SiPS 2019
SIGNAL PROCESSING SYSTEMS

The IEEE International Workshop on Signal Processing Systems

Oct. 20-23, 2019 · Nanjing, China

VENUE

Holiday Inn Nanjing Qinhuai South, Nanjing, China

Add.: NO.21 Mozhou East Road, Jiangning District, Nanjing, JS, 211111, China

CONTENTS

WELCOME MESSAGES

WELCOME MESSAGE FROM GENERAL CHAIRS

It is our great pleasure to welcome all our workshop delegates and distinguished speakers to the IEEE International Workshop on Signal Processing Systems 2019 (SiPS 2019) held in Nanjing, China, from October 20th to 23rd, 2019. IEEE SiPS is a premier international forum for discussion of new technology progress and innovations in the design and implementation of signal processing systems. It addresses all aspects of architecture and design methods of these systems. Emphasis is on current and future challenges in research and development in both academia and industry.

This year, SiPS's special theme is “Advanced Signal Processing for Smart Learning and Implementation”. The delegates of SiPS 2019 will discuss and present the latest advances in this field. In the workshop, a diverse range of featured topics will also be discussed, including machine learning, MIMO, error correction codes, signal processing, and so on. We are also delighted to have invited renowned scholars to give three Keynote Speeches and six Tutorials.

In addition to the excellent program of the workshop, Nanjing has long been a major center of culture, education, research, politics, economy, transport networks and tourism, and is an unforgettable place to visit. We hope that during SiPS 2019, you will have a chance to explore Nanjing’s many important heritage sites, including the Presidential Palace, Sun Yat-sen Mausoleum, Fuzimiao, Ming Palace, Chaotian Palace, Porcelain Tower, Drum Tower, Stone City, City Wall, Qinhuai River, Xuanwu Lake, and Purple Mountain.

SiPS 2019 will not succeed without the strong support of hundreds of volunteers who contributed to the various processes and it would not be possible for us to name them all in this short message. In particular, the Technical Program Committee, led by our TPC Chairs completed a thorough peer-review process with the help of reviewers. Our Organizing Committee, in particular, Special Session, Finance, Publication, Registration, Publicity, and Local Chairs, and Secretary, worked tirelessly to ensure the best quality experience of this workshop. Our thanks also go to Southeast University and Purple Mountain Laboratories for their support to the workshop. The Organizing Committee really appreciates Chengdu Young Academic Conference Co. Ltd. for the great help on detailed arrangements in organizing this workshop.

Finally, we would like to thank all the authors and attendees for participating in the workshop. We wish you all have a wonderful and fruitful time at SiPS 2019, and memorable experience in Nanjing.

SiPS 2019 General Chairs
Zhongfeng Wang, Nanjing University, China
Chuan Zhang, Southeast University, China

Zhongfeng Wang

Chuan Zhang

WELCOME MESSAGES

WELCOME MESSAGE FROM TPC CO-CHAIRS

Dear Friends and Colleagues,

On behalf of the Technical Program Committee, we would like to welcome all the participants of the 33rd IEEE Workshop on Signal Processing Systems (SiPS). It is an honor for us to welcome researchers, students, and industry representatives to attend this workshop in Nanjing, China.

This year, 96 papers were submitted to the regular sessions and 42 papers have been accepted and included in the proceedings, which corresponds to an acceptance rate of 43%; additionally, 19 papers were accepted from the special sessions for presentations and publication. The workshop program consists of six tutorials, six technical sessions, two poster sessions, and the following four special sessions:

- Practical Machine-Learning-Aided Communication Systems
- Emerging Computing Paradigms for Signal Processing and Smart Learning
- Hardware Security and Hardware Implementation of Emerging Cryptographic Primitives
- Improving the Performance of Autonomous Systems: Algorithm, Hardware, and Application

In addition, we have three keynotes by experts from academia and industry: Prof. Keshab K. Parhi from University of Minnesota, USA, Prof. Deming Chen from University of Illinois at Urbana-Champaign, USA, and Sunny Zhang from Intel Labs China.

Such a rich and diverse program has been made possible thanks to the hard work of the 53 members of the technical program committee, as well as the publication chairs, publicity chairs, tutorial chairs, and special session chairs, and our liaisons from around the globe.

We hope that you will enjoy this workshop as much as we enjoyed organizing it!

Yours sincerely,
Xinmiao Zhang, The Ohio State University
Tokunbo Ogunfunmi, Santa Clara University
Christoph Studer, Cornell Tech

Xinmiao Zhang

Tokunbo Ogunfunmi

Christoph Studer

CONFERENCE COMMITTEE

Honorary Chair

Xiaohu You, *Southeast University*

General Chairs

Zhongfeng Wang, *Nanjing University*

Chuan Zhang*, Southeast University*

TPC Co-Chairs

Xinmiao Zhang, *The Ohio State University*

Tokunbo Ogunfunmi, *Santa Clara University*

Christoph Studer, *Cornell Tech*

Local Chairs

Feng Yan, *Southeast University*

Jun Lin, *Nanjing University*

Financial Chair

Chunguo Li, *Southeast University*

Publication Chairs

Weiqiang Liu, *NUAA*

Yili Xia, *Southeast University*

Publicity Chairs

Yun Chen, *Fudan University*

Peiyi Zhao, *Chapman University*

Tutorial Chairs

Yeong-Luh Ueng, *National Tsing Hua University*

Yuan-Hao Huang, *National Tsing Hua University*

Special Session Chairs

Fei Qiao, *Tsinghua University*

Shan Cao, *Shanghai University*

Asia Liaison

Jienan Chen, *University of Electronic Science and Technology of China*

America Liaison

Bo Yuan, *Rutgers University*

Europe Liaison

Andreas Burg, *École Polytechnique Fédérale de Lausanne*

Industrial Liaison

Li Du, *University of California, Los Angeles*

Conference Secretary

Xiaosi Tan, *Southeast University*

Technical Program Committee

Alexios Balatsoukas-Stimming, *Eindhoven University of Technology, Netherlands*

Andreas Burg, *ETH Zurich*

Andy Wu, *National Taiwan University*

Bohan Yang, *Tsinghua University*

Chaitali Chakrabarti, *Arizona State University*

Chao-Tsung Huang, *National Tsing Hua University*

Charles Jeon, *Intel*

Che Lin, *National Tsing Hua University*

Ching-Te Chiu, *National Tsing Hua University*

Christophe Jego, *IMS Laboratory*

Daniel Menard, *UEB/INSA Rennes/IETR*

Fang Cai, *Credo Semiconductor*

Farhana Sheikh, *Intel*

Fei Qiao, *Tsinghua University*

Francesca Palumbo, *Information Eng. Unit - PolComIng - University of Sassari*

Gabriel Falcao, *Instituto de Telecomunicacoes*

Gerald Sobelman, *University of Minnesota*

Hanho Lee, *Inha University*

Hongmei Xie, *Western Digital*

Hoyoung Yoo, *Chungnam National University*

Hua Qian, *Shanghai Advanced Research Institute, Chinese Academy of Sciences*

Jani Boutellier, *University of Vaasa*

Jean Francois Nezan, *INSA Rennes, IETR laboratory*

Jiangli Zhu, *Micron Technolog*

John McAllister, *Queen's University Belfast*

Joseph Cavallaro, *Rice University*

Juergen Wassner, *Lucerne University*

Jun Lin, *Nanjing University*

Jungwook Choi, *IBM*

Laura Conde-Canencia, *Lab-STICC*

Lei Wang, *University of Connecticut*

Leonel Sousa, *Universidade de Lisboa*

Liang Liu, *Lund University*

Maire O'Neill, *Queen's University Belfast*

Marilyn Wolf, *Georgia Institute of Technology*

Markku Juntti, *University of Oulu*

Mohammad Mansour, *American University of Beirut*

Pei-Yun Tsai, *National Central University*

Philippe Coussy, *Universit de Bretagne Sud / Lab-STICC*

Roger Woods, *Queen's University Belfast*

Shuvra Bhattacharyya, *University of Maryland*

Tong Zhang, *Rensselaer Polytechnic Institute*

Vassilis Paliouras, *University of Patras*

Vincent Gaudet, *University of Waterloo*

Wonyong Sung, *Seoul National University*

Yang Sun, *Qualcomm*

Yeong-Luh Ueng, *National Tsing Hua University*

Yingjie Lao, *Clemson University*

Yingyan Lin, *Rice University; University of Illinois at Urbana-Champaign*

Youngjoo Lee, *Pohang University of Science and Technology (POSTECH)*

Yuan-Hao Huang, *National Tsing-Hua University*

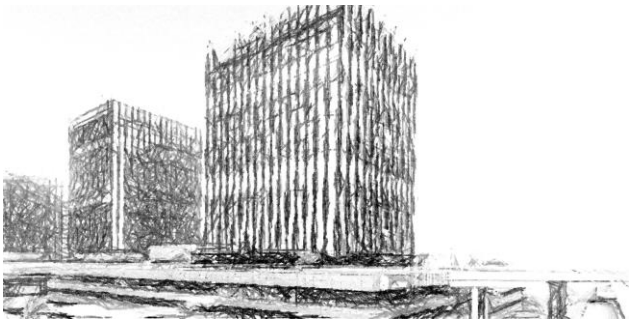
Yun Chen, *U.C.Berkeley*

Zhiyuan Yan, *Lehigh University*

CONFERENCE VENUE

Holiday Inn Nanjing Qinhuai South

Holiday Inn 南京上秦淮假日酒店



Add.: NO.21 Mozhou East Road, Jiangning District, Nanjing, JS, 211111, China
Tel: +86-400 830 2360
Email: rsvn@holidayinnjn.com

**Accommodation is not provided, and it's suggested to make an early reservation, since it's high season during the conference period.*

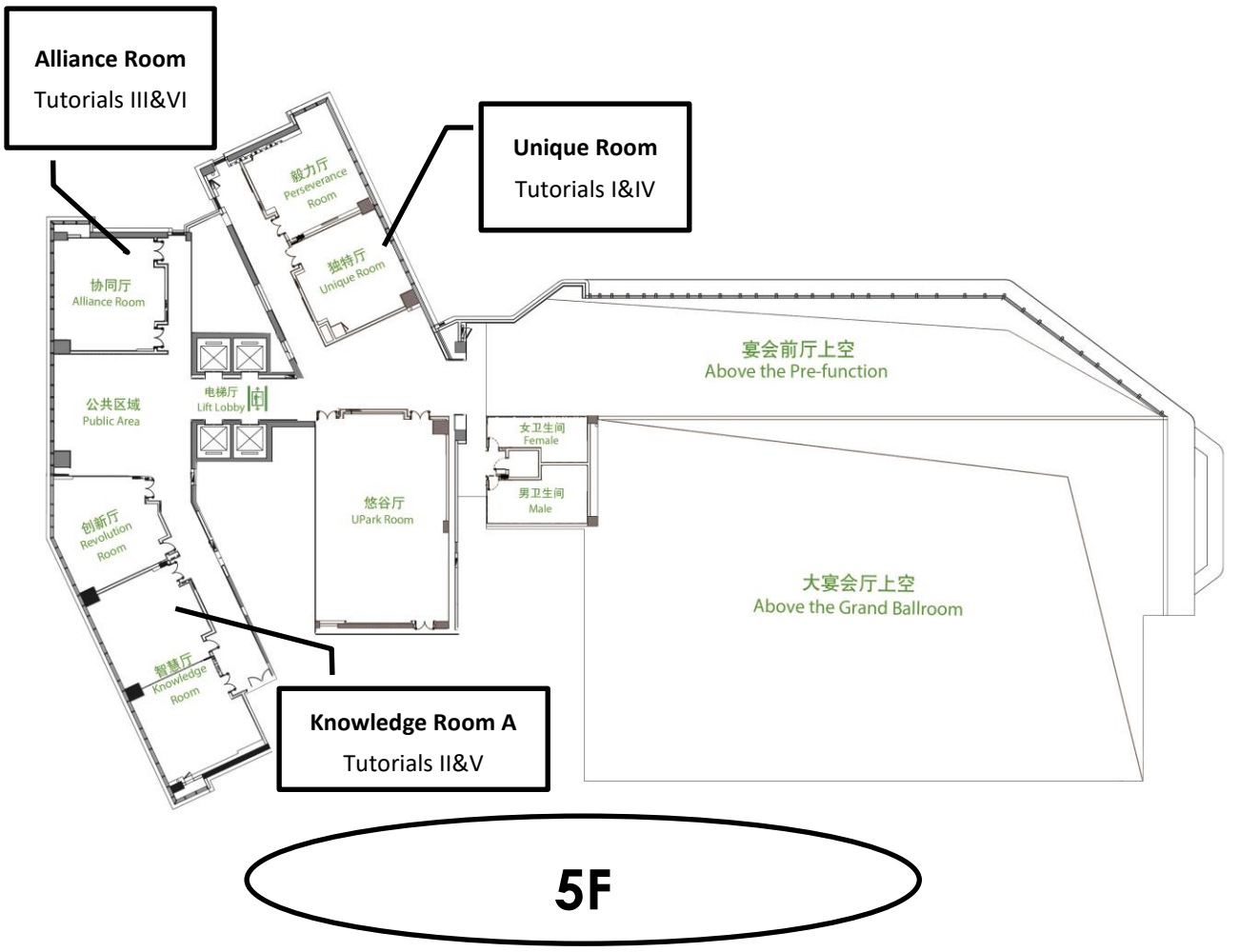
Negotiated price for SiPS 2019 participants:

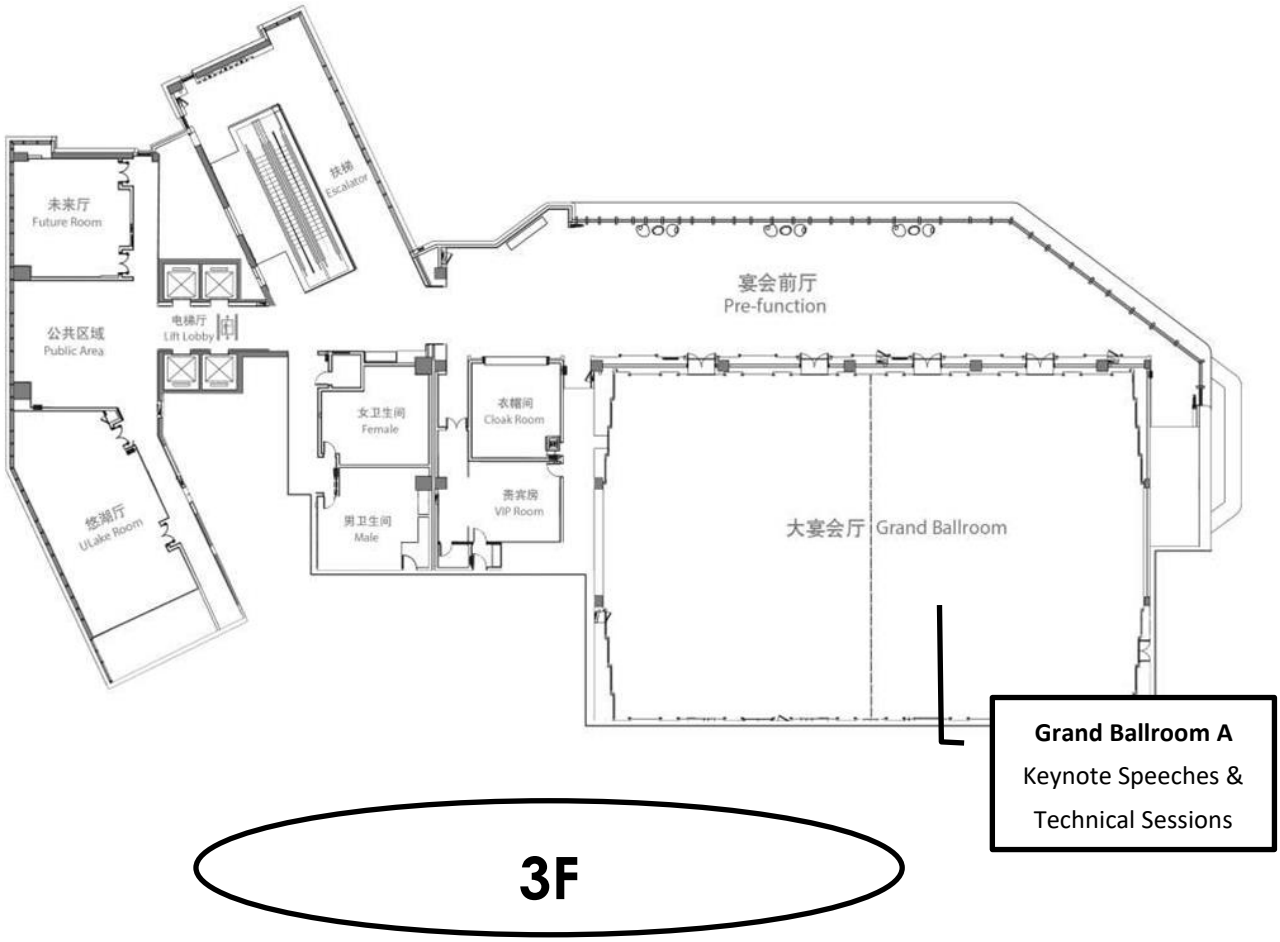
Standard Single Room: 490 CNY/night, including one breakfast, tax and service fee. Breakfast can be added at 60 CNY.

Standard Twin Room: 550 CNY/night, including two breakfasts, tax and service fee.

You are welcome to contact the hotel for room reservation by mentioning you are participating SiPS 2019, which is sponsored by Southeast University. If by email, please remark the room type, check-in and check-out dates in title, such as "SiPS 2019 Attendee, Single Room (check-in: 20/10/2019, check-out: 24/10/2019)".

地铁3号线秣周东路站 2号出口 距离：100米 行程：步行2分钟	Metro Line 3—Mozhou East Road Station, Exit 2 Distance : 100m Range : 2mins by walking
南京南站 距离：15公里 行程：地铁20分钟 驾车20分钟	Nanjing South Railway Station Distance : 15km Range : 20mins by Metro 20mins by driving
南京禄口国际机场 距离：19公里 行程：驾车20-25分钟	Nanjing Lukou International Airport Distance : 19km Range : 20-25mins by driving
南京站 距离：26公里 行程：地铁50分钟 驾车40分钟	Nanjing Railway Station Distance : 26km Range : 50mins by Metro 40mins by driving


Alliance Room
Tutorials III&VI
Unique Room
Tutorials I&IV
协同厅
Alliance Room
毅力厅
Perseverance Room
独特厅
Unique Room
公共区域
Public Area
电梯厅
Lift Lobby
宴会前厅上空
Above the Pre-function
女卫生间
Female
男卫生间
Male
悠谷厅
UPark Room
创新厅
Revolution Room
大宴会厅上空
Above the Grand Ballroom
智慧厅
Knowledge Room
Knowledge Room A
Tutorials II&V
5F


未来厅
Future Room
扶梯
Escalator
宴会前厅
Pre-function
公共区域
Public Area
电梯厅
Lift Lobby
女卫生间
Female
衣帽间
Cloak Room
贵宾房
VIP Room
悠湖厅
ULake Room
男卫生间
Male
大宴会厅 Grand Ballroom
Grand Ballroom A
Keynote Speeches &
Technical Sessions
3F

AGENDA OVERVIEW

OCTOBER 20, SUNDAY | SIGN-IN

09:00-20:00	**Sign-in & Materials Collection** @Hotel Lobby (1F)

OCTOBER 20, SUNDAY | TUTORIALS

09:00-12:00	**Tutorial Session | Morning**		
	Tutorial I @Unique Room 独特厅 (5F)	Tutorial II @Knowledge Room A 智慧厅 A (5F)	Tutorial III @Alliance Room 协同厅 (5F)
12:00-14:00	**Lunch** *--only served for tutorials participants* @Qinhuai Amazing Chinese Restaurant 尚秦淮中餐厅 (2F)		
14:00-17:00	**Tutorial Session | Afternoon**		
	Tutorial IV @Unique Room 独特厅 (5F)	Tutorial V @Knowledge Room A 智慧厅 A (5F)	Tutorial VI @ Alliance Room 协同厅 (5F)
18:00-20:00	**Welcome Reception** @Peppers All Day Dining Restaurant 百香西餐厅 (1F)		

Tutorial I

Title: Artificial Intelligence, Machine Learning, and Statistical Signal Processing in Financial Technology (FinTech)

Tutorial II

Title: Optical Mobile Communications

Tutorial III

Title: Internet of Things (IoT): Signals, Communications, Applications, Challenges, and Future Research

Tutorial IV

Title: Generative Adversarial Network and its Applications to Speech Signal and Natural Language Processing

Tutorial V

Title: Making Healthcare More Accessible via AI: Extension of Telemedicine

Tutorial VI

Title: Tensor Subspace Analysis in Signal Processing

AGENDA OVERVIEW

OCTOBER 21, MONDAY | MORNING

08:00-08:25	**Opening Ceremony |** *chaired by Prof. Chuan Zhang, Southeast University* **& Group Photo** @Grand Ballroom A 大宴会厅 A (3F)
08:25-09:25	**Keynote Speech I |** Machine Learning and Deep Learning Systems: Low-Energy VLSI Architectures and Applications *by Prof. Keshab K. Parhi, University of Minnesota, USA* @Grand Ballroom A 大宴会厅 A (3F)
09:25-09:40	**Coffee Break** @Pre-function 宴会前厅 (3F)
09:40-11:00	**Session M1 |** Special Session I-Practical Machine-Learning-Aided Communications Systems @Grand Ballroom A 大宴会厅 A (3F)
11:00-12:40	**Session M2 |** Machine Learning I @Grand Ballroom A 大宴会厅 A (3F)
12:40-14:00	**Lunch** @Peppers All Day Dining Restaurant 百香西餐厅 (1F)

Session M1 | Oral Presentations

#99 Neural Dynamic Successive Cancellation Flip Decoding of Polar Codes

#100 Deep Unfolding for Communications Systems: A Survey and Some New Directions

#110 A Channel-Blind Decoding for LDPC Based on Deep Learning and Dictionary Learning

#111 Structured Neural Network with Low Complexity for MIMO Detection

#112 Design and Implementation of a Neural Network Based Predistorter for Enhanced Mobile Broadband

Session M2 | Oral Presentations

#16 Neural Network-based Vehicle Image Classification for IoT Devices

#31 Learning to Design Constellation for AWGN Channel Using Auto-Encoders

#54 PRESS/HOLD/RELEASE Ultrasonic Gestures and Low Complexity Recognition Based on TCN

#15 Joint Image Deblur and Poisson Denoising based on Adaptive Dictionary Learning

#10 Semantic Segmentation of Retinal Vessel Images Via Dense Convolution and Depth Separable Convolution

AGENDA OVERVIEW

OCTOBER 21, MONDAY | AFTERNOON

14:00-15:40	**Session M3 |** Signal Processing @Grand Ballroom A 大宴会厅 A (3F)
15:40-16:30	**Poster Session I & Coffee Break** @ Pre-function 宴会前厅 (3F)
16:30-18:30	**Session M4 |** Polar Codes @Grand Ballroom A 大宴会厅 A (3F)
18:30-20:30	**Dinner** @Peppers All Day Dining Restaurant 百香西餐厅 (1F)

Session M3 | *Oral Presentations*

#84 Co-Design of Sparse Coding and Dictionary Learning for Real-Time Physiological Signals Monitoring

#67 A Data-driven Approach to Vibrotactile Data Compression

#86 A Low-Latency and Low-Complexity Hardware Architecture for CTC Beam Search Decoding

#40 A Root-RARE Algorithm for DOA Estimation with Partly Calibrated Uniform Linear Arrays

#37 Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection

Poster Session I | *Poster Presentations*

#1 CLA Formula Aided Fast Architecture Design for Clustered Look-Ahead Pipelined IIR Digital Filter

#2 SIR Beam Selector for Amazon Echo Devices Audio Front-End

#7 AdaBoost-assisted Extreme Learning Machine for Efficient Online Sequential Classification

#22 An FMCW Ranging Method with Identification Ability

#25 A Novel Approach to Angle-of-Arrival Estimation Based on Layered Ensemble Learning

#45 EAGLE: Exploiting Essential Address in Both Weight and Activation to Accelerate CNN Computing

#46 Sub-Spectrogram Segmentation for Environmental Sound Classification via Convolutional Recurrent Neural Network and Score Level Fusion

#83 On Secrecy Energy Efficiency of RF Energy Harvesting System

Session M4 | *Oral Presentations*

#65 Generation of Efficient Self-adaptive Hardware Polar Decoders using High-Level Synthesis

#79 DynExit: Dynamic Early-Exit Strategy in Deep Residual Networks

#73 Pipelined Implementations for Belief Propagation Polar Decoder: From Formula to Hardware

#28 Flexible and Simplified Multi-bit Successive-Cancellation List Decoding for Polar Codes

#77 Molecular Polar Belief Propagation Decoder and Successive Cancellation Decoder

#55 A Secure and Robust Key Generation Method Using Physical Unclonable Functions and Polar Codes

AGENDA OVERVIEW

OCTOBER 22, TUESDAY | MORNING

08:20-09:20	**Keynote Speech II |** Cognitive Computing on Heterogeneous Hardware Systems for the AI Revolution *by Prof. Deming Chen, University of Illinois at Urbana-Champaign, USA* @Grand Ballroom A 大宴会厅 A (3F)
09:20-09:40	**Coffee Break** @ Pre-function 宴会前厅 (3F)
09:40-11:00	**Session T1 |** Special Session II-Emerging Computing Paradigms for Signal Processing and Smart Learning @Grand Ballroom A 大宴会厅 A (3F)
11:00-12:20	**Session T2 |** Machine Learning II @Grand Ballroom A 大宴会厅 A (3F)
12:40-14:00	**Lunch** @Peppers All Day Dining Restaurant 百香西餐厅 (1F)

Session T1 | *Oral Presentations*

#97 Design and Evaluation of a Power-Efficient Approximate Systolic Array Architecture for Matrix Multiplication

#101 Ensemble Neural Network Method for Wind Speed Forecasting

#105 A Survey of Computation-Driven Data Encoding

#106 Parallel Convolutional Neural Network (CNN) Accelerators Based on Stochastic Computing

#109 A Data Structure-based Approximate Belief Propagation Decoder for Polar Codes

Session T2 | *Oral Presentations*

#82 RNN Models for Rain Detection

#91 Improving Reliability of ReRAM-based DNN Implementation through Novel Weight Distribution

#49 Exploration of On-device End-to-End Acoustic Modeling with Neural Networks

#52 Memory Reduction Through Experience Classification for Deep Reinforcement Learning with Prioritized Experience Replay

AGENDA OVERVIEW

OCTOBER 22, TUESDAY | AFTERNOON

14:00-15:40	**Session T3 |** MIMO @Grand Ballroom A 大宴会厅 A (3F)
15:40-16:30	**Poster Session II & Coffee Break** @ Pre-function 宴会前厅 (3F)
16:30-18:10	**Session T4 |** Special Session III- Hardware Security and Hardware Implementation of Emerging Cryptographic Primitives @Grand Ballroom A 大宴会厅 A (3F)
18:50-21:00	**Gala Dinner** @Grand Ballroom A 大宴会厅 A (3F)

Session T3 | *Oral Presentations*

#64 Lattice-Reduction-Aided Symbol-Wise Intra-Iterative Interference Cancellation Detector for Massive MIMO System

#27 A Distributed Detection Algorithm For Uplink Massive MIMO Systems

#20 Hybrid Preconditioned CG Detection with Sequential Update for Massive MIMO Systems

#5 Pilot-assistant methods for channel estimation in MIMO-V-OFDM systems

#41 A Unified and Flexible Eigen-Solver for Rank-Deficient Matrix in MIMO Precoding/Beamforming Applications

Poster Session II | *Poster Presentations*

#3 Modified Complementary Joint Sparse Representations: a novel post-filtering to MVDR beamforming

#6 FPGA Prototyping of A Millimeter-Wave Multiple Gigabit WLAN System

#34 An ISAR imaging algorithm based on RPCA for micro-Doppler effect suppression

#44 Data Driven Low-Complexity DOA Estimation for Ultra-Short Range Automotive Radar

#56 Towards Algebraic Modeling of GPU Memory Access for Bank Conflict Mitigation

#88 Nonlinear Functions in Learned Iterative Shrinkage-Thresholding Algorithm for Sparse Signal Recovery

Session T4 | *Oral Presentations*

#102 Dynamic Reconfigurable PUFs based on FPGA

#98 Side Channel Attack Resistant AES Design Based on Finite Field Construction Variation

#108 Ultra-Fast Modular Multiplication Implementation for Isogeny-Based Post-Quantum Cryptography

#107 An Efficient Polynomial Multiplier Architecture for the Bootstrapping Algorithm in A Fully Homomorphic Encryption Scheme

#103 Theoretical Analysis of Configurable RO PUFs and Strategies to Enhance Security

AGENDA OVERVIEW

OCTOBER 23, WEDNESDAY | MORNING

08:20-09:20	**Keynote Speech III |** Wireless Signal Processing at AI Era *by Sunny Zhang, Director of Communication Infrastructure Research, Intel Labs China, China* @Grand Ballroom A 大宴会厅 A (3F)
09:20-09:40	**Coffee Break** @ Pre-function 宴会前厅 (3F)
09:40-11:00	**Session W1 |** Special Session IV-Improving the Performance of Autonomous Systems: Algorithm, Hardware and Application @Grand Ballroom A 大宴会厅 A (3F)
11:00-12:00	**Session W2 |** Error-Correcting Codes @Grand Ballroom A 大宴会厅 A (3F)
12:00-13:30	**Lunch** @Peppers All Day Dining Restaurant 百香西餐厅 (1F)

Session W1 | *Oral Presentations*

#113 Efficiently Learning a Robust Self-Driving Model with Neuron Coverage Aware Adaptive Filter Reuse

#114 Autonomous UAV with Learned Trajectory Generation and Control

#115 A Hybrid GPU + FPGA System Design for Autonomous Driving Cars

#116 Accurate Congenital Heart Disease Model Generation for 3D Printing

Session W2 | *Oral Presentations*

#93 AVX-512 Based Software Decoding for 5G LDPC Codes

#32 A Low-Complexity Error-and-Erasure Decoding Algorithm for t=2 RS Codes

#80 A New Inversionless Berlekamp–Massey Algorithm with Efficient Architecture

OCTOBER 23, WEDNESDAY | AFTERNOON

13:30-14:30	**Academic Visit** Purple Mountain Laboratories *(every SIPS participant can join in)*
13:30-19:30	**Half Day Tour** A blessing trip to Niushou Mountain *(for prior sign-ups only)*

KEYNOTE SPEAKERS

OCTOBER 21 | MONDAY | 08:25-09:25

Prof. Keshab K. Parhi, IEEE Fellow, AAAS Fellow
University of Minnesota, USA

Bio: Keshab K. Parhi received the B.Tech. degree from the Indian Institute of Technology (IIT), Kharagpur, in 1982, the M.S.E.E. degree from the University of Pennsylvania, Philadelphia, in 1984, and the Ph.D. degree from the University of California, Berkeley, in 1988. He has been with the University of Minnesota, Minneapolis, since 1988, where he is currently Distinguished McKnight University Professor and Edgar F. Johnson Professor of Electronic Communication in the Department of Electrical and Computer Engineering. He has published 650 papers, is the inventor of 30 patents, and has authored the textbook VLSI Digital Signal Processing Systems (Wiley, 1999) and coedited the reference book Digital Signal Processing for Multimedia Systems (Marcel Dekker, 1999). His current research addresses VLSI architecture design of machine learning systems, hardware security, data-driven neuroscience and molecular/DNA computing. Dr. Parhi is the recipient of numerous awards including the 2017 Mac Van Valkenburg award and the 2012 Charles A. Desoer Technical Achievement award from the IEEE Circuits and Systems Society, the 2004 F. E. Terman award from the American Society of Engineering Education, the 2003 IEEE Kiyo Tomiyasu Technical Field Award, the 2001 IEEE W. R. G. Baker prize paper award, and a Golden Jubilee medal from the IEEE Circuits and Systems Society in 1999. He served as the Editor-in-Chief of the IEEE Trans. Circuits and Systems, Part-I during 2004 and 2005. He was elected a Fellow of IEEE in 1996 and a Fellow of the American Association for Advancement of Science (AAAS) in 2017.

TALK ON

Machine Learning and Deep Learning Systems: Low-Energy VLSI Architectures and Applications

Abstract: Machine learning and data analytics continue to expand the fourth industrial revolution and affect many aspects of our lives. This talk will explore machine learning applications in data-driven neuroscience, and low-energy implementations of machine learning and deep learning systems. Data-driven neuroscience can exploit machine learning approaches including deep learning to generate hypotheses associated with biomarkers for specific neuro-psychiatric disorders. In the first part, I will talk about use of machine learning to find biomarkers for epilepsy. In the second part of the talk, I will talk about approaches for energy-efficient implementations for both traditional machine learning and deep learning systems. I will talk about the roles of feature ranking and incremental-precision approaches in reducing energy consumption of traditional machine learning systems. I will then talk about reducing energy consumption in deep learning systems. I will describe our recent work on Perm-DNN based on permuted-diagonal interconnections in deep convolutional neural networks and how structured sparsity can reduce energy consumption associated with memory access in these systems.

KEYNOTE SPEAKERS

OCTOBER 22 | TUESDAY | 08:20-09:20

Prof. Deming Chen, IEEE Fellow

University of Illinois at Urbana-Champaign, USA

Bio: Dr. Deming Chen obtained his BS in computer science from University of Pittsburgh, Pennsylvania in 1995, and his MS and PhD in computer science from University of California at Los Angeles in 2001 and 2005 respectively. He joined the ECE department of University of Illinois at Urbana-Champaign in 2005 and has been a full professor in the same department since 2015. His current research interests include system-level and high-level synthesis, machine learning, GPU and reconfigurable computing, and hardware security. He has given more than 110 invited talks sharing these research results worldwide. He obtained the Arnold O. Beckman Research Award from UIUC in 2007, the NSF CAREER Award in 2008, and eight Best Paper Awards. He also received the ACM SIGDA Outstanding New Faculty Award in 2010, and IBM Faculty Award in 2014 and 2015. In 2017 and 2019 respectively, he led a team to win the first place of DAC International System Design Contest in the IoT domain. He is included in the List of Teachers Ranked as Excellent in 2008 and 2017. He is the Donald Biggar Willett Faculty Scholar of College of Engineering, an IEEE Fellow, an ACM Distinguished Speaker, and the Editor-in-Chief of ACM Transactions on Reconfigurable Technology and Systems (TRETS). He is also involved with several startup companies, including co-founding Inspirit IoT, Inc. in 2016.

TALK ON

Cognitive Computing on Heterogeneous Hardware Systems for the AI Revolution

Abstract: Many envision that AI (artificial intelligence) will usher in the next iteration of technology revolution, where humans and machines will work side-by-side to augment, enhance, or accelerate our ability to analyze, learn, create, and think. There are successful stories emerging fast already, such as IBM Watson, Microsoft HoloLens, and Google AlphaGo. One essential component to enable the new AI revolution is IoT (Internet of Things). Cognitive computing can learn from the rich IoT data, reason from models, and most importantly interact with us to perform complex tasks (ranging from healthcare to education to financial services) better than either humans or machines can do by themselves. Meanwhile, high-performance computing would be of paramount importance to help achieve the grand vision of cognitive computing. In this talk, Prof. Chen will share his recent research results on machine learning, reconfigurable computing, GPU computing, and cognitive application benchmarking. He will also present his recent work on extremely fast software and hardware modeling and the automated software/hardware co-design for accelerating cognitive computing workloads. Compelling AI applications will be introduced as well, such as autonomous driving and facial recognition.

KEYNOTE SPEAKERS

OCTOBER 23 | WEDNESDAY | 08:20-09:20

Sunny Zhang, Principle Engineer

Director of Communication Infrastructure Research, Intel Labs China

Bio: As principle engineer and director of communication infrastructure research in Intel labs China, Sunny is 15 years veteran on wireless communication system and wireless signal processing design. Sunny initialized and led Intel multi-radio coexistence, digital enhanced radio and Cloud Radio Access Network research effort, developed the first highly optimized LTE stacks on general purpose architecture, the first scalable and high performance Cloud RAN BBU pool reference design, which became Intel FlexRAN 4G and 5G product, widely used by cellular industry. Sunny leads Intel Lab China's 5G research on Radio Access Network architecture, proposed 5G low layer split massive MIMO solutions, widely adopted by industry. Sunny also leads Intel lab China team designing high performance programmable wireless processor to achieve efficiency and programmability at the same time. Sunny and his team received 1 Intel lab Gordy Award, and Intel Achievement Award in Intel, which are the highest award in Intel lab and Intel.

Prior joined Intel, Sunny was as the physical layer architect, designed the first passive optical network in 2001. From 2002 to 2004, he was at startup companies in Silicon Valley, and responsible for optical devices and RFID devices design.

Sunny got bachelor and master degree in Tsinghua University and Beijing University of post and telecom communication respectively. Published 10+ papers and filed 10+ patents.

TALK ON

Wireless Signal Processing at AI Era

Abstract: About 8 years ago, there were one paper asked the question and talked about "is PHY research dead?", of course, you can imagine that the conclusion from a wireless researcher definitely will be not, the paper also proposed quite a few innovation directions, such as short code, impairments of mmWave which are actually the key features of 5G now, which also talked about "wireless research is not as vibrate as before", "implementation" should be taken more attentions. 8 year later, if we revisit this question, we see the trend that wireless system become extremely complex on the standard and also implementation, industry wants to re-shape the wireless eco-system with open architecture and interface, power efficiency and higher frequency became the focal point of wireless implement, for signal processing side, AI technology is emerging which become another toolbox for wireless research but the directions are not clear on what can be done and what should done. In this talk, will talk about the problem in wireless system, where AI techniques can be applied such as on channel estimation, on MAC scheduler, on massive MIMO beamforming, the possibility and challenge to bring that into reality, as well as from design perspective, how could we apply wireless signal processing + AI processing together to solve practical problem on radar sensing, how can we combine wireless processing and AI processing together on same processor.

TIPS FOR PARTICIPANTS

ORAL PRESENTATION

- The duration of a presentation slot is ab. 20 minutes. Please target your lecture for a duration of about 17 minutes for the presentation plus about 3 minutes for questions from the audience.
- An LCD projector & computer will be available in every session room for regular presentations.
- Presentations MUST be uploaded at the session room at least 15 minutes before the session starts.
- The presentation software used at the session room supports the projection of the Windows: Office Power Point 2010 – .ppt .pptx, Office Excel, PDF.
- It is absolutely necessary that all files (audio / video / images) be stored in the same folder as the presentation.
- Please embed all fonts in your PowerPoint file and bring a backup PDF-version of your presentation.
- We suggest to include images in .gif, .jpg or .png format, in order to have light weight presentations.
- We suggest not to exceed 100MB (movies excluded) for each presentation.
- Movies created with proprietary codecs from professional machines, will not be viewable.

POSTER PRESENTATION

- We expect that at least one author stands by the poster for (most of the time of) the duration of the poster session. This is essential to present your work to anyone interest into it.
- If you wish to leave your poster to attend a lecture, please leave a message on the board to inform visitors.
- Posters should be set-up at least 15 minutes before the session starts and removed at the end of the session. Left behind posters at the end of the session will be disposed of.
- The size of poster is *A0 size (1189mm x 841mm, height > width) in Portrait mode.
- Materials to hang the poster will be made available on-site. Our staff will be available onsite to assist you.

SECURITY REMINDER

- Please keep all your belongings (laptop and camera etc.) with you in the public places, buses, metro; never leave them unattended in the conference rooms or hallways.
- You should wear the event card to access to each technical sessions and meals.
- Make sure you remove your USB each time after you use it in the conference laptops.

Session M1: Special Session I-Practical Machine-Learning-Aided Communications Systems

Session Chair: Alexios Balatsoukas-Stimming, Eindhoven University of Technology, Netherlands

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 09:40-11:00

09:40-9:56 | ID#99

Neural Dynamic Successive Cancellation Flip Decoding of Polar Codes

Nghia Doan, Seyyed Ali Hashemi, Furkan Ercan, Thibaud Tonnellier, Warren J. Gross

McGill University, Canada

Abstract: Dynamic successive cancellation flip (DSCF) decoding of polar codes is a powerful algorithm that can achieve the error correction performance of successive cancellation list (SCL) decoding, with a complexity that is close to that of successive cancellation (SC) decoding at practical signal-to-noise ratio (SNR) regimes. However, DSCF decoding requires costly transcendental computations which adversely affect its implementation complexity. In this paper, we first show that a direct application of common approximation schemes on the conventional DSCF decoding results in significant error-correction performance loss. We then introduce a training parameter and propose an approximation scheme which completely removes the need to perform transcendental computations in DSCF decoding, with almost no error-correction performance degradation.

09:56-10:12 | ID#100

Deep Unfolding for Communications Systems: A Survey and Some New Directions

Alexios Balatsoukas-Stimming and Christoph Studer

Eindhoven University of Technology, Netherlands

Abstract: Deep unfolding is a method of growing popularity that fuses iterative optimization algorithms with tools from neural networks to efficiently solve a range of tasks in machine learning, signal and image processing, and communication systems. This survey summarizes the principle of deep unfolding and discusses its recent use for communication systems with focus on detection and precoding in multi-antenna (MIMO) wireless systems and belief propagation decoding of error-correcting codes. To showcase the efficacy and generality of deep unfolding, we describe a range of other tasks relevant to communication systems that can be solved using this emerging paradigm. We conclude the survey by outlining a list of open research problems and future research directions.

10:12-10:28 | ID#110

A Channel-Blind Decoding for LDPC Based on Deep Learning and Dictionary Learning

Xu Pang, Chao Yang, Zaichen Zhang, **Xiaohu You**, and Chuan Zhang

Southeast University, China

Abstract: Low-density parity-check (LDPC) codes are a class, of linear block codes which provide near capacity performance, on a large collection of data transmission and storage channels., Existing decoders such as Min-sum decoding methods of belief, propagation (BP) are one-dimensional and require precise channel, estimation. This paper

proposes a blind decoder from twodimensional, perspective. Main work involves pattern construction, and pre-filtering. In this paper, a 8_4 Sudoku template is applied, for the pattern construction of two-dimensional LDPC signal., Dictionary learning is adopted for global pre-filtering. Image, training is adopted in BP to further reduce the environment noise., Moreover, the proposed network architecture can denoise and, decode without dimensional transformation. Numerical results, show that two-dimensional blind detection can compensate the, performance significantly when channel estimation is not perfect, and the combination of dictionary learning and image training, has a great improvement in performance and data size reduction.

10:28-10:44 | ID#111

Structured Neural Network with Low Complexity for MIMO Detection

Siyu Liao, Chunhua Deng, Lingjia Liuy and Bo Yuan

Rutgers University, United States

Abstract: Neural network has been applied into MIMO detection problem and has achieved the state-of-the-art performance. However, it is hard to deploy these large and deep neural network models to resource constrained platforms. In this paper, we impose the circulant structure inside neural network to generate a low complexity model for MIMO detection. This method can train the circulant structured network from scratch or convert from an existing dense neural network model. Experiments show that this algorithm can achieve half the model size with negligible performance drop.

10:44-11:00 | ID#112

Design and Implementation of a Neural Network Based Predistorter for Enhanced Mobile Broadband

Chance Tarver, Alexios Balatsoukas-Stimming and Joseph R. Cavallaro

Eindhoven University of Technology, Netherlands

Abstract: Digital predistortion is the process of using digital signal processing to correct nonlinearities caused by the analog RF front-end of a wireless transmitter. These nonlinearities contribute to adjacent channel leakage, degrade the error vector magnitude of transmitted signals, and often force the transmitter to reduce its transmission power into a more linear but less power-efficient region of the device. Most predistortion techniques are based on polynomial models with an indirect learning architecture which have been shown to be overly sensitive to noise. In this work, we use neural network based predistortion with a novel neural network training method that avoids the indirect learning architecture and that shows significant improvements in both the adjacent channel leakage ratio and error vector magnitude. Moreover, we show that, by using a neural network based predistorter, we are able to achieve a 42% reduction in latency and 9.6% increase in throughput on an FPGA accelerator with 15% fewer multiplications per sample when compared to a similarly performing memory-polynomial implementation.

Session M2: Machine learning I

Session Chair: Ching Te Chiu, National Tsing Hua University, Taiwan

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 11:00-12:40

11:00-11:20 | ID#16

Neural Network-based Vehicle Image Classification for IoT Devices

Saman Payvar, Mir Khan, Rafael Stahl, Daniel Mueller-Gritschneder,Jani Boutellier and **Luca Ferranti**

Tampere University, Finland

Abstract: Convolutional Neural Networks (CNNs) have previously provided unforeseen results in automatic image analysis and interpretation, an area which has numerous applications in both consumer electronics and industry. However, the signal processing related to CNNs is computationally very demanding, which has prohibited their use in the smallest embedded computing platforms, to which many Internet of Things (IoT) devices belong. Fortunately, in the recent years researchers have developed many approaches for optimizing the performance and for shrinking the memory footprint of CNNs. This paper presents a neural network- based image classifier that has been trained to classify vehicle images into four different classes. The neural network is optimized by a technique called binarization, and the resulting binarized network is placed to an IoT-class processor core for execution. Binarization reduces the memory footprint of the CNN by around 95% and increases performance by more than 6X. Furthermore, we show that by utilizing a custom instruction 'pop count' of the processor, the performance of the binarized vehicle classifier can still be increased by more than 2X, making the CNN-based image classifier suitable for the smallest embedded processors.

11:20-11:40 | ID#31

Learning to Design Constellation for AWGN Channel Using Auto-Encoders

Qisheng Huang, Ming Jiang and Chunming Zhao

Southeast University, China

Abstract: This paper proposes a novel constellation design in AWGN channel through learning based auto-encoder (AE). Additionally, this paper illustrates the reason why learning based constellation has better performance than the classical square-shaped QAM design by analyzing the Euclidean distance distribution and the bound of symbol error rate between learning designed symbols and other constellations. Moreover, the performance of learning based constellation will be compared to constellation based on convex optimization design. To solve the bit mapping problem of the learning based constellation, Q-ary LDPC encoding is applied to these specifically designed QAM modulation systems, where the soft decoding of Q-ary LDPC codes can be carried out with the symbol-level soft outputs of demodulation.

11:40-12:00 | ID#54

PRESS/HOLD/RELEASE Ultrasonic Gestures and Low Complexity Recognition Based on TCN

Emad A. Ibrahim, Min Li and Jose Pineda de Gyvez

Eindhoven University of Technology, Netherlands

Abstract: Targeting ultrasound-based gesture recognition, this paper proposes a new universal PRESS/HOLD/RELEASE approach that leverages the diversity of gestures performed on smart devices such as mobile phones and IoT nodes. The new set of gestures are generated by interleaving PRESS/HOLD/RELEASE patterns; abbreviated as P/H/R, with gestures like sweeps between a number of microphones. P/H/R patterns are constructed by a hand as it approaches a top of a microphone to generate a virtual Press. After that, the hand settles for an undefined period of time to generate a virtual Hold and finally departs to generate a virtual Release. The same hand can sweep to a 2nd microphone and perform another P/H/R. Interleaving the P/H/R patterns expands the number of performed gestures. Assuming an on-board speaker transmitting ultrasonic signals, the detection is performed on Doppler shift readings generated by a hand as it approaches and departs a top of a microphone. The Doppler shift readings are presented in a sequence of down-mixed ultrasonic spectrogram frames. We train a Temporal Convolutional Network (TCN) to classify the P/H/R patterns under different environmental noises. Our experimental results show that such P/H/R patterns at a top of a microphone can be achieved with 96.6% accuracy under different noise conditions. A group of P/H/R based gestures has been tested on commercially off-the-shelf (COTS) Samsung Galaxy S7 Edge. Different P/H/R interleaved gestures (such as sweeps, long taps, etc.) are designed using two microphones and a single speaker while using as low as ~ 5K parameters and as low as ~0.15 Million operations (MOPs) in compute power per inference. The P/H/R interleaved set of gestures are intuitive and hence are easy to learn by end users. This paves its way to be deployed by smartphones and smart speakers for mass production.

12:00-12:20 | ID#15

Joint Image Deblur and Poisson Denoising based on Adaptive Dictionary Learning

Xiangyang Zhang, **Hongqing Liu,** Zhen Luo and Yi Zhou

Chongqing University of Posts and Telecommunications, China

Abstract: This paper describes a blind image reconstruction algorithm for blurred image under Poisson noise. To that aim, in this work, the group sparse domain is explored to sparsely represent the image and blur kernel, and then l1-norm is utilized to enforce the sparse solutions. In doing so, a joint optimization framework is developed to estimate the blur kernel matrix while removing Poisson noise. To effectively solve the developed optimization, a two-step iteration scheme involving two sub-problems is proposed. For each subproblem, the alternating direction method of multipliers (ADMM) algorithm is devised to estimate the blur or denoise. The experimental simulations demonstrate that the proposed algorithm is superior to other approaches in terms of restoration quality and performance metrics.

12:20-12:40 | ID#10

Semantic Segmentation of Retinal Vessel Images Via Dense Convolution and Depth Separable Convolution

Zihui Zhu, Hengrui Gu, Zhengming Zhang, **Yongming Huang,** and Luxi Yang

Southeast University, China

Abstract: Semantic segmentation of retinal vessel images is of great value for clinical diagnosis. Due to the complex information of retinal vessel features, the existing algorithms have problems such as discontinuities of segmented

vessels. To achieve better semantic segmentation results, we propose an encoder-decoder structure combined with dense convolution and depth separable convolution. Firstly, the images are enhanced by extracting the original green channel, limiting contrast histogram equalization and sharpening, then data argumentation is performed to expand the data set. Secondly, the processed images are trained by the proposed network using a weighted loss function. Finally, the test images are segmented by the trained model. The proposed algorithm is tested on the DRIVE data set, and its average accuracy, sensitivity and specificity reached 96.83%, 83.71%, and 98.95%, respectively.

Lunch @ Peppers All Day Dining Restaurant 百香西餐厅（1F）
| 12:40-14:00

Session M3: Signal Processing

Session Chair: Yingjie Lao, Clemson University, USA

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 14:00-15:40

14:00-14:20 | ID#84

Co-Design of Sparse Coding and Dictionary Learning for Real-Time Physiological Signals Monitoring

Kuan-Chun Chen, Ching-Yao Chou and An-Yeu (Andy) Wu

National Taiwan University, Taiwan

Abstract: Compressive sensing (CS) is a novel technique to reduce overall transmission power in wireless sensors. For physiological signals telemonitoring of wearable devices, chip area and power efficiency need to be considered simultaneously. There are many prior studies aim to develop algorithms that applied to CS reconstruction chips with reconfigurable architecture. However, representative dictionaries are also important when these CS reconstruction chips are verified in real-time physiological signals monitoring tasks. That is, a more representative dictionary can not only enhance the reconstruction performance of these chips but also alleviate memory overhead. In this paper, we apply the concept of co-design between sparse coding algorithms and learned dictionaries. We also explore the representativeness and compatibility of each learned dictionary. In addition, the computational complexity of each reconstruction algorithm is provided through simulations. Our results show that the dictionaries trained by fast iterative shrinkage-thresholding algorithm (FISTA) are more representative according to the quality of reconstruction for physiological signals monitoring. Besides, FISTA reduces more than 90% of the computational time compared with other hardware-friendly reconstruction algorithms

14:20-14:40 | ID#67

A Data-driven Approach to Vibrotactile Data Compression

Xun Liu and Mischa Dohler

King's College London, United Kingdom

Abstract: The emerging Internet of Skills that exchanges tactile and other sensorial data, significantly augments traditional multimedia. The increase of data scale and modalities demands for codecs dedicated to these sensorial data. In this paper, we propose a codec for compression of vibrotactile data in the spirit of Weber's law. To be specific, a companding function is applied to the vibrotactile data, so that the quantisation step of high amplitude is larger than that of low amplitude. The curve of the companding function is optimised through a data-driven approach. To evaluate the performance of the vibrotactile codec in terms of human perceived quality, rigorous subjective tests are conducted. The results demonstrate that 75% compression of vibrotactile data is achieved without perceivable degradation. More importantly, the computational complexity is much lower and the latency performance is superior, compared with other vibrotactile codecs. The computational complexity of the proposed codec is about 1/20 of that of previous codecs, while the time delay is approximately 1/30 of that of previous codec.

14:40-15:00 | ID#86

A Low-Latency and Low-Complexity Hardware Architecture for CTC Beam Search Decoding

Siyuan Lu, Jinming Lu, Jun Lin, **Zhongfeng Wang** and Li Du

Nanjing University, Nanjing, China

Abstract: The recurrent neural networks (RNNs) along with connectionist temporal classification (CTC) have been widely used in many sequence to sequence tasks, including automatic speech recognition (ASR), lipreading, and scene text recognition (STR). In these systems, CTC-trained RNNs usually require specific CTC-decoders after their output layers. Many existing CTC-trained RNN inference systems use FPGA to do calculations of RNNs, and decode their outputs on CPU. However, with the development of FPGA-based RNN hardware accelerators, existing CPU-based CTC-decoder cannot meet the latency requirement of them. To resolve this issue, this paper proposes an efficient hardware architecture for the CTC beam search decoder based on the decoding method reported in our previous work. The experimental results show that the system latency per sample of the CTC-decoder is only 7.19us on Xilinx xc7vx1140tflg1930- 1 FPGA platform, which is lower than state-of-the-art RNNs. We also implement the origin algorithm on the same FPGA platform. Comparison results show that the improved one reduces the system latency per sample by 63.67%, the LUTRAMs by 97.44%, the FFs by 79.55%, and the DSPs by 50%. To the best of our knowledge, this is the first work on hardware implementation for CTC beam search decoder.

15:00-15:20 | ID#40

A Root-RARE Algorithm for DOA Estimation with Partly Calibrated Uniform Linear Arrays

Zhongchi Fang, Zheng Cao and Lan Wang

Jiangsu University, China

Abstract: The rank-reduction (RARE) algorithm is a wellknown class of algorithms for direction of arrival (DOA) estimation in the presence of imperfect array manifolds. Since the spectral peak search is inevitable for the current RARE algorithm, it may bring a huge occupational load for practical implementations. In order to reduce the computational complexity, in this paper, we propose a root-RARE algorithm for DOA estimation with partly calibrated uniform linear arrays (ULAs). Through replacing the spectral peak search with a polynomial root finding, our proposed method can get much higher efficiency than the original RARE method. Simulation results demonstrate that our method can significantly reduce the computational complexity and improve the DOA estimation performance in a low SNR case.

15:20-15:40 | ID#37

Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection

Cheng-Ping Hsieh, Yi-Ta Chen, Win-Ken Beh and An-Yeu (Andy) Wu

National Taiwan University, Taiwan

Abstract: Since stress has a strong influence on human's health, it is necessary to automatically detect stress in our daily life. In this paper, we aim to improve the performance and obtain the dominant features in stress detection based on Electrodermal Activity (EDA). Compared to the methods in Wearable Stress and Affect Dataset (WESAD), we propose several enhancements to get higher f1-scores, including less overlapped signal segmentation, more signal processing features, and extreme gradient boosting classification algorithm (XGBoost). Furthermore, we select dominant features according to their importance in classifier and correlation among other features while keeping high

performance. Experiment results show that with 9 dominant features in XGBoost, we can achieve 92.38% (+17.87%) and 89.92% (+14.58%) f1-scores compared to WESAD on chest- and wrist-based EDA signal respectively. The features we choose suggest that the magnitude of low frequency and the complexity of high frequency EDA signal contain the most significant information in stress detection.

Coffee Break @ Pre-function 宴会前厅| 15:40-16:30

Session M4: Polar Codes

Session Chair: Alexios Balatsoukas-Stimming, Eindhoven University of Technology, Netherlands

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 16:30-18:30

16:30-16:50 | ID#65

Generation of Efficient Self-adaptive Hardware Polar Decoders using High-Level Synthesis

Yann Delomier, **Bertrand Le Gal**, Jeremie Crenne and Christophe Jego

Bordeaux INP, Univ. Bordeaux, France

Abstract: Recent advances in 5G digital communication standard implementations advocate for the use of polar codes for the Enhanced Mobile Broad Band (EMBB) control channels. However, in many cases, implementing efficient hardware decoder over a short duration is very challenging. Specialized knowledge is required to facilitate testing, rapid design iterations and fast prototyping. In this paper, we present a model-based design methodology to generate efficient hardware SC polar decoders from high-level synthesis tools. The abstraction level flexibility is evaluated and generated decoders architectures are compared to competing approaches. It is shown that the fine-tuning of computation parallelism, bit width, pruning level and working frequency enable high throughput decoder designs with moderate hardware complexities. Decoding throughput between 10 to 310 Mbit/s and hardware complexity between 1,000 and 21,000 LUTs are reported for the generated architectures.

16:50-17:10 | ID#79

DynExit: Dynamic Early-Exit Strategy in Deep Residual Networks

Meiqi Wang, Jianqiao Mo, Jun Lin, **Zhongfeng Wang** and Li Du

Nanjing University, China

Abstract: Early-exit is a kind of technique to terminate a pre-specified computation at an early stage depending on the input samples and has been introduced to reduce energy consumption for Deep Neural Networks (DNNs). Previous early-exit approaches suffered from the burden of manually tuning early-exit loss-weights to find a good trade-off between complexity reduction and system accuracy. In this work, we first propose DynExit, a dynamic loss-weight modification strategy for ResNets, which adaptively modifies the ratio of different exit branches and searches for a proper spot for both accuracy and cost. Then, an efficient hardware unit for early-exit branches is developed, which can be easily integrated to existing hardware architectures of DNNs to reduce average computing latency and energy cost. Experimental results show that the proposed DynExit strategy can reduce up to 43.6% FLOPS compared to the state-of-the-arts approaches. On the other hand, it is able to achieve 1.2% accuracy improvement over the existing end-to-end fixed loss-weight training scheme with comparable computation reduction ratio. The proposed hardware architecture for DynExit is evaluated on the platform of Xilinx Zynq-7000 ZC706 development board. Synthesis results demonstrate that the architecture can achieve high speed with low hardware complexity. To the best of our knowledge, this is the first hardware implementation for early-exit techniques used for DNNs in open literature.

17:10-17:30 | ID#73

Pipelined Implementations for Belief Propagation Polar Decoder: From Formula to Hardware

Chao Ji, Zaichen Zhang, Xiaohu You, and **Chuan Zhang**

Southeast University, China

Abstract: This paper proposes a general design method for, belief propagation (BP) pipelined polar decoder. By associating, data flow graph (DFG) of polar encoder with factor graph (FG), of BP polar decoder, regular structure of FG helps to determine, the generation formula representing pipelined BP polar decoder., Using Python as a compiler, the generation formula is translated, into a series of synthesizable Verilog HDL files for various, code lengths and parallelisms. Considering the balance between, performance and cost, this formula-to-hardware design can be, extended to explore the design space, where we are able to make, tradeoffs according to specific application requirements. With, the evaluation of auto-generation system, implementation results, have shown that our design is reliable and practicable.

17:30-17:50 | ID#28

Flexible and Simplified Multi-bit Successive-Cancellation List Decoding for Polar Codes

Haojing Hu, **Rongke Liu** and Baoping Feng

Beihang University, China

Abstract: Polar codes, as the first channel code that can provably achieve the channel capacity, have received increasing attention these years. However, in practical application, the decoding of polar codes still has many aspects that need improvement. One of the key bottlenecks of polar codes decoding is the high latency of SC (Successive Cancellation)-based decoding algorithms. As one of the solutions to this problem, many SCL (Successive Cancellation List) decoding algorithms with the multi-bit decision are proposed. Despite of the reduction of decoding latency, the complexity spent for computation and sort of path candidates of these algorithms has significantly increased in contrast with the conventional SCL algorithm. In this paper, we propose a novel SCL decoding algorithm with multi-bit decision for polar codes, named as Flexible and Simplified Multi-bit Successive- Cancellation List (FSMSCL) decoding. The proposed algorithm further reduces the latency compared to other existing multi-bit decoding algorithms. On the other hand, we provide different path-splitting schemes for different code blocks to control the complexity of computation and sort of path metrics. In the analysis, the experiment results show that the proposed algorithm has similar FER performance compared to the conventional SCL algorithm but its decoding latency outperforms other peer decoding algorithms with multi-bit decision.

17:50-18:10 | ID#77

Molecular Polar Belief Propagation Decoder and Successive Cancellation Decoder

Zhiwei Zhong, Lulu Ge, Zaichen Zhang, Xiaohu You and **Chuan Zhang**

Southeast University, China

Abstract: Polar codes have been adopted for the enhanced, mobile broadband control channels for the 5th communication, system. By constructing chemical reaction networks (CRNs), this, paper proposes a method of synthesizing polar belief propagation, (BP) decoder and successive cancellation (SC) decoder. The, proposed method is suitable for polar codes with arbitrary code, length and code rate. Reactions in the proposed design could, be experimentally implemented with DNA strand displacement, reactions, making the proposed polar decoders

promising for, wide application in nanoscale devices. Theoretical analysis and, simulation results have validated the feasibility of this method.

18:10 -18:30 | ID#55

A Secure and Robust Key Generation Method Using Physical Unclonable Functions and Polar Codes

Yonghong Bai and **Zhiyuan Yan**

Lehigh University Bethlehem, USA

Abstract—In physical unclonable functions (PUFs) based key generation methods, the bias of PUF outputs would leak secrecy. A secure and robust key generation method based on PUFs and polar codes is proposed in this paper. First, a PUF-based key generation process is modeled as a wiretap channel. Then, two secure polar coding schemes are designed for the wiretap channel to improve the robustness of key generation and to reduce the secrecy leakage caused by the bias of PUF outputs. To construct the secure polar coding schemes, density evolution is used to evaluate the error probability of synthesized channels, which in turn is used to approximate both the error probability and the secrecy leakage of the system. To reduce the polar construction complexity, the channel independent polar construction method aids density evolution to select parameters of the secure polar coding schemes. Finally, we compare the key generation design with other works and find that our key generation scheme requires fewer PUF bits than other works when they generate the same length key with failure probability $\leq 10^{-6}$.

Dinner @ Peppers All Day Dining Restaurant 百香西餐厅（1F）
| 18:30-20:30

Session T1: Special Session II-Emerging Computing Paradigms for Signal Processing and Smart Learning

Co-Chair: Weikang Qian, Shanghai Jiao Tong University, China

Co-Chair: Weiqiang Liu, Nanjing University of Aeronautics and Astronautics, China

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 09:40-11:00

09:40-9:56 | ID#97

Design and Evaluation of a Power-Efficient Approximate Systolic Array Architecture for Matrix Multiplication

Haroon Waris, Chenghua Wang, **Weiqiang Liu** and Fabrizio Lombardi

Nanjing University of Aeronautics and Astronautics, China

Abstract: Matrix multiplication (MM) is a basic operation for many Digital Signal Processing applications. A Systolic Array (SA) is often considered as one of the most favorable architecture to achieve high performance for matrix multiplication. In this paper, the design exploration for an approximate SA is pursued; three design schemes are proposed by introducing approximation in multiple sub-modules. An approximation factor α is introduced; it is related to the inexact columns in the SA to explore the accuracy-efficiency trade-off present in the proposed designs. In the evaluation, an 8-bit input operand matrix multiplication is considered; the Synopsys Design Compiler at 45nm technology node is used to establish hardware-related metrics. The Error Rate (ER), Normalized Mean Error Distance (NMED) and Mean Relative Error Distance (MRED) are used as figures of merit for error analysis. Results show that the proposed architecture for 8-bit matrix multiplication with an approximation factor α = 7 has the lower power consumption compared to existing inexact designs found in the technical literature with comparable NMED. In addition, a power delay product vs NMED analysis shows the proposed designs have a lower PDP so applicable to low power applications. The practicality of the proposed architecture is established by computing the Discrete Cosine Transform.

09:56-10:12 | ID#101

Ensemble Neural Network Method for Wind Speed Forecasting

Binbin Yong, Fei Qiao, **Chen Wang,** Jun Shen, Yongqiang Wei and Qingguo Zhou

Lanzhou University, China

Abstract: Wind power generation has gradually developed into an important approach of energy supply. Meanwhile, due to the difficulty of electricity storage, wind power is greatly affected by the real-time wind speed in wind fields. Generally, wind speed has the characteristics of nonlinear, irregular, and non-stationary, which make accurate wind speed forecasting a difficult problem. Recent studies have shown that ensemble forecasting approaches combining different sub-models is an efficient way to solve the problem. Therefore, in this article, two single models are ensembled for wind speed forecasting. Meanwhile, four data pre-processing hybrid models are combined with the reliability weights. The proposed ensemble approaches are simulated on the real wind speed data in the Longdong

area of Loess Plateau in China from 2007 to 2015, the experimental results indicate that the ensemble approaches outperform individual models and other hybrid models with different pre-processing methods.

10:12-10:28 | ID#105

A Survey of Computation-Driven Data Encoding

Weikang Qian, Runsheng Wang, Yuan Wang, Marc Riedel and Ru Huang

Shanghai Jiao Tong University, China

Abstract: Although the metal-oxide-semiconductor field-effect transistor (MOSFET) has been the dominant device for modern very-large scale integration (VLSI) circuits for more than six decades, with the dawning of a post-Moore era, researchers are trying to find replacements. A foundation of modern digital computing is the encoding of digital values through a binary radix representation. However, as we enter into the post-Moore era, the challenges of increasing power density, signal noise, and device unreliability raise the question of whether this basic way of encoding data is still the best choice, particularly with novel electronic devices. Prior work has shown that binary radix encoding has some disadvantages. We argue that it is crucial to rethink the necessity of using this representation in the post-Moore era. In this paper, we review some recent development on computation-driven data encoding. We begin with stochastic encoding, a representation proposed a long time ago, discussing both its advantages and disadvantages. Then, we review several recent breakthroughs with variations of stochastic encoding that mitigate many of its disadvantages. Finally, we conclude the paper by extrapolating future directions for effective computation-driven data encoding.

10:28-10:44 | ID#106

Parallel Convolutional Neural Network (CNN) Accelerators Based on Stochastic Computing

Yawen Zhang, Xinyue Zhang, Jiahao Song, Yuan Wang, Ru Huang and Runsheng Wang

Peking University, China

Abstract: Stochastic computing (SC), which processes the data in the form of random bit streams, has been used in neural networks due to simple logic gates performing complex arithmetic and the inherent high error-tolerance. However, SC-based neural network accelerators suffer from high latency, random fluctuations, and large hardware cost of pseudo-random number generators (PRNG), thus diminishing the advantages of stochastic computing. In this paper, we address these problems with a novel technique of generating bit streams in parallel, which needs only one clock for conversion and significantly reduces the hardware cost. Based on this parallel bitstream generator, we further present two kinds of convolutional neural network (CNN) accelerator architectures with digital and analog circuits, respectively, showing great potential for low-power applications.

10:44-11:00 | ID#109

A Data Structure-based Approximate Belief Propagation Decoder for Polar Codes

Menghui Xu, Weikang Qian, Zaichen Zhang, **Xiaohu You** and Chuan Zhang

Southeast University, China

Abstract: Polar code has received great attention recently for, its capacity-achieving property for B-DMCs. Belief propagation, (BP) algorithm is one of the popular approaches for decoding, polar codes, which has unique advantage of high parallelism but, suffers from high hardware complexity. In this paper, we propose, a data structure-based

approximate BP decoder for polar code, for the first time. By reforming the data structure of the received, channel message and introducing the approximate computing, schemes, significant hardware reduction is achieved compared, to the conventional decoder. Simulation results show that the, proposed approximate decoder achieves nearly the same decoding, performance as the conventional one with much less hardware, consumption.

Session T2: Machine learning II

Session Chair: Chance Tarver, Rice University, USA

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 11:00-12:20

11:00-11:20 | ID#82

RNN Models for Rain Detection

Hai Victor Habi and Hagit Messer

Tel Aviv University, Israel

Abstract: The task of rain detection, also known as wet-dry classification, using recurrent neural networks (RNNs) utilizing data from commercial microwave links (CMLs) has recently gained attention. Whereas previous studies used long shortterm memory (LSTM) units, here we used gated recurrent units (GRUs). We compare the wet-dry classification performance of LSTM and GRU based network architectures using data from operational cellular backhaul networks and meteorological measurements in Israel and Sweden, and draw conclusions based on datasets consisting of actual measurements over two years in two different geological and climatic regions

11:20-11:40 | ID#91

Improving Reliability of ReRAM-based DNN Implementation through Novel Weight Distribution

Jingtao Li, Manqing Mao and Chaitali Chakrabarti

Arizona State University, USA

Abstract: Binary deep neural networks, that have been implemented in resistive random access memory (ReRAM) for storage efficiency, suffer from poor recognition performance in the presence of hardware errors. This paper addresses this problem by deriving a novel weight distribution and representation scheme that mitigates errors due to faulty ReRAM cells with minimal storage overhead. In the proposed scheme, the weight matrix is partitioned into grains, and each weight in a grain is represented by the sum of a multi-bit mean and a 1-bit deviation. The grain size as well as the mean to deviation ratio of the weights in a grain can be chosen such that the network is resilient to hardware errors. A hybrid processing-in-memory (PIM) architecture is proposed to support this scheme. The mean values are stored in a small SRAM and processed by a CMOS unit, and the deviations are stored and processed by the ReRAM unit. Compared to the baseline binary neural network which fails in the presence of severe hardware errors, the proposed hybrid scheme has only a mild recognition performance degradation. Simulation results show the proposed scheme achieves 97.84% test accuracy (a 0.84% accuracy drop) on a MNIST dataset, and 88.07% test accuracy (a 1.10% accuracy drop) on a CIFAR-10 dataset under 9.04% stuck-at-1 and 1.75% stuck-at-0 faults.

11:40-12:00 | ID#49

Exploration of On-device End-to-End Acoustic Modeling with Neural Networks

Wonyong Sung, Lukas Lee and Jinhwan Park

Seoul National University, Korea

Abstract: Real-time speech recognition on mobile and embedded devices is an important application of neural networks. Acoustic modeling is the fundamental part of speech recognition and is usually implemented with long short-term memory (LSTM)-based recurrent neural networks (RNNs). However, the single thread execution of an LSTM RNN is extremely slow in most embedded devices because the algorithm needs to fetch a large number of parameters from the DRAM for computing each output sample. We explore a few acoustic modeling algorithms that can be executed very efficiently on embedded devices. These algorithms reduce the overhead of memory accesses using multitimestep parallelization that computes multiple output samples at a time by reading the parameters only once from the DRAM. The algorithms considered are the quasi RNNs (QRNNs), Gated ConvNets, and diagonalized LSTMs. In addition, we explore neural networks that equip one-dimensional (1-D) convolution at each layer of these algorithms, and by which can obtain a very large performance increase in QRNNs and Gated ConvNets. The experiments were conducted using the connectionist temporal classification (CTC)-based end-to-end speech recognition on WSJ corpus. We not only significantly increase the execution speed but also obtain a much higher accuracy, compared to LSTM RNNbased modeling. Thus, this work can be applicable not only to embedded system-based implementations but also to server-based ones.

12:00-12:20 | ID#52

Memory Reduction Through Experience Classification for Deep Reinforcement Learning with Prioritized Experience Replay

Kai-Huan Shen and **Pei-Yun Tsai**

National Central University, Taiwan

Abstract: Prioritized experience replay has been widely used in many online reinforcement learning algorithms, providing high efficiency in exploiting past experiences. However, a large replay buffer consumes system storage significantly. Thus, in this paper, a segmentation and classification scheme is proposed. The distribution of temporal-difference errors (TD errors) is first segmented. The experience for network training is classified according to its updated TD error. Then, a swap mechanism for similar experiences is implemented to change the lifetimes of experiences in the replay buffer. The proposed scheme is incorporated in the Deep Deterministic Policy Gradient (DDPG) algorithm, and the Inverted Pendulum and Inverted Double Pendulum tasks are used for verification. From the experiments, our proposed mechanism can effectively remove the buffer redundancy and further reduce the correlation of experiences in the replay buffer. Thus, better learning performance with reduced memory size is achieved at the cost of additional computations of updated TD errors.

Lunch @ Peppers All Day Dining Restaurant 百香西餐厅（1F）
| 12:40-14:00

Session T3: MIMO

Session Chair: Zhiyuan Yan, Lehigh University, USA

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 14:00-15:40

14:00-14:20 | ID#64

Lattice-Reduction-Aided Symbol-Wise Intra-Iterative Interference Cancellation Detector for Massive MIMO System

Hsiao-Yu Yeh and **Yuan-Hao Huang**

National Tsing Hua University, Taiwan

Abstract: Massive multiple-input multiple-output (MIMO) system plays an important role of increasing spectral efficiency in the fifth-generation (5G) cellular communication. The MIMO detection complexity increases significantly along with the number of antennas. Thus, the design of high-performance low complexity detector for massive MIMO is a challenging design issue for the 5G system. This paper proposes a lattice-reduction aided (LRA) symbol-wise (SW) detection technique to enhance the performance of the intra-iterative interference cancellation (IIC) detector based on Newton's method. The proposed SW IIC detector has near minimum-mean-square-error performance with faster convergence speed and lower computational complexity than the original IIC detector. In a 64-QAM 128 _ 8 up-link MIMO system, the proposed LRA SW IIC detector reduces about 95.35% computational complexity of the original IIC detector under the same BER performance. Considering the preprocessing complexity of the LR in the time-varying channel, the proposed LRA SW IIC detector still has lower complexity when the coherent frame size is larger than 12 MIMO symbols.

14:20-14:40 | ID#27

A Distributed Detection Algorithm For Uplink Massive MIMO Systems

Qiufeng Liu, Hao Liu, Ying Yan and Peng Wu

University of Electronic Science and Technology of China, China

Abstract: Massive multiple-input multiple-output (MIMO) uplink detection algorithms usually rely on centralized base station (BS) architecture, which results in excessive amount of raw baseband data to be transmitted to central processing unit (CU) when the number of antennas is large. Considering the channel hardening characteristics occurs in massive MIMO channels, this paper develops a novel distributed algorithm based on a daisy chain architecture, where the BS antennas are divided into clusters and each owns independent computing hardware for signal processing. In distributed signal detection, only local channel state information (CSI), received data and some data exchange between clusters are needed on each cluster. It is demonstrated that the algorithm can achieve the tradeoff between complexity and performance better than other existing distributed methods.

14:40-15:00 | ID#20

Hybrid Preconditioned CG Detection with Sequential Update for Massive MIMO Systems

Jing Zeng, **Jun Lin**, Zhongfeng Wang, and Yun Chen

Nanjing University, China

Abstract: Massive Multi-Input Multi-Output (MIMO) is one of the key technologies for the fifth-generation communication systems. Conjugate Gradient (CG) algorithm approximates the minimum mean-square error (MMSE) in an iterative manner, which avoids full matrix inversion. Pre-conditioned CG (PCG) was presented to improve the robustness of CG method. However, for the PCG, a sparse matrix inversion is still required in preprocessing and the performance is only comparable to MMSE. In this paper, a hybrid PCG algorithm (HPCG) with sequential update is proposed with superior performance and low complexity. The preconditioned matrix is replaced by a diagonal matrix by exploring its characteristics, which avoids matrix inversion and incomplete Cholesky factorization. Besides, to improve the bit error performance, a sequential update strategy is employed for estimated signals after PCG detection. For a MIMO system with 128 receive antennas, simulation results show the proposed HPCG algorithm outperforms MMSE by 0:25dB to 1:5dB under different numbers of users. Based on the channel hardening theories, several signal vectors can be transmitted in the same channel condition. When 10 signal vectors are considered, compared to the other CG based algorithms, the overall complexity of HPCG can be reduced by 3:9% to 56%.

15:00-15:20 | ID#5

Pilot-assistant methods for channel estimation in MIMO-V-OFDM systems

Wei Zhang, Xuyang Gao and Yibing Shi

University of Electronic Science and Technology of China, China

Abstract: Multiple-input multiple-output (MIMO) with Orthogonal Frequency Division Multiplexing (OFDM) technology has both the advantages of MIMO and OFDM. Vector Orthogonal Frequency Division Multiplexing (V-OFDM) is an extension of OFDM, which makes data transmission flexible. In MIMO systems using V-OFDM technology, different novel schemes are proposed to improve channel estimation performance for different channel sparsity. The 2-D Kriging interpolation scheme is proposed for the non-sparse channels, which can significantly improve the performance of conventional Least Square (LS) and Minimum Mean Square Error (MMSE) algorithms. When the channel is sparse, the estimation process can be modeled as a sparse recovery problem using compressed sensing (CS) theory. In this paper, the measurement matrix is determined by pilot locations, and a pilot search algorithm based on random genetic algorithm (RGA) is proposed to minimize the cross-correlation value of the measurement matrix. Besides, a variable threshold sparsity adaptive matching pursuit (VTSAMP) algorithm is designed to obtain more accurate estimates, which achieves better Normalized Mean Square Error (NMSE) performance, higher calculation speed, and lower implementation complexity.

15:20-15:40 | ID#41

A Unified and Flexible Eigen-Solver for Rank-Deficient Matrix in MIMO Precoding/Beamforming Applications

Su-An Chou, Amalia Rakhmania and **Pei-Yun Tsai**

National Central University, Taiwan

Abstract: Eigenvalue decomposition (EVD) is a widely adopted technique to separate signal, interference, and noise subspaces. The paper presents a unified eigen-solver based on QR decomposition (QRD) to generate eigenpairs

associated with the largest eigenvalues or zero eigenvalues, which are required in the MIMO hybrid beamforming systems that need interference suppression. A non-uniformly constrained deflation is proposed, which forces the matrix to deflate in the beginning and efficiently allocates the computation power to the eigenpairs related with the largest eigenvalues. The computation complexity of generating interested eigenpairs is also evaluated for various matrix dimensions. The results demonstrate that the non-uniformly constrained deflation is effective and more computations can be saved if the desired number of eigenpairs is smaller than the rank of the matrix.

Coffee Break @ Pre-function 宴会前厅| 15:40-16:30

Session T4: Special Session III- Hardware Security and Hardware Implementation of Emerging Cryptographic Primitives

Co-Chair: Yingjie Lao, Clemson University, USA

Co-Chair: Weiqiang Liu, Nanjing University of Aeronautics and Astronautics, China

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 16:30-18:10

16:30-16:50 | ID#102

Dynamic Reconfigurable PUFs based on FPGA

Yijun Cui, Chenghua Wang, Yunpeng Chen, Ziwei Wei, Mengxian Chen and **Weiqiang Liu**

Nanjing University of Aeronautics and Astronautics, China

Abstract: Physical unclonable function (PUF) is a promising security primitive. Configurable ring oscillator (CRO) PUF is an evolvement of conventional RO PUF, which improves the entropy and decrease the hardware cost by introducing configurability. Compared with other types of PUF structures, CRO PUFs are FPGA friendly. In this paper, a dynamic reconfigurable mechanism is proposed for the CRO PUF in FPGA implementation. Three different CRO PUFs are implemented using the proposed reconfigurable method and each CRO can be implemented in a single configurable logic block (CLB) of FPGA. Based on the partial reconfigure functions provided by Xilinx FPGAs, the PUF structures can be configured to any of the three PUF structures. The experimental results show that the dynamic reconfigurable PUF structure has a higher hardware efficiency, reliability and stability compared with the previous works.

16:50-17:10 | ID#98

Side Channel Attack Resistant AES Design Based on Finite Field Construction Variation

Phillip Shvartsman and **Xinmiao Zhang**

The Ohio State University, USA

Abstract: The Advanced Encryption Standard (AES) is the current standard for symmetric key cipher and is algorithmically secure. Side channel attacks that target power consumption can reveal the secret key in AES implementations. Masking data with random variables is one of the main methods used to thwart power analysis attacks. Data can be masked with multiple random variables to prevent higher-order attacks at the cost of a large increase in area. A novel masking scheme for AES resistant to second-order attacks is proposed. Instead of using an additional mask, variation in finite field construction is exploited to increase resistance to second-order attacks. As a result, the area requirement is reduced. For an example AES encryptor, the proposed design is 12% smaller compared to the previous best design, with a very small drop in achievable security level.

17:10-17:30 | ID#108

Ultra-Fast Modular Multiplication Implementation for Isogeny-Based Post-Quantum Cryptography

Jing Tian, **Jun Lin** and Zhongfeng Wang

Nanjing University, China

Abstract: Supersingular isogeny key encapsulation (SIKE) protocol delivers promising public and secret key sizes over other post-quantum candidates. However, the huge computations form the bottleneck and limit its practical applications. The modular multiplication operation, which is one of the most computationally demanding operations in the fundamental arithmetics, takes up a large part of the computations in the protocol. In this paper, we propose an improved unconventional-radix finite-field multiplication (IFFM) algorithm which reduces the computational complexity by about 20% compared to previous algorithms. We then devise a new high-speed modular multiplier architecture based on the IFFM. It is shown that the proposed architecture can be extensively pipelined to achieve a very high clock speed due to its complete feedforward scheme, which demonstrates significant advantages over conventional designs. The FPGA implementation results show the proposed multiplier has about 67 times faster throughput than the state-of-the-art designs and more than 12 times better area efficiency than previous works. Therefore, we think that these achievements will greatly contribute to the practicability of this protocol.

17:30-17:50 | ID#107

An Efficient Polynomial Multiplier Architecture for the Bootstrapping Algorithm in A Fully Homomorphic Encryption Scheme

Weihang Tan, Gengran Hu, Benjamin Case, Shuhong Gao and **Yingjie Lao**

Clemson University, United States

Abstract: Bootstrapping algorithm, which is the intermediate refreshing procedure of a processed ciphertext, has been the performance bottleneck among various existing Fully Homomorphic Encryption (FHE) schemes. Specifically, the external product of polynomials is the most computationally expensive step of bootstrapping algorithms that are based on the Ring Learning With Error (RLWE) problem. In this paper, we design a novel and scalable polynomial multiplier architecture for a bootstrapping algorithm along with a conflict-free memory management scheme to reduce the latency, while achieving a full utilization of the processing elements (PEs). Each PE is a modified radix-2 butterfly unit from fast Fourier transform (FFT), which can be reconfigured to use in both the number theoretic transform (NTT) and the basic modular multiplication of polynomial multiplication in the external product step. The experimental results show that our design yields 33% less areatime product than prior designs.

17:50-18:10 | ID#103

Theoretical Analysis of Configurable RO PUFs and Strategies to Enhance Security

Jiang Li, Hao Gao, Yijun Cui, Chenghua Wang Yale Wang, Chongyan Gu and **Weiqiang Liu**

Nanjing University of Aeronautics and Astronautics, China

Abstract: Compared to traditional ring oscillator PUF (RO PUF), configurable RO PUF (CRO PUF) greatly increases the number of challenge response pairs (CRPs) and improves hardware utilization. However, in the conventional CRO PUF design, when a path is selected by the challenge to generate a response, the circuit characteristic information constituting the CRO PUF, such as the delay information of the configurable unit, the transmission model, and etc., can also be leaked. Once the adversary monitors and masters this information, they can use this information to attack

the CRO PUF circuits, such as modeling attacks. This paper establishes a theoretical model of CRO PUF and analyzes its unpredictability and security. Based on this model, a new mechanism to generate the proper challenges is proposed in this paper. In the proposed mechanism, the challenge is generated and utilized by a specific way, which can delay the feature leakage of the CRO PUF, thereby improving the security of the CRO PUF.

Gala Dinner @ Grand Ballroom A 大宴会厅 A (3F)
| 18:50-21:00

Session W1: Special Session IV-Improving the Performance of Autonomous Systems: Algorithm, Hardware and Application

Session Chair: Deming Chen, University of Illinois at Urbana-Champaign, USA

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 9:40-11:00

9:40-10:00 | ID#113

Efficiently Learning a Robust Self-Driving Model with Neuron Coverage Aware Adaptive Filter Reuse

Chunpeng Wu, Ang Li, **Bing Li** and Yiran Chen

Duke University, USA

Abstract: Human drivers learn driving skills from both regular (non-accidental) and accidental driving experiences, while most of current self-driving research focuses on regular driving only. We argue that learning from accidental driving data is necessary for robustly modeling driving behavior. A main challenge, however, is how accident data can be effectively used together with regular data to learn vehicle motion, since manually labeling accident data without expertise is significantly difficult. Our proposed solution for robust vehicle motion learning, in this paper, is to integrate layer-level discriminability and neuron coverage (neuron-level robustness) regulariziers into an unsupervised generative network for video prediction. Layer-level discriminability increases divergence of feature distribution between the regular data and accident data at network layers. Neuron coverage regulariziers enlarge interval span of neuron activation adopted by training samples, to reduce probability that a sample falls into untested interval regions. To accelerate training process, we propose adaptive filter reuse based on neuron coverage. Our strategies of filter reuse reduce structural network parameters, encourage memory reuse, and guarantee effectiveness of robust vehicle motion learning. Experiments results show that our model improves the inference accuracy by 1.1% compared to FCMLSTM, and cut down 10.2% training time over the traditional method with negligible accuracy loss.

10:00-10:20 | ID#114

Autonomous UAV with Learned Trajectory Generation and Control

Yilan Li, **Mingyang Li**, Amit Sanyal, Yanzhi Wang and Qinru Qiu

Syracuse University, USA

Abstract: Unmanned aerial vehicle (UAV) technology is a rapidly growing field with tremendous opportunities for research and applications. To achieve true autonomy for UAVs in the absence of remote control, external navigation aids like global navigation satellite systems and radar systems, a minimum energy trajectory planning that considers obstacle avoidance and stability control will be the key. Although this can be formulated as a constrained optimization problem, due to the complicated non-linear relationships between UAV trajectory and thrust control, it is almost impossible to be solved analytically. While deep reinforcement learning is known for its ability to provide model free optimization for complex system through learning, its state space, actions and reward functions must be designed carefully. This paper presents our vision of different layers of autonomy in a UAV system, and our effort in generating and tracking the trajectory both using deep reinforcement learning (DRL). The experimental results show that

compared to conventional approaches, the learned trajectory will need 20% less control thrust and 18% less time to reach the target. Furthermore, using the control policy learning by DRL, the UAV will achieve 58.14% less position error and 21.77% less system power.

10:20-10:40 | ID#115

A Hybrid GPU + FPGA System Design for Autonomous Driving Cars

Cong Hao, Atif Sarwari, Zhijie Jin, Husam Abu-Haimed, Daryl Sew, Yuhong Li, Xinheng Liu, Bryan Wu, Dongdong Fu, Junli Gu and **Deming Chen**

University of Illinois at Urbana-Champaign, USA

Abstract—Autonomous driving cars need highly complex hardware and software systems, which require high performance computing platforms in order to enable a real time AI-based perception and decision making pipeline. The industry has been exploring various in-vehicle accelerators such as GPUs, ASICs and FPGAs. Yet the autonomous driving platform design is far from mature when taking into account of system reliability, redundancy and higher level of autonomy. In this work, we propose a hybrid computing system design, which integrates a GPU as the primary computing system and a FPGA as a secondary system. This hybrid system architecture has multiple advantages: 1) The FPGA can be constantly running as a complementary system with very short latency, helping to detect main system failure and anomalous behavior, contributing to system functionality verification and reliability. 2) If the primary system fails (mostly from sensor or interconnection error), the FPGA will quickly detect the failure and run a safe-mode task with a subset of sensors. 3) The FPGA can be used as an independent computing system to run extra algorithm components to improve the overall system autonomy. For example, FPGA can handle driver monitoring tasks while GPU focuses on driving functions. Together they can boost the driving function from L2 (constantly requires driver's attention) to L3 (allows driver to mind off for 10 seconds). This paper defines how such a system works, discusses various use cases and potential design challenges, and shares some initial results and insights about how to make such a system deliver the maximum value for autonomous driving.

10:40-11:00 | ID#116

Accurate Congenital Heart Disease Model Generation for 3D Printing

Xiaowei Xu, Tianchen Wang, Dewen Zeng, Yiyu Shi, Qianjun Jia, Haiyun Yuan, Meiping Huang amd Jian Zhuang

Guangdong General Hospital, China

Abstract: 3D printing has been widely adopted for clinical decision making and interventional planning of Congenital heart disease (CHD), while whole heart and great vessel segmentation is the most significant but time-consuming step in model generation for 3D printing. While various automatic whole heart and great vessel segmentation frameworks have been developed in the literature, they are ineffective when applied to medical images in CHD, which have significant variations in heart structure and great vessel connections. To address the challenge, we leverage the power of deep learning in processing regular structures and that of graph algorithms in dealing with large variations, and propose a framework that combines both for whole heart and great vessel segmentation in CHD. Particularly, we first use deep learning to segment the four chambers and myocardium followed by blood pool, where variations are usually small. We then extract the connection information and apply graph matching to determine the categories of all the vessels. Experimental results using 68 3D CT images covering 14 types of CHD show that our method can increase Dice

score by 11.9% on average compared with the state-of-the-art whole heart and great vessel segmentation method in normal anatomy. The segmentation results are also printed out using 3D printers for validation.

Session W2: Error-correcting Codes

Session Chair: Wonyong Sung, Seoul National University, South Korea

Venue: Grand Ballroom A 大宴会厅 A (3F)

Duration: 11:00-12:00

11:00-11:20 | ID#93

AVX-512 Based Software Decoding for 5G LDPC Codes

Yi Xu, **Wenjin Wang,** Zhen Xu and Xiqi Gao

Southeast University, China

Abstract: In this paper, we investigate how the 5G NR LDPC codes can be decoded by GPP effectively with singleinstruction- multiple-data (SIMD) acceleration and evaluate the corresponding achievable throughput on newly released Intel Xeon CPUs. Firstly, a general software implementation architecture with SIMD acceleration for horizontal-layered LDPC decoding is presented, where the parallelism can be achieved in an intra-block manner. By utilizing Intel advanced vector extended 512 (AVX-512) instruction set, the efficiency of parallelism are maximized and therefore the capacity of x86 processors can be fully exploited. In addition, new features of AVX-512 are further exploited to optimize load and store operations as well as preprocessing to reduce the operation cost. Experiments results also show that Intel Xeon Gold 6154 processors can achieve 42 to 272 Mbps throughput with a single core for ten layered decoding iterations for various code rate and block length. The typical processing latency is below 100 μs. Consequently, an eighteen-core Intel Xeon CPU can achieve up to 5 Gbps decoding throughput

11:20-11:40 | ID#32

A Low-Complexity Error-and-Erasure Decoding Algorithm for t=2 RS Codes

Zengchao Yan, Wenjie Li, **Jun Lin**, and Zhongfeng Wang

Nanjing University, China

Abstract: Reed-Solomon (RS) codes are widely adopted in numerous digital communication systems to handle the possibly occurred errors and/or erasures during the data transmission. This paper focuses on the *t = 2 RS* codes and proposes a low-complexity error-and-erasure decoding algorithm for them. The proposed algorithm directly computes the errata location polynomial instead of performing the iterative Berlekmap-Massey (BM) algorithm which is usually adopted in the conventional RS decoding algorithm. Moreover, a method to directly compute the errata locations and errata magnitudes is also presented. For a (255, 251) RS code, the proposed error-and-erasure decoding algorithm can save over 90% multiplications and additions of the conventional decoding algorithm. In addition, the complexity reduction becomes more significant as code length increases.

11:40-12:00 | ID#80

A New Inversionless Berlekamp–Massey Algorithm with Efficient Architecture

Chao Chen, Yunghsiang S. Han, Zhongfeng Wang and Baoming Bai

Xidian University, China

Abstract: This paper presents a new inversionless BerlekampMassey (BM) algorithm as well as its efficient architecture. Starting with a lesser-known version of BM algorithm, we develop a serial of inversionless variants by successively applying algorithmic transformations. The final algorithm has a very compact description and a highly regular structure, which can be naturally mapped to a systolic architecture. Compared with the state-of-the-art architecture RiBM, the proposed one possesses a different cell structure and has slightly lower hardware requirements. More importantly, it enables us to establish a new architectural equivalence between the BM algorithm and the Euclidean algorithm.

Lunch @ Peppers All Day Dining Restaurant 百香西餐厅（1F）
| 12:00-13:30

POSTER SESSION I

Venue: Pre-function (3F)
15:40-16:30, October 21 | Monday

Poster Session I

Session Chair: Chuan Zhang, Southeast University, China

Venue: Pre-function 宴会前厅 (3F)

Duration: 15:40-16:30

ID#1

CLA Formula Aided Fast Architecture Design for Clustered Look-Ahead Pipelined IIR Digital Filter

Yuanyong Luo, Hongbing Pan, Qinghong Shen and Zhongfeng Wang

Nanjing University, China

Abstract: In VLSI design domain, Clustered Look-Ahead (CLA) technique is a promising method to further pipeline or accelerate IIR digital filters in the coming era of 5G network for mobile devices. However, much efforts are needed to acquire the stable CLA pipelined architecture. Therefore, this paper proposes a CLA Formula to aid the fast architecture design for CLA pipelined IIR digital filters. To obtain the stable architecture with the pipeline stage ranging from 6 to 96, comparison experiments show that when compared to the symbolic method with substitution, the proposed CLA Formula aided method can save more than half the software coding time for designers and reduce almost 168~30243 times the execution time for the programs.

ID#2

SIR Beam Selector for Amazon Echo Devices Audio Front-End

Xianxian Zhang, Trausti Kristjansson and Philip Hilmes

Amazon Inc., USA

Abstract: The Audio Front-End (AFE) is a key component in mitigating acoustic environmental challenges for far-field automatic speech recognition (ASR) on Amazon Echo family of products. A critical component of the AFE is the Beam Selector, which identifies which beam points to the target user. In this paper, we proposed a new SIR beam selector that utilizes subband-based signal-to-interference ratios to learn the locations of the audio sources and therefore further improve the beam selection accuracy for multi-microphone based AFE system. We analyzed the performance of a Signal to Interference Ratio (SIR) beam selector with a comparison to classic beam selector using the datasets collected under various conditions. This method is evaluated and shown to simultaneously decrease word-error-rate (WER) for speech recognition by up to 46.20% and improve barge-in performance via FRR by up to 39.18%.

ID#7

AdaBoost-assisted Extreme Learning Machine for Efficient Online Sequential Classification

Yi-Ta Chen, Yu-Chuan Chuang, and An-Yeu (Andy) Wu

National Taiwan University, Taiwan

Abstract: In this paper, we propose an AdaBoost-assisted extreme learning machine for efficient online sequential classification (AOS-ELM). In order to achieve better accuracy in online sequential learning scenarios, we utilize the cost-sensitive algorithm-AdaBoost, which diversifying the weak classifiers, and adding the forgetting mechanism, which stabilizing the performance during the training procedure. Hence, AOS-ELM adapts better to sequentially

arrived data compared with other voting based methods. The experiment results show AOS-ELM can achieve 94.41% accuracy on MNIST dataset, which is the theoretical accuracy bound performed by original batch learning algorithm, AdaBoost-ELM. Moreover, with the forgetting mechanism, the standard deviation of accuracy during the online sequential learning process is reduced to 8.26x.

ID#22

A FMCW Ranging Method with Identification Ability

Meiqing Liu, Shang Ma, Boen Chi, Kai Long, Qiu Huang and Bixin Zhu

University of Electronic Science and Technology of China, China

Abstract: Frequency modulated continuous wave (FMCW) radar has been widely used and thoroughly studied in high precision ranging. However, FMCW radar cannot identify target type while ranging. To address this problem, this paper presents a high-precision ranging method based on FMCW, which enables the FMCW radar to identify target type while ranging. The proposed method improves the ranging accuracy by performing frequency sweep in a specific frequency range and realizes target identification through the orthogonal spread spectrum. The designed prototype verification system operates at a frequency of 5.8 GHz and the sweep bandwidth is 200 MHz. Exemplary experiment results are presented to illustrate that the system has a ranging accuracy of 20 cm and a range of up to 12 m in an indoor scenario. Furthermore, multiple user identification can also be realized.

ID#25

A Novel Approach to Angle-of-Arrival Estimation Based on Layered Ensemble Learning

Rui Li, Tingqiang Deng, Yongming Huang, **Chuan Zhang** and Luxi Yang

Southeast University, China

Abstract: Hybrid beamforming is a promising solution for, multiple-input multiple-output (MIMO) systems with large scale, antennas due to its low-cost and good performance compared, with pure digital and analog beamforming. Unfortunately, conventional, angle-of-arrival (AoA) estimation methods, such as, MUSIC and ESPRIT algorithms, need a lot of calculations, and must solve the issue of phase ambiguity. Therefore, this, paper proposes a novel AoA estimation method based on layered, ensemble learning. Because the training process can be completed, off-line, only estimating complexity is taken into account which, make the AoA detection process low complexity and meet realtime, requirements. The simulation results indicate that the, accuracy of the proposed AOA estimation method is higher than, that of traditional algorithms. In addition, our proposed method, is robust to the phase error.

ID#45

EAGLE: Exploiting Essential Address in Both Weight and Activation to Accelerate CNN Computing

Wenjian Liu, Xiayuan Wen, Jun Lin, **Zhongfeng Wang** and Li Du

Nanjing University, Nanjing, China

Abstract: Efficient quantization techniques can compress Convolutional Neural Networks (CNNs) with less bit-width while maintaining the accuracy on large extent. However, the quantized CNN models hardly boost the computation performance of CNN accelerators with the conventional bit-parallel Multiply- Accumulate (MAC) operations. Previous works proposed a shifting-based bit-serial operation, which can be called as Shift- Accumulate (SAC) operation, to take

advantage of the reduced bit-width. However, it is also found that there are many invalid computations in both MAC and SAC operation, caused by zero bits in activations and weights, which are not optimized. To fully exploit the computations in CNN models, we proposed a Essential Address only GAC based Low-latency Efficient (EAGLE) architecture that can further accelerate the CNN computation through bypassing zero bits computation in the activations and weights. An essential address is adopted to encode the nonzero bits in activations and weights in this architecture. Furthermore, to support the essential address-only computations, Generate- Accumulate (GAC), an operation which produces partial sums with essential addresses, is implemented. The architecture is implemented with a TSMC 28nm CMOS technology. Based on the results, if scaled in a 65nm technology, the EAGLE only requires 63.6% area and 43.1% power consumption compare to that of the Pragmatic. The EAGLE reaches an average speedup of 2.08× and 1.43× on six CNN models over the Stripe and Pragmatic at a similar frequency, respectively. It also improves energy efficiency by 3.69× on average over the DaDianNao baseline.

ID#46

Sub-Spectrogram Segmentation for Environmental Sound Classification via Convolutional Recurrent Neural Network and Score Level Fusion

Tianhao Qiao, Shunqing Zhang, Zhichao Zhang, Shan Cao and Gong Xu

Shanghai University, China

Abstract: Environmental Sound Classification (ESC) is an important and challenging problem, and feature representation is a critical and even decisive factor in ESC. Feature representation ability directly affects the accuracy of sound classification. Therefore, the ESC performance is heavily dependent on the effectiveness of representative features extracted from the environmental sounds. In this paper, we propose a subspectrogram segmentation based ESC classification framework. In addition, we adopt the proposed Convolutional Recurrent Neural Network (CRNN) and score level fusion to jointly improve the classification accuracy. Extensive truncation schemes are evaluated to find the optimal number and the corresponding band ranges of sub-spectrograms. Based on the numerical experiments, the proposed framework can achieve 81.9% ESC classification accuracy on the public dataset ESC-50, which provides 9.1% accuracy improvement over traditional baseline schemes.

ID#83

On Secrecy Energy Efficiency of RF Energy Harvesting System

Zhengxia Ji, Mengyun Nie, Lingquan Meng, Qingran Wang, Chunguo Li and **Kang Song**

Qingdao University, China

Abstract: The increasing use of information source in unreliable wireless communication is a driving force to explore the networks' energy efficiency and security. To fully improve the performance of the system, in this paper, we combine these two directions and investigate the secrecy energy efficiency (SEE) of the network in which the information can be eavesdropped consisting of an energy source, an information source, a destination and an eavesdrop node, all of which are equipped with single antenna. The system model is based on ST (savethen- transmit) protocol. The information source node harvests energy from the received signal power to charge its battery, which is used to retransmit the received signal to the destination. Under the limited transmit power mode, we get the expression for SEE, which depends on energy absorption rate and time. Our analytical results reveal that the secrecy efficiency has a maximum. The optimal energy absorption rate was further calculated by Newton iterative algorithm.

Then we propose optimal energy source selection method. Simulation results finally verify the accuracy and efficiency of our proposed algorithm for secrecy energy efficiency maximization.

POSTER SESSION II

Venue: Pre-function (3F)
15:40-16:30, October 22 | Tuesday

Poster Session II

Session Chair: Yeong-Luh Ueng, National Tsing Hua University, Taiwan

Venue: Pre-function 宴会前厅 (3F)

Duration: 15:40-16:30

ID#3

Modified Complementary Joint Sparse Representations: a novel post-filtering to MVDR beamforming

Yuanyuan Zhu, Jiafei Fu, Xu Xu and Zhongfu Ye

University of Science and Technology of China, China

Abstract: Post-filtering is a popular technique for multichannel speech enhancement system, in order to further improve the speech quality and intelligibility after beamforming. This paper presents a novel post-filtering to a minimum variance distortion less response (MVDR) beamforming which is a single channel modified complementary joint sparse representations (M-CJSR) method. First, MVDR beamformer is used to suppress interference and noise. Subsequently, the proposed M-CJSR approach based on joint dictionary learning is applied as a single microphone post-filter to process the beamformer output. Different from the existing post-filtering techniques which rely on the assumptions about the noise field, this algorithm considers a more generalized signal model including the ambient noise, like diffuse noise or white noise, as well as the point-source interference. Moreover, the original CJSR method is extended to jointly learn dictionaries for not only the mappings from mixture to speech and noise, but also the mapping from mixture to interference. In order to take the complementary advantages of different sparse representations, we design the weighting parameters based on the residual components of the estimated signals. An experimental study which consists of objective evaluations under various conditions verifies the superiority of the proposed algorithm compared to other state-of-the-art methods.

ID#6

FPGA Prototyping of A Millimeter-Wave Multiple Gigabit WLAN System

Dongming Ren, Kang Chen, Shengheng Liu and **Yongming Huang**

Southeast University, China

Abstract: IEEE 802.11aj (45-GHz) standard is recently proposed for wireless local area network operating in an undefined millimeter-wave (mmWave) band. In this work, an ultra-highspeed mmWave orthogonal frequency division multiplexing transmission prototype is developed and some primary amendments in this standard are verified using NI-PXIe mmWave software defined-radio platform. A mixed parallel processing scheme is devised to meet the clock requirements of field programmable gate arrays baseband processing. A queue-based synchronization mechanism is designed to facilitate the implementation of data transporting. Data transmission test indicates that the system is able to achieve an extremely high data rate of multi-gigabits per second with a low bit error rate.

POSTER SESSION II

Venue: Pre-function (3F)
15:40-16:30, October 22 | Tuesday

ID#34

An ISAR imaging algorithm based on RPCA for micro-Doppler effect suppression

Xinbo Xu, Xinfei Jin and Fulin Su

Harbin Institute of Technology, China

Abstract: In Inverse Synthetic Aperture Radar (ISAR) imaging, the micro-Doppler (m-D) effect caused by micromotion parts of the target will not only make parameter extraction and motion compensation difficult but also cause image defocusing. It will appear as azimuth interference sidebands and decrease image quality seriously. Therefore, studying the micro-Doppler suppression problem in practical applications is of great importance in high-quality imaging of ISAR. In this paper, a reasonable and effective mathematical model is established, and the m-D suppression algorithm inspired by the robust principal component analysis (RPCA) matrix reconstruction theory is proposed. Our algorithm transforms the problem of separating radar echoes into the decomposition of a low rank rotating components m-D signal matrix and a sparse main body ISAR image signal matrix. Moreover, experimental results based on simulated and real measured data are utilized to verify the effectiveness of our method.

ID#44

Data Driven Low-Complexity DOA Estimation for Ultra-Short Range Automotive Radar

Yixin Song, Yang Li, Cheng Zhang and **Yongming Huang**

Southeast University, China

Abstract: In recent applications of millimeter wave automotive radars, the short range detection and estimation performance becomes an important design metric. Due to the sphere rather than plane form of array incoming signals, direct use of conventional spectrum or direction of arrival (DOA) estimators generally result in large performance degradation. In this paper, a naive look-up table based solution is first introduced. To solve its involved large storage requirement problem, we further transform the DOA estimation problem into the DOA classification problem, and utilize the support vector machine (SVM) framework to propose a data-driven low-complexity DOA estimator. Simulations validate the effectiveness of the propose SVM solution especially for small sample set and high storage limit.

ID#56

Towards Algebraic Modeling of GPU Memory Access for Bank Conflict Mitigation

Luca Ferranti and Jani Boutellier

Tampere University, Finland

Abstract: Graphics Processing Units (GPU) have been widely used in various fields of scientific computing, such as in signal processing. GPUs have a hierarchical memory structure with memory layers that are shared between GPU processing elements. Partly due to the complex memory hierarchy, GPU programming is non-trivial, and several aspects must be taken into account, one being memory access patterns. One of the fastest GPU memory layers, shared memory, is grouped into banks to enable fast, parallel access for processing elements. Unfortunately, it may happen that multiple threads of a GPU program may access the same shared memory bank simultaneously causing a bank conflict. If this happens, program execution slows down as memory accesses have to be rescheduled to determine

which instruction to execute first. Bank conflicts are not taken into account automatically by the compiler, and hence the programmer must detect and deal with them prior to program execution. In this paper, we present an algebraic approach to detect bank conflicts and prove some theoretical results that can be used to predict when bank conflicts happen and how to avoid them. Also, our experimental results illustrate the savings in computation time.

ID#88

Nonlinear Functions in Learned Iterative Shrinkage-Thresholding Algorithm for Sparse Signal Recovery

Elaine Crespo Marques, **Nilson Maciel**, Lirida Naviner and Hao Cai and Jun Yang

Télécom Paris, France

Abstract: Compressive sensing requires fewer measurements than Nyquist rate to recover sparse signals, leading to processing and energy saving. The efficiency of this technique strongly depends on the quality of the considered sparse recovery algorithm. This work focuses on a learned iterative shrink-age-thresholding algorithm where iterations are related to layers of a neural network. We analyze the performance of this algorithm for different shrinkage functions. A decrease up to *9dB* in the NMSE value is achieved by choosing appropriate shrinkage function. Moreover, the estimation performance can be close to the theoretical performance bound, showing deep learning as a promising tool for sparse signal estimation. This work can be applied in several areas such as image processing, Internet of Things (IoT), cognitive radio networks, and sparse channel estimation for wireless communications.

TUTORIAL I

Chaired by Prof. Yeong-Luh Ueng

Title- Artificial Intelligence, Machine Learning, and Statistical Signal Processing in Financial Technology (FinTech)

Abstract-Financial technology (Fintech), a broad category that refers to the innovative use of ICT technology in the design and delivery of financial services and products, as revolutionized the financial industry or even more broadly the service industry. Artificial intelligence (AI) is the intelligence demonstrated by machines, in contrast to the natural intelligence (NI) displayed by humans and other animals. The tremendous success of AI and machine learning algorithms in the area of computer vision and natural language processing has demonstrated great potential of further applications into other disciplines. One of the most promising areas is the use of AI, machine learning, and financial data analytics in FinTech. In this tutorial, we will first introduce the financial concepts in the context of ICT and then aim to provide an overview of the recent trends in Fintech, machine learning and statistical signal processing approaches for tackling important challenges in FinTech, and how such approaches can be applied to solve real-world problems. We will discuss the challenges, constraints, and opportunities in this fascinating research area.

SPEAKER I

Prof. Wei-Ho Chung, National Tsing Hua University

Bio: Dr. Wei-Ho Chung received the B.Sc. and M.Sc. degrees in Electrical Engineering from the National Taiwan University, Taipei, Taiwan, in 2000 and 2002, respectively, and the Ph.D. degree in Electrical Engineering from the University of California, Los Angeles, in 2009. From 2002 to 2005, he was with ChungHwa Telecommunications Company. In 2008, he worked on CDMA systems at Qualcomm, Inc., San Diego, CA. His research interests include communications, signal processing, and networks. Dr. Chung received the Ta-You Wu Memorial Award from Ministry of Science and Technology in 2016, Best Paper Award in IEEE WCNC 2012, and Taiwan Merit Scholarship from 2005 to 2009. He has published over 50 journal articles and over 50 conference papers. Since January 2010, Dr. Chung had been an assistant research fellow, and promoted to the rank of associate research fellow in 2014 in Academia Sinica. Since 2018, he has been a full Professor in Electrical Engineering, National Tsing Hua University. He leads the Wireless Communications Lab in National Tsing Hua University, Taiwan.

SPEAKER II

Prof. Che Lin, National Taiwan University

Bio: Che Lin received the B.S. degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, in 1999. He received the M.S. degree in Electrical and Computer Engineering in 2003, the M.S. degree in Math in 2008, and the Ph.D. degree in Electrical and Computer Engineering in 2008, all from the University of Illinois at Urbana-Champaign, IL. In 2008, he joined the Department of Electrical Engineering at National Tsing Hua University as an assistant professor, and has been an associate professor since August 2014. Dr. Lin received a two-year Vodafone graduate fellowship in 2006, the E. A. Reid fellowship award in 2008, and holds a U.S. patent, which has been included in the 3GPP LTE standard. In 2012, he received the Excellent Teaching Award for the college of EECS, NTHU. He won the best paper award for 2014 GIW-ISCB-ASIA conference. In 2015, he received the CIEE outstanding young electrical engineer Award. In 2017, he received the Young Scholar Innovation Award from Foundation For The Advancement Of Outstanding Scholarship. He is a senior member of IEEE. His research interests include deep learning, data mining and analytic, signal processing in wireless communications, optimization theory, systems biology, and FinTech.

TUTORIAL II

Chaired by Prof. Fei Qiao

Title-Optical Mobile Communications

Abstract-The 6G mobile communication system will be an integrated information system covering deep space, air, terrestrial, sea surface, and undersea communications. All frequency spectrum will be exploited to cater the needs of high-speed and full-coverage information transmission. The optical spectrum, which has huge frequency spectrum resources, will play a more and more important role in the development of the next generation mobile communication systems. In this tutorial, we will introduce state-of-the-art optical wireless communication (OWC) technologies, including free space optical (FSO) communications, visible light communications (VLC), as well as a newly proposed optical mobile communication (OMC) technology. Further development of OWC technologies in 6G scenario will be addressed, with emphasis on how the OWC technologies will be incorporated into the mobile communication architecture.

SPEAKER

Prof. Zaichen Zhang, Southeast University

Bio: Professor Zaichen Zhang received B.S. and M.S. degrees in electrical and information engineering from Southeast University, Nanjing in 1996 and 1999, respectively, and Ph.D. degree in electrical and electronic engineering from the University of Hong Kong in 2002. From 2002 to 2004, he was a Postdoctoral Fellow at the National mobile Communications Research Laboratory (NCRL), Southeast University. He is currently the executive Dean of School of Information Science and Engineering, Southeast University. He has authored over 200 papers and issued over 30 patents. He is senior member of the IEEE. He served as IEEE ICC 2015 Keynote Chair, IEEE ICNC 2015 and PIMRC Symposium Chairs, and IEEE ICC 2019 Operation Chair. He was the Distinguished Visiting Fellow of Royal Academy of Engineering, UK, 2017 and the invited speaker of IEEE ICCC 2017. His current research interests include 6G mobile information systems, optical wireless communications, and quantum information technologies.

TUTORIAL III

Chaired by Prof. Pei-Yun Tsai

Title-Internet of Things (IoT): Signals, Communications, Applications, Challenges, and Future Research

Abstract-Internet of Things (IoT) is the network of physical objects or "things" embedded with electronics, software, sensors, and network connectivity. It enables the objects to collect, share, and analyze data. The IoT has become an integral part of our daily lives through applications such as public safety, intelligent tracking in transportation, industrial wireless automation, personal health monitoring, and health care for the aged community. IoT is one of the latest technologies that will change our lifestyle in coming years. Experts estimate that as of now, there are 23 billon connected devices, and by 2020 it would reach to 30 billion devices. This tutorial aims to introduce the design and implementation of IoT systems. The foundations of IoT will be discussed throughout real applications. Challenges and constrains for the future research in IoT will be discussed. In addition, research opportunities and collaboration will be offered for the attendees.

SPEAKER

Prof. Ahmed Abdelgawad, Central Michigan University

Bio: Dr. Ahmed Abdelgawad received his M.S. and a Ph.D. degree in Computer Engineering from University of Louisiana at Lafayette in 2007 and 2011 and subsequently joined IBM as a Design Aids & Automation Engineering Professional at Semiconductor Research and Development Center. In Fall 2012 he joined Central Michigan University as a Computer Engineering Assistant Professor. In Fall 2017, Dr. Abdelgawad was early promoted as a Computer Engineering Associate Professor. He is a senior member of IEEE. His area of expertise is distributed computing for Wireless Sensor Network (WSN), Internet of Things (IoT), Structural Health Monitoring (SHM), data fusion techniques for WSN, low power embedded system, video processing, digital signal processing, Robotics, RFID, Localization, VLSI, and FPGA design. He has published two books and more than 82 articles in related journals and conferences. Dr. Abdelgawad served as a reviewer for several conferences and journals, including IEEE WF-IoT, IEEE ISCAS, IEEE SAS, IEEE IoT Journal, IEEE Communications Magazine, Springer, Elsevier, IEEE Transactions on VLSI, and IEEE Transactions on I&M. He severed in the technical committees of IEEE ISCAS 2007, IEEE ISCAS 2008, and IEEE ICIP 2009 conferences. He served in the administration committee of IEEE SiPS 2011. He also served in the organizing committee of ICECS2013 and 2015. Dr. Abdelgawad was the publicity chair in North America of the IEEE WF-IoT 2016/18/19 conferences. He was the finance chair of the IEEE ICASSP 2017. He is the TPC Co-Chair of I3C'17, the TPC Co-Chair of GIoTS 2017, and the technical program chair of IEEE MWSCAS 2018. He was the keynote speaker for many international conferences and conducted many webinars. He is currently the IEEE Northeast Michigan section chair and IEEE SPS Internet of Things (IoT) SIG Member. In addition, Dr. Abdelgawad served as a PI and Co-PI for several funded grants from NSF.

TUTORIAL IV

Chaired by Prof. Che Lin

Title-Generative Adversarial Network and its Applications to Speech Signal and Natural Language Processing

Abstract-Generative adversarial network (GAN) is a new idea for training models, in which a generator and a discriminator compete against each other to improve the generation quality. Recently, GAN has shown amazing results in image generation, and a large amount and a wide variety of new ideas, techniques, and applications have been developed based on it. Although there are only few successful cases, GAN has great potential to be applied to text and speech generations to overcome limitations in the conventional methods. There are three parts in this tutorial. In the first part, I will give an introduction of GAN and provide a thorough review about this technology. In the second part, I will focus on the applications of GAN to speech signal processing. In the third part, I will describe the major challenge of sentence generation by GAN and review a series of approaches dealing with the challenge.

SPEAKER

Prof. Hung-yi Lee, National Taiwan University

Bio: Hung-yi Lee received the M.S. and Ph.D. degrees from National Taiwan University (NTU), Taipei, Taiwan, in 2010 and 2012, respectively. From September 2012 to August 2013, he was a postdoctoral fellow in Research Center for Information Technology Innovation, Academia Sinica. From September 2013 to July 2014, he was a visiting scientist at the Spoken Language Systems Group of MIT Computer Science and Artificial Intelligence Laboratory (CSAIL). He is currently an assistant professor of the Department of Electrical Engineering of National Taiwan University, with a joint appointment at the Department of Computer Science & Information Engineering of the university. His research focuses on machine learning (especially deep learning), spoken language understanding and speech recognition. He owns a YouTube channel teaching deep learning (in Mandarin) with more than 3M views and 39k subscribers (https://www.youtube.com/channel/UC2ggjtuuWvxrHHHiaDH1dlQ/playlists).

TUTORIAL V

Chaired by Prof. Yuan-Hao Huang

Title-Making Healthcare More Accessible via AI: Extension of Telemedicine

Abstract-This tutorial will focus on innovative digital health ecosystem and analytics system which fosters extension of telemedicine through the transfer of comprehensive medical expertise and experiences via Artificial Intelligence (AI) from tertiary medical centers to remote care facilities in making healthcare more accessible! Cancer is among the most important issues of healthcare worldwide. However, under current medical systems, diagnosis of these severe diseases is commonly delayed, especially in remote locations with limited medical resources. Hence it is necessary to facilitate early screening at these distant care units using Computer-Aided-Diagnostic (CAD) tools possessing tertiary centers' experiences accumulated through AI. In attempts to reform and advance the digital health environment, using skin care as example, this tutorial introduces an ecosystem, by which integration of remote care facilities is substantiated, through utilization of high accuracy and efficiency of AI as extension of telemedicine! Being an exemplary, this AI medical networking model is readily extensible to global medical and biotech communities! This tutorial will also introduce the speaker's reconfigurable edge system for the detection of Skin Cancer on mobile devices with more than 95% accuracy.

SPEAKER

Prof. Gwo Giun (Chris) Lee, National Cheng Kung University

Bio: Chris Gwo Giun Lee is an investigator in the field of signal processing systems including multimedia and bioinformatics. His endeavors in system design, based on analytics of algorithm concurrently with analytics architecture, has made possible computations on System-on-Chip and cloud platforms in resolving complex problems with both accuracy and efficiency. Having previously held leading and managerial positions in the industry such as System Architect in former Philips Semiconductor in Silicon Valley, Lee was recruited to NCKU in 2003 where he found and is currently directing the Bioinfotronics Research Center.

Lee received his B.S. degree in electrical engineering from National Taiwan University and both his M.S. and Ph.D. degrees in electrical engineering from University of Massachusetts. He has contributed more than 130 original research and technical publications with the invention of 60+ patents worldwide.

Lee serves as the AE for IEEE TSP and Journal of Signal Processing Systems. He was formerly the AE for IEEE TCSVT for which he received the Best Associate Editor's Award in 2011.

TUTORIAL VI

Chaired by Prof. Wei-Ho Chung

Title-Tensor Subspace Analysis in Signal Processing

Abstract-The standard matrix computation can not fully exploit the global data structure in higher-order signal processing. The recent advances in tensor computation allow us to move from classical matrix based methods to tensor based methods for many signal processing techniques. This tutorial focuses on different tensor decompositions for tensor subspace analysis in signal processing. Firstly, a basic coverage of tensor notations, preliminary operations, and main tensor decompositions is briefly provided. Based on them, a series of tensor subspace analysis methods are presented, as the multi-linear extensions of classical sparse signal recovery, principle component analysis, matrix completion, non-negative matrix factorization, linear regression, subspace cluster, etc. The experimental results for some signal processing applications are given, e.g., image reconstruction, denoising, illumination normalization, background extraction, and classification. Finally, some deep tensor networks are discussed for possible tensor signal processing applications.

SPEAKER

Prof. Yipeng Liu, University of Electronic Science and Technology of China

Bio: Yipeng Liu received the B.Sc. degree in biomedical engineering and the Ph.D. degree in information and communication engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, in 2006 and 2011, respectively. From 2010 to 2011, he was a visiting PhD student in Tsinghua University, Beijing, China. In 2011, he was a research engineer at Huawei Technologies, Chengdu, China. From 2011 to 2014, he was a postdoctoral research fellow at University of Leuven, Leuven, Belgium. Since 2014, he has been an associate professor with University of Electronic Science and Technology of China (UESTC), Chengdu, China.

His research interest is tensor signal processing. He has authored or co-authored over 50 publications on these areas. One of the co-authored papers received the ISMRM MERIT AWARD of Magna cum laude at ISMRM 2015. He also holds more than 10 patents. He has been a principal investigator (PI) or Co-PI for a number of R&D projects (funded by government and industry) on tensor signal processing theory and its applications. He has given and will give tutorials on a number of international conferences, such as ISCAS 2019, APSIPA ASC 2019.

He serves as managing guest editor of the Special Issue on Tensor Image Processing of the journal Signal Processing: Image Communication. As an expert on tensor signal processing, he has served 4 international conferences as a technical/program committee member. He is an IEEE senior member, a member of the Multimedia Technology Technical Committee of Chinese Computer Federation, and a member of China Society of Image and Graphics on Youth Working Committee. He is the scientific advisor of Beiton AI. He has been teaching the course optimization theory and applications for graduates since 2015, and got the first prize of the 8th University Teaching Achievement Award in 2016.

ACADEMIC VISIT TO PURPLE MOUNTAIN LABORATORIES

Purple **M**ountain **L**aboratories is a major technological innovation platform jointly promoted and built by Jiangsu Province and Nanjing Municipality to thoroughly implement Xi Jinping's thought on socialism with Chinese characteristics in a new era and create a famous innovative city with global influence. Oriented toward meeting the national major strategic needs in the field of network communication and security, with the mission of leading the global information technology development direction and solving major scientific and technological problems in the industry, Purple Mountain Laboratories conducts prospective and basic research through gathering high-end talents from around the world. By breaking through the core technologies and conducting major demonstration applications, the Purple Mountain Laboratories strives to facilitate applying the research findings in the construction of national economy and national defense. Purple Mountain Laboratories is committed to grow into a national-level laboratory and become a vital force in national scientific and technological innovation.

Address: No. 9, Mozhou East Road, Jiangning, Nanjing

Research Areas: focuses of future network, new communications, cyber security and other research directions.

Future Network: network operating system and multi-cloud exchange basic network platform; service customization network, intelligent security, artificial intelligence and other applications in the industrial internet; protocol-independent forwarding and other applications in civil-military integration networks.

New Communications: millimeter wave communications, perception and detection, air-space-ground fusion communication system, terahertz communication, artificial intelligence application in communications and energy-efficient mobile communication.

Cyber Security: internet endogenous security mechanism, wireless communication physical layer security mechanism, artificial intelligence application in cyber security.

Date: October 23, 2019
Time of Gathering: 13:30
Time for Departure: 13:40
Place of Gathering: Hotel Lobby
Estimated Duration: 30 minutes
Walking Distance: 15 minutes

紫金山实验室
Purple Mountain Laboratories

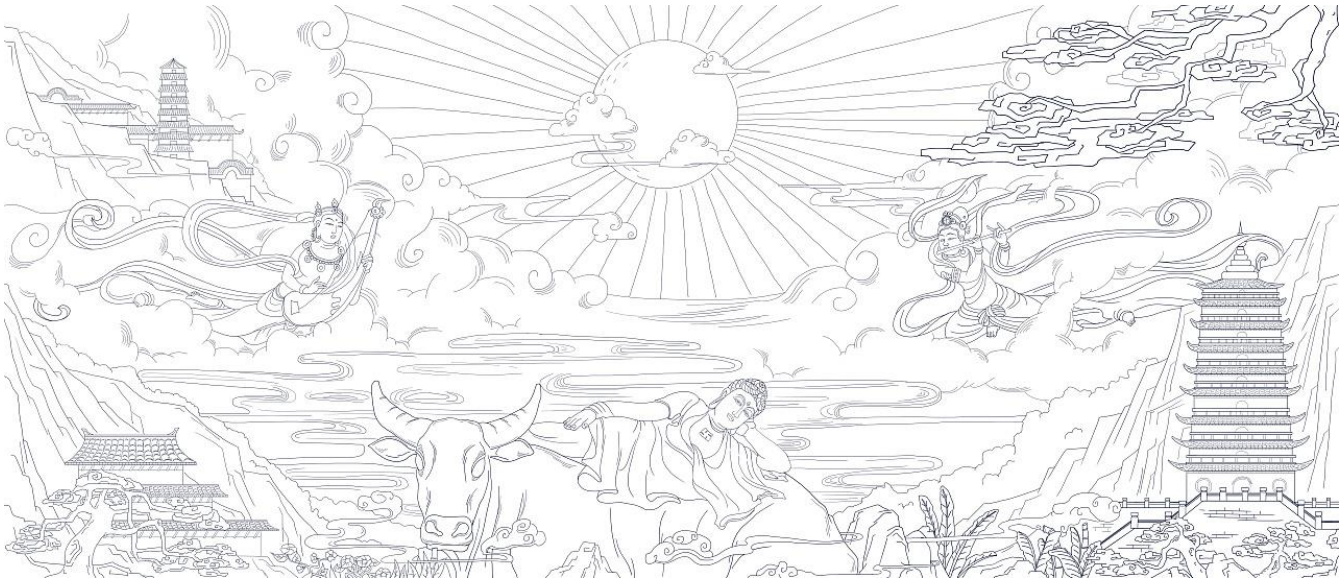
HALF DAY TOUR

A BLESSING TRIP-NIUSHOU MOUNTAIN (HALF DAY)

Niushou Mountain is located in Jiangning District of Nanjing. The mountain twin peaks look like ox horns, hence the name. 1700 years ago, the first founder of the Eastern Jin Dynasty, emperor Yuandi, wanted to set up two magnificent gate towers to symbolize imperial power. However, Prime Minister, Wang Dao, considered it inappropriate to do so. One day, when they looked afar at the east and west peak of Niushou Mountain, the Minister suggested that the two peaks were the most suitable heavenly-made gate towers. Hence, the name Tianque Mountain.

The core of scenic area construction is as follows: fixing the heavenly peaks, meaning restoring mountain massif and recovering twin peaks; building underground palace, meaning making use of the natural pit to build Usnisa Palace for enshrining Usnisa; renovating the sacred road, meaning expanding original tunnel for exhibiting Sakyamuni's life story and understanding the way to become Buddha; revealing twin pagodas, meaning constructing Usnisa Pagoda and reproducing the grand pattern of twin pagodas; spreading five levels of Chan(Zen Buddhism) and understanding the atmosphere of Chan in rivers and mountains.

Date: October 23, 2019
Time of Gathering: 13:30
Time for Departure: 13:40
Estimated Duration: 6 hours
Place of Gathering: Hotel Lobby
Note: Please bring your ID/Passport with you when you out for the tour.



Police Call: 110
Emergency Center: 120

Averages in October
13 °- 22°