx.doi.org/10.1109/TASLP.2015.2479041 ..... X. Zhong and J. R. Hopgood 2356
Analysis of Phase Spectrum of Speech Signals Using Allpass Modeling http://dx.doi.org/10.1109/TASLP.2015.2479045 ..... K. Vijayan and K. S. R. Murty 2371
Theoretical Analysis of Linearly Constrained Multi-Channel Wiener Filtering Algorithms for Combined Noise Reduction and Binaural Cue Preservation in Binaural Hearing Aids http://dx.doi.org/10.1109/TASLP.2015.2479940 ..... D. Marquardt, E. Hadad, S. Gannot, and S. Doclo 2384
Representation Learning for Single-Channel Source Separation and Bandwidth Extension http://dx.doi.org/10.1109/TASLP.2015.2470560 ..... M. Zöhrer, R. Peharz, and F. Pernkopf 2398
Exponential Language Modeling Using Morphological Features and Multi-Task Learning http://dx.doi.org/10.1109/TASLP.2015.2482118 ..... H. Fang, M. Ostendorf, P. Baumann, and J. Pierrehumbert 2410
A Framework for Speech Activity Detection Using Adaptive Auditory Receptive Field http://dx.doi.org/10.1109/TASLP.2015.2481179S ..... M. A. Carlin and M. Elhilali 2422
Convolutive Blind Source Separation Using an Iterative Least-Squares Algorithm for Non-Orthogonal Approximate Joint Diagonalization http://dx.doi.org/10.1109/TASLP.2015.2485663 ..... S. Saito, K. Oishi, and T. Furukawa 2434
Theoretical Analysis of Binaural Transfer Function MVDR Beamformers with Interference Cue Preservation Constraints http://dx.doi.org/10.1109/TASLP.2015.2486381 ..... E. Hadad, D. Marquardt, S. Doclo, and S. Gannot 2449
Psychophysical Evaluation of An Ultra-Low Power, Analog Biomimetic Cochlear Implant Processor Filterbank Architecture With Across Channels AGC http://dx.doi.org/10.1109/TASLP.2015.2488290 ..... G. Yang, R. F. Lyon, and E. M. Drakakis 2465

List of Reviewers http://dx.doi.org/10.1109/TASLP.2015.2488318 2474

EDICS—Editor's Information and Classification Scheme http://dx.doi.org/10.1109/TASLP.2015.2502080 2477
Information for Authors http://dx.doi.org/10.1109/TASLP.2015.2502081 2479

2015 INDEX http://dx.doi.org/10.1109/TASLP.2015.2502076 2481

CALL FOR PAPERS

IEEE/ACM Transactions on Audio, Speech and Language Processing

Special issue on **Sound Scene and Event Analysis**

It is still difficult for a machine listening system to demonstrate the same capabilities as human listeners in the analysis of realistic acoustic scenes. Besides speech and music, the analysis of other types of sounds, generally referred to as environmental sounds, is the subject of growing interest from the community and is targeting an ever increasing set of audio categories. In realistic environments, multiple sources are often present simultaneously, and in reverberant conditions, which makes the computational scene analysis challenging.

Typical tasks on audio scene analysis are audio-based scene classification and audio event detection and recognition targeting categories such as "door knocks", "gunshots", "crowds", "car engine noise", as well as marine mammal and bird species, etc. The wide heterogeneity of possible sounds means that novel types of signal processing and machine learning methods should be developed including novel concepts for audio source segmentation and separation. Beyond recognizing sound scenes and sources of interest, a key task of complex audio scene analysis is sound-source localization.

Further, most of the methods developed until now are probably not tractable on big data so there is also a need for new approaches that are, by design, efficient on large scale problems. Acquiring large scale labelled databases is still problematic and such datasets are most likely collected on heterogeneous sets of acoustic conditions (mobile phone recordings, urban/domestic audio,…) most of which are usually offering a degraded version of the signal of interest with potential variable annotation strategies. Therefore methods to tackle large scale problems also have to be robust against signal degradation, acoustic variability, and annotation variability.

We invite papers on various topics on Sound Scene and Event Analysis, including but not limited to :

* Audio scene classification;
* Sound event detection and classification
* Large-scale environmental audio data sets;
* Acoustic features for environmental sound analysis;
* Source localization methods for environmental audio scene analysis
* Source separation for environmental audio scene analysis
* Big data in environmental audio;
* Environmental sound recognition;
* Computational auditory scene analysis;

The authors are required to follow the Author's Guide for manuscript submission to the IEEE /ACM Transactions on Audio, Speech, and Language Processing at
http://www.signalprocessingsociety.org/publications/periodicals/taslp/taslp-author-information

Important Dates:

Manuscript submission due: July 1st, 2016
First review completed: Sept. 30th 2016
Revised manuscript due: October 20th, 2016
Second review completed: Dec. 1st, 2016
Final manuscript due: Dec. 31st, 2016
Publication date: February 2017

Guest Editors:

Gaël Richard, Télécom ParisTech, France (lead guest editor)
Tuomas Virtanen, Tampere University of Technology, Finland
Juan Pablo Bello, New York University, USA
Nobutaka Ono, National Institute of Informatics, Japan

The Ninth IEEE Sensor Array and Multichannel Signal Processing Workshop

10th-13th July 2016,Rio de Janeiro, Brazil

IEEE Signal Processing Society

Call for Papers

General Chairs
Rodrigo C. de Lamare,
PUC-Rio, Brazil and University of York, United Kingdom

Martin Haardt,
TU Ilmenau, Germany

Technical Chairs
Aleksandar Dogandzic,
Iowa State University, USA

Vítor Nascimento,
University of São Paulo, Brazil

Special Sessions Chair
Cédric Richard,
University of Nice, France

Publicity Chair
Maria Sabrina Greco,
Universityof Pisa, Italy

Important Dates

Special Session Proposals
29th January , 2016

Submission of Papers
26th February, 2016

Notification of Acceptance
29th April , 2016

Final Manuscript Submission
16th May, 2016

Advance Registration
16th May, 2016

Technical Program
The SAM Workshop is an important IEEE Signal Processing Society event dedicated to sensor array and multichannel signal processing. The organizing committee invites the international community to contribute with state-of-the-art developments in the field. SAM 2016 will feature plenary talks by leading researchers in the field as well as poster and oral sessions with presentations by the participants.

Welcome to Rio de Janeiro! – The workshop will be held at the Pontifical Catholic University of Rio de Janeiro, located in Gávea, in a superb area surrounded by beaches, mountains and the Tijuca National Forest, the world's largest urban forest. Rio de Janeiro is a world renowned city for its culture, beautiful landscapes, numerous tourist attractions and international cuisine. The workshop will take place during the first half of July about a month before the 2016 Summer Olympic Games when Rio will offer plenty of cultural activities and festivities, which will make SAM 2016 a memorable experience.

Research Areas
Authors are invited to submit contributions in the following areas:
- Adaptive beamforming
- Array processing for biomedical applications
- Array processing for communications
- Blind source separation and channel identification
- Computational and optimization techniques
- Compressive sensing and sparsity-based signal processing
- Detection and estimation
- Direction-of-arrival estimation
- Distributed and adaptive signal processing
- Intelligentsystems and knowledge-based signal processing
- Microphone and loudspeaker array applications
- MIMO radar
- Multi-antenna systems: multiuser MMO, massive MIMO and space-time coding
- Multi-channel imaging and hyperspectral processing
- Multi-sensor processing for smart grid and energy
- Non-Gaussian, nonlinear, and non-stationary models
- Performance evaluations with experimental data
- Radar and sonar array processing
- Sensor networks
- Source Localization, Classification and Tracking
- Synthetic aperture techniques
- Space-time adaptive processing
- Statistical modelling for sensor arrays
- Waveform diverse sensors and systems

Submission of papers – Full-length four-page papers will be accepted only electronically.

Special session proposals – They should be submitted by e-mail to the Technical Program Chairs and the Special Sessions Chair and include a topical title, rationale, session outline, contact information, and list of invited speakers.

IEEE TRANSACTIONS ON

IMAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

PubMed MEDLINE U.S. National Library of Medicine

DECEMBER 2015 VOLUME 24 NUMBER 12 IIPRE4 (ISSN 1057-7149)

PAPERS

Objective Quality Assessment for Color-to-Gray Image Conversion http://dx.doi.org/10.1109/TIP.2015.2460015 ..... K. Ma, T. Zhao, K. Zeng, and Z. Wang 4673
Robust Visual Tracking via Sparsity-Induced Subspace Learning http://dx.doi.org/10.1109/TIP.2015.2462076 ..... Y. Sui, S. Zhang, and L. Zhang 4686
Robust and Non-Negative Collective Matrix Factorization for Text-to-Image Transfer Learning http://dx.doi.org/10.1109/TIP.2015.2465157 ..... L. Yang, L. Jing, and M. K. Ng 4701
Spline Driven: High Accuracy Projectors for Tomographic Reconstruction From Few Projections http://dx.doi.org/10.1109/TIP.2015.2466083 ..... F. Momey, L. Denis, C. Burnier, É. Thiébaut, J.-M. Becker, and L. Desbat 4715
Face Spoofing Detection Through Visual Codebooks of Spectral Temporal Cubes http://dx.doi.org/10.1109/TIP.2015.2466088 ..... A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha 4726
Relevance Metric Learning for Person Re-Identification by Exploiting Listwise Similarities http://dx.doi.org/10.1109/TIP.2015.2466117 ..... J. Chen, Z. Zhang, and Y. Wang 4741
Adaptive Skin Classification Using Face and Body Detection http://dx.doi.org/10.1109/TIP.2015.2467209 ..... S. Bianco, F. Gasparini, and R. Schettini 4756
Bit-Scalable Deep Hashing With Regularized Similarity Learning for Image Retrieval and Person Re-Identification http://dx.doi.org/10.1109/TIP.2015.2467315 ..... R. Zhang, L. Lin, R. Zhang, W. Zuo, and L. Zhang 4766
Spartans: Single-Sample Periocular-Based Alignment-Robust Recognition Technique Applied to Non-Frontal Scenarios http://dx.doi.org/10.1109/TIP.2015.2468173 ..... F. Juefei-Xu, K. Luu, and M. Savvides 4780
Cubic Convolution Scaler Optimized for Local Property of Image Data http://dx.doi.org/10.1109/TIP.2015.2468176 ..... J.-K. Chae, J.-Y. Lee, M.-H. Lee, J.-K. Han, T. Q. Nguyen, and W.-Y. Yeo 4796
Nonlinear Hyperspectral Unmixing With Robust Nonnegative Matrix Factorization http://dx.doi.org/10.1109/TIP.2015.2468177 ..... C. Févotte and N. Dobigeon 4810
Local Multi-Grouped Binary Descriptor With Ring-Based Pooling Configuration and Optimization http://dx.doi.org/10.1109/TIP.2015.2469093 ..... Y. Gao, W. Huang, and Y. Qiao 4820
Image Outlier Detection and Feature Extraction via L1-Norm-Based 2D Probabilistic PCA http://dx.doi.org/10.1109/TIP.2015.2469136 ..... F. Ju, Y. Sun, J. Gao, Y. Hu, and B. Yin 4834

IEEE

Subjective and Objective Video Quality Assessment of 3D Synthesized Views With Texture/Depth Compression Distortion http://dx.doi.org/10.1109/TIP.2015.2469140 ..... X. Liu, Y. Zhang, S. Hu, S. Kwong, C.-C. J. Kuo, and Q. Peng 4847
Graph Matching Based on Stochastic Perturbation http://dx.doi.org/10.1109/TIP.2015.2469153 ..... C. Leng, W. Xu, I. Cheng, and A. Basu 4862
Fast Translation Invariant Multiscale Image Denoising http://dx.doi.org/10.1109/TIP.2015.2470601 ..... M. Li and S. Ghosal 4876
Photometric Stereo for General BRDFs via Reflection Sparsity Modeling http://dx.doi.org/10.1109/TIP.2015.2471081 .... T.-Q. Han and H.-L. Shen 4888
Unsupervised Unmixing of Hyperspectral Images Accounting for Endmember Variability http://dx.doi.org/10.1109/TIP.2015.2471182 ..... A. Halimi, N. Dobigeon, and J.-Y. Tourneret 4904
Dual Graph Regularized Latent Low-Rank Representation for Subspace Clustering http://dx.doi.org/10.1109/TIP.2015.2472277 ..... M. Yin, J. Gao, Z. Lin, Q. Shi, and Y. Guo 4918
Joint Group Sparse PCA for Compressed Hyperspectral Imaging http://dx.doi.org/10.1109/TIP.2015.2472280 .... Z. Khan, F. Shafait, and A. Mian 4934
2D Non-Separable Block-Lifting Structure and Its Application to M-Channel Perfect Reconstruction Filter Banks for Lossy-to-Lossless Image Coding http://dx.doi.org/10.1109/TIP.2015.2472294 ..... T. Suzuki and H. Kudo 4943
Stroke Detector and Structure Based Models for Character Recognition: A Comparative Study http://dx.doi.org/10.1109/TIP.2015.2473105 ..... C.-Z. Shi, S. Gao, M.-T. Liu, C.-Z. Qi, C.-H. Wang, and B.-H. Xiao 4952
A Probabilistic Method for Image Enhancement With Simultaneous Illumination and Reflectance Estimation http://dx.doi.org/10.1109/TIP.2015.2474701 ..... X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding 4965
Deformed Palmprint Matching Based on Stable Regions http://dx.doi.org/10.1109/TIP.2015.2478386 ..... X. Wu and Q. Zhao 4978
Efficient Nonnegative Tucker Decompositions: Algorithms and Uniqueness http://dx.doi.org/10.1109/TIP.2015.2478396 ..... G. Zhou, A. Cichocki, Q. Zhao, and S. Xie 4990
Rayleigh-Rice Mixture Parameter Estimation via EM Algorithm for Change Detection in Multispectral Images http://dx.doi.org/10.1109/TIP.2015.2474710 ..... M. Zanetti, F. Bovolo, and L. Bruzzone 5004
PCANet: A Simple Deep Learning Baseline for Image Classification? http://dx.doi.org/10.1109/TIP.2015.2475625 ..... T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma 5017
Visual Quality Evaluation of Image Object Segmentation: Subjective Assessment and Objective Measure http://dx.doi.org/10.1109/TIP.2015.2473099 ..... R. Shi, K. N. Ngan, S. Li, R. Paramesran, and H. Li 5033
Efficient and Robust Image Restoration Using Multiple-Feature L2-Relaxed Sparse Analysis Priors http://dx.doi.org/10.1109/TIP.2015.2478405 ..... J. Portilla, A. Tristán-Vega, and I. W. Selesnick 5046
Digital Image Watermarking via Adaptive Logo Texturization http://dx.doi.org/10.1109/TIP.2015.2476961 ..... M. Andalibi and D. M. Chandler 5060
From Local Similarities to Global Coding: A Framework for Coding Applications http://dx.doi.org/10.1109/TIP.2015.2465171 ..... A. Shaban, H. R. Rabiee, M. Najibi, and S. Yousefi 5074
Robust, Error-Tolerant Photometric Projector Compensation http://dx.doi.org/10.1109/TIP.2015.2478388 ..... A. Grundhöfer and D. Iwai 5086
A Surface Approximation Method for Image and Video Correspondences http://dx.doi.org/10.1109/TIP.2015.2462029 ..... J. Huang, B. Wang, W. Wang, and P. Sen 5100
Compact and Discriminative Descriptor Inference Using Multi-Cues http://dx.doi.org/10.1109/TIP.2015.2479917 ..... Y. Han, Y. Yang, F. Wu, and R. Hong 5114
Variational Dirichlet Blur Kernel Estimation http://dx.doi.org/10.1109/TIP.2015.2478407 ..... X. Zhou, J. Mateos, F. Zhou, R. Molina, and A. K. Katsaggelos 5127
Weighted Part Context Learning for Visual Tracking http://dx.doi.org/10.1109/TIP.2015.2479460 ..... G. Zhu, J. Wang, C. Zhao, and H. Lu 5140
Compositional Dictionaries for Domain Adaptive Face Recognition http://dx.doi.org/10.1109/TIP.2015.2479456 ..... Q. Qiu and R. Chellappa 5152
Fast and Robust Object Tracking via Probability Continuous Outlier Model http://dx.doi.org/10.1109/TIP.2015.2478399 ..... D. Wang, H. Lu, and C. Bo 5166
Local Sparse Structure Denoising for Low-Light-Level Image http://dx.doi.org/10.1109/TIP.2015.2447735 ..... J. Han, J. Yue, Y. Zhang, and L. Bai 5177
Horror Image Recognition Based on Context-Aware Multi-Instance Learning http://dx.doi.org/10.1109/TIP.2015.2479400 ..... B. Li, W. Xiong, O. Wu, W. Hu, S. Maybank, and S. Yan 5193
Dynamical System Approach for Edge Detection Using Coupled FitzHugh-Nagumo Neurons http://dx.doi.org/10.1109/TIP.2015.2467206 ..... S. Li, S. Dasmahapatra, and K. Maharatna 5206
Multiscale Tikhonov-Total Variation Image Restoration Using Spatially Varying Edge Coherence Exponent http://dx.doi.org/10.1109/TIP.2015.2479471 ..... V. B. S. Prasath, D. Vorotnikov, R. Pelapur, S. Jose, G. Seetharaman, and K. Palaniappan 5220
Visual Tracking Based on the Adaptive Color Attention Tuned Sparse Generative Object Model http://dx.doi.org/10.1109/TIP.2015.2479409 ..... C. Tian, X. Gao, W. Wei, and H. Zheng 5236
Optimized Quasi-Interpolators for Image Reconstruction http://dx.doi.org/10.1109/TIP.2015.2478385 ..... L. Sacht and D. Nehab 5249
Video Stabilization Based on Feature Trajectory Augmentation and Selection and Robust Mesh Grid Warping http://dx.doi.org/10.1109/TIP.2015.2479918 ..... Y. J. Koh, C. Lee, and C.-S. Kim 5260
Tracking Human Pose Using Max-Margin Markov Models http://dx.doi.org/10.1109/TIP.2015.2473662 ..... L. Zhao, X. Gao, D. Tao, and X. Li 5274
Gaussian Process Regression-Based Video Anomaly Detection and Localization With Hierarchical Feature Representation http://dx.doi.org/10.1109/TIP.2015.2479561 ..... K.-W. Cheng, Y.-T. Chen, and W.-H. Fang 5288
A Class of Manifold Regularized Multiplicative Update Algorithms for Image Clustering http://dx.doi.org/10.1109/TIP.2015.2457033 ..... S. Yang, Z. Yi, X. He, and X. Li 5302

Automatic Liver Segmentation Based on Shape Constraints and Deformable Graph Cut in CT Images http://dx.doi.org/10.1109/TIP.2015.2481326 ..... G. Li, X. Chen, F. Shi, W. Zhu, J. Tian, and D. Xiang 5315
MToS: A Tree of Shapes for Multivariate Images http://dx.doi.org/10.1109/TIP.2015.2480599 ..... E. Carlinet and T. Géraud 5330
Unsupervised Feature Selection via Nonnegative Spectral Analysis and Redundancy Control http://dx.doi.org/10.1109/TIP.2015.2479560 ..... Z. Li and J. Tang 5343
Cost-Sensitive Local Binary Feature Learning for Facial Age Estimation http://dx.doi.org/10.1109/TIP.2015.2481327 ..... J. Lu, V. E. Liong, and J. Zhou, 5356
Variational Depth From Focus Reconstruction http://dx.doi.org/10.1109/TIP.2015.2479469 ..... M. Moeller, M. Benning, C. Schönlieb, and D. Cremers 5369
Texture Classification Using Local Pattern Based on Vector Quantization http://dx.doi.org/10.1109/TIP.2015.2476955 ..... Z. Pan, H. Fan, and L. Zhang 5379
Designing Robust Sensing Matrix for Image Compression http://dx.doi.org/10.1109/TIP.2015.2479474 ..... G. Li, X. Li, S. Li, H. Bai, Q. Jiang, and X. He 5389
Extremal Regions Detection Guided by Maxima of Gradient Magnitude http://dx.doi.org/10.1109/TIP.2015.2477215 ..... M. Faraji, J. Shanbehzadeh, K. Nasrollahi, and T. B. Moeslund 5401
Robust Matching Cost Function for Stereo Correspondence Using Matching by Tone Mapping and Adaptive Orthogonal Integral Image http://dx.doi.org/10.1109/TIP.2015.2481702 ..... V. Q. Dinh, V. D. Nguyen, and J. W. Jeon 5416
Edge-Preserving Decomposition-Based Single Image Haze Removal http://dx.doi.org/10.1109/TIP.2015.2482903 ..... Z. Li and J. Zheng 5432
Common Visual Pattern Discovery via Nonlinear Mean Shift Clustering http://dx.doi.org/10.1109/TIP.2015.2481701 ..... L. Wang, D. Tang, Y. Guo, and M. N. Do 5442
A General-Thresholding Solution for lp(0 < p < 1) Regularized CT Reconstruction http://dx.doi.org/10.1109/TIP.2015.2468175 ..... C. Miao and H. Yu 5455
Image-Specific Prior Adaptation for Denoising http://dx.doi.org/10.1109/TIP.2015.2473098 ..... X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang 5469
DASH-N: Joint Hierarchical Domain Adaptation and Feature Learning http://dx.doi.org/10.1109/TIP.2015.2479405 ..... H. V. Nguyen, H. T. Ho, V. M. Patel, and R. Chellappa 5479
Layered Compression for High-Precision Depth Data http://dx.doi.org/10.1109/TIP.2015.2481324 .... D. Miao, J. Fu, Y. Lu, S. Li, and C. W. Chen 5492
Transforms for Intra Prediction Residuals Based on Prediction Inaccuracy Modeling http://dx.doi.org/10.1109/TIP.2015.2481328 ..... X. Cai and J. S. Lim 5505
Modeling, Prediction, and Reduction of 3D Crosstalk in Circular Polarized Stereoscopic LCDs http://dx.doi.org/10.1109/TIP.2015.2466114 ..... M. Zeng, A. E. Robinson, and T. Q. Nguyen 5516
Inpainting for Fringe Projection Profilometry Based on Geometrically Guided Iterative Regularization http://dx.doi.org/10.1109/TIP.2015.2481707 ..... Budianto and D. P. K. Lun 5531
Cross-Modal Subspace Learning via Pairwise Constraints http://dx.doi.org/10.1109/TIP.2015.2466106 ..... R. He, M. Zhang, L. Wang, Y. Ji, and Q. Yin 5543
Image Segmentation With Cage Active Contours http://dx.doi.org/10.1109/TIP.2015.2472298 ..... L. Garrido, M. Guerrieri, and L. Igual 5557
RISM: Single-Modal Image Registration via Rank-Induced Similarity Measure http://dx.doi.org/10.1109/TIP.2015.2479462 ..... A. Ghaffari and E. Fatemizadeh 5567
The STOne Transform: Multi-Resolution Image Enhancement and Compressive Video http://dx.doi.org/10.1109/TIP.2015.2474697 ..... T. Goldstein, L. Xu, K. F. Kelly, and R. Baraniuk 5581
Compressed Image Quality Metric Based on Perceptually Weighted Distortion http://dx.doi.org/10.1109/TIP.2015.2481319 ..... S. Hu, L. Jin, H. Wang, Y. Zhang, S. Kwong, and C.-C. J. Kuo 5594
High-Resolution Light Field Capture With Coded Aperture http://dx.doi.org/10.1109/TIP.2015.2468179 ..... Y.-P. Wang, L.-C. Wang, D.-H. Kong, and B.-C. Yin 5609
A Boosted Multi-Task Model for Pedestrian Detection With Occlusion Handling http://dx.doi.org/10.1109/TIP.2015.2483376 ..... C. Zhu and Y. Peng 5619
Encoding Color Information for Visual Tracking: Algorithms and Benchmark http://dx.doi.org/10.1109/TIP.2015.2482905 ..... P. Liang, E. Blasch, and H. Ling 5630
Kernelized Saliency-Based Person Re-Identification Through Multiple Metric Learning http://dx.doi.org/10.1109/TIP.2015.2487048 ..... N. Martinel, C. Micheloni, and G. L. Foresti 5645
Multimodal Deep Autoencoder for Human Pose Recover http://dx.doi.org/10.1109/TIP.2015.2487860y ..... C. Hong, J. Yu, J. Wan, D. Tao, and M. Wang 5659
Normalized Cut-Based Saliency Detection by Adaptive Multi-Level Region Merging http://dx.doi.org/10.1109/TIP.2015.2485782 ..... K. Fu, C. Gong, I. Y.-H. Gu, and J. Yang 5671
Dimensionality Reduction by Integrating Sparse Representation and Fisher Criterion and Its Applications http://dx.doi.org/10.1109/TIP.2015.2479559 ..... Q. Gao, Q. Wang, Y. Huang, X. Gao, X. Hong, and H. Zhang 5684
Interface Detection Using a Quenched-Noise Version of the Edwards-Wilkinson Equation http://dx.doi.org/10.1109/TIP.2015.2484061 ..... T. Turpeinen, M. Myllys, P. Kekäläinen, and J. Timonen 5696
Salient Object Detection: A Benchmark http://dx.doi.org/10.1109/TIP.2015.2487833 ..... A. Borji, M.-M. Cheng, H. Jiang, and J. Li 5706

Single Object Tracking With Fuzzy Least Squares Support Vector Machine http://dx.doi.org/10.1109/TIP.2015.2484068 ..... S. Zhang, S. Zhao, Y. Sui, and L. Zhang 5723
Topical Video Object Discovery From Key Frames by Modeling Word Co-Occurrence Prior http://dx.doi.org/10.1109/TIP.2015.2487834 ..... G. Zhao, J. Yuan, G. Hua, and J. Yang 5739
Local Inverse Tone Curve Learning for High Dynamic Range Image Scalable Compression http://dx.doi.org/10.1109/TIP.2015.2483899 ..... M. Le Pendu, C. Guillemot, and D. Thoreau 5753
Rough Sets and Stomped Normal Distribution for Simultaneous Segmentation and Bias Field Correction in Brain MR Images http://dx.doi.org/10.1109/TIP.2015.2488900 ..... A. Banerjee and P. Maji 5764
Beyond Explicit Codebook Generation: Visual Representation Using Implicitly Transferred Codebooks http://dx.doi.org/10.1109/TIP.2015.2485783 ..... C. Zhang, J. Cheng, J. Liu, J. Pang, Q. Huang, and Q. Tian 5777
Personal Object Discovery in First-Person Videos http://dx.doi.org/10.1109/TIP.2015.2487868 ..... C. Lu, R. Liao, and J. Jia 5789
Collaborative Sparse Regression Using Spatially Correlated Supports-Application to Hyperspectral Unmixing http://dx.doi.org/10.1109/TIP.2015.2487862 ..... Y. Altmann, M. Pereyra, and J. Bioucas-Dias 5800
Multi-View Learning With Incomplete Views http://dx.doi.org/10.1109/TIP.2015.2490539 ..... C. Xu, D. Tao, and C. Xu 5812
Joint Sparse Representation and Robust Feature-Level Fusion for Multi-Cue Visual Tracking http://dx.doi.org/10.1109/TIP.2015.2481325 ..... X. Lan, A. J. Ma, P. C. Yuen, and R. Chellappa 5826
Global Color Sparseness and a Local Statistics Prior for Fast Bilateral Filtering http://dx.doi.org/10.1109/TIP.2015.2492822 ..... M. G. Mozerov and J. van de Weijer 5842
Random Walk and Graph Cut for Co-Segmentation of Lung Tumor on PET-CT Images http://dx.doi.org/10.1109/TIP.2015.2488902 ..... W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen 5854
Robust Face Recognition via Minimum Error Entropy-Based Atomic Representation http://dx.doi.org/10.1109/TIP.2015.2492819 ..... Y. Wang, Y. Y. Tang, and L. Li 5868
Color Image Demosaicing Using Iterative Residual Interpolation http://dx.doi.org/10.1109/TIP.2015.2482899 ..... W. Ye and K.-K. Ma 5879
Local Wavelet Pattern: A New Feature Descriptor for Image Retrieval in Medical CT Databases http://dx.doi.org/10.1109/TIP.2015.2493446 ..... S. R. Dubey, S. K. Singh, and R. K. Singh 5892
Temporal Variance Analysis for Action Recognition http://dx.doi.org/10.1109/TIP.2015.2490551 ..... J. Miao, X. Xu, S. Qiu, C. Qing, and D. Tao 5904
STARE: Spatio-Temporal Attention Relocation for Multiple Structured Activities Detection http://dx.doi.org/10.1109/TIP.2015.2487837 ..... K. Lee, D. Ognibene, H. J. Chang, T.-K. Kim, and Y. Demiris 5916
Closed-Loop Restoration Approach to Blurry Images Based on Machine Learning and Feedback Optimization http://dx.doi.org/10.1109/TIP.2015.2492825 ..... S. Yousaf and S. Qin 5928
Segmentation of Overlapping Elliptical Objects in Silhouette Images http://dx.doi.org/10.1109/TIP.2015.2492828 ..... S. Zafari, T. Eerola, J. Sampo, H. Kälviäinen, and H. Haario 5942
Depth Analogy: Data-Driven Approach for Single Image Depth Estimation Using Gradient Samples http://dx.doi.org/10.1109/TIP.2015.2495261 ..... S. Choi, D. Min, B. Ham, Y. Kim, C. Oh, and K. Sohn 5953
A Benchmark and Comparative Study of Video-Based Face Recognition on COX Face Database http://dx.doi.org/10.1109/TIP.2015.2493448 ..... Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kuerban, and X. Chen 5967
Simultaneous Camera Path Optimization and Distraction Removal for Improving Amateur Video http://dx.doi.org/10.1109/TIP.2015.2493959 ..... F.-L. Zhang, J. Wang, H. Zhao, R. R. Martin, and S.-M. Hu 5982
Matching Images With Multiple Descriptors: An Unsupervised Approach for Locally Adaptive Descriptor Selection http://dx.doi.org/10.1109/TIP.2015.2496305 ..... Y.-T. Hu, Y.-Y. Lin, H.-Y. Chen, K.-J. Hsu, and B.-Y. Chen 5995
Distributed Signal Decorrelation and Detection in Multi View Camera Networks Using the Vector Sparse Matrix Transform http://dx.doi.org/10.1109/TIP.2015.2481709 ..... L. R. Bachega, S. Hariharan, C. A. Bouman, and N. B. Shroff 6011
Image Denoising With Edge-Preserving and Segmentation Based on Mask NHA http://dx.doi.org/10.1109/TIP.2015.2494461 ..... F. Hosotani, Y. Inuzuka, M. Hasegawa, S. Hirobayashi, and T. Misawa 6025
Micro-Expression Recognition Using Color Spaces http://dx.doi.org/10.1109/TIP.2015.2496314 ..... S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, C.-G. Zhou, X. Fu, M. Yang, and J. Tao 6034
Video Compression Artifact Reduction via Spatio-Temporal Multi-Hypothesis Prediction http://dx.doi.org/10.1109/TIP.2015.2485780 ..... X. Zhang, R. Xiong, W. Lin, S. Ma, J. Liu, and W. Gao 6048
An Underwater Color Image Quality Evaluation Metric http://dx.doi.org/10.1109/TIP.2015.2491020 ..... M. Yang and A. Sowmya 6062

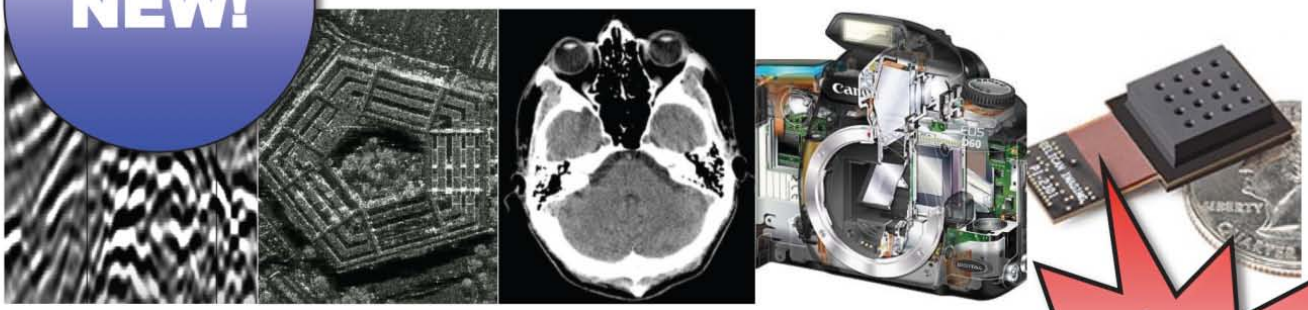
EDICS-Editor's Information Classification Scheme http://dx.doi.org/10.1109/TIP.2015.2499788 6072
Information for Authors http://dx.doi.org/10.1109/TIP.2015.2499787 6073

List of Reviewers http://dx.doi.org/10.1109/TIP.2015.2488780 Available online at http://ieeexplore.ieee.org

2015 INDEX http://dx.doi.org/10.1109/TIP.2015.2506438 Available online at http://ieeexplore.ieee.org

NEW!

IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING



30% FASTER TURN AROUND THAN TIP!

Editor-in-Chief
W. Clem Karl
Boston University

Technical Committee
Charles Bouman
Eric Miller
Peter Corcoran
Jong Chul Ye
Dave Brady
William Freeman

The IEEE Transactions on Computational Imaging publishes research results where computation plays an integral role in the image formation process. All areas of computational imaging are appropriate, ranging from the principles and theory of computational imaging, to modeling paradigms for computational imaging, to image formation methods, to the latest innovative computational imaging system designs. Topics of interest include, but are not limited to the following:

Computational Imaging Methods and Models

- Coded image sensing
- Compressed sensing
- Sparse and low-rank models
- Learning-based models, dictionary methods
- Graphical image models
- Perceptual models

Computational Image Formation

- Sparsity-based reconstruction
- Statistically-based inversion methods
- Multi-image and sensor fusion
- Optimization-based methods; proximal iterative methods, ADMM

Computational Photography

- Non-classical image capture
- Generalized illumination
- Time-of-flight imaging
- High dynamic range imaging
- Plenoptic imaging

Computational Consumer Imaging

- Mobile imaging, cell phone imaging
- Camera-array systems
- Depth cameras, multi-focus imaging
- Pervasive imaging, camera networks

Computational Acoustic Imaging

- Multi-static ultrasound imaging
- Photo-acoustic imaging
- Acoustic tomography

Computational Microscopy

- Holographic microscopy
- Quantitative phase imaging
- Multi-illumination microscopy
- Lensless microscopy
- Light field microscopy

Imaging Hardware and Software

- Embedded computing systems
- Big data computational imaging
- Integrated hardware/digital design

Tomographic Imaging

- X-ray CT
- PET
- SPECT

Magnetic Resonance Imaging

- Diffusion tensor imaging
- Fast acquisition

Radar Imaging

- Synthetic aperture imaging
- Inverse synthetic aperture imaging

Geophysical Imaging

- Multi-spectral imaging
- Ground penetrating radar
- Seismic tomography

Multi-spectral Imaging

- Multi-spectral imaging
- Hyper-spectral imaging
- Spectroscopic imaging

For more information on the **IEEE Transactions on Computational Imaging** see *http://www.signalprocessingsociety.org/publications/periodicals/tci/*

IEEE Signal Processing Society ® EMB GRSS

NOW WITH NO PAGE CHARGES!

General Chair
Lina Karam
Arizona State University

General Co-Chair
Aggelos Katsaggelos
Northwestern University

Technical Program Chairs
Fernando Pereira
Instituto Superior Técnico
Gaurav Sharma
University of Rochester

Innovation Program Chairs
Haohong Wang
TCL Research America
Jeff Bier
BDTI & Embedded Vision Alliance

Finance Chair
Sohail Dianat
Rochester Institute of Technology

Plenary Chairs
Michael Marcellin
University of Arizona
Sethuraman Panchanathan
Arizona State University

Special Sessions Chairs
Dinei Florencio
Microsoft Research
Chaker Larabi
Poitiers University
Zhou Wang
University of Waterloo

Tutorials Chairs
Ghassan AlRegib
Georgia Tech
Rony Ferzli
Intel

Publicity Chair
Michel Sarkis
Qualcomm Technologies Inc.

Awards Chairs
Vivek Goyal
Boston University
Ivana Tosic
Ricoh Innovations

Exhibits Chair
David Frakes
Arizona State University & Google

Publication Chairs
Patrick Le Callet
Nantes University
Baoxin Li
Arizona State University

Local Arrangement Chairs
Jorge Caviedes
Intel
Pavan Turaga
Arizona State University

Registration Chair
Ricardo De Queiroz
Universidade de Brasilia

Conference Management
Conference Management Services

The 23rd IEEE International Conference on Image Processing (ICIP) will be held in the Phoenix Convention Centre, Phoenix, Arizona, USA, on September 25 - 28, 2016. ICIP is the world's largest and most comprehensive technical conference focused on image and video processing and computer vision. In addition to the Technical Program, ICIP 2016 will feature **an Innovation Program focused on innovative vision technologies and fostering innovation, entrepreneurship, and networking**. The conference will feature world-class speakers, tutorials, exhibits, and a vision technology showcase.

Topics in the ICIP 2016 Technical Program include but are not limited to the following:

Filtering, Transforms, Multi-Resolution Processing
Restoration, Enhancement, Super-Resolution
Computer Vision Algorithms and Technologies
Compression, Transmission, Storage, Retrieval
Computational Imaging
Color and Multispectral Processing
Multi-View and Stereoscopic Processing
Multi-Temporal and Spatio-Temporal Processing
Video Processing and Analytics
Authentication and Biometrics
Biological and Perceptual-based Processing
Visual Quality Assessment
Scanning, Display, and Printing
Document and Synthetic Visual Processing
Applications to various fields (e.g., biomedical, Advanced Driving Assist Systems, assistive living, security, learning, health and environmental monitoring, manufacturing, consumer electronics)

The ICIP 2016 innovation program will feature a **vision technology showcase** of state-of-the-art vision technologies, innovation challenges, talks by innovation leaders and entrepreneurs, tutorials, and networking.

Paper Submission: Prospective authors are invited to submit full-length papers at the conference website, with up to four pages for technical content including figures and references, and with one additional optional 5th page for references only. Submission instructions, templates for the required paper format, and information on "no show" policy are available at www.icip2016.com.

Tutorials and Special Sessions Proposals: Tutorials will be held on September 25, 2016. Tutorial proposals should be submitted to *tutorials@icip2016.com* and must include title, outline, contact information, biography and selected publications for the presenter(s), and a description of the tutorial and material to be distributed to participants. Special Sessions proposals should be submitted to *specialsessions@icip2016.com* and must include a topical title, rationale, session outline, contact information, and a list of invited papers. For detailed submission guidelines, please refer the ICIP 2016 website at **www.icip2016.com**.

Important Deadlines:
Special Session and Tutorial Proposals: November 16, 2015
Notification of Special Session and Tutorial Acceptance: December 18, 2015

Paper Submissions: January 25, 2016
Notification of Paper Acceptance: April 30, 2016

Visual Technology Innovator Award Nomination: March 30, 2016

Revised Paper Upload Deadline: May 30, 2016
Authors' Registration Deadline: May 30, 2016

http://www.facebook.com/icip2016
https://twitter.com/icip2016/
https://www.linkedin.com/groups/ICIP-2016-6940658

IEEE TRANSACTIONS ON

INFORMATION FORENSICS AND SECURITY

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

DECEMBER 2015 VOLUME 10 NUMBER 12 ITIFA6 (ISSN 1556-6013)

PAPERS

Tamper Detection of JPEG Image Due to Seam Modifications http://dx.doi.org/10.1109/TIFS.2015.2464776 ..... *K. Wattanachote, T. K. Shih, W.-L. Chang, and H.-H. Chang* 2477

Wireless Device Identification Based on RF Oscillator Imperfections http://dx.doi.org/10.1109/TIFS.2015.2464778 ..... *A. C. Polak and D. L. Goeckel* 2492

AAC-OT: Accountable Oblivious Transfer With Access Control http://dx.doi.org/10.1109/TIFS.2015.2464781 ..... *J. Han, W. Susilo, Y. Mu, M. H. Au, and J. Cao* 2502

TorWard: Discovery, Blocking, and Traceback of Malicious Traffic Over Tor http://dx.doi.org/10.1109/TIFS.2015.2465934 ..... *Z. Ling, J. Luo, K. Wu, W. Yu, and X. Fu* 2515

On the Continuity of the Secrecy Capacity of Compound and Arbitrarily Varying Wiretap Channels http://dx.doi.org/10.1109/TIFS.2015.2465937 ..... *H. Boche, R. F. Schaefer, and H. V. Poor* 2531

Reliable and Trustworthy Memory Acquisition on Smartphones http://dx.doi.org/10.1109/TIFS.2015.2467356 ..... *H. Sun, K. Sun, Y. Wang, and J. Jing* 2547

An Effective Address Mutation Approach for Disrupting Reconnaissance Attacks http://dx.doi.org/10.1109/TIFS.2015.2467358 ..... *J. H. Jafarian, E. Al-Shaer, and Q. Duan* 2562

User-Centric View of Jamming Games in Cognitive Radio Networks http://dx.doi.org/10.1109/TIFS.2015.2467593 ..... *L. Xiao, J. Liu, Q. Li, N. B. Mandayam, and H. V. Poor* 2578

Employing Program Semantics for Malware Detection http://dx.doi.org/10.1109/TIFS.2015.2469253 ..... *S. Naval, V. Laxmi, M. Rajarajan, M. S. Gaur, and M. Conti* 2591

Subband Time-Frequency Image Texture Features for Robust Audio Surveillance http://dx.doi.org/10.1109/TIFS.2015.2469254 ..... *R. V. Sharan and T. J. Moir* 2605

K-Center: An Approach on the Multi-Source Identification of Information Diffusion http://dx.doi.org/10.1109/TIFS.2015.2469256 ..... *J. Jiang, S. Wen, S. Yu, Y. Xiang, and W. Zhou* 2616
Modeling and Extending the Ensemble Classifier for Steganalysis of Digital Images Using Hypothesis Testing Theory http://dx.doi.org/10.1109/TIFS.2015.2470220 ..... *R. Cogranne and J. Fridrich* 2627
BL-MLE: Block-Level Message-Locked Encryption for Secure Large File Deduplication http://dx.doi.org/10.1109/TIFS.2015.2470221 ..... *R. Chen, Y. Mu, G. Yang, and F. Guo* 2643
Improved Generation of Identifiers, Secret Keys, and Random Numbers From SRAMs http://dx.doi.org/10.1109/TIFS.2015.2471279 ..... *I. Baturone, M. A. Prada-Delgado, and S. Eiroa* 2653
Using Statistical Image Model for JPEG Steganography: Uniform Embedding Revisited http://dx.doi.org/10.1109/TIFS.2015.2473815 ..... *L. Guo, J. Ni, W. Su, C. Tang, and Y.-Q. Shi* 2669
An Efficient Identity-Based Conditional Privacy-Preserving Authentication Scheme for Vehicular Ad Hoc Networks http://dx.doi.org/10.1109/TIFS.2015.2473820 ..... *D. He, S. Zeadally, B. Xu, and X. Huang* 2681
Camera Model Identification With Unknown Models http://dx.doi.org/10.1109/TIFS.2015.2474836 ..... *Y. Huang, J. Zhang, and H. Huang* 2692

List of Reviewers http://dx.doi.org/10.1109/TIFS.2015.2486198 2705

EDICS-Editor's Information Classification Scheme http://dx.doi.org/10.1109/TIFS.2015.2495921 2715
Information for Authors http://dx.doi.org/10.1109/TIFS.2015.2495919 2716

ANNOUNCEMENTS

Call for Papers-IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS Special Issue on Distributed Information Processing in Social Networks http://dx.doi.org/10.1109/TIFS.2015.2498461 2718
Call for Papers-IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS Special Issue on Inference and Learning Over Networks http://dx.doi.org/10.1109/TIFS.2015.2498469 2719
Introducing IEEE Collabratec http://dx.doi.org/10.1109/TIFS.2015.2498378 2720

2015 INDEX http://dx.doi.org/10.1109/TIFS.2015.2508578 Available online at http://ieeexplore.org

IEEE TRANSACTIONS ON

INFORMATION FORENSICS AND SECURITY

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.signalprocessingsociety.org

JANUARY 2016 VOLUME 11 NUMBER 1 (ISSN 1556-6013)

PAPERS

Manifold Learning and Spectral Clustering for Image Phylogeny Forests http://dx.doi.org/10.1109/TIFS.2015.2442527 ..... *M. A. Oikawa, Z. Dias, A. de Rezende Rocha, and S. Goldenstein* 5

JPGcarve: An Advanced Tool for Automated Recovery of Fragmented JPEG Files http://dx.doi.org/10.1109/TIFS.2015.2475238 ..... *J. De Bock and P. De Smet* 19

Hidden Ciphertext Policy Attribute-Based Encryption Under Standard Assumptions http://dx.doi.org/10.1109/TIFS.2015.2475723 ..... *T. V. X. Phuong, G. Yang, and W. Susilo* 35

On the Fast Fading Gaussian Wiretap Channel With Statistical Channel State Information at the Transmitter http://dx.doi.org/10.1109/TIFS.2015.2476464 ..... *P.-H. Lin and E. Jorswieck* 46

Unconditionally Secure, Universally Composable Privacy Preserving Linear Algebra http://dx.doi.org/10.1109/TIFS.2015.2476783 ..... *B. David, R. Dowsley, J. van de Graaf, D. Marques, A. C. A. Nascimento, and A. C. B. Pinto* 59

Large System Secrecy Rate Analysis for SWIPT MIMO Wiretap Channels http://dx.doi.org/10.1109/TIFS.2015.2477050 ..... *J. Zhang, C. Yuen, C.-K. Wen, S. Jin, K.-K. Wong, and H. Zhu* 74

Cooperative Change Detection for Voltage Quality Monitoring in Smart Grids http://dx.doi.org/10.1109/TIFS.2015.2477796 ..... *S. Li and X. Wang* 86

Designing Secure and Dependable Mobile Sensing Mechanisms With Revenue Guarantees http://dx.doi.org/10.1109/TIFS.2015.2478739 ... *Y. Zhang, H. Zhang, S. Tang, and S. Zhong* 100

Analyzing Android Encrypted Network Traffic to Identify User Actions http://dx.doi.org/10.1109/TIFS.2015.2478741 ..... *M. Conti, L. V. Mancini, R. Spolaor, and N. V. Verde* 114

Preprocessing Reference Sensor Pattern Noise via Spectrum Equalization http://dx.doi.org/10.1109/TIFS.2015.2478748 ..... *X. Lin and C.-T. Li* 126

A Joint Source-Channel Adaptive Scheme for Wireless H.264/AVC Video Authentication http://dx.doi.org/10.1109/TIFS.2015.2481366 .... *X. Zhu and C. W. Chen* 141

Energy Harvesting for Physical-Layer Security in OFDMA Networks http://dx.doi.org/10.1109/TIFS.2015.2481797 ... *M. Zhang and Y. Liu* 154

On the Permanence of EEG Signals for Biometric Recognition http://dx.doi.org/10.1109/TIFS.2015.2481870 ..... *E. Maiorana, D. La Rocca, and P. Campisi* 163

Profiling Online Social Behaviors for Compromised Account Detection http://dx.doi.org/10.1109/TIFS.2015.2482465 ..... *X. Ruan, Z. Wu, H. Wang, and S. Jajodia* 176

Optimized Search-and-Compute Circuits and Their Application to Query Evaluation on Encrypted Data http://dx.doi.org/10.1109/TIFS.2015.2483486 ..... *J. H. Cheon, M. Kim, and M. Kim* 188

Robust Image Hashing With Ring Partition and Invariant Vector Distance http://dx.doi.org/10.1109/TIFS.2015.2485163 ..... *Z. Tang, X. Zhang, X. Li, and S. Zhang* 200

The 2016 IEEE Image, Video, and Multidimensional Signal Processing (IVMSP) Workshop is the 12th of a series of unique meetings that bring together researchers in academia and industry to share the most recent and exciting advances in image, video, and multidimensional signal processing and analysis. The main theme of the 2016 IVMSP Workshop is **low-dimensional models for image and video processing and analysis**.

The scientific program of IVMSP 2016 will include plenary talks, invited sessions, as well as regular sessions with contributed research papers. Papers are solicited in (but not limited to) the following topics:

- Image and video modeling on manifolds
- Manifold learning for classification and recognition
- Manifold sparse coding
- Graph-based manifold ranking
- Information geometry on probabilistic manifolds
- Low dimensional models and low rank methods
- Sparse and low-rank models in learning and pattern recognition
- Graphical models for image and video analysis
- Other related topics

PAPER SUBMISSION

Papers cannot be longer than 5 pages (double-column IEEE conference format), including all text, figures, and references. The 5th page cannot contain any text other than references. See the website for additional information regarding the submission process: www.ivmsp2016.org.

BEST STUDENT PAPER AWARDS

The IVMSP Best Student Paper Awards will be granted to the first, second, and third best contributed papers in the regular sessions for which a student is the principal author and presenter. The selection will be based on the technical quality, originality, and clarity of the submission.

IMPORTANT DATES

• Submission of full papers	March 10, 2016
• Notification of acceptance	May 10, 2016
• Author advance registration	June 1, 2016
• Camera-ready paper submission	June 1, 2016

General chair

Yannick Berthoumieu
Univ. of Bordeaux, France

Technical program co-chairs

Christine Guillemot, INRIA, France

Mujdat Cetin, Sabanci Univ., Turkey

Plenary and invited session chairs

Béatrice Pesquet-Popescu, Telecom-paristech, France

Pascal Frossard, EPFL, Switzerland

Financial chair

Christian Germain, Univ. of Bordeaux, France

Publication co-chairs

Marc Donias, Univ. of Bordeaux, France

Nelly Pustelnik, ENS Lyon, France

Local arrangement co-chairs

Nathalie Toulon, Univ. of Bordeaux, France

Jean-Pierre Da Costa, Univ. of Bordeaux, France

IEEE Signal Processing Society

IEEE TRANSACTIONS ON

MULTIMEDIA

A PUBLICATION OF
THE IEEE SIGNAL PROCESSING SOCIETY
THE IEEE CIRCUITS AND SYSTEMS SOCIETY
THE IEEE COMMUNICATIONS SOCIETY

IEEE Signal Processing Society

CAS IEEE CIRCUITS AND SYSTEMS SOCIETY

IEEE COMMUNICATIONS SOCIETY

http://www.signalprocessingsociety.org/tmm/

TECHNICALLY COSPONSORED BY THE IEEE COMPUTER SOCIETY

IEEE computer society

NOVEMBER 2015 VOLUME 17 NUMBER 11 ITMUF8 (ISSN 1520-9210)

SPECIAL ISSUE ON DEEP LEARNING

GUEST EDITORIAL

Guest Editorial: Deep Learning for Multimedia Computing http://dx.doi.org/10.1109/TMM.2015.2485538 ..... *G.-J. Qi, H. Larochelle, B. Huet, J. Luo, and K. Yu* 1873

SPECIAL ISSUE PAPERS

Describing Multimedia Content Using Attention-Based Encoder-Decoder Networks *(Invited Paper)* http://dx.doi.org/10.1109/TMM.2015.2477044 ..... *K. Cho, A. Courville, and Y. Bengio* 1875

Large-Margin Multi-Modal Deep Learning for RGB-D Object Recognition http://dx.doi.org/10.1109/TMM.2015.2476655 ..... *A. Wang, J. Lu, J. Cai, T.-J. Cham, and G. Wang* 1887

RGB-D Object Recognition via Incorporating Latent Data Structure and Prior Knowledge http://dx.doi.org/10.1109/TMM.2015.2476660 ..... *J. Tang, L. Jin, Z. Li, and S. Gao* 1899

A Continuous Learning Framework for Activity Recognition Using Deep Hybrid Feature Models http://dx.doi.org/10.1109/TMM.2015.2477242 ..... *M. Hasan and A. K. Roy-Chowdhury* 1909

Unconstrained Multimodal Multi-Label Learning http://dx.doi.org/10.1109/TMM.2015.2476658 ..... *Y. Huang, W. Wang, and L. Wang* 1923

IEEE

Heterogeneous Feature Selection With Multi-Modal Deep Neural Networks and Sparse Group LASSO http://dx.doi.org/10.1109/TMM.2015.2477058 ..... *L. Zhao, Q. Hu, and W. Wang* 1936
Multi-Task CNN Model for Attribute Prediction http://dx.doi.org/10.1109/TMM.2015.2477680 ..... *A. H. Abdulnabi, G. Wang, J. Lu, and K. Jia* 1949
Learning Representative Deep Features for Image Set Analysis http://dx.doi.org/10.1109/TMM.2015.2477681 ..... *Z. Wu, Y. Huang, and L. Wang* 1960
Cross Indexing With Grouplets http://dx.doi.org/10.1109/TMM.2015.2478055 ..... *S. Zhang, X. Wang, Y. Lin, and Q. Tian* 1969
Understanding Blooming Human Groups in Social Networks http://dx.doi.org/10.1109/TMM.2015.2476657 ..... *R. Hong, Z. Hu, L. Liu, M. Wang, S. Yan, and Q. Tian* 1980
Weakly Supervised Deep Metric Learning for Community-Contributed Image Retrieval http://dx.doi.org/10.1109/TMM.2015.2477035 ..... *Z. Li and J. Tang* 1989
Learning Cross Space Mapping via DNN Using Large Scale Click-Through Logs http://dx.doi.org/10.1109/TMM.2015.2480340 ..... *W. Yu, K. Yang, Y. Bai, H. Yao, and Y. Rui* 2000
Deep Multimodal Learning for Affective Analysis and Retrieval http://dx.doi.org/10.1109/TMM.2015.2482228 ..... *L. Pang, S. Zhu, and C.-W. Ngo* 2008
Rating Image Aesthetics Using Deep Learning http://dx.doi.org/10.1109/TMM.2015.2477040 ..... *X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang* 2021
Query-Dependent Aesthetic Model With Deep Learning for Photo Quality Assessment http://dx.doi.org/10.1109/TMM.2015.2479916 ..... *X. Tian, Z. Dong, K. Yang, and T. Mei* 2035
Robust Face Recognition via Multimodal Deep Face Representation http://dx.doi.org/10.1109/TMM.2015.2477042 ..... *C. Ding and D. Tao* 2049
Deep Learning and Music Adversaries http://dx.doi.org/10.1109/TMM.2015.2478068 ..... *C. Kereliuk, B. L. Sturm, and J. Larsen* 2059
DeepBag: Recognizing Handbag Models http://dx.doi.org/10.1109/TMM.2015.2480228 ..... *Y. Wang, S. Li, and A. C. Kot* 2072
Predicting Eye Fixations on Webpage With an Ensemble of Early Features and High-Level Representations from Deep Network http://dx.doi.org/10.1109/TMM.2015.2483370 ..... *C. Shen, X. Huang, and Q. Zhao* 2084
Deep Head Pose: Gaze-Direction Estimation in Multimodal Video http://dx.doi.org/10.1109/TMM.2015.2482819 ..... *S. S. Mukherjee and N. M. Robertson* 2094

CALLS FOR PAPERS

IEEE TRANSACTIONS ON MULTIMEDIA Special Issue on Multimedia-Based Healthcare http://dx.doi.org/10.1109/TMM.2015.2483038 2108
IEEE International Conference on Multimedia and Expo http://dx.doi.org/10.1109/TMM.2015.2483039 2109

IEEE TRANSACTIONS ON MULTIMEDIA

A PUBLICATION OF
THE IEEE SIGNAL PROCESSING SOCIETY
THE IEEE CIRCUITS AND SYSTEMS SOCIETY
THE IEEE COMMUNICATIONS SOCIETY

IEEE Signal Processing Society · CAS · IEEE COMMUNICATIONS SOCIETY

http://www.signalprocessingsociety.org/tmm/

TECHNICALLY COSPONSORED BY THE IEEE COMPUTER SOCIETY

IEEE computer society

DECEMBER 2015 VOLUME 17 NUMBER 12 ITMUF8 (ISSN 1520-9210)

PAPERS

Audio/Video Analysis and Synthesis

Real-Time Piano Music Transcription Based on Computer Vision http://dx.doi.org/10.1109/TMM.2015.2473702 ..... *M. Akbari and H. Cheng* 2113
Content-Aware Video2Comics With Manga-Style Layout http://dx.doi.org/10.1109/TMM.2015.2474263 ... *G. Jing, Y. Hu, Y. Guo, Y. Yu, and W. Wang* 2122

Compression and Coding

Rate Distortion Optimized Inter-View Frame Level Bit Allocation Method for MV-HEVC http://dx.doi.org/10.1109/TMM.2015.2477682 ..... *H. Yuan, S. Kwong, X. Wang, W. Gao, and Y. Zhang* 2134
Fast HEVC Inter CU Decision Based on Latent SAD Estimation http://dx.doi.org/10.1109/TMM.2015.2491018 ..... *J. Xiong, H. Li, F. Meng, Q. Wu, and K. N. Ngan* 2147

3D Audio/Video Processing

Faithful Disocclusion Filling in Depth Image Based Rendering Using Superpixel-Based Inpainting http://dx.doi.org/10.1109/TMM.2015.2476372 ..... *M. Schmeing and X. Jiang* 2160

Quality Assessment and User Experience

Perceptual Quality Assessment for 3D Triangle Mesh Based on Curvature http://dx.doi.org/10.1109/TMM.2015.2484221 ..... *L. Dong, Y. Fang, W. Lin, and H. S. Seah* 2174
Multiple Emotion Tagging for Multimedia Data by Exploiting High-Order Dependencies Among Emotions http://dx.doi.org/10.1109/TMM.2015.2484966 ..... *S. Wang, J. Wang, Z. Wang, and Q. Ji* 2185
Transition of Visual Attention Assessment in Stereoscopic Images With Evaluation of Subjective Visual Quality and Discomfort http://dx.doi.org/10.1109/TMM.2015.2493367 ..... *H. Kim and S. Lee* 2198

Subjective Test Methodology and Databases

The Accuracy of Subjects in a Quality Experiment: A Theoretical Subject Model http://dx.doi.org/10.1109/TMM.2015.2484963 ..... *L. Janowski and M. Pinson* 2210

Content Description and Annotation

Video Object Segmentation Via Dense Trajectories http://dx.doi.org/10.1109/TMM.2015.2481711 ..... *L. Chen, J. Shen, W. Wang, and B. Ni* 2225

IEEE

Knowledge and Semantics Modeling for Multimedia Databases

Interactive Multimodal Learning for Venue Recommendation http://dx.doi.org/10.1109/TMM.2015.2480007 ..... *J. Zahálka, S. Rudinac, and M. Worring* 2235

Multimedia Search and Retrieval

Biased Discriminant Analysis With Feature Line Embedding for Relevance Feedback-Based Image Retrieval http://dx.doi.org/10.1109/TMM.2015.2492926 ..... *Y.-C. Wang, C.-C. Han, C.-T. Hsieh, Y.-N. Chen, and K.-C. Fan* 2245

Social and Web Multimedia

Cross-OSN User Modeling by Homogeneous Behavior Quantification and Local Social Regularization http://dx.doi.org/10.1109/TMM.2015.2486524 ..... *J. Sang, Z. Deng, D. Lu, and C. Xu* 2259

A Multifaceted Approach to Social Multimedia-Based Prediction of Elections http://dx.doi.org/10.1109/TMM.2015.2487863 ..... *Q. You, L. Cao, Y. Cong, X. Zhang, and J. Luo* 2271

Word-of-Mouth Understanding: Entity-Centric Multimodal Aspect-Opinion Mining in Social Media http://dx.doi.org/10.1109/TMM.2015.2491019 ..... *Q. Fang, C. Xu, J. Sang, M. S. Hossain, and G. Muhammad* 2281

Multimedia Streaming and Transport

Optimizing HTTP-Based Adaptive Streaming in Vehicular Environment Using Markov Decision Process http://dx.doi.org/10.1109/TMM.2015.2494458 ..... *A. Bokani, M. Hassan, S. Kanhere, and X. Zhu* 2297

Economic and Financial Models

Profit Optimization for Wireless Video Broadcasting Systems Based on Polymatroidal Analysis http://dx.doi.org/10.1109/TMM.2015.2479860 ..... *W. Ji, P. Frossard, B.-W. Chen, and Y. Chen* 2310

Multimedia Algorithms, Systems, and Interfaces

Controlling a Robotic Fish via a Natural User Interface for Informal Science Education http://dx.doi.org/10.1109/TMM.2015.2480226 ..... *P. Phamduy, M. DeBellis, and M. Porfiri* 2328

CORRESPONDENCES

3D Audio Signal Processing

Reduced Reference Stereoscopic Image Quality Assessment Based on Binocular Perceptual Information http://dx.doi.org/10.1109/TMM.2015.2493781 ..... *F. Qi, D. Zhao, and W. Gao* 2338

Wireless/Mobile Multimedia

Adaptive Scalable Video Transmission Strategy in Energy Harvesting Communication System http://dx.doi.org/10.1109/TMM.2015.2476662 ..... *X. Wu, J. Yang, Y. Ran, and H. Xi* 2345

COMMENTS AND CORRECTIONS

Correction to "An Energy-Efficient Coarse-Grained Reconfigurable Processing Unit for Multiple-Standard Video Decoding" http://dx.doi.org/10.1109/TMM.2015.2492738 ..... *L. Liu, D. Wang, M. Zhu, Y. Wang, S. Yin, P. Cao, J. Yang, and S. Wei* 2354

List of Reviewers http://dx.doi.org/10.1109/TMM.2015.2492718 2356

ANNOUNCEMENTS

IEEE TRANSACTIONS ON MULTIMEDIA Special Issue on Multimedia-Based Healthcare http://dx.doi.org/10.1109/TMM.2015.2483040 2361

IEEE International Conference on Multimedia and Expo http://dx.doi.org/10.1109/TMM.2015.2483041 2362

Introducing IEEE Collabratec http://dx.doi.org/10.1109/TMM.2015.2500079 2363

Information for Authors http://dx.doi.org/10.1109/TMM.2015.2500118 2367

2015 INDEX http://dx.doi.org/10.1109/TMM.2015.2509798 ..... Available online at http://ieeexplore.ieee.org

SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags

IEEE JOURNAL OF
SELECTED TOPICS IN SIGNAL PROCESSING

IEEE Signal Processing Society

www.ieee.org/sp/index.html

OCTOBER 2015 VOLUME 9 NUMBER 7 IJSTGY (ISSN 1932-4553)

ISSUE ON SIGNAL AND INFORMATION PROCESSING FOR PRIVACY

EDITORIAL

Introduction to the Issue on Signal and Information Processing for Privacy http://dx.doi.org/10.1109/JSTSP.2015.2462391 ..... *W. Trappe, L. Sankar, R. Poovendran, H. Lee, and S. Capkun* 1173

PAPERS

The Staircase Mechanism in Differential Privacy http://dx.doi.org/10.1109/JSTSP.2015.2425831 ..... *Q. Geng, P. Kairouz, S. Oh, and P. Viswanath* 1176

Optical Signal Processing and Stealth Transmission for Privacy http://dx.doi.org/10.1109/JSTSP.2015.2424690 ..... *B. Wu, B. J. Shastri, P. Mittal, A. N. Tait, and P. R. Prucnal* 1185

Achieving Undetectable Communication http://dx.doi.org/10.1109/JSTSP.2015.2421477 ..... *S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst* 1195

Distributed Secret Dissemination Across a Network http://dx.doi.org/10.1109/JSTSP.2015.2422682 ..... *N. B. Shah, K. V. Rashmi, and K. Ramchandran* 1206

Secure Comparison Protocols in the Semi-Honest Model http://dx.doi.org/10.1109/JSTSP.2015.2429117 ..... *T. Veugen, F. Blom, S. J. A. de Hoogh, and Z. Erkin* 1217

Efficient Private Information Retrieval Over Unsynchronized Databases http://dx.doi.org/10.1109/JSTSP.2015.2432740 ..... *G. Fanti and K. Ramchandran* 1229

IEEE

SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags

Managing Your Private and Public Data: Bringing Down Inference Attacks Against Your Privacy http://dx.doi.org/10.1109/JSTSP.2015.2442227 ..... S. Salamatian, A. Zhang, F. du Pin Calmon, S. Bhamidipati, N. Fawaz, B. Kveton, P. Oliveira, and N. Taft 1240
Privacy or Utility in Data Collection? A Contract Theoretic Approach http://dx.doi.org/10.1109/JSTSP.2015.2425798 ..... L. Xu, C. Jiang, Y. Chen, Y. Ren, and K. J. R. Liu 1256
A Study of Online Social Network Privacy Via the TAPE Framework http://dx.doi.org/10.1109/JSTSP.2015.2427774 ..... Y. Zeng, Y. Sun, L. Xing, and V. Vokkarane 1270
Enabling Data Exchange in Two-Agent Interactive Systems Under Privacy Constraints http://dx.doi.org/10.1109/JSTSP.2015.2427775 ..... E. V. Belmega, L. Sankar, and H. V. Poor 1285
A Secure Radio Environment Map Database to Share Spectrum http://dx.doi.org/10.1109/JSTSP.2015.2426132 ..... S. Sodagari 1298
A Belief Propagation Approach to Privacy-Preserving Item-Based Collaborative Filtering http://dx.doi.org/10.1109/JSTSP.2015.2426677 ..... J. Zou and F. Fekri 1306
The Price of Privacy in Untrusted Recommender Systems http://dx.doi.org/10.1109/JSTSP.2015.2423254 ..... S. Banerjee, N. Hegde, and L. Massoulié 1319
PPDM: A Privacy-Preserving Protocol for Cloud-Assisted e-Healthcare Systems http://dx.doi.org/10.1109/JSTSP.2015.2427113 ..... J. Zhou, Z. Cao, X. Dong, and X. Lin 1332
Privacy-Aware Distributed Bayesian Detection http://dx.doi.org/10.1109/JSTSP.2015.2429123 ..... Z. Li and T. J. Oechtering 1345

CALL FOR PAPERS

IEEE Signal Processing Society

IEEE Journal of Selected Topics in Signal processing

Special Issue on Advances in Time/Frequency Modulated Array Signal Processing

Aims and Scope

In recent years, time/frequency modulated array techniques have developed beyond the focus of antenna design to address promising applications in radar, navigation, communications, microwave imaging, and biomedical engineering. Specifically, a time modulated array (TMA), by connecting and disconnecting the elements from the feed network, can create a beampattern with low sidelobes. Two typical frequency modulated arrays are the frequency diverse array (FDA) and multiple-input multiple-output (MIMO). The former produces a range-dependent pattern, whereas the latter provides increased degrees-of-freedom.

There are still many technical challenges for time/frequency modulated arrays, which include: (a) optimization of array parameters involving waveform design is required to further assess their affects on performance; (b) although linear array geometry is predominantly considered in the literature, theoretical analysis and applications of time/frequency modulated arrays should expose possible benefits of alternative geometries, and direct further studies toward advanced concepts for novel array structures; and (c) optimal time/frequency modulated array signal processing algorithms should be devised to address the issues of range, time, angle, and frequency dependent responses for target localization and dimensionality reduction. There is thus a huge demand for developing innovative, effective and efficient signal processing algorithms for time/frequency modulated array techniques and systems.

This special issue will touch on a wide variety of signal processing topics for optimal time/frequency modulated array design and their potential applications. It aims to compile relevant research contributions from various disciplines including signal processing, radar, wireless communications, antenna array design, geophysics, biomedical engineering, and applied mathematics to foster future research in this emerging area.

Topics of Interest include (but not limited to):

Signal Processing for Optimal Array Design	**Potential Applications**
– Space, time / range and frequency signal processing	– Cognitive radar / communications
– Dimensionality reduction algorithms	– Physical security communications
– Range-coupled adaptive signal processing	– Low probability of identification (LPI) radar
– Biomimetic spatial processing	– Adaptive interference/clutter suppression
– Array manifold analysis	– Ground moving target indication
– Low-sidelobe beampattern synthesis	– Microwave imaging
– Array parameter optimization and design	– Source detection and estimation
– Multidimensional waveform design / optimization	– Target localization and tracking
– Information fusion and knowledge extraction	– MIMO / hybrid array / stealth radar
– Array calibration and diagnosis	– Biomedical engineering

Important Dates:

Manuscript submission due: 1 April 2016
First review completed: 30 June 2016
Revised manuscript due: 15 August 2016
Second review completed: 1 October 2016
Final manuscript due: 15 November 2016
Publication: March 2017

Prospective authors should visit http://www.signalprocessingsociety.org/publications/periodicals/jstsp/ for information on paper submission. Manuscripts should be submitted at http://mc.manuscriptcentral.com/jstsp-ieee.

Guest Editors:

Hing Cheung So, City University of Hong Kong, Hong Kong, email: hcso@ee.cityu.edu.hk
Moeness G. Amin, Villanova University, USA, email: moeness.amin@villanova.edu
Shannon Blunt, University of Kansas, USA, email: sdblunt@ittc.ku.edu
Fulvio Gini, University of Pisa, Italy, email: f.gini@ing.unipi.it
Wen-Qin Wang, University of Electronic Science and Technology of China, China, email: wqwang@uestc.edu.cn

Call for Papers
IEEE Journal of Selected Topics in Signal Processing

Special Issue on Measuring Quality of Experience for Advanced Media Technologies and Services

Progress in technology has led to a stunning increase in the quality of multimedia content over the past few decades: Ultra-HD or 3D displays have become affordable nowadays, and high-definition streaming is quickly replacing conventional media libraries at home as well as on mobile devices. At the same time, quality considerations have become a lot more intricate because of the additional complexities in content generation, processing, distribution, and display. The challenge of Quality of Experience (QoE) measurement is taking into account not only system performance parameters and content quality metrics, but also notions such as user perception, personality, satisfaction, expectations, and context. With the exponentially growing amount of media being produced, shared, and consumed, as well as emerging new forms of media technologies and services (e.g. light-field imaging, or advanced spatial audio), many challenges remain in developing effective and practical subjective and objective QoE measurement methods.

We solicit original papers describing innovative techniques for measuring the quality of experience (QoE) for multimedia content. Topics of interest include:

- New Technologies: spatial/3D audio quality; stereo/multi-view video quality; high-dynamic range imaging; light-field imaging; holographic imaging; quality in immersive environments (virtual/ augmented/ mixed realities).
- QoE for Mobile Devices: Quality evaluation for mobile devices; adaptive media streaming; impact of viewing conditions, context and device properties, user behavior.
- Interactive Systems QoE: Tele-conferencing/tele-presence; multimedia-based group inter-action; online and cloud gaming; Web and social media applications.
- Big data QoE analytics: Media streaming platforms such as Netflix, YouTube, etc.; crowdsourcing studies; massive open online courses (MOOCs); online gaming; machine learning for QoE.
- QoE Fundamentals: Understanding experience and quality formation; alternatives to mean opinion score (MOS); quality vs. user satisfaction vs. acceptance; long-term quality measurement; physiological QoE assessment; sensory user experiences.
- Reproducible QoE Research: Benchmarking and certification; multimedia quality databases; testing conditions and methods; standardization efforts; open-source QoE tools.

Prospective authors should visit http://www.signalprocessingsociety.org/publications/periodicals/jstsp/ for information on paper submission. Manuscripts should be submitted at http://mc.manuscriptcentral.com/jstsp-ieee. Manuscripts will be peer reviewed according to the standard IEEE process.

Important Dates:

Manuscript submission due:	March 1st, 2016
First review completed:	May 31st, 2016
Revised manuscript due:	July 15th, 2016
Second review completed:	Sept. 1st, 2016
Final manuscript due:	Oct. 15th, 2016
Publication date:	February 2017

Guest Editors:

Stefan Winkler, Advanced Digital Sciences Center (ADSC)/UIUC, Singapore (stefan.winkler@adsc.com.sg)
Chang Wen Chen, State University of New York (SUNY) at Buffalo, USA (chencw@buffalo.edu)
Alexander Raake, Deutsche Telekom Labs, TU Berlin, Germany (Alexander.Raake@telekom.de)
Peter Schelkens, Vrije Universiteit Brussel (VUB) - iMinds, Belgium (Peter.Schelkens@vub.ac.be)
Lea Skorin-Kapov, Faculty of EE and Computing, University of Zagreb, Croatia (Lea.Skorin-Kapov@fer.hr)

CALL FOR PAPERS

IEEE Journal of Selected Topics in Signal Processing

Special Issue on Exploiting Interference towards Energy Efficient and Secure Wireless Communications

Interference has long been the central focus for meeting the ever increasing requirements on quality of service (QoS) in modern and future wireless communication systems. Traditional approaches aim to minimise, cancel or avoid interference. Contrary to this traditional view, which treats interference as a detrimental phenomenon, recent interest has emerged on innovative approaches that consider interference as a useful resource for developing energy efficient and secure 5G communication systems. These include exploiting constructive interference as a source of useful signal power at the modulation level by use of practical multiuser downlink precoding, and also the use of radio frequency radiation for energy harvesting that handles interference and unintended signals as a source of green energy. These techniques open new exciting opportunities in wireless communications by enabling energy self-sustainable and environmentally friendly networks with extended lifetimes, untethered mobility and independence from the power grid, and joint distribution of information and energy within networks. Interference is also being used for physical (PHY) layer secrecy, as an efficient means to jam potential eavesdroppers. This is particularly useful in networks without infrastructure to secure wireless links without the computational overhead imposed by standard cryptographic techniques. These research streams introduce a new vision about interference in wireless networks and motivate a plethora of potential new applications and services. The purpose of this special issue is to re-examine the notion of interference in communications networks and introduce a new paradigm that considers interference as a useful resource in the context of 5G communications.

This special issue seeks to bring together contributions from researchers and practitioners in the area of signal processing for wireless communications with an emphasis on new methods for exploiting interference including symbol level precoding, physical layer security, radiated energy harvesting and wireless power transfer. We solicit high-quality original research papers on topics including, but not limited to:

- Fundamental limits of communication by interference exploitation,
- Modulation level precoding for interference exploitation, interference exploitation in 5G techniques,
- Interference exploitation in the presence of channel state information errors, limited feedback and hardware imperfections,
- Energy harvesting, cooperation and relaying in wireless networks, and in conjunction with 5G methods,
- Time switching, power splitting and antenna switching for simultaneous energy and information transfer,
- Interference exploitation and management in coexisting wireless communications and power transfer systems,
- Joint optimisation of the baseband processing and RF circuit design for energy harvesting,
- Joint interference exploitation and wireless power transfer techniques at the transmitter,
- Security concerns in energy harvesting networks,
- Signal processing for information-theoretic privacy,
- PHY layer secrecy and jamming, PHY secrecy in 5G technologies,
- Introducing artificial and controlled interference for enhancing wireless security,
- Beamforming for PHY-layer secrecy and energy harvesting,
- Interference exploitation and management in coexisting wireless communications and power transfer systems

In addition to technical research results, we invite very high quality submissions of a tutorial or overview nature; we also welcome creative papers outside of the areas listed here but related to the overall scope of the special issue. Prospective authors can contact the Guest Editors to ascertain interest on topics that are not listed.

Prospective authors should visit http://www.signalprocessingsociety.org/publications/periodicals/jstsp/ for information on paper submission. Manuscripts should be submitted using the Manuscript Central system at http://mc.manuscriptcentral.com/jstsp-ieee. Manuscripts will be peer reviewed according to the standard IEEE process.

Manuscript Submission:	January 30, 2016
First review completed:	March 31, 2016
Revised manuscript due:	May 15, 2016
Second review completed:	July 1, 2016
Final manuscript due:	August 15, 2016
Publication date:	December 2016

Guest Editors
Dr. Ioannis Krikidis, University of Cyprus, Cyprus, email: krikidis.ioannis@ucy.ac.cy
Dr. Christos Masouros, University College London, UK, email: c.masouros@ucl.ac.uk
Dr. Gan Zheng, University of Essex, UK, email: ganzheng@essex.ac.uk
Prof. Rui Zhang, National University of Singapore, Singapore, email: elezhang@nus.edu.sg
Prof. Robert Schober, Universität Erlangen-Nürnberg, Germany, email: robert.schober@fau.de

IEEE

SIGNAL PROCESSING LETTERS

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.ieee.org/sp/index.html

DECEMBER 2015 VOLUME 22 NUMBER 12 ISPLEM (ISSN 1070-9908)

LETTERS

Similarity Validation Based Nonlocal Means Image Denoising http://dx.doi.org/10.1109/LSP.2015.2465291 ..... *M. Sharifymoghaddam, S. Beheshti, P. Elahi, and M. Hashemi* 2185
An Efficiency Maximization Design for SWIPT http://dx.doi.org/10.1109/LSP.2015.2464082 ..... *Q.-D. Vu, L.-N. Tran, R. Farrell, and E.-K. Hong* 2189
Lossless Compression of Hyperspectral Imagery via Clustered Differential Pulse Code Modulation with Removal of Local Spectral Outliers http://dx.doi.org/10.1109/LSP.2015.2443913 ..... *J. Wu, W. Kong, J. Mielikainen, and B. Huang* 2194
New Conditions on Achieving the Maximal Possible Dynamic Range for a Generalized Chinese Remainder Theorem of Multiple Integers http://dx.doi.org/10.1109/LSP.2015.2469537 ..... *L. Xiao, X.-G. Xia, and H. Huo* 2199
Double Strobe Technique for Unambiguous Tracking of TMBOC Modulated Signal in GPS http://dx.doi.org/10.1109/LSP.2015.2470240 ..... *Z. Liu, J. Pang, Y. Liu, and F. Wang* 2204
Edge Propagation KD-Trees: Computing Approximate Nearest Neighbor Fields http://dx.doi.org/10.1109/LSP.2015.2469738 ..... *X. Bai, X. Dong, and Y. Su* 2209
Image Matching with Multi-order Features http://dx.doi.org/10.1109/LSP.2015.2469297 ..... *Y. Li, S. Zeng, and Y. Yang* 2214
Anti-Forensics of Lossy Predictive Image Compression http://dx.doi.org/10.1109/LSP.2015.2472561 ..... *Y. Li and J. Zhou* 2219
QoS Constrained Optimization for Multi-Antenna AF Relaying With Multiple Eavesdroppers http://dx.doi.org/10.1109/LSP.2015.2472972 ..... *H. Shen, W. Xu, and C. Zhao* 2224
Von Mises Mixture PHD Filter http://dx.doi.org/10.1109/LSP.2015.2472962 ..... *I. Marković, J. Ćesić, and I. Petrović* 2229
Cooperative Precoding and Artificial Noise Design for Security Over Interference Channels http://dx.doi.org/10.1109/LSP.2015.2472275 ..... *A. Özçelikkale and T. M. Duman* 2234
Generalized Wald Test for Binary Composite Hypothesis Test http://dx.doi.org/10.1109/LSP.2015.2472991 ..... *M. Naderpour, A. Ghobadzadeh, A. Tadaion, and S. Gazor* 2239

IEEE

The Random Walk Model Kalman Filter in Multichannel Active Noise Control http://dx.doi.org/10.1109/LSP.2015.2475357 ..... P. A. C. Lopes, J. A. B. Gerald, and M. S. Piedade 2244
Leader-Follower Consensus in Mobile Sensor Networks http://dx.doi.org/10.1109/LSP.2015.2474134 ..... S. Safavi and U. A. Khan 2249
Single-Photon Depth Imaging Using a Union-of-Subspaces Model http://dx.doi.org/10.1109/LSP.2015.2475274 ..... D. Shin, J. H. Shapiro, and V. K. Goyal 2254
Multiple-Input Multiple-Output OFDM with Index Modulation http://dx.doi.org/10.1109/LSP.2015.2475361 ..... E. Başar 2259
Proximal Mappings Involving Almost Structured Matrices http://dx.doi.org/10.1109/LSP.2015.2476381 ..... İ. Bayram 2264
Fitting Skeletal Object Models Using Spherical Harmonics Based Template Warping http://dx.doi.org/10.1109/LSP.2015.2476366 ..... L. Tu, D. Yang, J. Vicory, X. Zhang, S. M. Pizer, and M. Styner 2269
Iterative Dynamic Power Splitting for Multi-relay Networks with Wireless Energy Harvesting http://dx.doi.org/10.1109/LSP.2015.2477167 ..... C. Zhang and L. Hu 2274
Novel Graph Cuts Method for Multi-Frame Super-Resolution http://dx.doi.org/10.1109/LSP.2015.2477079 ..... D. Zhang, P.-M. Jodoin, C. Li, Y. Wu, and G. Cai 2279
Joint Time Switching and Power Allocation for Multicarrier Decode-and-Forward Relay Networks with SWIPT http://dx.doi.org/10.1109/LSP.2015.2477424 ..... G. Huang, Q. Zhang, and J. Qin 2284
Alpha-Stable Matrix Factorization http://dx.doi.org/10.1109/LSP.2015.2477535 ..... U. Şimşekli, A. Liutkus, and A. T. Cemgil 2289
Low-Overhead Distributed Jamming for SIMO Secrecy Transmission with Statistical CSI http://dx.doi.org/10.1109/LSP.2015.2477883 ..... C. Wang, H.-M. Wang, and B. Wang 2294
Reconstruction of Sparse Wavelet Signals From Partial Fourier Measurements http://dx.doi.org/10.1109/LSP.2015.2478007 ..... Y. Chen, C. Cheng, and Q. Sun 2299
Lightweight Lossy Compression of Biometric Patterns via Denoising Autoencoders http://dx.doi.org/10.1109/LSP.2015.2476667 ..... D. Del Testa and M. Rossi 2304
Adaptive Bayesian Estimation with Cluster Structured Sparsity http://dx.doi.org/10.1109/LSP.2015.2477440 .... L. Yu, C. Wei, and G. Zheng 2309
A New Riemannian Averaged Fixed-Point Algorithm for MGGD Parameter Estimation http://dx.doi.org/10.1109/LSP.2015.2478803 ..... Z. Boukouvalas, S. Said, L. Bombrun, Y. Berthoumieu, and T. Adalı 2314
Block Iterative Reweighted Algorithms for Super-Resolution of Spectrally Sparse Signals http://dx.doi.org/10.1109/LSP.2015.2478854 ..... M. Cho, K. V. Mishra, J.-F. Cai, and W. Xu 2319
Locating Facial Landmarks Using Probabilistic Random Forest http://dx.doi.org/10.1109/LSP.2015.2480758 ..... C. Luo, Z. Wang, S. Wang, J. Zhang, and J. Yu 2324
Sensitivity Analysis of Likelihood Ratio Test in K Distributed and/or Gaussian Noise http://dx.doi.org/10.1109/LSP.2015.2480426 ..... O. Besson and Y. Abramovich 2329
An Optimum Architecture for Continuous-Flow Parallel Bit Reversal http://dx.doi.org/10.1109/LSP.2015.2470519 ..... C. Cheng and F. Yu 2334
DeepPano: Deep Panoramic Representation for 3-D Shape Recognition http://dx.doi.org/10.1109/LSP.2015.2480802 ..... B. Shi, S. Bai, Z. Zhou, and X. Bai 2339
Structural SVM with Partial Ranking for Activity Segmentation and Classification http://dx.doi.org/10.1109/LSP.2015.2480097 ..... G. Zhang and M. Piccardi 2344
Fusion of Gait and Facial Features using Coupled Projections for People Identification at a Distance http://dx.doi.org/10.1109/LSP.2015.2481930 ..... X. Xing, K. Wang, and Z. Lv 2349
A Zero-Crossing Rate Property of Power Complementary Analysis Filterbank Outputs http://dx.doi.org/10.1109/LSP.2015.2481720 ..... R. R. Shenoy and C. S. Seelamantula 2354
3-D Object Recognition via Aspect Graph Aware 3-D Object Representation http://dx.doi.org/10.1109/LSP.2015.2482489 ..... M. Hu, Z. Wei, M. Shao, and G. Zhang 2359
Refined Composite Multiscale Permutation Entropy to Overcome Multiscale Permutation Entropy Length Dependence http://dx.doi.org/10.1109/LSP.2015.2482603 ..... A. Humeau-Heurtier, C.-W. Wu, and S.-D. Wu 2364
Long-Term Stationary Object Detection Based on Spatio-Temporal Change Detection http://dx.doi.org/10.1109/LSP.2015.2482598 ..... D. Ortego, J. C. SanMiguel, and J. M. Martínez 2368
LBP Encoding Schemes Jointly Utilizing the Information of Current Bit and Other LBP Bits http://dx.doi.org/10.1109/LSP.2015.2481435 ..... J. Ren, X. Jiang, and J. Yuan 2373
Outage Region Characterization for Beamforming in MISO Interference Networks with Imperfect CSI http://dx.doi.org/10.1109/LSP.2015.2484382 ..... E. Stathakis, J. Jaldén, L. K. Rasmussen, and M. Skoglund 2378
Recovery of Sparse Positive Signals on the Sphere from Low Resolution Measurements http://dx.doi.org/10.1109/LSP.2015.2485281 ..... T. Bendory and Y. C. Eldar 2383

A Patch-Structure Representation Method for Quality Assessment of Contrast Changed Images http://dx.doi.org/10.1109/LSP.2015.2487369 ..... *S. Wang, K. Ma, H. Yeganeh, Z. Wang, and W. Lin* 2387
Multiple Spectral Peak Tracking for Heart Rate Monitoring from Photoplethysmography Signal During Intensive Physical Exercise http://dx.doi.org/10.1109/LSP.2015.2486681 ..... *N. K. Lakshminarasimha Murthy, P. C. Madhusudana, P. Suresha, V. Periyasamy, and P. K. Ghosh* 2391
Robust Image Hashing Based on Selective Quaternion Invariance http://dx.doi.org/10.1109/LSP.2015.2487824 ..... *Y. N. Li, P. Wang, and Y. T. Su* 2396
Robust Model-Free Multi-Object Tracking with Online Kernelized Structural Learning http://dx.doi.org/10.1109/LSP.2015.2488678 ..... *R. Yao* 2401
An Efficient SVD Shrinkage for Rank Estimation http://dx.doi.org/10.1109/LSP.2015.2487600 ..... *S. K. Yadav, R. Sinha, and P. K. Bora* 2406
Bilinear Embedding Label Propagation: Towards Scalable Prediction of Image Labels http://dx.doi.org/10.1109/LSP.2015.2488632 ..... *Y. Liang, Z. Zhang, W. Jiang, M. Zhao, and F. Li* 2411
Translation on Graphs: An Isometric Shift Operator http://dx.doi.org/10.1109/LSP.2015.2488279 .... *B. Girault, P. Gonçalves, and É. Fleury* 2416
Probabilistic Class Histogram Equalization Based on Posterior Mean Estimation for Robust Speech Recognition http://dx.doi.org/10.1109/LSP.2015.2490202 ..... *Y. Suh and H. Kim* 2421
A Novel Sampling Theorem on the Rotation Group http://dx.doi.org/10.1109/LSP.2015.2490676 ..... *J. D. McEwen, M. Büttner, B. Leistedt, H. V. Peiris, and Y. Wiaux* 2425
Joint Channel Estimation for Three-Hop MIMO Relaying Systems http://dx.doi.org/10.1109/LSP.2015.2490251 ..... *I. V. Cavalcante, A. L. F. de Almeida, and M. Haardt* 2430
Cooperative Secrecy Beamforming in Wiretap Interference Channels http://dx.doi.org/10.1109/LSP.2015.2490602 ..... *L. Li, C. Huang, and Z. Chen* 2435
Clock Skew Estimation of Listening Nodes with Clock Correction upon Every Synchronization in Wireless Sensor Networks http://dx.doi.org/10.1109/LSP.2015.2491320 ..... *H. Wang, H. Zeng, and P. Wang* 2440
Target Localization from Bistatic Range Measurements in Multi-Transmitter Multi-Receiver Passive Radar http://dx.doi.org/10.1109/LSP.2015.2491961 ..... *A. Noroozi and M. A. Sebt* 2445
Approximate Bayesian Smoothing with Unknown Process and Measurement Noise Covariances http://dx.doi.org/10.1109/LSP.2015.2490543 ..... *T. Ardeshiri, E. Özkan, U. Orguner, and F. Gustafsson* 2450
Face Recognition Using Dual Difference Regression Classification http://dx.doi.org/10.1109/LSP.2015.2492980 .... *N. Piao and R.-H. Park* 2455
A Genetic Algorithm-Based Feature Selection for Kinship Verification http://dx.doi.org/10.1109/LSP.2015.2490805 ..... *P. Alirezazadeh, A. Fathi, and F. Abdali-Mohammadi* 2459
Robust Multi-Dimensional Harmonic Retrieval Using Iteratively Reweighted HOSVD http://dx.doi.org/10.1109/LSP.2015.2493521 ..... *F. Wen and H. C. So* 2464
Cost-Constrained Feature Optimization in Kernel Machine Classifiers http://dx.doi.org/10.1109/LSP.2015.2492238 ..... *C. R. Ratto, C. A. Caceres, and H. C. Schoeberlein* 2469
Detection Probability of a CFAR Matched Filter with Signal Steering Vector Errors http://dx.doi.org/10.1109/LSP.2015.2494013 ..... *J. Liu, W. Liu, B. Chen, H. Liu, and H. Li* 2474
Noncontact Vital Sign Detection based on Stepwise Atomic Norm Minimization http://dx.doi.org/10.1109/LSP.2015.2494604 ..... *L. Sun, H. Hong, Y. Li, C. Gu, F. Xi, C. Li, and X. Zhu* 2479
Quasi-Factorial Prior for i-vector Extraction http://dx.doi.org/10.1109/LSP.2015.2459059 ..... *L. Chen, K. A. Lee, L. R. Dai, and H. Li* 2484
A Fine-Resolution Frequency Estimator in the Odd-DFT Domain http://dx.doi.org/10.1109/LSP.2015.2496276 ..... *Y. Dun and G. Liu* 2489
Distributed Detection Over Channels with Memory http://dx.doi.org/10.1109/LSP.2015.2495205 ..... *N. Biswas, P. Ray, and P. K. Varshney* 2494

List of Reviewers http://dx.doi.org/10.1109/LSP.2015.2496824 2499

Information for Authors http://dx.doi.org/10.1109/LSP.2015.2498419 2510
EDICS—Editors' Information Classification Scheme http://dx.doi.org/10.1109/LSP.2015.2498418 2512

2015 Index http://dx.doi.org/10.1109/LSP.2015.2499818 2513

IEEE
SignalProcessing
MAGAZINE
[VOLUME 33 NUMBER 1 JANUARY 2016]
TAKING SPS TO NEW HEIGHTS
BAYESIAN MACHINE LEARNING FOR EEG/MEG
DECISION LEARNING
BLOCK-STRUCTURED OPTIMIZATION FOR BIG DATA
COMPRESSIVE COVARIANCE SENSING
GAME THEORY FOR NETWORKS
COMBINATIONS OF ADAPTIVE FILTER
IEEE Signal Processing Society
IEEE

CONTENTS

[VOLUME 33 NUMBER 1]

FEATURES

LEARNING

14 **BAYESIAN MACHINE LEARNING**
Wei Wu, Srikantan Nagarajan, and Zhe Chen

37 **DECISION LEARNING**
Yan Chen, Chunxiao Jiang, Chih-Yu Wang, Yang Gao, and K.J. Ray Liu

BIG DATA

57 **A UNIFIED ALGORITHMIC FRAMEWORK FOR BLOCK-STRUCTURED OPTIMIZATION INVOLVING BIG DATA**
Mingyi Hong, Meisam Razaviyayn, Zhi-Quan Luo, and Jong-Shi Pang

THEORY AND METHODS

78 **COMPRESSIVE COVARIANCE SENSING**
Daniel Romero, Dyonisius Dony Ariananda, Zhi Tian, and Geert Leus

94 **GAME THEORY FOR NETWORKS**
Giacomo Bacci, Samson Lasaulce, Walid Saad, and Luca Sanguinetti

120 **COMBINATIONS OF ADAPTIVE FILTERS**
Jerónimo Arenas-García, Luis A. Azpicueta-Ruiz, Magno T.M. Silva, Vítor H. Nascimento, and Ali H. Sayed

Digital Object Identifier 10.1109/MSP.2015.2479476

COLUMNS

4 **FROM THE EDITOR**
Women in Science, Engineering, and Signal Processing
Min Wu

6 **PRESIDENT'S MESSAGE**
We Need Your Help to Take the Society to New Heights
Alex Acero

11 **SPECIAL REPORTS**
Signal Processing Research Resonates with Hearing Loss Sufferers
John Edwards

141 **REFLECTIONS**
James L. Flanagan: A Scholar and a True Gentleman
Lawrence Rabiner

144 **LECTURE NOTES**
Polyphase Channelizer Demystified
P. Murali Krishna and T.P. Sameer Babu

151 **SP TIPS&TRICKS**
Cascading Tricks for Designing Composite Filters with Sharp Transition Bands
David Shiung, Ya-Yin Yang, and Chu-Sing Yang

158 **STANDARDS IN A NUTSHELL**
Next-Generation Broadcast Television: ATSC 3.0
Richard Chernock and Jerry C. Whitaker

163 **BOOK REVIEWS**
Foundations of Signal Processing
Andres Kwasinski

Handbook of Digital Forensics of Multimedia Data and Devices
Luisa Verdoliva

166 **BOOK DIGEST**

DEPARTMENTS

7 **SOCIETY NEWS**
2016 Class of Distinguished Lecturers

IEEE Signal Processing Society Young Professionals Meet During ICIP 2015

168 **DATES AHEAD**

ICIP 2016 2016 IEEE International Conference on Image Processing

September 25-28, 2016 • Phoenix, Arizona, USA

New for ICIP 2016:

- maximize the visibility of your work via open preview: Papers accepted to ICIP 2016 will (upon author approval) be available on IEEE Xplore, **freely accessible and downloadable by all** in their final format from August 20, 2016 through September 30, 2016.
- nominate an individual or team for the Visual Innovation Award by March 31, 2016: The Visual Innovation Award was created to recognize pioneers of transformative technologies and business models in areas within the technical scope of the IEEE ICIP conference series.
- maximize the visibility of your work through reproducible research: ICIP 2016 supports reproducible research by allowing authors to submit supplementary material, including code and data.
- maximize your networking and career connections: attendees will be given the opportunity to upload their CVs to be shared among interested recruiters for full-time, part-time, and consulting job opportunities.
- experience state-of-the-art visual technology products and prototypes at the ICIP 2016 Visual Technology Showcase.

For more details on this and other new initiatives at ICIP 2016, visit **2016.ieeeicip.org** and connect on social media.

Important Deadlines:

Challenge Session Proposals: October 31, 2015 Special Session/Tutorial Proposals: November 16, 2015
Paper Submissions: January 25, 2016 Visual Innovation Award Nomination: March 31, 2016
Visual Technology Showcase Submission: May 15, 2016

Call for Papers

http://ssp2016.tsc.uc3m.es

2016 IEEE Statistical Signal Processing Workshop

26-29 June 2016, Palma de Mallorca, Spain

The 2016 IEEE Workshop on Statistical Signal Processing (SSP 2016) is the 19th of a series of unique meetings that bring members of the IEEE Signal Processing Society together with researchers from allied fields such as bioinformatics, communications, machine learning, and statistics.

The scientific program of SSP 2016 will include invited plenary talks, as well as regular and special sessions with contributed research papers. All submitted papers will be reviewed by experts and only a proportion will be accepted to maintain a high quality workshop. All accepted papers will be published on IEEE Xplore. The scope of the workshop includes basic theory, methods and algorithms, and applications in the following areas:

Theoretical Topics

- Adaptive systems and signal processing
- Detection and estimation theory
- Learning theory and pattern recognition
- Multivariate statistical analysis
- System identification and calibration
- Monte Carlo methods
- Network and graph analysis
- Random matrix theory
- Time-frequency and time-scale analysis
- Compressed sensing
- Point process estimation
- Stochastic filtering

Application Areas

- Bioinformatics and genomics
- Array processing, radar and sonar
- Communication systems and networks
- Sensor networks
- Information forensics and security
- Medical imaging
- Biomedical signal processing
- Preventive, social network analysis
- Smart grids and industrial applications
- Geoscience
- Astrophysics
- New methods, directions and applications

IEEE Signal Processing Society IEEE

Venue

Es Baluard Museu d'Art Modern i Contemporani, Palma de Mallorca, Spain

Paper Submission

Prospective authors are invited to submit full-length papers, with up to four pages for technical content including figures and references, using the templates and formatting guidelines posted on the website. All accepted papers must be presented at the workshop in order to be published in the proceedings. Best student paper awards, selected by a SSP committee, will be presented at the workshop.

Special Sessions

In addition to regular sessions, the workshop will also have a number of special sessions on topics of particular relevance. Prospective organizers of special sessions are invited to submit a proposal form, available on the workshop website, by e-mail to the Special Sessions Chair.

Important Dates

Submission of proposals for special sessions	**Nov 09, 2015**
Notification of acceptance of special sessions	**Nov 30, 2015**
Full paper submission deadline	**Feb 08, 2016**
Notificacion of acceptance	**April 04, 2016**
Camera ready papers due on	**April 18, 2016**

Organization

General Chairs:

Antonio Artés-Rodríguez (Universidad Carlos III de Madrid, Spain)

Joaquín Miguez (Universidad Carlos III de Madrid, Spain)

Technical Program Chairs:

Sergios Theodoridis (National and Kapodistrian University of Athens, Greece)

Konstantinos Slavakis (University of Minnesota, USA)

Finance Chair:

Matilde Sánchez-Fernández (Universidad Carlos III de Madrid, Spain)

Special Sessions Chair:

Mónica F. Bugallo (Stony Brook University, USA)

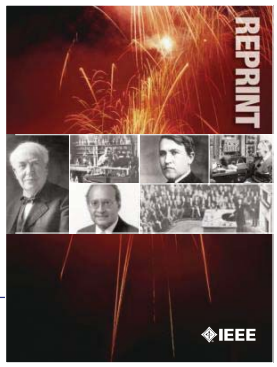
Local Arrangements Chairs:

Guillem Femenias (Universitat de les Illes Balears, Spain)

Felip Riera-Palou (Universitat de les Illes Balears, Spain)

Publications Chair

Pau Closas (CTTC, Spain)



IEEE ORDER FORM FOR REPRINTS

Purchasing IEEE Papers in Print is easy, cost-effective and quick.

Complete this form, send via our secure fax (24 hours a day) to 732-981-8062 or mail it back to us.

PLEASE FILL OUT THE FOLLOWING

Author: _____

Publication Title: _____

Paper Title: _____

_____

RETURN THIS FORM TO:
IEEE Publishing Services
445 Hoes Lane
Piscataway, NJ 08855-1331

Email the Reprint Department at reprints@ieee.org for questions regarding this form

PLEASE SEND ME

■ 50 ■ 100 ■ 200 ■ 300 ■ 400 ■ 500 or _____(in multiples of 50) reprints.

■ YES ■ NO Self-covering/title page required. COVER PRICE: $74 per 100, $39 per 50.

■ $58.00 Air Freight must be added for all orders being shipped outside the U.S.

■ $21.50 must be added for all USA shipments to cover the cost of UPS shipping and handling.

PAYMENT

■ Check enclosed. Payable on a bank in the USA.

■ Charge my: ■ Visa ■ Mastercard ■ Amex ■ Diners Club

Account # _____ Exp. date _____

Cardholder's Name (please print): _____

_____

■ Bill me (you must attach a purchase order) Purchase Order Number _____

Send Reprints to:

_____

_____

_____

_____

Bill to address, if different:

_____

_____

_____

_____

Because information and papers are gathered from various sources, there may be a delay in receiving your reprint request. This is especially true with postconference publications. Please provide us with contact information if you would like notification of a delay of more than 12 weeks.

Telephone: _____ Fax: _____ Email Address: _____

2012 REPRINT PRICES (without covers)

Number of Text Pages

	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40	41-44	45-48
50	$129	$213	$245	$248	$288	$340	$371	$408	$440	$477	$510	$543
100	$245	$425	$479	$495	$573	$680	$742	$817	$885	$953	$1021	$1088

Larger quantities can be ordered. Email reprints@ieee.org with specific details.

Tax Applies on shipments of regular reprints to CA, DC, FL, MI, NJ, NY, OH and Canada (GST Registration no. 12534188).
Prices are based on black & white printing. Please call us for full color price quote, if applicable.

Authorized Signature: _____ Date: _____

2016 IEEE MEMBERSHIP APPLICATION

(students and graduate students must apply online)

IEEE
Advancing Technology for Humanity

Start your membership immediately: Join online www.ieee.org/join

Please complete both sides of this form, typing or **printing in capital letters**. Use only English characters and abbreviate only if more than 40 characters and spaces per line. We regret that incomplete applications cannot be processed.

1 Name & Contact Information

Please PRINT your name as you want it to appear on your membership card and IEEE correspondence. As a key identifier for the IEEE database, circle your last/surname.

☐ Male ☐ Female Date of birth (Day/Month/Year) ____/____/____

Title First/Given Name Middle Last/Family Surname

▼ Primary Address ☐ Home ☐ Business (All IEEE mail sent here)

Street Address

City State/Province

Postal Code Country

Primary Phone

Primary E-mail

▼ Secondary Address ☐ Home ☐ Business

Company Name Department/Division

Street Address City State/Province

Postal Code Country

Secondary Phone

Secondary E-mail

To better serve our members and supplement member dues, your postal mailing address is made available to carefully selected organizations to provide you with information on technical services, continuing education, and conferences. Your e-mail address is not rented by IEEE. Please check box only if you do not want to receive these postal mailings to the selected address. ☐

2 Attestation

I have graduated from a three- to five-year academic program with a university-level degree.
☐ Yes ☐ No

This program is in one of the following fields of study:
☐ Engineering
☐ Computer Sciences and Information Technologies
☐ Physical Sciences
☐ Biological and Medical Sciences
☐ Mathematics
☐ Technical Communications, Education, Management, Law and Policy
☐ Other (please specify): _____

This academic institution or program is accredited in the country where the institution is located. ☐ Yes ☐ No ☐ Do not know

I have _____ years of professional experience in teaching, creating, developing, practicing, or managing within the following field:

☐ Engineering
☐ Computer Sciences and Information Technologies
☐ Physical Sciences
☐ Biological and Medical Sciences
☐ Mathematics
☐ Technical Communications, Education, Management, Law and Policy
☐ Other (please specify): _____

3 Please Tell Us About Yourself

Select the numbered option that best describes yourself. This information is used by IEEE magazines to verify their annual circulation. Please enter numbered selections in the boxes provided.

A. Primary line of business → ☐

1. Computers
2. Computer peripheral equipment
3. Software
4. Office and business machines
5. Test, measurement and instrumentation equipment
6. Communications systems and equipment
7. Navigation and guidance systems and equipment
8. Consumer electronics/appliances
9. Industrial equipment, controls and systems
10. ICs and microprocessors
11. Semiconductors, components, sub-assemblies, materials and supplies
12. Aircraft, missiles, space and ground support equipment
13. Oceanography and support equipment
14. Medical electronic equipment
15. OEM incorporating electronics in their end product (not elsewhere classified)
16. Independent and university research, test and design laboratories and consultants (not connected with a mfg. co.)
17. Government agencies and armed forces
18. Companies using and/or incorporating any electronic products in their manufacturing, processing, research or development activities
19. Telecommunications services, telephone (including cellular)
20. Broadcast services (TV, cable, radio)
21. Transportation services (airline, railroad, etc.)
22. Computer and communications and data processing services
23. Power production, generation, transmission and distribution
24. Other commercial users of electrical, electronic equipment and services (not elsewhere classified)
25. Distributor (reseller, wholesaler, retailer)
26. University, college/other educational institutions, libraries
27. Retired
28. Other_____

B. Principal job function → ☐

1. General and corporate management
2. Engineering management
3. Project engineering management
4. Research and development management
5. Design engineering management—analog
6. Design engineering management—digital
7. Research and development engineering
8. Design/development engineering—analog
9. Design/development engineering—digital
10. Hardware engineering
11. Software design/development
12. Computer science
13. Science/physics/mathematics
14. Engineering (not elsewhere specified)
15. Marketing/sales/purchasing
16. Consulting
17. Education/teaching
18. Retired
19. Other _____

C. Principal responsibility → ☐

1. Engineering and scientific management
2. Management other than engineering
3. Engineering design
4. Engineering
5. Software: science/mngmnt/engineering
6. Education/teaching
7. Consulting
8. Retired
9. Other _____

D. Title → ☐

1. Chairman of the Board/President/CEO
2. Owner/Partner
3. General Manager
4. VP Operations
5. VP Engineering/Dir. Engineering
6. Chief Engineer/Chief Scientist
7. Engineering Management
8. Scientific Management
9. Member of Technical Staff
10. Design Engineering Manager
11. Design Engineer
12. Hardware Engineer
13. Software Engineer
14. Computer Scientist
15. Dean/Professor/Instructor
16. Consultant
17. Retired
18. Other _____

Are you now or were you ever a member of IEEE?
☐ Yes ☐ No If yes, provide, if known:

Membership Number Grade Year Expired

4 Please Sign Your Application

I hereby apply for IEEE membership and agree to be governed by the IEEE Constitution, Bylaws, and Code of Ethics. I understand that IEEE will communicate with me regarding my individual membership and all related benefits. **Application must be signed.**

Signature Date

Over Please

5 Add IEEE Society Memberships (Optional)

The 39 IEEE Societies support your technical and professional interests. Many society memberships include a personal subscription to the core journal, magazine, or newsletter of that society. **For a complete list of everything included with your IEEE Society membership, visit www.ieee.org/join.** All prices are quoted in US dollars.

Please check ✔ the appropriate box.		BETWEEN 16 AUG 2015-28 FEB 2016 PAY	BETWEEN 1 MAR 2016-15 AUG 2016 PAY
IEEE Aerospace and Electronic Systems ⌘ 🖂	AES010	25.00 ☐	12.50 ☐
IEEE Antennas and Propagation ⌘ 🖂	AP003	15.00 ☐	7.50 ☐
IEEE Broadcast Technology ▪ ⌘ 🖂	BT002	15.00 ☐	7.50 ☐
IEEE Circuits and Systems ▪ ⌘ 🖂	CAS004	22.00 ☐	11.00 ☐
IEEE Communications ⌘ 🖂	COM019	30.00 ☐	15.00 ☐
IEEE Components, Packaging, & Manu. Tech. ▪ ⌘ 🖂	CPMT021	15.00 ☐	7.50 ☐
IEEE Computational Intelligence ⌘ 🖂	CIS011	29.00 ☐	14.50 ☐
IEEE Computer ▪ ⌘ 🖂	C016	56.00 ☐	28.00 ☐
IEEE Consumer Electronics ⌘ 🖂	CE008	20.00 ☐	10.00 ☐
IEEE Control Systems ⌘ 🖂	CS023	25.00 ☐	12.50 ☐
IEEE Dielectrics and Electrical Insulation ⌘ 🖂	DEI032	26.00 ☐	13.00 ☐
IEEE Education ⌘ 🖂	E025	20.00 ☐	10.00 ☐
IEEE Electromagnetic Compatibility ▪ ⌘ 🖂	EMC027	31.00 ☐	15.50 ☐
IEEE Electron Devices ▪ ⌘ 🖂	ED015	18.00 ☐	9.00 ☐
IEEE Engineering in Medicine and Biology ⌘ 🖂	EMB018	40.00 ☐	20.00 ☐
IEEE Geoscience and Remote Sensing ⌘ 🖂	GRS029	19.00 ☐	9.50 ☐
IEEE Industrial Electronics ⌘ 🖂	IE013	9.00 ☐	4.50 ☐
IEEE Industry Applications ⌘ 🖂	IA034	20.00 ☐	10.00 ☐
IEEE Information Theory ▪ ⌘ 🖂	IT012	30.00 ☐	15.00 ☐
IEEE Instrumentation and Measurement ⌘ 🖂	IM009	29.00 ☐	14.50 ☐
IEEE Intelligent Transportation Systems ▪ ⌘ 🖂	ITSS038	35.00 ☐	17.50 ☐
IEEE Magnetics ▪ 🖂 ♦⌘	MAG033	26.00 ☐	13.00 ☐
IEEE Microwave Theory and Techniques ⌘ 🖂	MTT017	17.00 ☐	8.50 ☐
IEEE Nuclear and Plasma Sciences ▪ ⌘ 🖂	NPS005	35.00 ☐	17.50 ☐
IEEE Oceanic Engineering ▪ ⌘ 🖂	OE022	19.00 ☐	9.50 ☐
IEEE Photonics ▪ ⌘ 🖂	PHO036	34.00 ☐	17.00 ☐
IEEE Power Electronics ⌘ 🖂	PEL035	25.00 ☐	12.50 ☐
IEEE Power & Energy ▪ ⌘ 🖂	PE031	35.00 ☐	17.50 ☐
IEEE Product Safety Engineering ▪	PSE043	35.00 ☐	17.50 ☐
IEEE Professional Communication ▪ ⌘ 🖂	PC026	31.00 ☐	15.50 ☐
IEEE Reliability ▪ ⌘ 🖂	RL007	35.00 ☐	17.50 ☐
IEEE Robotics and Automation ⌘ 🖂	RA024	9.00 ☐	4.50 ☐
IEEE Signal Processing ▪ ⌘ 🖂	SP001	22.00 ☐	11.00 ☐
IEEE Social Implications of Technology ▪ ⌘ 🖂	SIT030	33.00 ☐	16.50 ☐
IEEE Solid-State Circuits ⌘ 🖂	SSC037	22.00 ☐	11.00 ☐
IEEE Systems, Man, & Cybernetics ⌘ 🖂	SMC028	12.00 ☐	6.00 ☐
IEEE Technology & Engineering Management ⌘ 🖂	TEM014	35.00 ☐	17.50 ☐
IEEE Ultrasonics, Ferroelectrics, & Frequency Control ▪ ⌘ 🖂	UFFC020	20.00 ☐	10.00 ☐
IEEE Vehicular Technology ▪ ⌘ 🖂	VT006	18.00 ☐	9.00 ☐

Legend—Society membership includes:

⌘ One or more Society publications
▪ Society newsletter
🖂 Online access to publication
♦ CD-ROM of selected society publications

Complete both sides of this form, sign, and return to:
IEEE MEMBERSHIP APPLICATION PROCESSING
445 HOES LN, PISCATAWAY, NJ 08854-4141 USA
or fax to +1 732 981 0225
or join online at www.ieee.org/join

Please reprint your full name here

6 2016 IEEE Membership Rates (student rates available online)

IEEE member dues and regional assessments are based on where you live and when you apply. Membership is based on the calendar year from 1 January through 31 December. All prices are quoted in US dollars.

Please check ✔ the appropriate box. RESIDENCE	BETWEEN 16 AUG 2015-28 FEB 2016 PAY	BETWEEN 1 MAR 2016-15 AUG 2016 PAY
United States	$197.00 ☐	$98.50 ☐
Canada (GST)*	$173.35 ☐	$86.68 ☐
Canada (NB, NF and ON HST)*	$185.11 ☐	$92.56 ☐
Canada (Nova Scotia HST)*	$188.05 ☐	$94.03 ☐
Canada (PEI HST)*	$186.58 ☐	$93.29 ☐
Canada (GST and QST Quebec)	$188.01 ☐	$94.01 ☐
Africa, Europe, Middle East	$160.00 ☐	$80.00 ☐
Latin America	$151.00 ☐	$75.50 ☐
Asia, Pacific	$152.00 ☐	$76.00 ☐

*IEEE Canada Business No. 125634188

Minimum Income or Unemployed Provision

Applicants who certify that their prior year income did not exceed US$14,700 (or equivalent) or were not employed are granted 50% reduction in: full-year dues, regional assessment and fees for one IEEE Membership plus one Society Membership. If applicable, please check appropriate box and adjust payment accordingly. Student members are not eligible.

☐ I certify I earned less than US$14,700 in 2015
☐ I certify that I was unemployed in 2015

7 More Recommended Options

Proceedings of the IEEE ..... print $47.00 ☐ or online $41.00 ☐
Proceedings of the IEEE (print/online combination) ..... $57.00 ☐
IEEE Standards Association (IEEE-SA) ..... $53.00 ☐
IEEE Women in Engineering (WIE) ..... $25.00 ☐

8 Payment Amount

Please total the Membership dues, Society dues, and other amounts from this page:

IEEE Membership dues ⑥ ..... $_____
IEEE Society dues (optional) ⑤ ..... $_____
IEEE-SA/WIE dues (optional) ⑦ ..... $_____
Proceedings of the IEEE (optional) ⑦ ..... $_____
Canadian residents pay 5% GST or appropriate HST (BC—12%; NB, NF, ON-13%;NS-15%) on Society payments & publications only ..... TAX $_____
AMOUNT PAID ..... **TOTAL** $_____

Payment Method

All prices are quoted in US dollars. You may pay for IEEE membership by credit card (see below), check, or money order payable to IEEE, drawn on a US bank.

☐ Check ☐ VISA ☐ MasterCard ☐ Diners Club International ☐ AMERICAN EXPRESS ☐ DISCOVER

Credit Card Number

MONTH / YEAR EXPIRATION DATE

CARDHOLDER'S 5-DIGIT ZIPCODE (BILLING STATEMENT ADDRESS) USA ONLY

Name as it appears on card

Signature

Auto Renew my Memberships and Subscriptions (available when paying by credit card).
☐ I agree to the Terms and Conditions located at www.ieee.org/autorenew

9 Were You Referred to IEEE?

☐ Yes ☐ No If yes, provide the following:
Member Recruiter Name _____
IEEE Recruiter's Member Number (Required) _____

CAMPAIGN CODE	PROMO CODE

15-MEM-385 P 6/15

Information for Authors

(Updated/Effective January 2015)

For Transactions and Journals:

Authors are encouraged to submit manuscripts of Regular papers (papers which provide a complete disclosure of a technical premise), or Comment Correspondences (brief items that provide comment on a paper previously published in these TRANSACTIONS).

Submissions/resubmissions must be previously unpublished and may not be under consideration elsewhere.

Every manuscript must:

i. provide a clear statement of the problem and what the contribution of the work is to the relevant research community;
ii. state why this contribution is significant (what impact it will have);
iii. provide citation of the published literature most closely related to the manuscript; and
iv. state what is distinctive and new about the current manuscript relative to these previously published works.

By submission of your manuscript to these TRANSACTIONS, all listed authors have agreed to the authorship list and all the contents and confirm that the work is original and that figures, tables and other reported results accurately reflect the experimental work. In addition, the authors all acknowledge that they accept the rules established for publication of manuscripts, including agreement to pay all overlength page charges, color charges, and any other charges and fees associated with publication of the manuscript. Such charges are not negotiable and cannot be suspended. The corresponding author is responsible for obtaining consent from all co-authors and, if needed, from sponsors before submission.

In order to be considered for review, a paper must be within the scope of the journal and represent a novel contribution. A paper is a candidate for an Immediate Rejection if it is of limited novelty, e.g. a straightforward combination of theories and algorithms that are well established and are repeated on a known scenario. Experimental contributions will be rejected without review if there is insufficient experimental data. These TRANSACTIONS are published in English. Papers that have a large number of typographical and/or grammatical errors will also be rejected without review.

In addition to presenting a novel contribution, acceptable manuscripts must describe and cite related work in the field to put the contribution in context. Do not give theoretical derivations or algorithm descriptions that are easily found in the literature; merely cite the reference.

New and revised manuscripts should be prepared following the "Manuscript Submission" guidelines below, and submitted to the online manuscript system, ScholarOne Manuscripts. Do not send original submissions or revisions directly to the Editor-in-Chief or Associate Editors; they will access your manuscript electronically via the ScholarOne Manuscript system.

Manuscript Submission. Please follow the next steps.

1. *Account in ScholarOne Manuscripts.* If necessary, create an account in the on-line submission system ScholarOne Manuscripts. Please check first if you already have an existing account which is based on your e-mail address and may have been created for you when you reviewed or authored a previous paper.
2. *Electronic Manuscript.* Prepare a PDF file containing your manuscript in double-column, single-spaced format using a font size of 10 points or larger, having a margin of at least 1 inch on all sides. Upload this version of the manuscript as a PDF file "double.pdf" to the ScholarOne-Manuscripts site. Since many reviewers prefer a larger font, you are strongly encouraged to also submit a single-column, double-spaced version (11 point font or larger), which is easy to create with the templates provided **IEEE Author Digital Toolbox (http://www.ieee.org/publications_standards/publications/authors/authors_journals.html)**. Page length restrictions will be determined by the double-column version. Proofread your submission, confirming that all figures and equations are visible in your document before you "SUBMIT" your manuscript. Proofreading is critical; once you submit your manuscript, the manuscript cannot be changed in any way. You may also submit your manuscript as a .PDF or MS Word file. The system has the capability of converting your files to PDF, however it is your responsibility to confirm that the conversion is correct and there are no font or graphics issues prior to completing the submission process.
3. *EDICS (Not applicable to Journal of Selected Topics in Signal Processing).* All submissions must be classified by the author with an EDICS (Editors' Information Classification Scheme) selected from the list of EDICS published online at the at the publication's EDICS webpage (*please see the list below). Upon submission of a new manuscript, please choose the EDICS categories that best suit your manuscript. Failure to do so will likely result in a delay of the peer review process.
4. *Additional Documents for Review.* Please upload pdf versions of all items in the reference list that are not publicly available, such as unpublished (submitted) papers. Graphical abstracts and supplemental materials intended to appear with the final paper (see below) must also be uploaded for review at the time of the initial submission for consideration in the review process. Use short filenames without spaces or special characters. When the upload of each file is completed, you will be asked to provide a description of that file.
5. *Supplemental Materials.* IEEE Xplore can publish multimedia files (audio, images, video), datasets, and software (e.g. Matlab code) along with your paper. Alternatively, you can provide the links to such files in a README file that appears on Xplore along with your paper. For details, please see IEEE Author Digital Toolbox under "Multimedia." To make your work reproducible by others, these TRANSACTIONS encourages you to submit all files that can recreate the figures in your paper.
6. *Submission.* After uploading all files and proofreading them, submit your manuscript by clicking "Submit." A confirmation of the successful submission will open on screen containing the manuscript tracking number and will be followed with an e-mail confirmation to the corresponding and all contributing authors. Once you click "Submit," your manuscript cannot be changed in any way.
7. *Copyright Form and Consent Form.* By policy, IEEE owns the copyright to the technical contributions it publishes on behalf of the interests of the IEEE, its authors, and their employers; and to facilitate the appropriate reuse of this material by others. To comply with the IEEE copyright policies, authors are required to sign and submit a completed "IEEE Copyright and Consent Form" prior to publication by the IEEE. The IEEE recommends authors to use an effective electronic copyright form (eCF) tool within the ScholarOne Manuscripts system. You will be redirected to the "IEEE Electronic Copyright Form" wizard at the end of your original submission; please simply sign the eCF by typing your name at the proper location and click on the "Submit" button.

Comment Correspondence. Comment Correspondences provide brief comments on material previously published in these TRANSACTIONS. These items may not exceed 2 pages in double-column, single spaced format, using 9 point type, with margins of 1 inch minimum on all sides, and including: title, names and contact information for authors, abstract, text, references, and an appropriate number of illustrations and/or tables. Correspondence items are submitted in the same way as regular manuscripts (see "Manuscript Submission" above for instructions).

Authors may also submit manuscripts of overview articles, but note that these include an additional white paper approval process http://www.signalprocessingsociety.org/publications/overview-articles/. *[This does not apply to the Journal of Selected Topics in Signal Processing. Please contact the Editor-in-Chief.]*

Digital Object Identifier

Manuscript Length. For the initial submission of a regular paper, the manuscript may not exceed 13 double-column pages (10 point font), including title; names of authors and their complete contact information; abstract; text; all images, figures and tables, appendices and proofs; and all references. Supplemental materials and graphical abstracts are not included in the page count. For regular papers, the revised manuscript may not exceed 16 double-column pages (10 point font), including title; names of authors and their complete contact information; abstract; text; all images, figures and tables, appendices and proofs; and all references. For Overview Papers, the maximum length is double that for regular submissions at each stage (please reference http://www.signalprocessingsociety.org/publications/overview-articles/ for more information).

Note that any paper in excess of 10 pages will be subject to mandatory overlength page charges. Since changes recommended as a result of peer review may require additions to the manuscript, it is strongly recommended that you practice economy in preparing original submissions. Note: Papers submitted to the TRANSACTIONS ON MULTIMEDIA in excess of 8 pages will be subject to mandatory overlength page charges.

Exceptions to manuscript length requirements may, under extraordinary circumstances, be granted by the Editor-in-Chief. However, such exception does not obviate your requirement to pay any and all overlength or additional charges that attach to the manuscript.

Resubmission of Previously Rejected Manuscripts. Authors of manuscripts rejected from any journal are allowed to resubmit their manuscripts only once. At the time of submission, you will be asked whether your manuscript is a new submission or a resubmission of an earlier rejected manuscript. If it is a resubmission of a manuscript previously rejected by any journal, you are expected to submit supporting documents identifying the previous submission and detailing how your new version addresses all of the reviewers' comments. Papers that do not disclose connection to a previously rejected paper or that do not provide documentation as to changes made may be immediately rejected.

Author Misconduct. Author misconduct includes plagiarism, self-plagiarism, and research misconduct, including falsification or misrepresentation of results. All forms of misconduct are unacceptable and may result in sanctions and/or other corrective actions. Plagiarism includes copying someone else's work without appropriate credit, using someone else's work without clear delineation of citation, and the uncited reuse of an author's previously published work that also involves other authors. Self-plagiarism involves the verbatim copying or reuse of an authors own prior work without appropriate citation, including duplicate submission of a single journal manuscript to two different journals, and submission of two different journal manuscripts which overlap substantially in language or technical contribution. For more information on the definitions, investigation process, and corrective actions related to author misconduct, see the Signal Processing Society Policies and Procedures Manual, Section 6.1. http://www.signalprocessingsociety.org/about-sps/governance/policy-procedure/part-2. Author misconduct may also be actionable by the IEEE under the rules of Member Conduct.

Extensions of the Author's Prior Work. It is acceptable for conference papers to be used as the basis for a more fully developed journal submission. Still, authors are required to cite their related prior work; the papers cannot be identical; and the journal publication must include substantively novel aspects such as new experimental results and analysis or added theoretical work. The journal publication should clearly specify how the journal paper offers novel contributions when citing the prior work. Limited overlap with prior journal publications with a common author is allowed only if it is necessary for the readability of the paper, and the prior work must be cited as the primary source.

Submission Format. Authors are required to prepare manuscripts employing the on-line style files developed by IEEE, which include guidelines for abbreviations, mathematics, and graphics. All manuscripts accepted for publication will require the authors to make final submission employing these style files. The style files are available on the web at the **IEEE Author Digital Toolbox** under "Template for all TRANSACTIONS." (LaTeX and MS Word). Please note the following requirements about the abstract:

- The abstract must be a concise yet comprehensive reflection of what is in your article.
- The abstract must be self-contained, without abbreviations, footnotes, displayed equations, or references.
- The abstract must be between 150-250 words.
- The abstract should include a few keywords or phrases, as this will help readers to find it. Avoid over-repetition of such phrases as this can result in a page being rejected by search engines.

In addition to written abstracts, papers may include a graphical abstract; see http://www.ieee.org/publications_standards/publications/authors/authors_journals.html. for options and format requirements.

IEEE supports the publication of author names in the native language alongside the English versions of the names in the author list of an article. For more information, see "Author names in native languages" (http://www.ieee.org/publications_standards/publications/authors/auth_names_native_lang.pdf) on the IEEE Author Digital Toolbox page.

Open Access. The publication is a hybrid journal, allowing either Traditional manuscript submission or Open Access (author-pays OA) manuscript submission. Upon submission, if you choose to have your manuscript be an Open Access article, you commit to pay the discounted $1,750 OA fee if your manuscript is accepted for publication in order to enable unrestricted public access. Any other application charges (such as overlength page charge and/or charge for the use of color in the print format) will be billed separately once the manuscript formatting is complete but prior to the publication. If you would like your manuscript to be a Traditional submission, your article will be available to qualified subscribers and purchasers via IEEE Xplore. No OA payment is required for Traditional submission.

Page Charges.

Voluntary Page Charges. Upon acceptance of a manuscript for publication, the author(s) or his/her/their company or institution will be asked to pay a charge of $110 per page to cover part of the cost of publication of the first ten pages that comprise the standard length (two pages, in the case of Correspondences).

Mandatory Page Charges The author(s) or his/her/their company or institution will be billed $220 per each page in excess of the first ten published pages for regular papers and six published pages for correspondence items. (**NOTE: Papers accepted to IEEE TRANSACTIONS ON MULTIMEDIA in excess of 8 pages will be subject to mandatory overlength page charges.) These are mandatory page charges and the author(s) will be held responsible for them. They are not negotiable or voluntary. The author(s) signifies his willingness to pay these charges simply by submitting his/her/their manuscript to the TRANSACTIONS. The Publisher holds the right to withhold publication under any circumstance, as well as publication of the current or future submissions of authors who have outstanding mandatory page charge debt. No mandatory overlength page charges will be applied to overview articles in the Society's journals.

Color Charges. Color figures which appear in color only in the electronic (Xplore) version can be used free of charge. In this case, the figure will be printed in the hardcopy version in grayscale, and the author is responsible that the corresponding grayscale figure is intelligible. Color reproduction charges for print are the responsibility of the author. Details of the associated charges can be found on the IEEE Publications page.

Payment of fees on color reproduction is not negotiable or voluntary, and the author's agreement to publish the manuscript in these TRANSACTIONS is considered acceptance of this requirement.

*EDICS Webpages:

IEEE TRANSACTIONS ON SIGNAL PROCESSING:
http://www.signalprocessingsociety.org/publications/periodicals/tsp/TSP-EDICS/
IEEE TRANSACTIONS ON IMAGE PROCESSING:
http://www.signalprocessingsociety.org/publications/periodicals/image-processing/tip-edics/
IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE / ACM:
http://www.signalprocessingsociety.org/publications/periodicals/taslp/taslp-edics/
IEEE TRANSACTIONS ON INFORMATION, FORENSICS AND SECURITY:
http://www.signalprocessingsociety.org/publications/periodicals/forensics/forensics-edics/
IEEE TRANSACTIONS ON MULTIMEDIA:
http://www.signalprocessingsociety.org/tmm/tmm-edics/
IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING:
http://www.signalprocessingsociety.org/publications/periodicals/tci/tci-edics/
IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS:
http://www.signalprocessingsociety.org/publications/periodicals/tsipn/tsipn-edics/

2016 IEEE SIGNAL PROCESSING SOCIETY MEMBERSHIP APPLICATION

Mail to: **IEEE OPERATIONS CENTER, ATTN: Matthew Plotner, Member and Geographic Activities, 445 Hoes Lane, Piscataway, New Jersey 08854 USA**
or Fax to (732) 981-0225 (credit card payments only.)
For info call (732) 981-0060 or 1 (800) 678-IEEE or E-mail: new.membership@ieee.org

IEEE

1. PERSONAL INFORMATION

NAME AS IT SHOULD APPEAR ON IEEE MAILINGS: **SEND MAIL TO**: ☐ **Home Address** OR ☐ **Business/School Address**
If not indicated, mail will be sent to home address. Note: Enter your name as you wish it to appear on membership card and all correspondence.
PLEASE PRINT Do not exceed 40 characters or spaces per line. Abbreviate as needed. Please circle your last/surname as a key identifier for the IEEE database.

TITLE | FIRST OR GIVEN NAME | MIDDLE NAME | SURNAME/LAST NAME

HOME ADDRESS

CITY | STATE/PROVINCE | POSTAL CODE | COUNTRY

2.

Are you now or were you ever a member of IEEE? ☐ **Yes** ☐ **No**
If yes, please provide, if known:

MEMBERSHIP NUMBER_____

Grade_____ **Year Membership Expired:**_____

3. BUSINESS/PROFESSIONAL INFORMATION

Company Name

Department/Division

Title/Position | Years in Current Position

Years in the Profession Since Graduation | ☐ PE State/Province

Street Address

City | State/Province | Postal Code | Country

4. EDUCATION

A baccalaureate degree from an IEEE recognized educational program assures assignment of "Member" grade. For others, additional information and references may be necessary for grade assignment.

A.
Baccalaureate Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

B.
Highest Technical Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

5.

Full signature of applicant

6. DEMOGRAPHIC INFORMATION – ALL APPLICANTS -

Date Of Birth _____
Day Month Year

☐ Male ☐ Female

7. CONTACT INFORMATION

Office Phone/Office Fax | Home Phone/Home Fax

Office E-Mail | Home E-Mail

8. 2016 IEEE MEMBER RATES

IEEE DUES Residence	16 Aug 14-28 Feb 15 Pay Full Year	1 Mar -15 Aug 15 Pay Half Year**
United States	$197.00 ☐	$98.50 ☐
Canada (incl. GST)	$173.35 ☐	$86.68 ☐
Canada (incl. HST for PEI)	$186.58 ☐	$93.29 ☐
Canada (incl. HST for Nova Scotia)	$188.05 ☐	$94.03 ☐
Canada (incl. HST for NB, NF and ON)	$185.11 ☐	$92.56 ☐
Canada (incl. GST and QST Quebec)	$188.01 ☐	$94.01 ☐
Africa, Europe, Middle East	$160.00 ☐	$80.00 ☐
Latin America	$151.00 ☐	$75.50 ☐
Asia, Pacific	$152.00 ☐	$76.00 ☐

Canadian Taxes (GST/HST): All supplies, which include dues, Society membership fees, online products and publications (except CD-ROM and DVD media), shipped to locations within Canada are subject to the GST of 5% or the HST of 12%, 13% or 15%, depending on the Province to which the materials are shipped. GST and HST do not apply to Regional Assessments. (IEEE Canadian Business Number 12563 4188 RT0001)
Value Added Tax (VAT) in the European Union: In accordance with the European Union Council Directives 2002/38/EC and 77/388/EEC amended by Council Regulation (EC)792/2002, IEEE is required to charge and collect VAT on electronic/digitized products sold to private consumers that reside in the European Union. The VAT rate applied is the EU member country standard rate where the consumer is resident. (IEEE's VAT registration number is EU826000081)
U.S. Sales Taxes: Please add applicable state and local sales and use tax on orders shipped to **Alabama, Arizona, California, Colorado, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, West Virginia, Wisconsin.** Customers claiming a tax exemption must include an appropriate and properly completed tax-exemption certificate with their first order.

IEEE Signal Processing Society

2016 SPS MEMBER RATES

		16 Aug-28 Feb Pay Full Year	1 Mar-15 Aug Pay Half Year
Signal Processing Society Membership Fee*		$ 22.00 ☐	$ 11.00 ☐
Fee includes: **IEEE Signal Processing Magazine** (electronic and digital), **Inside Signal Proc. eNewsletter** (electronic) and **IEEE Signal Processing Society Content Gazette** (electronic).			
Add $17 to enhance SPS Membership and also receive:		$17.00 ☐	$ 8.50 ☐
IEEE Signal Processing Magazine (print) and **SPS Digital Library**: online access to Signal Processing Magazine, Signal Processing Letters, Journal of Selected Topics in Signal Processing, Trans. on Audio, Speech, and Language Processing, Trans. on Image Processing, Trans. on Information Forensics and Security and Trans. on Signal Processing.			
Publications available only with SPS membership:			
Signal Processing, IEEE Transactions on:	Print	$209.00 ☐	$104.50 ☐
Audio, Speech, and Lang. Proc., IEEE/ACM Trans. on:	Print	$160.00 ☐	$ 80.00 ☐
Image Processing, IEEE Transactions on:	Print	$207.00 ☐	$103.50 ☐
Information Forensics and Security, IEEE Trans. on:	Print	$179.00 ☐	$ 89.50 ☐
IEEE Journal of Selected Topics in Signal Processing:	Print	$176.00 ☐	$ 88.00 ☐
Affective Computing, IEEE Transactions on:	Electronic	$ 36.00 ☐	$ 18.00 ☐
Biomedical and Health Informatics, IEEE Journal of:	Print	$ 55.00 ☐	$ 27.50 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$ 65.00 ☐	$ 32.50 ☐
IEEE Cloud Computing	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
IEEE Trans. on Cognitive Comm. &Networking	Electronic	$ 27.00 ☐	$ 13.50 ☐
IEEE Trans. on Computational Imaging	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Trans. on Big Data	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Molecular, Biological, & Multi-scale Communications	Electronic	$ 25.00 ☐	$ 12.50 ☐
IEEE Internet of Things Journal	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Cloud Computing	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE Trans. on Computational Social Systems	Electronic	$ 30.00 ☐	$ 15.00 ☐
IEEE Trans. on Signal & Info Proc. Over Networks	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Biometrics Compendium:	Online	$ 30.00 ☐	$ 15.00 ☐
Computing in Science & Engrg. Mag.:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Medical Imaging, IEEE Transactions on:	Print	$ 74.00 ☐	$ 37.00 ☐
	Electronic	$ 53.00 ☐	$ 26.50 ☐
	Print & Electronic	$ 89.00 ☐	$ 44.50 ☐
Mobile Computing, IEEE Transactions on:	ELE/Print Abstract/CD-ROM	$ 41.00 ☐	$ 20.50 ☐
Multimedia, IEEE Transactions on:	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE MultiMedia Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Network Science and Engrg., IEEE Trans. on:	Electronic	$ 34.00 ☐	$ 17.00 ☐
IEEE Reviews in Biomedical Engineering:	Print	$ 25.00 ☐	$ 12.50 ☐
	Print & Electronic	$ 40.00 ☐	$ 20.00 ☐
IEEE Security and Privacy Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
IEEE Sensors Journal:	Electronic	$ 40.00 ☐	$ 20.00 ☐
Smart Grid, IEEE Transactions on:	Print	$100.00 ☐	$ 50.00 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$120.00 ☐	$ 60.00 ☐
Wireless Communications, IEEE Transactions on:	Print	$127.00 ☐	$ 63.50 ☐
	Electronic	$ 49.00 ☐	$ 24.50 ☐
	Print & Electronic	$127.00 ☐	$ 63.50 ☐
IEEE Wireless Communications Letters:	Electronic	$ 19.00 ☐	$ 9.50 ☐

*IEEE membership required or requested
Affiliate application to join SP Society only.

Amount Paid $_____

9.

IEEE Membership Affiliate Fee (See pricing in Section 8) $_____

Signal Processing Society Fees $_____

Canadian residents pay 5% GST or 13% HST
Reg. No. 125634188 on Society payment(s) & pubs only Tax $_____

AMOUNT PAID WITH APPLICATION TOTAL $_____

Prices subject to change without notice.

☐ **Check or money order enclosed Payable to IEEE on a U.S. Bank**
☐ **American Express** ☐ **VISA** ☐ **MasterCard**
☐ **Diners Club**

Exp. Date/ Mo./Yr.

Cardholder Zip Code Billing Statement Address/USA Only

Full signature of applicant using credit card Date

10. WERE YOU REFERRED?

☐ Yes ☐ No If yes, please provide the follow information:
Member Recruiter Name:_____
IEEE Recruiter's Member Number (Required): _____

2016 IEEE SIGNAL PROCESSING SOCIETY STUDENT MEMBERSHIP APPLICATION

(Current and reinstating IEEE members joining SPS complete areas 1, 2, 8, 9.)

Mail to: **IEEE OPERATIONS CENTER, ATTN: Matthew Plotner, Member and Geographic Activities, 445 Hoes Lane, Piscataway, New Jersey 08854 USA**
or Fax to (732) 981-0225 (credit card payments only.)
For info call (732) 981-0060 or 1 (800) 678-IEEE or E-mail: new.membership@ieee.org

1. PERSONAL INFORMATION

NAME AS IT SHOULD APPEAR ON IEEE MAILINGS: **SEND MAIL TO**: ☐ **Home Address** OR ☐ **Business/School Address**
If not indicated, mail will be sent to home address. Note: Enter your name as you wish it to appear on membership card and all correspondence.
PLEASE PRINT Do not exceed 40 characters or spaces per line. Abbreviate as needed. Please circle your last/surname as a key identifier for the IEEE database.

TITLE | FIRST OR GIVEN NAME | MIDDLE NAME | SURNAME/LAST NAME

HOME ADDRESS

CITY | STATE/PROVINCE | POSTAL CODE | COUNTRY

2.

Are you now or were you ever a member of IEEE? ☐ **Yes** ☐ **No**
If yes, please provide, if known:

MEMBERSHIP NUMBER_____

Grade_____**Year Membership Expired:**_____

3. BUSINESS/PROFESSIONAL INFORMATION

Company Name

Department/Division

Title/Position | Years in Current Position

Years in the Profession Since Graduation | ☐ PE State/Province

Street Address

City | **State/Province** | **Postal Code** | **Country**

4. EDUCATION

A baccalaureate degree from an IEEE recognized educational program assures assignment of "Member" grade. For others, additional information and references may be necessary for grade assignment.

A.

Baccalaureate Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

B.

Highest Technical Degree Received | Program/Course of Study

College/University | Campus

State/Province | Country | Mo./Yr. Degree Received

5.

Full signature of applicant

6. DEMOGRAPHIC INFORMATION – ALL APPLICANTS -

Date Of Birth _____ ☐ Male ☐ Female
Day Month Year

7. CONTACT INFORMATION

Office Phone/Office Fax | Home Phone/Home Fax

Office E-Mail | Home E-Mail

8. 2016 IEEE STUDENT MEMBER RATES

IEEE DUES Residence	16 Aug 14-28 Feb 15 Pay Full Year	1 Mar -15 Aug 15 Pay Half Year**
United States	$32.00 ☐	$16.00 ☐
Canada (incl. GST)	$33.60 ☐	$16.80 ☐
Canada (incl. HST for NB, NF, and ON)	$36.16 ☐	$18.08 ☐
Canada (incl. HST for Nova Scotia)	$36.80 ☐	$18.40 ☐
Canada (incl. HST for PEI)	$36.48 ☐	$18.24 ☐
Canada (incl. GST and QST Quebec)	$36.79 ☐	$18.40 ☐
Africa, Europe, Middle East, Latin America, Asia, Pacific	$27.00 ☐	$13.50 ☐

Canadian Taxes (GST/HST): All supplies, which include dues, Society membership fees, online products and publications (except CD-ROM and DVD media), shipped to locations within Canada are subject to the GST of 5% or the HST of 12%,13% or 15%, depending on the Province to which the materials are shipped. GST and HST do not apply to Regional Assessments. (IEEE Canadian Business Number 12563 4188 RT0001)

Value Added Tax (VAT) in the European Union: In accordance with the European Union Council Directives 2002/38/EC and 77/388/EEC amended by Council Regulation (EC)792/2002, IEEE is required to charge and collect VAT on electronic/digitized products sold to private consumers that reside in the European Union. The VAT rate applied is the EU member country standard rate where the consumer is resident. (IEEE's VAT registration number is EU826000081)

U.S. Sales Taxes: Please add applicable state and local sales and use tax on orders shipped to **Alabama, Arizona, California, Colorado, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, West Virginia, Wisconsin.** Customers claiming a tax exemption must include an appropriate and properly completed tax-exemption certificate with their first order.

IEEE Signal Processing Society

2016 SPS STUDENT MEMBER RATES

	16 Aug-28 Feb Pay Full Year	1 Mar-15 Aug Pay Half Year
Signal Processing Society Membership Fee*	$11.00 ☐	$ 5.50 ☐

Fee includes: **IEEE Signal Processing Magazine** (electronic and digital), **Inside Signal Processing eNewsletter** (electronic) and **IEEE Signal Processing Society Content Gazette** (electronic).

Add $8 to enhance SPS Membership and also receive: $ 9.00 ☐ $ 4.50 ☐
IEEE Signal Processing Society Magazine (print) and **SPS Digital Library**: online access to Signal Processing Magazine, Signal Processing Letters, Journal of Selected Topics in Signal Processing, Trans. on Audio, Speech, and Language Processing, Trans. on Image Processing, Trans. on Information Forensics and Security and Trans. on Signal Processing.

Publications available only with SPS membership:

Signal Processing, IEEE Transactions on:	Print	$105.00 ☐	$ 52.50 ☐
Audio, Speech, and Lang. Proc., IEEE/ACM Trans. on:	Print	$ 80.00 ☐	$ 40.00 ☐
Image Processing, IEEE Transactions on:	Print	$104.00 ☐	$ 52.00 ☐
Information Forensics and Security, IEEE Trans. on:	Print	$ 90.00 ☐	$ 45.00 ☐
IEEE Journal of Selected Topics in Signal Processing:	Print	$ 88.00 ☐	$ 44.00 ☐
Affective Computing, IEEE Transactions on:	Electronic	$ 18.00 ☐	$ 9.00 ☐
Biomedical and Health Informatics, IEEE Journal of:	Print	$ 28.00 ☐	$ 14.00 ☐
	Electronic	$ 20.00 ☐	$ 10.00 ☐
	Print & Electronic	$ 33.00 ☐	$ 16.50 ☐
IEEE Cloud Computing	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
IEEE Trans. on Cognitive Comm. &Networking	Electronic	$ 14.00 ☐	$ 7.00 ☐
IEEE Trans. on Computational Imaging	Electronic	$ 14.00 ☐	$ 7.00 ☐
IEEE Trans. on Big Data	Electronic	$ 13.00 ☐	$ 6.50 ☐
IEEE Trans. on Molecular, Biological, & Multi-Scale Communications	Electronic	$ 13.00 ☐	$ 6.50 ☐
IEEE Internet of Things Journal	Electronic	$ 13.00 ☐	$ 6.50 ☐
IEEE Trans. on Cloud Computing	Electronic	$ 22.00 ☐	$ 11.00 ☐
IEEE Trans. on Computational Social Systems	Electronic	$ 15.00 ☐	$ 7.50 ☐
IEEE Trans. on Signal & Info Proc. Over Networks	Electronic	$ 14.00 ☐	$ 7.00 ☐
IEEE Biometrics Compendium:	Online	$ 15.00 ☐	$ 7.50 ☐
Computing in Science & Engrg. Mag.:	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
	Print	$ 35.00 ☐	$ 17.50 ☐
Medical Imaging, IEEE Transactions on:	Print	$ 37.00 ☐	$ 18.50 ☐
	Electronic	$ 27.00 ☐	$ 13.50 ☐
	Print & Electronic	$ 45.00 ☐	$ 22.50 ☐
Mobile Computing, IEEE Transactions on:	ELE/Print Abstract/CD-ROM	$ 21.00 ☐	$ 10.50 ☐
Multimedia, IEEE Transactions on:	Electronic	$ 22.00 ☐	$ 11.00 ☐
IEEE MultiMedia Magazine:	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
	Print	$ 35.00 ☐	$ 17.50 ☐
Network Science and Engrg., IEEE Trans. on:	Electronic	$ 17.00 ☐	$ 8.50 ☐
IEEE Reviews in Biomedical Engineering:	Print	$ 13.00 ☐	$ 6.50 ☐
	Print & Electronic	$ 20.00 ☐	$ 10.00 ☐
IEEE Security and Privacy Magazine:	Electronic and Digital	$ 20.00 ☐	$ 10.00 ☐
	Print	$ 35.00 ☐	$ 17.50 ☐
IEEE Sensors Journal:	Electronic	$ 20.00 ☐	$ 10.00 ☐
Smart Grid, IEEE Transactions on:	Print	$ 50.00 ☐	$ 25.00 ☐
	Electronic	$ 20.00 ☐	$ 10.00 ☐
	Print & Electronic	$ 60.00 ☐	$ 30.00 ☐
Wireless Communications, IEEE Transactions on:	Print	$ 64.00 ☐	$ 32.00 ☐
	Electronic	$ 25.00 ☐	$ 12.50 ☐
	Print & Electronic	$ 64.00 ☐	$ 32.00 ☐
IEEE Wireless Communications Letters:	Electronic	$ 10.00 ☐	$ 5.00 ☐

*IEEE membership required or requested
Affiliate application to join SP Society only.

Amount Paid $_____

9.

IEEE Membership Fee (See pricing in Section 8) $_____
Signal Processing Society Fees $_____

Canadian residents pay 5% GST or 13% HST
Reg. No. 125634188 on Society payment(s) & pubs only Tax $_____

AMOUNT PAID WITH APPLICATION TOTAL $_____
Prices subject to change without notice.

☐ **Check or money order enclosed Payable to IEEE on a U.S. Bank**
☐ **American Express** ☐ **VISA** ☐ **MasterCard** ☐ **Diners Club**

Exp. Date/ Mo./Yr.

Cardholder Zip Code Billing Statement Address/USA Only

Full signature of applicant using credit card | Date

10. WERE YOU REFERRED?

☐ Yes ☐ No If yes, please provide the follow information:
Member Recruiter Name:_____
IEEE Recruiter's Member Number (Required): _____

2016 IEEE SIGNAL PROCESSING SOCIETY AFFILIATE MEMBERSHIP APPLICATION

Mail to: **IEEE OPERATIONS CENTER, ATTN: Matthew Plotner, Member and Geographic Activities, 445 Hoes Lane, Piscataway, New Jersey 08854 USA**
or Fax to (732) 981-0225 (credit card payments only.)
For info call (732) 981-0060 or 1 (800) 678-IEEE or E-mail: new.membership@ieee.org

IEEE

1. PERSONAL INFORMATION

NAME AS IT SHOULD APPEAR ON IEEE MAILINGS: **SEND MAIL TO**: ☐ **Home Address** OR ☐ **Business/School Address**
If not indicated, mail will be sent to home address. Note: Enter your name as you wish it to appear on membership card and all correspondence.
PLEASE PRINT Do not exceed 40 characters or spaces per line. Abbreviate as needed. Please circle your last/surname as a key identifier for the IEEE database.

TITLE FIRST OR GIVEN NAME MIDDLE NAME SURNAME/LAST NAME

HOME ADDRESS

CITY STATE/PROVINCE POSTAL CODE COUNTRY

2.

Are you now or were you ever a member of IEEE? ☐ **Yes** ☐ **No**
If yes, please provide, if known:

MEMBERSHIP NUMBER_____

Grade_____**Year Membership Expired:**_____

3. BUSINESS/PROFESSIONAL INFORMATION

Company Name

Department/Division

Title/Position Years in Current Position

Years in the Profession Since Graduation ☐ PE State/Province

Street Address

City State/Province Postal Code Country

4. EDUCATION

A baccalaureate degree from an IEEE recognized educational program assures assignment of "Member" grade. For others, additional information and references may be necessary for grade assignment.

A.

Baccalaureate Degree Received Program/Course of Study

College/University Campus

State/Province Country Mo./Yr. Degree Received

B.

Highest Technical Degree Received Program/Course of Study

College/University Campus

State/Province Country Mo./Yr. Degree Received

5.

Full signature of applicant

6. DEMOGRAPHIC INFORMATION – ALL APPLICANTS -

Date Of Birth _____ ☐ Male ☐ Female
Day Month Year

7. CONTACT INFORMATION

Office Phone/Office Fax Home Phone/Home Fax

Office E-Mail Home E-Mail

8. 2016 IEEE AFFILIATE RATES

IEEE DUES Residence	16 Aug 14-28 Feb 15 Pay Full Year	1 Mar -15 Aug 15 Pay Half Year**
United States	$36.50 ☐	$36.75 ☐
Canada (incl. GST)	$77.18 ☐	$38.59 ☐
Canada (incl. HST for PEI)	$83.79 ☐	$41.89 ☐
Canada (incl. HST for Nova Scotia)	$84.58 ☐	$42.26 ☐
Canada (incl. HST for NB, NF and ON)	$83.06 ☐	$41.53 ☐
Canada (incl. GST and QST Quebec)	$84.51 ☐	$42.25 ☐
Africa, Europe, Middle East	$73.50 ☐	$36.75 ☐
Latin America	$73.50 ☐	$36.75 ☐
Asia, Pacific	$73.50 ☐	$36.75 ☐

Canadian Taxes (GST/HST): All supplies, which include dues, Society membership fees, online products and publications (except CD-ROM and DVD media), shipped to locations within Canada are subject to the GST of 5% or the HST of 13%, 14% or 15%, depending on the Province to which the materials are shipped. GST and HST do not apply to Regional Assessments. (IEEE Canadian Business Number 12563 4188 RT0001)

Value Added Tax (VAT) in the European Union: In accordance with the European Union Council Directives 2002/38/EC and 77/388/EEC amended by Council Regulation (EC)792/2002, IEEE is required to charge and collect VAT on electronic/digitized products sold to private consumers that reside in the European Union. The VAT rate applied is the EU member country standard rate where the consumer is resident. (IEEE's VAT registration number is EU826000081)

U.S. Sales Taxes: Please add applicable state and local sales and use tax on orders shipped to **Alabama, Arizona, California, Colorado, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Massachusetts, Maryland, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, West Virginia, Wisconsin.** Customers claiming a tax exemption must include an appropriate and properly completed tax-exemption certificate with their first order.

IEEE Signal Processing Society

2016 SPS MEMBER RATES

		16 Aug-28 Feb Pay Full Year	1 Mar-15 Aug Pay Half Year
Signal Processing Society Membership Fee*		$ 22.00 ☐	$ 11.00 ☐
Fee includes: **IEEE Signal Processing Magazine** (electronic and digital), **Inside Signal Proc. eNewsletter** (electronic) and **IEEE Signal Processing Society Content Gazette** (electronic).			
Add $17 to enhance SPS Membership and also receive:		$17.00 ☐	$ 8.50 ☐
IEEE Signal Processing Magazine (print) and **SPS Digital Library**: online access to Signal Processing Magazine, Signal Processing Letters, Journal of Selected Topics in Signal Processing, Trans. on Audio, Speech, and Language Processing, Trans. on Image Processing, Trans. on Information Forensics and Security and Trans. on Signal Processing.			
Publications available only with SPS membership:			
Signal Processing, IEEE Transactions on:	Print	$209.00 ☐	$104.50 ☐
Audio, Speech, and Lang. Proc., IEEE/ACM Trans. on:	Print	$160.00 ☐	$ 80.00 ☐
Image Processing, IEEE Transactions on:	Print	$207.00 ☐	$103.50 ☐
Information Forensics and Security, IEEE Trans. on:	Print	$179.00 ☐	$ 89.50 ☐
IEEE Journal of Selected Topics in Signal Processing:	Print	$176.00 ☐	$ 88.00 ☐
Affective Computing, IEEE Transactions on:	Electronic	$ 36.00 ☐	$ 18.00 ☐
Biomedical and Health Informatics, IEEE Journal of:	Print	$ 55.00 ☐	$ 27.50 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$ 65.00 ☐	$ 32.50 ☐
IEEE Cloud Computing	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
IEEE Trans. on Cognitive Comm. &Networking	Electronic	$ 27.00 ☐	$ 13.50 ☐
IEEE Trans. on Computational Imaging	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Trans. on Big Data	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Molecular, Biological, & Multi-scale Communications	Electronic	$ 25.00 ☐	$ 12.50 ☐
IEEE Internet of Things Journal	Electronic	$ 26.00 ☐	$ 13.00 ☐
IEEE Trans. on Cloud Computing	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE Trans. on Computational Social Systems	Electronic	$ 30.00 ☐	$ 15.00 ☐
IEEE Trans. on Signal & Info Proc. Over Networks	Electronic	$ 28.00 ☐	$ 14.00 ☐
IEEE Biometrics Compendium:	Online	$ 30.00 ☐	$ 15.00 ☐
Computing in Science & Engrg. Mag.:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Medical Imaging, IEEE Transactions on:	Print	$ 74.00 ☐	$ 37.00 ☐
	Electronic	$ 53.00 ☐	$ 26.50 ☐
	Print & Electronic	$ 89.00 ☐	$ 44.50 ☐
Mobile Computing, IEEE Transactions on:	ELE/Print Abstract/CD-ROM	$ 41.00 ☐	$ 20.50 ☐
Multimedia, IEEE Transactions on:	Electronic	$ 43.00 ☐	$ 21.50 ☐
IEEE MultiMedia Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
Network Science and Engrg., IEEE Trans. on:	Electronic	$ 34.00 ☐	$ 17.00 ☐
IEEE Reviews in Biomedical Engineering:	Print	$ 25.00 ☐	$ 12.50 ☐
	Electronic	$ 25.00 ☐	$ 12.50 ☐
	Print & Electronic	$ 40.00 ☐	$ 20.00 ☐
IEEE Security and Privacy Magazine:	Electronic and Digital	$ 39.00 ☐	$ 19.50 ☐
	Print	$ 69.00 ☐	$ 34.50 ☐
IEEE Sensors Journal:	Electronic	$ 40.00 ☐	$ 20.00 ☐
Smart Grid, IEEE Transactions on:	Print	$100.00 ☐	$ 50.00 ☐
	Electronic	$ 40.00 ☐	$ 20.00 ☐
	Print & Electronic	$120.00 ☐	$ 60.00 ☐
Wireless Communications, IEEE Transactions on:	Print	$127.00 ☐	$ 63.50 ☐
	Electronic	$ 49.00 ☐	$ 24.50 ☐
	Print & Electronic	$127.00 ☐	$ 63.50 ☐
IEEE Wireless Communications Letters:	Electronic	$ 19.00 ☐	$ 9.50 ☐

*IEEE membership required or requested
Affiliate application to join SP Society only. Amount Paid $_____

9.

IEEE Membership Affiliate Fee (See pricing in Section 8) $_____

Signal Processing Society Fees $_____

Canadian residents pay 5% GST or 13% HST
Reg. No. 125634188 on Society payment(s) & pubs only Tax $_____

AMOUNT PAID WITH APPLICATION TOTAL $_____

Prices subject to change without notice.

☐ **Check or money order enclosed Payable to IEEE on a U.S. Bank**
☐ **American Express** ☐ **VISA** ☐ **MasterCard**
☐ **Diners Club**

Exp. Date/ Mo./Yr.

Cardholder Zip Code Billing Statement Address/USA Only

Full signature of applicant using credit card Date

10. WERE YOU REFERRED?

☐ Yes ☐ No If yes, please provide the follow information:
Member Recruiter Name:_____
IEEE Recruiter's Member Number (Required): _____

SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®

IEEE Signal Processing Society®

IEEE

SignalProcessing Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page Qmags THE WORLD'S NEWSSTAND®