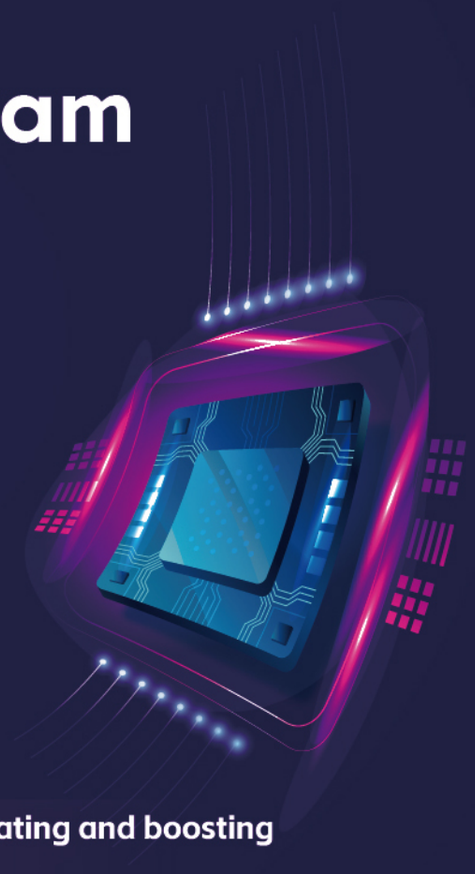
AICAS 2023

IEEE International Conference on Artificial Intelligence Circuits and Systems

Program

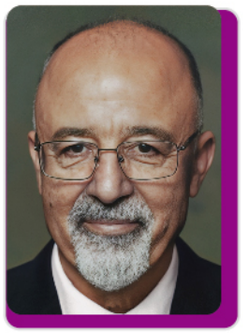


Circuits and Systems for accelerating and boosting AI performance and impact

June 11-13, 2023
Hangzhou, China

CONTENT

Welcome Message from the General Co-Chairs



Mohamad Sawan
Westlake University, China

Shaojun Wei
Tsinghua University, China

Dear AICAS 2023 Attendees

On behalf of all members of the Organizing Committee, we would like to welcome you all in Hangzhou for the 2023 5th IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS 2023) this week of June 11-13, 2023.

AICAS 2023 is one of the most highly acclaimed annual conferences in the field of Circuits and Systems for Artificial Intelligence applications. We are truly delighted to host such a prestigious event in Hangzhou, Zhejiang, China. Since its inception, AICAS has been continuously leading the latest innovations and trends in artificial intelligence circuits and systems. with active participation of worldwide researchers from both academia and industry.

The organizing committee has made all possible efforts to plan high quality conference program, including keynote talks, industrial lectures, tutorials, life demos, regular contribution lecture and poster sessions on a variety of AICAS topics, and various social activities. The conference program presented by experts on leading-edge topics cover the entire spectrum of AI circuits and systems - from algorithms, applications, and hardware implementations for artificial intelligence algorithms. We hope that all participants enjoy memorable experience and high-quality presentations by authors from all over the world.

We hope that AICAS 2023 serves as an excellent forum for discussion and collaboration where innovative ideas can be shared among participants to lead technologies in the fields of AI Circuits and Systems. As such, we strongly encourage all participants of AICAS 2023 to engage discussions with their fellow participants at this conference for fruitful collaborations in this emerging and quickly growing technical field.

Along with the strong technical program of AICAS 2023, We hope you will enjoy the many cultural and recreational activities in Hangzhou.

Mohamad Sawan and Shaojun Wei
General Co-Chairs, IEEE AICAS 2023

Welcome Message from the TPC Co-Chairs

Yong Lian
York University, Canada

Yen-Kuang Chen
Princeton AI Group, USA

Abe Elfadel
Khalifa University, UAE

On behalf of the Technical Program Committee of the 2023 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS 2023), it is our great pleasure and excitement to extend a warm welcome to all participants. This year, we are thrilled to host the conference in the vibrant city of Hangzhou, China, renowned for its rich cultural heritage and technological advancements.

Under the theme "Circuits and Systems for accelerating and boosting AI performance and impact," AICAS 2023 aims to gather leading researchers, scholars, and industry professionals to explore the latest innovations and advancements in the field of artificial intelligence circuits and systems. In response to the unprecedented demands, this conference will provide a platform for discussing and developing cutting-edge research that enhances and amplifies AI's performance and influence.

We are delighted to inform you that the AICAS 2023 technical program committee received 187 papers from 29 countries across various tracks, including special sessions and live-demo sessions. With the assistance of 44 Review Committee Members and 337 reviewers, we managed to collect an average of 3.8 reviews per paper, resulting in an acceptance rate of 58.39%.

The technical program of AICAS 2023 will consist of oral and poster sessions, and a live demo session, providing ample opportunities for knowledge sharing and interactive discussions. In addition to the regular program, we have curated a series of captivating keynote speeches to be delivered by esteemed visionaries from renowned organizations. These thought-provoking presentations will offer valuable insights into the latest trends, challenges, and prospects in the AI circuits and systems domain.

Furthermore, we are excited to offer a diverse range of educational tutorials to enhance participants' understanding of emerging technologies and foster professional growth. These tutorials, conducted by esteemed experts, will cover a broad spectrum of topics, enabling attendees to gain practical knowledge and stay at the forefront of AICAS research and development.

We would like to express our deepest gratitude to all the authors, reviewers, review committee members, special session chairs, live-demo session chairs, and technical program committee members for their unwavering dedication and invaluable contributions. AICAS 2023 would not be possible without your tireless efforts and commitment to advancing the field of AI circuits and systems.

As we embark on this enriching journey together, we hope that the technical program at AICAS 2023 proves to be both enjoyable and rewarding for all participants. We eagerly anticipate fruitful networking, lively discussions, and the exchange of ground-breaking ideas that will shape the future of AI.

Thank you once again for joining us at AICAS 2023 in Hangzhou, China. Let us come together to create a memorable conference experience that will have a lasting impact on the AI circuits and systems community.

Warm regards,

Yong Lian, Yen-Kuang Chen and Abe Elfadel
Technical Program Co-Chairs, AICAS 2023

IEEE AICAS 2023 Organizing Committee

Honorary Chair

Qionghai Dai, Tsinghua University, China

General Co-Chairs

Mohamad Sawan, Westlake University, China
Shao-Jun Wei, Tsinghua University, China

Technical Program Co-Chairs

Yong Lian, York University, Canada
Yen-Kuang Chen, Princeton AI Group, USA
Abe Elfadel, Khalifa University, UAE

Plenary Sessions Co-Chairs

Tobi Delbruck, UZH-ETH Zurich, Switzerland
Yong-Pan Liu, Tsinghua University, China
Hai Li, Duke University, USA
Wen-Hsiao Peng, National Chiao Tung University, Taiwan, China

Special Sessions Co-Chairs

Robert Chen-Hao Chang, National Chung Hsing University, Taiwan, China
Shih-Chii Liu, INI-UZH, Switzerland
Maurizio Valle, University of Genova, Italy

Tutorials Co-Chairs

Eduard Alarcon, Technical University of Catalunya, Spain
Mounir Boukadoum, University of Quebec in Montreal, Canada
Feng Wu, University of Science and Technology of China, China
David Brooks, Harvard University, USA

Publications Co-Chairs

Kea-Tiong (Samuel) Tang, National Tsing Hua University, Taiwan, China
Krishnendu Chakrabarty, Duke University, USA
Jun Zhou, University of Electronic Science and Technology of China, China

Demo sessions Co-Chairs

Guoxing Wang, Shanghai Jiao Tong University, China
Hyuk-Jae Lee, Seoul National University, Korea
Hao Yu, Southern University of Science and Technology, China
Deepu John, University College Dublin, Ireland

Industrial Sessions Co-Chairs

Chiara Bartolozzi, Italian Institute of Technology, Italy
Gwo-Giun Lee, National Cheng Kung University, Taiwan, China
Damien Querlioz, University Paris-Saclay, France
Shi-Peng Li, Chinese University of Hong Kong, Shenzhen, China
Hao Zhang, Ocean University, China

Local Arrangement Co-Chairs

Jie Yang, Westlake University, China
Nanjian Wu, Chinese Academy of Sciences, China
Le Ye, Peking University, China

Tutorials Co-Chairs

Amara Amara, TDH, Switzerland
Andreas Andreou, Johns Hopkins University, USA
Magdy Bayoumi, University of Louisiana, USA
Giacomo Indiveri, ETH Zurich, Switzerland
Rajiv Joshi, IBM, USA
SeokbumKo, University of Saskatchewan, Canada
Martin Kumm, Fulda University of Applied Sciences, Germany
Herbert Ho-Ching Lu, Western Australia
Manuel Roveri, Politecnico di Milano, Italy
Kaushik Roy, Purdue, USA
Myung Hoon Sunwoo, Ajou University, Korea
Atsushi Takahashi, TIT, Japan.

Conference Secretariat

Yitian (Claire) Zhang, Westlake University, Hangzhou, China
(secretariat@aicas2023.org)

Program at a Glance

Sunday, June 11, 2023 Grand Hyatt Hangzhou			
MORNING	**West Lake II**	**West Lake III**	**West Lake IV**
8:30-10:00	**Tutorial 1** Carbon Neutral Computing for Engineerable AI	**Tutorial 2** Energy-Efficient Recurrent Neural Network Accelerator Design for Real-Time Inference	**Tutorial 5** Analog Matrix Computing with Resistive Memory: Circuit Designs and Applications
10:00-10:30	**BREAK**		
10:30-12:00	**Tutorial 1** Carbon Neutral Computing for Engineerable AI		**Tutorial 3** Near-Sensor Analytics and Machine Learning for Long-Term Wearable Brain-Machine Interfaces
12:00-13:30	**BREAK**		
AFTERNOON	**West Lake II**	**West Lake III**	**West Lake IV**
13:30-15:00	**Tutorial 1** Carbon Neutral Computing for Engineerable AI	**Tutorial 4** In-memory computing for accelerating deep neural networks and neuro-vector-symbolic architectures	**Tutorial 5** Analog Matrix Computing with Resistive Memory: Circuit Designs and Applications
15:00-17:00	**Grand Challenges Contest**		**WiCAS & YPP Event**
17:00-18:00	**Welcome Reception**		

Monday, June 12, 2023 Grand Hyatt Hangzhou						
MORNING	**West Lake II**		**West Lake III**		**West Lake IV**	
8:30-8:45	**Compute-in-Memory (A1L-1)**	**6048**	**Methodologies of Neuromorphic Design (A1L-2)**	**6030**	**Bio Applications (A1L-3)**	**6163**
8:45-9:00		**6065**		**6112**		**6085**
9:00-9:15		**6106**		**6079**		**6007**
9:15-9:30		**6061**		**6081**		**6021**
9:30-9:45		**6141**		**6157**		**6107**
9:45-10:00	**BREAK**					
10:00-10:15	**Opening Ceremony**					
10:15-11:05	**Keynote 1 - Prof. Moncef Gabbouj** *The Super Neuron Model - A new generation of ANN-based Machine Learning and Applications*					
11:05-11:55	**Keynote 2 - Prof. Giacomo Indiveri** *Neuromorphic Intelligence: mixed signal analog/digital implementations of spiking neural networks for real-time signal processing*					
12:00-12:20	**Group Photo, Grand Ballroom II (B2)**					
12:20-13:30	**LUNCH, Grand Café (Buffet), Grand Hyatt**					

AFTERNOON	**West Lake II**		**West Lake III**		**West Lake IV**	
13:30-13:45	**Computing-in-Memory Circuits & Systems (A4L-1)**	**6121**	**Neuromorphic Circuits & Systems 1 (A4L-2)**	**6023**	**Visual Algorithms (A4L-3)**	**6045**
13:45-14:00		**6086**		**6147**		**6115**
14:00-14:15		**6118**		**6135**		**6114**
14:15-14:30		**6113**		**6151**		**6049**
14:30-14:45		**6143**		**6172**		**6111**
14:45-16:00	**Live Demos (A5P-4), West Lake II & III & IV Lobby**					
16:00-16:15	**BREAK**					
16:15-16:30	**Processing in Memory Techniques as AI Acccelerators (A6L-1)**	**6202**	**Neuromorphic Circuits & Systems 2 (A6L-2)**	**6131**	**Visual Applications (A6L-3)**	**6095**
16:30-16:45		**6200**		**6174**		**6126**
16:45-17:00		**6201**		**6154**		**6101**
17:00-17:15		**6190**		**6020**		**6171**
17:15-17:30		**6189**		**6152**		**6003**
18:30-20:30	**BANQUET Grand Ballroom II(B2), Grand Hyatt**					

Tuesday, June 13, 2023 Grand Hyatt Hangzhou						
MORNING	**West Lake II**		**West Lake III**		**West Lake IV**	
8:30-8:45	**Computer Arithmetic for ML (B1L-1)**	**6188**	**Emerging Neuromorphic Paradigms (B1L-2)**	**6039**	**Efficient Algorithms (B1L-3)**	**6145**
8:45-9:00		**6197**		**6070**		**6056**
9:00-9:15		**6195**		**6063**		**6072**
9:15-9:30		**6044**		**6130**		**6116**
9:30-9:45		**6002**		**6096**		**6068**
9:45-10:00	**BREAK**					
10:00-10:50	**Keynote 3 - Prof. Yuchao Yang** *Integrated Memristor Networks for Higher-complexity Neuromorphic Computing*					
10:50-12:20	**Industrial Session (T-head, Ali-Cloud, ARM, SynSense, PIMChip, KeySight)**					
12:20-13:30	**LUNCH, Grand Café (Buffet), Grand Hyatt**					
AFTERNOON	**West Lake II**		**West Lake III**		**West Lake IV**	
13:30-13:45	**Energy-efficient Circuits & Systems for AI-enabled Biomedical Sensors (B5L-1)**	**6199**	**Hardware Accelerator (B5L-2)**	**6034**	**Learning Algorithms (B5L-3)**	**6094**
13:45-14:00		**6176**		**6024**		**6059**
14:00-14:15		**6178**		**6013**		**6132**
14:15-14:30		**6179**		**6058**		**6014**
14:30-14:45		**6187**		**6136**		**6066**
14:45-15:45	**Poster Session (Tiny, Efficient and Engineerable Machine Learning)(B4P-4) and other posters (B4P-5), West Lake II & III & IV Lobby**					
15:45-16:00	**BREAK**					
16:00-16:15	**Devices, Circuits & Systems, Algorithms & Applications (B6L-1)**	**6082**	**Algorithm-Hardware Co-design (B6L-2)**	**6037**	**General Applications (B6L-3)**	**6177**
16:15-16:30		**6142**		**6137**		**6146**
16:30-16:45		**6193**		**6055**		**6125**
16:45-17:00		**6028**		**6074**		**6026**
17:00-17:15		**6109**		**6071**		**6167**
17:15-18:00	**Farewell Celebration, West Lake II & III & IV**					

*** Red color identifies special sessions**

Grand Challenge

on Software and Hardware Co-Optimization for E-Commerce Recommendation System

Introduction:

The AICAS Grand Challenge 2023 focuses on a fast and accurate recommender system for e-commerce applications. A reliable recommender system can not only contribute to an enjoyable shopping experience by suggesting the most suitable products or services to customers, but also to higher profits for retail companies. Software and hardware co-optimization for e-commerce recommender systems has gained increasing attention in academia and industry in recent years. AICAS Grand Challenge consists of two rounds: the preliminary round encourages participants to develop accurate algorithms for the E-commerce recommendation scenario. Then 32 qualified teams are selected to enter the final round and compete to design fast and accurate acceleration schemes for the on-chip algorithm inference process. The thesis of the AICAS Grand Challenge session is to invite the 6 winning teams to share their unique solutions. We hope this session inspires more students and researchers to focus on our software and hardware co-optimized competition for AI applications.

Session Agenda:

- **Award Ceremony of the AICAS Grand Challenge**
- **Keynote #1 Introduction to Yitian Hardware Platform**
- **Keynote #2 Overview of the AICAS Grand Challenge**
- **Coffee break**
- **Presentations from the winning teams**:
  - ◆Title: Building a fast and accurate recommender for E-Commerce
    Team Member: Xiaoran Yang
  - ◆Title: How to optimize the efficiency of a recommender system algorithm
    Team Member: He Wang, Zexin Xu
  - ◆Title: Powerful CatBoost and Specialized Features
    Team Member: Qianli Ma, Yan Wang, Zhiyong Deng
  - ◆ More from other winning teams

Organizing Committee

- **Li Du, Nanjing University**
- **Xiaohan Ma, Alibaba Group**
- **Wei Mao, Southern University of Science and Technology**
- **Yuan Du, Nanjing University**
- **Yongfu Li, Shanghai Jiaotong University**

Sponsors

- **T-Head Semiconductor Co. https://www.t-head.cn/**
- **Arm https://www.armchina.com/**
- **2023 AICAS http://www.aicas2023.org/**
- **Alibaba Cloud https://www.aliyun.com**

Tutorials

Tutorial 1

Time: 8:30-10:00 & 10:30-12:00 & 13:30-15:00, SUNDAY JUNE 11, 2023
Venue: West Lake II (2F)

Title: Carbon Neutral Computing for Engineerable AI

Abstract:

The emerging technologies for ultra-efficient computing circuits and systems (CAS) are introduced in this tutorial on application-, architecture-, circuit-, and physics-level. For powering "engineerable AI" which is expected to be wide range programmable, low-cost, sustainably developable, and easy-use, the user-friendly computing platforms are instructed by the four lectures along with a hand-to-hand setup/start-up hour. The first lecture demonstrates a hardware/software co-optimization framework of neural networks (seen as hardware-aware neural architecture search (NAS)). The architecture level designs of massive core computers are lectured in the second slot with a novel prototype series of coarse grained re-configurable accelerator (CGRA) seen as IMAX-1/2/3. The third lecture illustrates the approximate and parallel computing processors by a novel series of elastic many core calculator-array known as "DiaNet-1/2/3/4". The final lecture is to help engineers understanding and developing the super-conductive circuits and systems. All four lectures are supplemented by releasing start-up and run-up resources such as IDEs, prototype/Verilog codes, videos, manuals, or other necessary materials.

Speaker 1: Dr. Yiyu Shi

Biography:

Dr. Yiyu Shi is currently a professor in the Department of Computer Science and Engineering at the University of Notre Dame, the site director of National Science Foundation I/UCRC Alternative and Sustainable Intelligent Computing, and the director of the Sustainable Computing Lab (SCL). He is also a visiting scientist at Boston Children's Hospital, the primary pediatric program of Harvard Medical School. He received his B.S. in Electronic Engineering from Tsinghua University, Beijing, China in 2005, the M.S and Ph.D. degree in Electrical Engineering from the University of California, Los Angeles in 2007 and 2009 respectively. His current research interests focus on hardware intelligence and biomedical applications.

Speaker 2: Dr. Yasuhiko Nakashima

Biography:

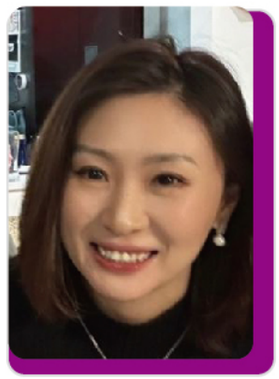
Yasuhiko Nakashima received B.E., M.E., and Ph.D. degrees in Computer Engineering from Kyoto University in 1986, 1988 and 1998, respectively. He was a computer architect in the Computer and System Architecture Department, FUJITSU Limited from 1988 to 1999. From 1999 to 2005, he was an associate professor in the Graduate School of Economics, Kyoto University. Since 2006, he has been a professor in the Graduate School of Information Science, Nara Institute of Science and Technology. His research interests include computer architecture, emulation, circuit design, and accelerators. He is a fellow of IEICE and a senior member of IEEE.

Speaker 3: Dr. Renyuan Zhang

Biography:

Renyuan Zhang (Senior Member, IEEE) received the B.E degree from Tongji University in 2007; the M.E. degree from Waseda University in 2010 and the Ph.D. degree from the University of Tokyo in 2013. He was an assistant professor with Japan Advanced Institute of Science and Technology from 2013 to 2017. He has been an assistant professor with Nara Institute of Science and Technology (NAIST) and the PRESTO researcher with Japan Science and Technology Agency since 2017 and 2018, respectively. From 2021, he serves NAIST (Japan) as an associate professor. His research interests include analog–digital-mixed circuits, approximate computing, high performance computing architectures, and hardware implementation of AI. He is experienced in organizing international conferences by serving the IEEE SOCC, CoolChips, ASP-DAC etc. as a TPC member. In 2022, he proposed and successfully organized the special session titled “Non-Deterministic Computing ”on IEEE SOCC, which included five technical papers. He also organized a dual special session with 11 technical papers titled “Emerging Computational Mechanisms ”on IEEE GCCE, 2022.

Speaker 4: Dr. Olivia Chen



Biography:

Olivia Chen is an associate professor in the Department of Computer Science at Tokyo City University, and a visiting associate professor in Yokohama National University, Japan. In 2017, she received her PhD degree in Electronic Engineering from Yokohama National University, Yokohama, Japan. She has been engaged in research on high performance computing systems using superconducting integrated circuits for eight years. She was invited as young plenary speakers to give a lecture at Applied Superconductivity Conference (ASC2018), the world’s largest international conference on superconducting applications. Her research results have also been published in numerous journals (Scientific Report, IEEE Transaction on Applied Superconductivity, Superconductor Science and Technology) and major superconducting conferences (ISEC, ASC, EUCAS).

Tutorial 2

Time: 8:30-10:00, SUNDAY JUNE 11, 2023
Venue: West Lake III (2F)

Title: Energy-Efficient Recurrent Neural Network Accelerator Design for Real-Time Inference

Abstract:

Gated recurrent neural networks (RNNs) are effective in speech recognition and machine translation that involves time-sequential data processing. Typical RNNs include Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). RNNs are memory-bounded algorithms, making it challenging to run them in real-time on resource-constrained edge devices. This tutorial gives a survey of state-of-the-art methods used to reduce the memory and computes of the network by capitalizing on spatial weight sparsity and skipping over zero activations. We will cover model compression techniques that reduce the arithmetic and memory cost of RNNs and co-designed hardware architectures to efficiently run the RNNs optimized by the software methods. The tutorial will focus on a particular network architecture called the delta network inspired by the spiking neuron model. It can induce temporal sparsity in RNNs, potentially reducing the compute cost of the RNNs by 10 times. The delta network algorithm can be further combined with model compression to induce spatio-temporal sparsity in RNNs and reduce the cost by 100 times. On the hardware side, the tutorial will cover two GRU-RNN accelerators, DeltaRNN and EdgeDRNN, that exploit temporal sparsity and a Long Short-Term Memory (LSTM) RNN accelerator, Spartus, that exploits spatio-temporal sparsity. The tutorial also presents related edge applications of the accelerators, including edge audio processing and robotic control, to showcase their capability of solving real-world problems.

Speaker 1: Dr. Chang Gao

Biography:

Chang Gao is an Assistant Professor at the Department of Microelectronics, Delft University of Technology, Delft, Netherlands. He is leading the Lab of Efficient circuit & system for Machine Intelligence (EMI, https://www.tudemi.com). He obtained his Ph.D. degree with distinction from the University of Zurich to design energy-efficient recurrent neural network accelerators for real-time inference. His current research interest is developing efficient deep-learning-specific hardware for edge inference and training.

Speaker 2: Dr. Shih-Chii Liu

Biography:

Shih-Chii Liu co-directs the Sensors group (http://sensors.ini.uzh.ch) at the Institute of Neuroinformatics, University of Zurich and ETH Zurich. Her group works on event-driven deep networks and bio-inspired auditory sensors, and the real-time implementation of intelligent hardware systems with state-of-art power efficiency, latency, and throughput; and applications in machine learning tasks, including speech recognition.

Speaker 3: Dr. Tobi Delbruck

Biography:

Tobi Delbruck is a Professor of Physics and Electrical Engineering at the Institute of Neuroinformatics, the University of Zurich and ETH Zurich, Zürich, Switzerland. He codirects the Sensors group (http://-sensors.ini.uzh.ch) at the Institute. His group focuses on neuromorphic vision processing and efficient deep learning architectures.

Tutorial 3

Time: 10:30-12:00, SUNDAY JUNE 11, 2023
Venue: West Lake IV (2F)

Title: Near-Sensor Analytics and Machine Learning for Long-Term Wearable Brain-Machine Interfaces

Abstract:

A brain-machine interface (BMI) provides a communication and control pathway between the human brain and external devices. It plays a crucial role in stroke rehabilitation and prosthetic control. A BMI recognizes a subject's intention based on brain activities that are often recorded by non-invasive electroencephalographic (EEG) devices. Traditional data processing workflow consists of sending EEG data to a remote processing engine where resource-demanding algorithms are executed to accurately extract useful information. Remote data transmission causes privacy concerns, risks of long latency, and high-power consumption yielding short battery life. To mitigate these issues, recent developments in smart wearable BMIs bring the processing near the sensors directly at the edge device where the data is collected and processed in real time. However, the resources available on wearable BMIs are limited posing big challenges in embedding accurate machine learning models at the edge.

In this tutorial, we introduce various BMI paradigms, summarize state-of-the-art models especially in the context of tiny machine learning, present optimization techniques to successfully embed accurate models on resource-limited edge devices, and finally show demos of EEG acquisition using BMIs and control of devices such as drones.

Speaker: Dr. Xiaying Wang

Biography:

Xiaying Wang received her B.Sc. and M.Sc. degrees in biomedical engineering from Politecnico di Milano, Italy and ETH Zurich, Switzerland in 2016 and 2018, respectively. She continued with a doctoral study in the Integrated Systems Laboratory, ETH Zurich, under the supervision of Prof. Dr. Luca Benini and obtained the Dr. Sc. degree in 2022. She received an Excellent paper award in IEEE International Conference on E-health Networking, Application and Service (HEALTHCOM) in 2018 and the Ph.D. Fellowship from the Swiss Data Science Center in 2019. Her research interests include biosignal processing, low-power embedded systems, energy-efficient smart sensors, brain–machine interfaces, and machine learning on microcontrollers.

Tutorial 4

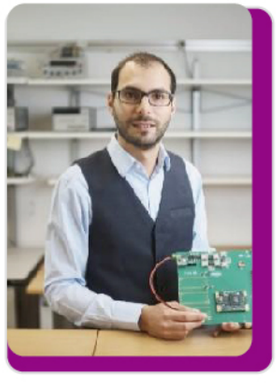
Time: 13:30-15:00, SUNDAY JUNE 11, 2023
Venue: West Lake III (2F)

Title: In-memory computing for accelerating deep neural networks and neuro-vector-symbolic architectures

Abstract:

The computing systems that run today's AI algorithms are based on the von Neumann architecture which is inefficient at the task of shuttling huge amounts of data back and forth at high speeds. Thus, to build efficient cognitive computers, we need to transition to novel architectures where memory and processing are better collocated. In-memory computing is one such approach where the physical attributes and state dynamics of memory devices are exploited to perform certain computational tasks in place with very high areal and energy efficiency. This tutorial aims at exploring in-memory computing for accelerating deep neural networks (DNNs) and neuro-vector-symbolic architectures (NeuroVSA) and is organized in three parts. The first part will provide an overview to phase-change memory-based in-memory computing for neural network inference and training, and a state-of-the art 64-core inference chip will be introduced. The second part will focus on heterogeneous in-memory computing architectures for systems with different power budgets and performance targets for DNN inference. Motivated by key properties of VSA (vector-symbolic architecture), including its amenability for implementation on in-memory computing, the last part presents a summary of recently developed methods on the integration of VSA with deep neural networks, dubbed NeuroVSA, that enabled impactful applications from few-shot continual learning to abstract reasoning to computationally hard problems such as vector factorization.

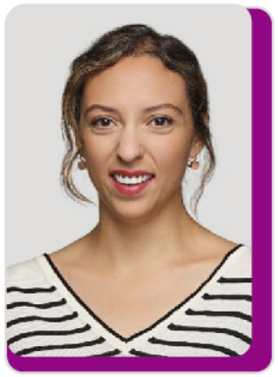
Speaker 1: Dr. Manuel Le Gall



Biography:

Manuel Le Gallo joined IBM Research Europe in 2013, where he is currently employed as a Staff Research Scientist in the In-Memory Computing group of the Zurich laboratory. His main research interest is in using phase-change memory devices for non-von Neumann computing. He has co-authored more than 50 scientific papers in journal and conferences and holds more than 20 granted patents. He was appointed IBM Master Inventor in 2019 for significant contributions to intellectual property and is a recipient of the MIT Technology Review's 2020 Innovators Under 35 award.

Speaker 2: Dr. Irem Boybat



Biography:

Irem Boybat is a Research Staff Member at IBM Research Europe, Zurich, Switzerland. She received her Ph.D. degree in Electrical Engineering from Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland, in 2020. Previously, she had obtained an M.Sc. degree in Electrical Engineering from EPFL, Switzerland, in 2015, and a B.Sc. degree in Electronics Engineering from Sabanci University, Turkey, in 2013. Her research interests include in-memory computing for AI, neuromorphic computing, and emerging resistive memory. She has co-authored over 45 scientific papers in journals and conferences, received three best conference presentation/paper/poster awards and holds 7 granted patents. She was a co-recipient of the 2018 IBM Pat Goldberg Memorial Best Paper Award and 2020 EPFL PhD Thesis Distinction in Electrical Engineering.

Speaker 3: Dr. Abbas Rahimi

Biography:

Abbas Rahimi received the B.S. degree in computer engineering from the University of Tehran in 2010, and the M.S. and Ph.D. degrees in computer science and engineering from the University of California San Diego in 2015, followed by postdoctoral researches at the University of California Berkeley, and at the ETH Zürich. In 2020, he has joined the IBM Research-Zürich laboratory as a Research Staff Member.

He has received the 2015 Outstanding Dissertation Award in the area of "New Directions in Embedded System Design and Embedded Software" from the European Design and Automation Association, and the ETH Zürich Postdoctoral Fellowship in 2017. He was a co-recipient of the Best Paper Nominations at DAC (2013) and DATE (2019), and the Best Paper Awards at BICT (2017), BioCAS (2018), and IBM's Pat Goldberg Memorial Best Paper Award (2020).

Tutorial 5

Time: 8:30-10:00 & 13:30-15:00, SUNDAY JUNE 11, 2023
Venue: West Lake IV (2F)

Title: Analog Matrix Computing with Resistive Memory: Circuit Designs and Applications

Abstract:

In this Tutorial, we will lecture on the designs and applications of analog matrix computing (AMC) circuits based on crosspoint resistive memory arrays. Typical AMC circuits, including those for matrix multiplication, matrix inversion, generalized inverse, and eigenvector computations will be introduced. The underlying design principles, such as global or local connection, negative or positive feedback, row-wise or column-wise splitting, and conductance compensation strategy will be elaborated. The aforementioned matrix computations constitute the foundations of many important algorithms in a wide range of applications, where the AMC circuits could be a highly promising acceleration method. Here, typical application of AMC including signal processing, machine learning, scientific computing, wireless communications will be showcased, demonstrating orders of magnitude improvement of equivalent throughput and energy efficiency over conventional digital computers.

Speaker: Dr. Zhong Sun

Biography:

Zhong Sun is an Assistant Professor at Peking University, China. He received his Ph.D. in 2016 from Tsinghua University, China. His research interests include emerging resistive memory, analog computing, in-memory computing, and their applications to matrix problems. He has published ~20 papers in journals such as PNAS, Science Advances, and IEEE TCAS-I/II on the topic of this Tutorial. He was awarded the Intellectual Property Award (Italy) in 2019, and he was recognized as a Distinguished Young Scholar by NSFC in 2021.

Opening Ceremony

Time: 10:00-10:15, June 12, 2023
Venue: West Lake II & III & IV

Chair: Mohamad Sawan

Welcome Address

Mohamad Sawan, General Co-Chair

Shao-Jun Wei, General Co-Chair

Conference Statistics

Yong Lian, Technical Program Co-Chair

WiCAS & Young Professionals (YP) Event

Time: 15:00-17:00, Sunday June 11, 2023
Venue: West Lake IV(2F)

Fostering Excellence: Empowering Women in Engineering and Young Professionals for the Future

Welcome Speech (5 mins)

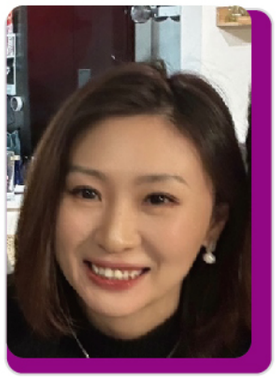
Speakers from Academia and Industry (60 ~ 80 mins)

Roundtable talk (20 ~ 30 mins)

INVITED SPEAKERS

Speaker 1: Dr. Olivia Chen

Up Against the Tide: Cultivating Change for Women and Early-Career Professionals – Japan's Narrative



Biography:

Olivia Chen is an associate professor in the Department of Computer Science at Tokyo City University, where she is also a distinguished researcher within the JST PRESTO and FOREST Program. She laid her academic foundation at Yokohama National University, Japan, where she received her PhD in Electrical and Computer Engineering in 2017. Chen's scholarly pursuits push the boundaries of understanding in several advanced areas, notably superconducting electronics, energy-efficient computing, approximate computing, deep learning hardware accelerators, and design automation for superconducting VLSI implementation. Her research findings have been published in major applied superconductivity journals, such as IEEE TAS and SUST, and presented at top tier EDA/CAS conferences including ISCA, DAC, DATE, ICCD, and ICCAD.

Abstract:

Across the globe, engineering academia strives for a balanced representation of women and young professionals. The journey is not without its trials, but each challenge surmounted is a step closer to achieving diversity and fostering innovation. This presentation explores the intricate factors contributing to this disparity, using Japan's unique experiences as a case study. Our focus will be on societal norms, inflexible career paths, and the lack of readily available mentorship opportunities that often hinder progress. However, our narrative isn't one of barriers alone—it's also about overcoming them. By discussing Japan's recent initiatives, such as implementing mentorship programs, promoting flexible academic career paths, and advocating for supportive policies, we aim to provide a beacon of hope for aspiring academics. Our aim is to instil a sense of resilience and inspire action among women and young professionals navigating their academic careers in engineering. By promoting a diverse, innovative academic community, we lay the groundwork for an exciting future in engineering research and education.

Speaker 2: Dr. Chen Chao

Integrated Circuits and Systems for Smart Ultrasound

Biography:

Chen Chao received the B.Sc. degree in microelectronics from Tsinghua University, Beijing, China in 2010, and the M.Sc. (cum laude) and Ph.D. degree in microelectronics from the TU Delft, Delft, The Netherlands, in 2012 and 2018, respectively. From 2017 to 2022, Dr. Chen was with Butterfly Network Inc., New York, USA, where he helped the company commercialize the world's first CMOS/CMUT-based handheld ultrasound probe and became public in 2021. Since 2021, he has also been with Delft University of Technology, where he holds a researcher position focusing on the development of next-generation functional ultrasound imagers. In 2023, Dr. Chen founded Sonosilicon Co., Ltd., in both Hangzhou, China and Delft, the Netherlands, focusing on the development of smart integrated circuits and systems for next-generation ultrasound imaging devices.

Dr. Chen was a recipient of the 2021 ISSCC Anantha P. Chandrakasan Distinguished Technical Paper Award and Demonstration Award, the 2017 IEEE International Ultrasonics Symposium (IUS) Best Student Paper Award and the 2013 ISSCC STGA Award. He was also a co-recipient of the 2020 ISSCC Technology Innovation Award and the 2017 A-SSCC Best Student Paper Award. For his PhD research on front-end ASICs for 3D ultrasound imagers, he received the 2019 Else Kooi Award, a recognition for the annual best Dutch PhD dissertation in microelectronics. Dr. Chen has authored and co-authored more than 30 publications and holds 3 US patents.

Abstract:

As medical ultrasound imaging moves from conventional cart-based scanners to new form factors such as 3D imaging catheters, hand-held point-of-care probes and smart wearable patches, there is an increasing need for integrated circuits that can be closely integrated with the transducer to provide channel-count reduction, improved signal quality, in-probe digitization and edge AI. This talk will start with a brief overview of ultrasound transducer technologies and highlight the significant changes in the system architecture required for moving towards miniatured and wearable 3D imaging devices. The potential of next-generation in-probe electronics will be illustrated by means of examples of state-of-the-art designs featuring transducer-on-CMOS integration and pitch-matched circuits for high-voltage pulsing, beamforming, digitization, and edge computing.

Keynotes

Time: 10:15-11:05, MONDAY JUNE 12, 2023
Venue: Westlake II&III&IV (2F)

The Super Neuron Model – A new generation of ANN-based Machine Learning and Applications

Prof. Moncef Gabbouj / Tampere university

Biography:

Moncef Gabbouj received his BS degree in 1985 from Oklahoma State University, and his MS and PhD degrees from Purdue University, in 1986 and 1989, respectively, all in electrical engineering. Dr. Gabbouj is a Professor of Information Technology at the Department of Computing Sciences, Tampere University, Tampere, Finland. He was Academy of Finland Professor during 2011-2015. His research interests include Big Data analytics, multimedia content-based analysis, indexing and retrieval, artificial intelligence, machine learning, pattern recognition, nonlinear signal and image processing and analysis, voice conversion, and video processing and coding. Dr. Gabbouj is a Fellow of the IEEE and member of the Academia Europaea and the Finnish Academy of Science and Letters. He is the past Chairman of the IEEE CAS TC on DSP and committee member of the IEEE Fourier Award for Signal Processing. He served as associate editor and guest editor of many IEEE, and international journals and Distinguished Lecturer for the IEEE CASS. Dr. Gabbouj served as General Co-Chair of IEEE ISCAS 2019, ICIP 2020, ICIP 2024 and ICME 2021. Gabbouj is Finland Site Director of the USA NSF IUCRC funded Center for Visual and Decision Informatics (CVDI) and led the Artificial Intelligence Research Task Force of Finland's Ministry of Economic Affairs and Employment funded Research Alliance on Autonomous Systems (RAAS).

Abstract:

Operational Neural Networks (ONNs) are new generation network models targeting to address two major drawbacks of conventional Convolutional Neural Networks (CNNs): the homogenous network configuration and the "linear" neuron model that can only perform linear transformations over previous layer outputs. ONNs can perform any linear or non-linear transformation with a proper combination of "nodal" and "pool" operators. This is a great leap towards expanding the neuron's learning capacity in CNNs, which thus far required the use of a single nodal operator for all synaptic connections for each neuron. This restriction has recently been lifted by introducing a superior neuron called the "generative neuron" where each nodal operator can be customized during the training in order to maximize learning. As a result, the network is able to self-organize the nodal operators of its neurons' connections. Self-Organized ONNs (Self-ONNs) equipped with superior generative neurons can achieve diversity even with a compact configuration. We shall explore several signal processing applications of neural network models equipped with the superior neuron.

Time: 11:05-11:55, MONDAY JUNE 12, 2023
Venue: Westlake II&III&IV (2F)

Neuromorphic Intelligence: mixed signal analog/digital implementations of spiking neural networks for real-time signal processing

Prof. Giacomo Indiveri / University of Zurich and ETH Zurich

Biography:

Giacomo Indiveri is a dual professor at the University of Zurich and ETH Zurich, and the director of the Institute of Neuroinformatics, Zurich, Switzerland. He obtained an M.Sc. degree in electrical engineering in 1992 and a Ph.D. degree in computer science from the University of Genoa, Italy in 2004. Engineer by training, Indiveri has also expertise in neuroscience, computer science, and machine learning. He has been combining these disciplines by studying natural and artificial intelligence in neural processing systems and in neuromorphic cognitive agents. His latest research interests lie in the study of spike-based learning mechanisms and recurrent networks of biologically plausible neurons, and in their integration in real-time closed-loop sensory-motor systems designed using analog/digital circuits and emerging memory technologies. Indiveri is senior member of the IEEE society, and a recipient of the 2021 IEEE Biomedical Circuits and Systems Best Paper Award. He is also an ERC fellow, recipient of three European Research Council grants.

Abstract:

Artificial Intelligence (AI) neural networks and machine learning inference accelerators represent a successful technology for solving a wide range of complex tasks. However for many practical purposes that involve fast real-time interactions with the environment these systems still cannot match the performance and efficiency of their biological counterparts. One possible reason lies in the differences between the principles of computation used by nervous systems and those used by conventional time-multiplexed computing systems. In this talk I will present neuromorphic electronic circuits that directly emulate the physics of computation used in animal brains to build neural processing systems which use spike-based representations and brain-inspired adaptation and learning mechanisms. I will show how large-scale multi-core architectures can be built by combining these circuits with asynchronous digital logic ones, and I will present examples of chips that are ideally suited for real-world sensory-processing edge-computing applications.

Time: 10:00-10:50, TUESDAY JUNE 13, 2023
Venue: Westlake II&III&IV (2F)

Integrated Memristor Networks for Higher-complexity Neuromorphic Computing

Prof. YuchaoYang / Peking University

Biography:

Yuchao Yang is a Boya Distinguished Professor at School of Integrated Circuits, Peking University. He serves as Deputy Dean for School of Electronic and Computer Engineering, and Director of Center for Brain Inspired Chips. His research interests include memristors, neuromorphic computing, and inmemory computing. He has published over 130 papers in high-profile journals and conferences such as Nature Electronics, Nature Reviews Materials, Nature Communications, Nature Nanotechnology, Science Advances, Advanced Materials, Nano Letters, IEDM, etc. as well as 5 book chapters. He was invited to give >40 keynote/invited talks on international conferences and serves as TPC chair or member for 9 international conferences. Yuchao Yang serves as the Associate Editor for 3 journals including Microelectronic Engineering, APL Machine Learning and Nano Select, and editorial board member of National Science Review, Chip, Scientific Reports and Science China Information Sciences. He was invited to guest edit 5 special issues and write 12 News & Views, review articles, etc. He is a recipient of the National Outstanding Youth Science Fund, Qiu Shi Outstanding Young Scholar Award, Wiley Young Researcher Award, MIT Technology Review Innovators Under 35 in China, and the EXPLORER PRIZE. He was recognized as Highly Cited Chinese Researchers by Elsevier in 2020 and 2021.

Abstract:

As Moore's law slows down and memory-intensive tasks get prevalent, digital computing becomes increasingly capacity- and power-limited. In order to meet the requirement for increased computing capacity and efficiency in the post-Moore era, emerging computing architectures, such as in-memory computing and neuromorphic computing architectures based on memristors, have been extensively pursued and become an important candidate for new-generation non-von Neumann computers. Since the connection of the theoretical memristor concept with resistive switching devices in 2008, tremendous progress has been made in their applications in-memory and computing systems. Here, we report an optoelectronic synapse that has controllable temporal dynamics under electrical and optical stimuli. Tight coupling between ferroelectric and optoelectronic processes in the synapse can be used to realize heterosynaptic plasticity, with relaxation timescales that are tunable via light intensity or back-gate voltage. We use the synapses to create a multimode reservoir computing system with adjustable nonlinear transformation and multisensory fusion, which is demonstrated using a multimode handwritten digit recognition task and a QR code recognition task. We also realize a multiscale reservoir computing system via the tunable relaxation timescale, which is tested using a temporal signal prediction task.

Industrial Session

Location: West Lake II & III & IV Lobby
Date & Time: Tuesday June 13, 2023 (10:50 - 12:20)
Chair: Jie Yang

Time	Presenter	Title/Abstracts
10:50-11:05	Jing Chou	**T-head: RISC-V Innovations with Matrix Extensions for Neural Networks** Abstract：The RISC-V offers a high degree of flexibility to customise design for AI applications. We proposal an extension that provides a powerful and scalable solution for implementing matrix operations in AI applications, enabling more efficient execution of neural network models across a range of data types and operations.The matrix extension is scalable and efficient to all reasonable design points, and is binary code portable across all possible matrix register sizes. The extension supports multiple data types, for AI training and inference, and is easily extensible for future AI extensions. In addition to hardware designs, the RISC-V Matrix Extension also provides a set of software tools and libraries to help developers optimise their AI applications, further enhancing the performance and efficiency of the extension.
11:05-11:20	Xianguo Zhang, Alibaba Cloud architecture	**Ali-Cloud：Cloud Native CPU New era, Yitian Boost HPC, Bigdata, and AI** Abstract: The new era of Cloud Native CPU, Yitian ARM Boost HPC, Bigdata, and AI. Let's learn together, cloud-native processor how to power the most heave application, AI, HPC, Bigdata. Bruce ZHANG, Alibaba Cloud architecture will show you the technic of hardware and software.
11:20-11:35	Milos Puzovic	**ARM: Accelerating Machine Learning Inference on Arm Neoverse CPUs** Abstract: Over the last few years, Machine Learning (ML), more specifically Deep Learning (DL), has become important workload that runs on wide range of hardware. DL uses frameworks like TensorFlow and PyTorch, which use underlying hardware features for better performance. In this talk, we will talk about how DL frameworks leverage ML-specific Arm Architecture features like BFloat16 for faster inference on Arm Neoverse cores as compared to other platforms.
11:35-11:50	Qiao Ning, CEO	**SynSense: Speck: A fully asynchronous neuromorphic smart dynamic vision SoC** Abstract: With the increasing demand for edge computing solutions and the rise of smart devices relying on sensor processing at the edge, the need for efficient and low-power solutions is paramount. In

		response to this challenge, we introduce Speck, an innovative neuromorphic System-on-Chip (SoC) solution for dynamic vision processing. Speck represents a breakthrough as the world's first neuromorphic chip that seamlessly integrates a dynamic vision sensor and a low-power asynchronous Spiking Neural Network (SNN) ASIC onto a single die. This integration offers an efficient all-in-one architecture, combining sensing and computing capabilities, which results in significant reductions in production costs and power consumption. Furthermore, the event-driven nature of the vision sensor and asynchronous processor enables a high-speed, sparse, and streaming information propagation pipeline. The fully configurable and highly optimized processor within Speck is designed with the sCNN structure, empowering a broad range of dynamic vision tasks. Speck excels at providing intelligent visual processing for various applications, operating at power levels as low as milli-Watts, with response latencies as low as a few milliseconds. This level of performance makes Speck an ideal choice for power-constrained edge computing environments.
11:50-12:05	Bonan Yan, Chief Scientist	**PIMChip: Processing-In-memory (PIM) – A Key Technology for Next-Generation AI** Abstract: Artificial Intelligence of Things (AIoT) systems has been widely used in modern society and is always focused on the trend of power and performance improvement. However, limited by what so as called “memory wall”, most part of energy consumption comes from data movement, edge devices based on conventional von-Neumann structures are insufficient to handle the cost of AIoT computation. In this work we present the structure of SRAM based Processing-In-Memory (SRAM-PIM) due to its high power efficiency. We explain the inside out of SRAM-PIM, from circuit design, SoC implementation, and software development collaboration. Furthermore, we also demonstrate the PIM-Chip ‘Series S’ which aim to achieve multimodal fusion perception and decision-making. Our PIM-Chip ‘Series S’ fulfills the AIoT applications like intelligent education, smart wearables, healthcare, data encryption, and so on. PIM-Chip will be able to build the ecosystem based on the great performance and efficiency of the PIM technology.
12:05-12:20	Renping Yang, Senior Marketing Manager	**KeySight: Test Solution Overview for High-Speed Interface of Computing Chip** Abstract: The wave of ChatGPT is sweeping the world, and the chipset is key support of high arithmetic power. System performance is constantly improving, and high-speed interconnection and storage are the key technologies to support the composition of high arithmetic system. As a global head test and measurement enterprise, Keysight has a deep understanding of chip testing and comprehensive solutions. Sharing will be included as: ① Development and characteristics of arithmetic and AI chips②How Keysight is testing high-speed interconnect(PCIe/CXL) and memory chipset(DDR).

Technical Program: Special Sessions

Computing-in-Memory Circuits & Systems Special Session

Session Type: Lecture
Session Code: A4L-1
Location: West Lake II
Date & Time: Monday June 12, 2023 (13:30 - 14:45)
Chairs: Wei Mao, Jiuren Zhou

Time	Paper ID	Title/Authors
13:30-13:45	6121	A Novel Transpose 2T-DRAM Based Computing-in-Memory Architecture for On-Chip DNN Training and Inference *Yuansheng Zhao, Zixuan Shen, Jiarui Xu, Kevin C.T. Chai, Yanqing Wu, Chao Wang*
13:45-14:00	6086	A Column-Parallel Time-Interleaved SAR/Ss ADC for Computing in Memory with 2-8bit Reconfigurable Resolution *Yuandong Li, Li Du, Yuan Du*
14:00-14:15	6118	RRAM-Based Precision-Scaleable Computing-in-Memory Scheme and its Error Correction Approach *Wenling Ma, Lianzheng Li, Ziye Li, Guangchao Zhao, Xiaojing Long, Mingqiang Huang*
14:15-14:30	6113	Deep Learning Compiler Optimization on Multi-Chiplet Architecture *Huiqing Xu, Kuang Mao, Quihong Pan, Zhaorong Tang, Mengdi Wang, Ying Wang*
14:30-14:45	6143	RISC-V Based Fully-Parallel SRAM Computing-in-Memory Accelerator with High Hardware Utilization and Data Reuse Rate *Haoxiang Zhou, Haiqiao Hong, Dingbang Liu, Hang Liu, Yu Xia, Kai Li, Jun Liu, Shaobo Luo, Wei Mao, Hao Yu*

Processing in Memory Techniques as AI Acccelerators Special Session

Session Type: Lecture
Session Code: A6L-1
Location: West Lake II
Date & Time: Monday June 12, 2023 (16:15 - 17:30)
Chairs: Won Woo Ro, Minsik Kim

Time	Paper ID	Title/Authors
16:15-16:30	6202	In-Memory Activation Compression for GPT Training *Seungyong Lee, Geonu Yun, Hyuk-Jae Lee*
16:30-16:45	6200	Architecture-Aware Optimization of Layer Fusion for Latency-Optimal CNN Inference *Minyong Yoon, Jungwook Choi*
16:45-17:00	6201	Context Swap: Multi-PIM System Preventing Remote Memory Access for Large Embedding Model Acceleration *Hongju Kal, Cheolhwan Kim, Minjae Kim, Won Woo Ro*
17:00-17:15	6190	TRIO: A Novel 10T Ternary SRAM Cell for Area-Efficient in-Memory Computing of Ternary Neural Networks *Thanh Dat Nguyen, Minh Son Le, Thi Nhan Pham, Ik Joon Chang*
17:15-17:30	6189	Ai Processor Based Data Correction for Enhancing Accuracy of Ultrasonic Sensor *Jin Young Shin, Sang Ho Lee, Kwanghyun Go, Soohee Kim, Seung Eun Lee*

Computer Arithmetic for ML Special Session

Session Type: Lecture
Session Code: B1L-1
Location: West Lake II
Date & Time: Tuesday June 13, 2023 (08:30 - 09:45)
Chairs: Thanos Stouraitis, Vassilis Paliouras

Time	Paper ID	Title/Authors
8:30-8:45	6188	Machine Learning Using Logarithmic Arithmetic with Preconditioned Input to Mitchell's Method *Mark Arnold*
8:45-9:00	6197	Modified Logarithmic Multiplication Approximation for Machine Learning *Ioannis Kouretas, Vassilis Paliouras, Thanos Stouraitis*
9:00-9:15	6195	Reduced-Precision Floating-Point Arithmetic in Systolic Arrays with Skewed Pipelines *Dionysios Filippas, Christodoulos Peltekis, Giorgos Dimitrakopoulos, Chrysostomos Nicopoulos*
9:15-9:30	6044	F-CNN: Faster CNN Exploiting Data Re-Use with Statistical Analysis *Fatmah Alantali, Yasmin Halawani, Baker Mohammad, Mahmoud Al-Qutayri*
9:30-9:45	6002	Task-Aware Scheduling and Performance Optimization on Yitian710 SoC for GEMM-Based Workloads on the Cloud *Guosheng Yu, Zhihong Lv, Haijiang Wang, Zilong Huang, Jicheng Chen*

Energy-efficient Circuits & Systems for AI-enabled Biomedical Sensors Special Session

Session Type: Lecture
Session Code: B5L-1
Location: West Lake II
Date & Time: Tuesday June 13, 2023 (13:30 - 14:45)
Chairs: Yang Zhao, Xiaochen Tang

Time	Paper ID	Title/Authors
13:30-13:45	6199	A Lightweight Convolutional Neural Network for Atrial Fibrillation Detection Using Dual-Channel Binary Features from Single-Lead Short ECG *Jiahao Liu, Xiao Liu, Liang Zhou, Liang Chang, Jun Zhou*
13:45-14:00	6176	Classification of ECG Based on Hybrid Features Using CNNs for Wearable Applications *Xiaolin Li, Fang Xiang, Rajesh Panicker, Barry Cardiff, Deepu John*
14:00-14:15	6178	Integrating Delta Modulation and Stochastic Computing for Real-Time Machine Learning Based Heartbeats Monitoring in Wearable Systems *Xiaochen Tang, Shanshan Liu, Farzad Niknia, Wei Tang, Pedro Reviriego, Fabrizio Lombardi*
14:15-14:30	6179	A 12-Lead ECG Delineation Algorithm Based on a Quantized CNN-BiLSTM Auto-Encoder with 1-12 Mapping *Xinzi Xu, Qiao Cai, Hongqian Wang, Yanxing Suo, Yang Zhao, Wan Tianwei, Guoxing Wang, Yong Lian*
14:30-14:45	6187	LungHeart-AtMe: Adventitious Cardiopulmonary Sounds Classification Using MMoE with STFT and MFCCs Spectrograms *Changyan Chen, Qing Zhang, Shirui Sheng, Huajie Huang, Yuhang Zhang, Yongfu Li*

Devices, Circuits & Systems, Algorithms, & Applications

Session Type: Lecture
Session Code: B6L-1
Location: West Lake II
Date & Time: Tuesday June 13, 2023 (16:00 - 17:15)
Chairs: Shuenn-Yuh Lee, Xin Zi Xu

Time	Paper ID	Title/Authors
16:00-16:15	6082	CSwin2SR: Circular Swin2SR for Compressed Image Super-Resolution *Honggui Li, Maria Trocan, Mohamad Sawan, Dimitri Galayko*
16:15-16:30	6142	Efficiency Comparison of Machine Learning Algorithms for EEG Interpretation *Xia Han, Frédéric Amiel, Xun Zhang, Kunni Wei, Cong Yan, Wenjun Hu, Zefeng Wang*
16:30-16:45	6193	A High Performance Accelerating CNN Inference on FPGA with Arrhythmia Classification *Ming-Yueh Ku, Tai-Siang Zhong, Yi-Ting Hsieh, Shuenn-Yuh Lee, Ju-Yi Chen*
16:45-17:00	6028	Hardware-Friendly Activation Function Designs and its Efficient VLSI Implementations for Transformer-Based Applications *Yu Hsiang Huang, Pei Hsuan Kuo, Juinn Dar Huang*
17:00-17:15	6109	A Memristor-Inspired Computation for Epileptiform Signals in Spheroids *Iván Díez de Los Ríos, John Wesley Ephraim, Gemma Palazzolo, Teresa Serrano-Gotarredona, Gabriella Panuccio,* Bernabé Linares-Barranco

Technical Program: Lectures

Compute-in-Memory

Session Type: Lecture
Session Code: A1L-1
Location: West Lake II
Date & Time: Monday June 12, 2023 (08:30 - 09:45)
Chairs: Xueqing Li, Bing Li

Time	Paper ID	Title/Authors
8:30-8:45	6048	A Fully Differential 4-Bit Analog Compute-in-Memory Architecture for Inference Application *Dinesh Kushwaha, Rajat Kohli, Jwalant Mishra, Rajiv V. Joshi, Sudeb Dasgupta, Anand Bulusu*
8:45-9:00	6065	A 115.1 TOPS/W, 12.1 TOPS/mm2 Computation-in-Memory Using Ring-Oscillator Based ADC for Edge Ai *Abhairaj Singh, Rajendra Bishnoi, Ali Kaichouhi, Sumit Shaligram Diware, Rajiv V. Joshi, Said Hamdioui*
9:00-9:15	6106	Memory-Immersed Collaborative Digitization for Area-Efficient Compute-in-Memory Deep Learning *Shamma Nasrin, Maeesha Binte Hashem, Nastaran Darabi, Benjamin Parpillon, Farah Fahim, Wilfred Gomes, Amit Ranjan Trivedi*
9:15-9:30	6061	Optimization Strategies for Digital Compute-in-Memory from Comparative Analysis with Systolic Array *Wantong Li, Junmo Lee, Shimeng Yu*
9:30-9:45	6141	A Systolic Computing-in-Memory Array Based Accelerator with Predictive Early Activation for Spatiotemporal Convolutions *Xiaofeng Chen, Ruiqi Guo, Zhiheng Yue, Yang Hu, Leibo Liu, Shaojun Wei, Shouyi Yin*

Methodologies of Neuromorphic Design

Session Type: Lecture
Session Code: A1L-2
Location: West Lake III
Date & Time: Monday June 12, 2023 (08:30 - 09:45)
Chairs: Xinfei Guo, Chao Wang

Time	Paper ID	Title/Authors
8:30-8:45	6030	RC-GNN: Fast and Accurate Signoff Wire Delay Estimation with Customized Graph Neural Networks *Linyu Zhu, Yue Gu, Xinfei Guo*
8:45-9:00	6112	FEEP: Functional ECO Synthesis with Efficient Patch Minimization *Yaotian Liu, Yuhang Zhang, Qing Zhang, Rui Chen, Yongfu Li*
9:00-9:15	6079	Reducing Overhead of Feature Importance Visualization via Static GradCAM Computation *Ashwin Bhat, Arijit Raychowdhury*
9:15-9:30	6081	WeightLock: A Mixed-Grained Weight Encryption Approach Using Local Decrypting Units for Ciphertext Computing in DNN Accelerators *Jianfeng Wang, Zhonghao Chen, Yiming Chen, Yixin Xu, Tianyi Wang, Yao Yu, Vijaykrishnan Narayanan, Sumitha George, Huazhong Yang , Xueqing Li.*
9:30-9:45	6157	Securing Decision Tree Inference Using Order-Preserving Cryptography *Rupesh Karn, Kashif Nawaz, Ibrahim Elfadel*

Bio Applications

Session Type: Lecture
Session Code: A1L-3
Location: West Lake IV
Date & Time: Monday June 12, 2023 (08:30 - 09:45)
Chairs: Yang Zhao, Xin Zi Xu

Time	Paper ID	Title/Authors
8:30-8:45	6163	Binary Is All You Need: Ultra-Efficient Arrhythmia Detection with a Binary-Only Compressive System *Fengshi Tian, Xiaomeng Wang, Jinbo Chen, Jie Yang, Mohamad Sawan, Chi-Ying Tsui, Kwang-Ting Tim Cheng*
8:45-9:00	6085	A Convolved Self-Attention Model for IMU-Based Gait Detection and Human Activity Recognition *Shuailin Tao, Wang Ling Goh, Yuan Gao*
9:00-9:15	6007	EpilepsyNet: Interpretable Self-Supervised Seizure Detection for Low-Power Wearable Systems *Baichuan Huang, Renato Zanetti, Azra Abtahi, David Atienza, Amir Aminifar*
9:15-9:30	6021	Bandit-Supported Care Planning for Older People with Complex Health and Care Needs *Gi-Soo Kim, Young Suh Hong, Tae Hoon Lee, Myunghee Cho Paik, Hongsoo Kim*
9:30-9:45	6107	A Convolutional Spiking Network for Gesture Recognition in Brain-Computer Interfaces *Yiming Ai, Bipin Rajendran*

Neuromorphic Circuits & Systems 1

Session Type: Lecture
Session Code: A4L-2
Location: West Lake III
Date & Time: Monday June 12, 2023 (13:30 - 14:45)
Chairs: Zhuo Zou, Jian Zhao

Time	Paper ID	Title/Authors
13:30-13:45	6023	Integrated System-on-Module for Design-Space Exploration of Spiking Neural Networks *Moamen El-Masry, Taher Kourany, Rex Kho, Thilo Werner, Amir Zjajo, Robert Weigel*
13:45-14:00	6147	Validation of a CMOS SNN Network Based on a Time-Domain Threshold Neuron Circuit Achieving 114.90 pJ/Inference on MNIST *Diego Antolin Garcia Soto, Javier Granizo Cuadrado, Luis Hernandez Corporales*
14:00-14:15	6135	Neuromorphic Analog Circuits for Robust On-Chip Always-on Learning in Spiking Neural Networks *Arianna Rubino, Matteo Cartiglia, Melika Payvand, Giacomo Indiveri*
14:15-14:30	6151	Unsupervised Learning of Spike-Timing-Dependent Plasticity Based on a Neuromorphic Implementation *Yi Zhong, Zilin Wang, Xiaoxin Cui, Jian Cao, Yuan Wang*
14:30-14:45	6172	FrameFire: Enabling Efficient Spiking Neural Network Inference for Video Segmentation *Qinyu Chen, Congyi Sun, Chang Gao, Xinyuan Fang, Haitao Luan*

Visual Algorithms

Session Type: Lecture
Session Code: A4L-3
Location: West Lake IV
Date & Time: Monday June 12, 2023 (13:30 - 14:45)
Chairs: Shiqi Wang, Jainaveen Sundaram

Time	Paper ID	Title/Authors
13:30-13:45	6045	Landmark-Based Adversarial Network for RGB-D Pose Invariant Face Recognition *Wei-Jyun Chen, Ching-Te Chiu, Ting-Chun Lin*
13:45-14:00	6115	An Interpretable Pixel Intensity Reconstruction Model for Asynchronous Event Camera *Hongwei Shan, Lichen Feng, Yueqi Zhang, Zhangming Zhu*
14:00-14:15	6114	Object-Augmented Skeleton-Based Action Recognition *Zhengyu Li, Heng Guo, Lap Pui Chau, Cheen Hau Tan, Xiaoxi Ma, Dan Lin, Kim-Hui Yap*
14:15-14:30	6049	CPGAN: Collective Punishment Generative Adversarial Network for Dry Fingerprint Image Enhancement *Yu-Chi Su, Ching-Te Chiu, Chih-Han Cheng, Kuan-Hsien Liu, Tsung-Chan Lee, Jia-Lin Chen, Jie-Yu Luo, Wei-Chang Chung, Yao-Ren Chang, Kuan-Ying Ho*
14:30-14:45	6111	Image Recovery Through Scattering Media via GAN Reconstruction and SNES Optimization *Pengfei Qi, Yuanjin Zheng*

Neuromorphic Circuits & Systems 2

Session Type: Lecture
Session Code: A6L-2
Location: West Lake III
Date & Time: Monday June 12, 2023 (16:15 - 17:30)
Chairs: Jian Zhao, Zhuo Zou

Time	Paper ID	Title/Authors
16:15-16:30	6131	PN-TMS: Pruned Node-Fusion Tree-Based Multicast Scheme for Efficient Neuromorphic Systems *Ziyang Shen, Chaoming Fang, Fengshi Tian, Jie Yang, Mohamad Sawan*
16:30-16:45	6174	SNNOpt: An Application-Specific Design Framework for Spiking Neural Networks *Jingyu He, Ziyang Shen, Fengshi Tian, Jinbo Chen, Jie Yang, Mohamad Sawan, Kwang-Ting Tim Cheng, Paul Bogdan, Chi-Ying Tsui*
16:45-17:00	6154	Embedded Neuromorphic Attention Model Leveraging a Novel Low-Power Heterogeneous Platform *Amélie Gruel, Alfio Di Mauro, Robin Hunziker, Luca Benini, Jean Martinet, Michele Magno*
17:00-17:15	6020	Synaptic Metaplasticity with Multi-Level Memristive Devices *Simone D'Agostino, Filippo Moro, Tifenn Hirtzlin, Julien Arcamone, Niccolò Castellani, Damien Querlioz, Melika Payvand, Elisa Vianello*
17:15-17:30	6152	Mapping-Aware Biased Training for Accurate Memristor-Based Neural Networks *Sumit Shaligram Diware, Anteneh Gebregiorgis, Rajiv V. Joshi, Said Hamdioui, Rajendra Bishnoi*

Visual Applications

Session Type: Lecture
Session Code: A6L-3
Location: West Lake IV
Date & Time: Monday June 12, 2023 (16:15 - 17:30)
Chair: Shiqi Wang, Qing Song Xie

Time	Paper ID	Title/Authors
16:15-16:30	6095	E-Track: Eye Tracking with Event Camera for Extended Reality (XR) Applications *Nealson Li, Ashwin Bhat, Arijit Raychowdhury*
16:30-16:45	6126	KP2Dtiny: Quantized Neural Keypoint Detection and Description on the Edge *Thomas Rüegg, Marco Giordano, Michele Magno*
16:45-17:00	6101	FPGA-Based High-Speed and Resource-Efficient 3D Reconstruction for Structured Light System *Feng Bao, Zehua Dong, Jie Yu, Songping Mai*
17:00-17:15	6171	AI-Assisted ISP Hyperparameter Auto Tuning *Fa Xu, Zihao Liu, Yanheng Lu, Sicheng Li, Susong Xu, Yibo Fan, Yen-Kuang Chen*
17:15-17:30	6003	Image Frequency Separation Residual Network for End-to-End Raw to RGB Mapping *Mengchuan Dong, Wei Zhou, Cong Pang, Xiangyu Zhang, Xin Lou*

Emerging Neuromorphic Paradigms

Session Type: Lecture
Session Code: B1L-2
Location: West Lake III
Date & Time: Tuesday June 13, 2023 (08:30 - 09:45)
Chairs: Chao Wang, Yuan Du

Time	Paper ID	Title/Authors
8:30-8:45	6039	Efficient Algorithms for Accelerating Spiking Neural Networks on MAC Array of SpiNNaker 2 *Jiaxin Huang, Florian Kelber, Bernhard Vogginger, Binyi Wu, Felix Kreutz, Pascal Gerhards, Daniel Scholz, Klaus Knobloch, Christian Georg Mayr*
8:45-9:00	6070	Read-Disturb Detection Methodology for RRAM-Based Computation-in-Memory Architecture *Mohammad Amin Yaldagard, Sumit Shaligram Diware, Rajiv V. Joshi, Said Hamdioui, Rajendra Bishnoi*
9:00-9:15	6063	Online Low-Power Large-Scale Real-Time Decision-Making All at Once *Thomas Pontoizeau, Éric Jacopin*
9:15-9:30	6130	NeuroBMI: A New Neuromorphic Implantable Wireless Brain Machine Interface with a 0.48 µW Event-Driven Noise-Tolerant Spike Detector *Jinbo Chen, Hui Wu, Xing Liu, Razieh Eskandari, Fengshi Tian, Wenjun Zou, Chaoming Fang, Jie Yang, Mohamad Sawan*
9:30-9:45	6096	Bringing Touch to the Edge: A Neuromorphic Processing Approach for Event-Based Tactile Systems *Harshil Patel, Anup Vanarse, Kristofor D. Carlson, Adam Osseiran*

Efficient Algorithms

Session Type: Lecture
Session Code: B1L-3
Location: West Lake IV
Date & Time: Tuesday June 13, 2023 (08:30 - 09:45)
Chairs: Yongfu Li, Yuhang Zhang

Time	Paper ID	Title/Authors
8:30-8:45	6145	Temporal Similarity-Based Computation Reduction for Video Transformers in Edge Camera Nodes *Udari De Alwis, Zhongheng Xie, Massimo Alioto*
8:45-9:00	6056	Free Bits: Latency Optimization of Mixed-Precision Quantized Neural Networks on the Edge *Georg Rutishauser, Francesco Conti, Luca Benini*
9:00-9:15	6072	TinyissimoYOLO: A Quantized, Low-Memory Footprint, TinyML Object Detection Network for Low Power Microcontrollers *Julian Moosmann, Marco Giordano, Christian Vogt, Michele Magno*
9:15-9:30	6116	Searching Tiny Neural Networks for Deployment on Embedded FPGA *Haiyan Qin, Yejun Zeng, Jinyu Bai, Wang Kang*
9:30-9:45	6068	Energy Efficient Software-Hardware Co-Design of Quantized Recurrent Convolutional Neural Network for Continuous Cardiac Monitoring *Jinhai Hu, Cong Sheng Leow, Wang Ling Goh, Yuan Gao*

Hardware Accelerator

Session Type: Lecture
Session Code: B5L-2
Location: West Lake III
Date & Time: Tuesday June 13, 2023 (13:30 - 14:45)
Chairs: Guanghui He, Xueqing Li

Time	Paper ID	Title/Authors
13:30-13:45	6034	Bit-Offsetter: A Bit-Serial DNN Accelerator with Weight-Offset MAC for Bit-Wise Sparsity Exploitation *Siqi He, Hongyi Zhang, Mengjie Li, Haozhe Zhu, Chixiao Chen, Qi Liu, Xiaoyang Zeng*
13:45-14:00	6024	A 40nm Area-Efficient Effective-Bit-Combination-Based DNN Accelerator with the Reconfigurable Multiplier *Yanghan Zheng, Zhaofang Li, Kaihang Sun, Kuan-Pei Lee, Kea-Tiong Tang*
14:00-14:15	6013	Low-Power Convolutional Neural Network Accelerator on FPGA *Kasem Khalil, Ashok Kumar, Magdy Bayoumi*
14:15-14:30	6058	Group Vectored Absolute-Value-Subtraction Cell Array for the Efficient Acceleration of AdderNet *Jiahao Chen, Wanbo Hu, Wenling Ma, Zhilin Zhang, Mingqiang Huang*
14:30-14:45	6136	Configurable Multi-Precision Floating-Point Multiplier Architecture Design for Computation in Deep Learning *Pei Hsuan Kuo, Yu Hsiang Huang, Juinn Dar Huang*

Learning Algorithms Special Session

Session Type: Lecture
Session Code: B5L-3
Location: West Lake IV
Date & Time: Tuesday June 13, 2023 (13:30 - 14:45)
Chairs: Yongfu Li, Yuhang Zhang

Time	Paper ID	Title/Authors
13:30-13:45	6094	GPIL: Gradient with PseudoInverse Learning for High Accuracy Fine-Tuning *Gilha Lee, Nam Joon Kim, Hyun Kim*
13:45-14:00	6059	SLIM-Net: Rethinking How Neural Networks Use Systolic Arrays *Thomas Dalgaty, Maria Lepecq*
14:00-14:15	6132	Online Spatio-Temporal Learning with Target Projection *Thomas Ortner, Lorenzo Pes, Joris Gentinetta, Charlotte Frenkel, Angeliki Pantazi*
14:15-14:30	6014	LG-LSQ: Learned Gradient Linear Symmetric Quantization for Low-Precision Integer Hardware *Shih-Ting Lin, Zhaofang Li, Yu-Hsiang Cheng, Hao-Wen Kuo, Rui-Hsuan Wang, Nai-Jen Sung, Chih-Cheng Lu, Kea-Tiong Tang*
14:30-14:45	6066	HNSG – A SNN Training Method Ultilizing Hidden Network *Chunhui Wu, Wenbing Fang, Yi Kang*

Algorithm-Hardware Co-design Special Session

Session Type: Lecture
Session Code: B6L-2
Location: West Lake III
Date & Time: Tuesday June 13, 2023 (16:00 - 17:15)
Chairs: Bing Li, Guanghui He

Time	Paper ID	Title/Authors
16:00-16:15	6037	SALSA: Simulated Annealing Based Loop-Ordering Scheduler for DNN Accelerators *Victor Jung, Arne Symons, Linyan Mei, Marian Verhelst, Luca Benini*
16:15-16:30	6137	Simulation-Driven Latency Estimations for Multi-Core Machine Learning Accelerators *Yannick Braatz, Dennis Sebastian Rieber, Taha Soliman, Oliver Bringmann*
16:30-16:45	6055	A Systolic Array with Activation Stationary Dataflow for Deep Fully-Connected Networks *Haochuan Wan, Chaolin Rao, Yueyang Zheng, Pingqiang Zhou, Xin Lou*
16:45-17:00	6074	TPE : A High-Performance Edge-Device Inference with Multi-Level Transformational Mechanism *Zhou Wang, Jingchuan Wei, Xiaonan Tang, Boxiao Han, Hongjun He, Leibo Liu, Shaojun Wei, Shouyi Yin*
17:00-17:15	6071	A Hardware-Centric Approach to Increase and Prune Regular Activation Sparsity in CNNs *Tim Hotfilter, Julian Hoefer, Fabian Kreß, Fabian Kempf, Leonhard Kraft, Tanja Harbaum, Jürgen Becker*

General Applications Special Session

Session Type: Lecture
Session Code: B6L-3
Location: West Lake IV
Date & Time: Tuesday June 13, 2023 (16:00 - 17:15)
Chairs: Wei Mao, Qing Song Xie

Time	Paper ID	Title/Authors
16:00-16:15	6177	An Integrated CPU-GPU Frequency Scaling Governor Based on Deep Recurrent Q-Network for Partially Observable Rendering Applications *Qinxin Zhou, Yang Zhao, Wenkai Zhang*
16:15-16:30	6146	A Byte Sequence Is Worth an Image: CNN for File Fragment Classification Using Bit Shift and n-Gram Embeddings *Wenyang Liu, Yi Wang, Kejun Wu, Kim-Hui Yap, Lap Pui Chau*
16:30-16:45	6125	Deep-Learning-Based X-Ray CT Slice Analysis for Layout Verification in Printed Circuit Boards *Deruo Cheng, Yiqiong Shi, Yee-Yang Tee, Jingsi Song, Xue Wang, Bihan Wen, Bah-Hwee Gwee*
16:45-17:00	6026	Multi-Head Attention Based Bi-LSTM for Anomaly Detection in Multivariate Time-Series of WSN *Mustafa Matar, Tian Xia, Kimberly Huguenard, Dryver Huston, Safwan Wshah*
17:00-17:15	6167	PCB Identification Based on Machine Learning Utilizing Power Consumption Variability *Anupam Golder, Arijit Raychowdhury*

Technical Program: Posters

Tiny, Efficient, and Engineerable Machine Learning Special Session

Session Code: B4P-4
Location: West Lake II & III & IV Lobby
Date & Time: Tuesday June 13, 2023 (14:45 - 15:45)
Chairs: Renyuan Zhang, Yirong Kan

Time	Paper ID	Title/Authors
14:45-15:00	6138	Performance Assessment of an Extremely Energy-Efficient Binary Neural Network Using Adiabatic Superconductor Devices *Olivia Chen, Zhengang Li, Tomoharu Yamauchi, Yanzhi Wang, Nobuyuki Yoshikawa*
15:00-15:15	6078	Efficient Parameter Learning of Bayesian Network with Latent Variables from High-Dimensional Data *Xinran Wu, Xiang Chen, Kun Yue*
15:15-15:30	6162	Three Challenges in ReRAM-Based Process-in-Memory for Neural Network *Ziyi Yang, Kehan Liu, Yiru Duan, Mingjia Fan, Qiyue Zhang, Zhou Jin*
15:30-15:45	6089	Computer-Aided-Prediction of Body Constitution with Efficient Cock-Tail Learning *Guang Shi, Yirong Kan, Renyuan Zhang*

Poster Session

Session Type: Poster
Session Code: B4P-5
Location: West Lake II & III & IV Lobby
Date & Time: Tuesday June 13, 2023 (14:45 - 15:45)
Chairs: Yuhang Zhang, Xin Zi Xu

Paper ID	Title/Authors
6017	Novel Knowledge Distillation to Improve Training Accuracy of Spin-Based SNN *Hanrui Li, Aijaz Lone, Fengshi Tian, Jie Yang, Mohamad Sawan, Nazek El-Atab*
6022	4b/4b/8b Precision Charge-Domain 8T-SRAM Based CiM for CNN Processing *Qibang Zang, Wang Ling Goh, Yi Sheng Chong, Anh Tuan Do*
6035	A Low-Power Hardware Accelerator of MFCC Extraction for Keyword Spotting in 22nm FDSOI *Liyuan Guo, Matthias Jobst, Johannes Partzsch, Stefan Scholze, Andreas Dixius, Matthias Lohrmann, Seyed Mohammad Ali Zeinolabedin, Christian Georg Mayr*
6064	High-Accuracy and Energy-Efficient Acoustic Inference Using Hardware-Aware Training and a 0.34 nW/Ch Full-Wave Rectifier *Sheng Zhou, Xi Chen, Kwantae Kim, Shih-Chii Liu*
6091	A Ternary Weight Mapping and Charge-Mode Readout Scheme for Energy Efficient FeRAM Crossbar Compute-in-Memory System *Tiancheng Cao, Zhongyi Zhang, Wang Ling Goh, Chen Liu, Yao Zhu, Yuan Gao*
6164	A 1W8R 20T SRAM Codebook for 20% Energy Reduction in Mixed-Precision Deep-Learning Inference Processor System *Ryotaro Ohara, Kabuto Masaya, Masakazu Taichi, Astushi Fukunaga, Yuto Yasuda, Riku Hamabe, Shintaro Izumi, Hiroshi Kawaguchi*
6042	An Efficient Design Framework for 2×2 CNN Accelerator Chiplet Cluster with SerDes Interconnects *Yajie Wu, Tianze Li, Zhuang Shao, Li Du, Yuan Du*
6062	MF-DSNN:An Energy-Efficient High-Performance Multiplication-Free Deep Spiking Neural Network Accelerator *Yue Zhang, Shuai Wang, Yi Kang*
6117	A Hierarchically Reconfigurable SRAM-Based Compute-in-Memory Macro for Edge Computing *Runxi Wang, Xinfei Guo*
6047	SpatialHD: Spatial Transformer Fused with Hyperdimensional Computing for AI Applications *Meriem Bettayeb, Eman Hassan, Baker Mohammad, Hani Saleh*
6076	Multi-Agent Cooperative Control in Neural MMO Environment Based on MAPPO Algorithm *Gengcheng Lyu, Meng Li*
6087	PPT-KP: Pruning Point Training-Based Kernel Pruning for Deep Convolutional Neural Networks *Kwanghyun Koo, Hyun Kim*
6161	Convergent Waveform Relaxation Schemes for the Transient Analysis of Associative ReLU Arrays *Ibrahim Elfadel*
6102	Enhancing Fault Resilience of QNNs by Selective Neuron Splitting *Mohammad Hasan Ahmadilivani, Mahdi Taheri, Jaan Raik, Masoud Daneshtalab, Maksim Jenihhin*
6104	Grand Challenge on Software and Hardware Co-Optimization for E-Commerce Recommendation System *Jianing Li, Jiabin Liu, Xingyuan Hu, Yuhang Zhang, Guosheng Yu, Shimeng Qian, Wei Mao, Li Du, Yongfu Li, Yuan Du*

Live Demos

Session Type: Live Demos
Session Code: A5P-4
Location: West Lake II & III & IV Lobby
Date & Time: Monday June 12, 2023 (14:45 - 16:00)
Chairs: Guoxing Wang, Feng Wei An

Paper ID	Title/Authors
6108	Live Demonstration: An Efficient Neural Network Processor with Reduced Data Transmission and On-Chip Shortcut Mapping *Yichuan Bai, Zhuang Shao, Chenshuo Zhang, Aojie Jiang, Yuan Du, Li Du*
6123	Live Demonstration: an Integrated Computing and Communication Platform for Vehicle-Infrastructure Cooperative Autonomous Driving *Yuhang Gu, Wei Zhang, Yi Shi, Limin Jiang, Shan Li, Shan Cao, Zhiyuan Jiang, Ruiqing Mao, Zhewen Lou, Sheng Zhou*
6160	Live Demonstration: Supervised-Learning-Based Visual Quantification for Image Enhancement *Wei Zhang, Junfeng Chang, Zizhao Peng, Lei Chen, Fengwei An*
6180	Live Demonstration: Real-Time Analyses of Biosignals Based on a Dedicated CMOS Configurable Deep Learning Engine *Junzhe Wang, Shiqi Zhao, Chaoming Fang, Jie Yang, Mohamad Sawan*
6181	Live Demonstration: Efficient Organic Photodetector Based Active Matrix Imager for Real-Time Optical Character Recognition *Tong Shan, Jun Li, Xiao Hou, Peijin Huang, Xiaojun Guo*
6182	Live Demonstration: Face Recognition at the Edge Using Fast On-Chip Deep Learning Neuromorphic Chip *Zhengqing Zhong, Tengxiao Wang, Haibing Wang, Zhihua Zhou, Junxian He, Fang Tang, Xichuan Zhou, Shuang-Ming Yu, Liyuan Liu, Nanjian Wu, Min Tian, Cong Shi*
6184	An Energy-Efficient and Reconfigurable CNN Accelerator Applied to Lung Cancer Detection *Yihsin Liao, Shuyou Lin, Yuchiao Chen, Dingxiao Wu, Hong Wen Luo, Hsinhan Chen, Kea-Tiong Tang*
6185	Live Demonstration: SRAM Compute-in-Memory Based Visual & Aural Recognition System *Anjunyi Fan, Bo Hu, Zhonghua Jin, Haiyue Han, Yaojun Zhang, Yue Yang, Yuchao Yang, Bonan Yan, Ru Huang*
6186	A Demonstration Platform for Large-Scaled Point Cloud Network Based on 28nm 2D/3D Unified Sparse Convolution Accelerator *Xiaoyu Feng, Wenyu Sun, Shupei Fan, Chen Tang, Yixiong Yang, Jinshan Yue, Qingmin Liao, Huazhong Yang, Yongpan Liu*
6183	Live Demonstration: A Smart Ring for Continuous Health Data Monitoring Based on Photoplethysmography *Bin Liu, Hao Wu, Guoxing Wang*

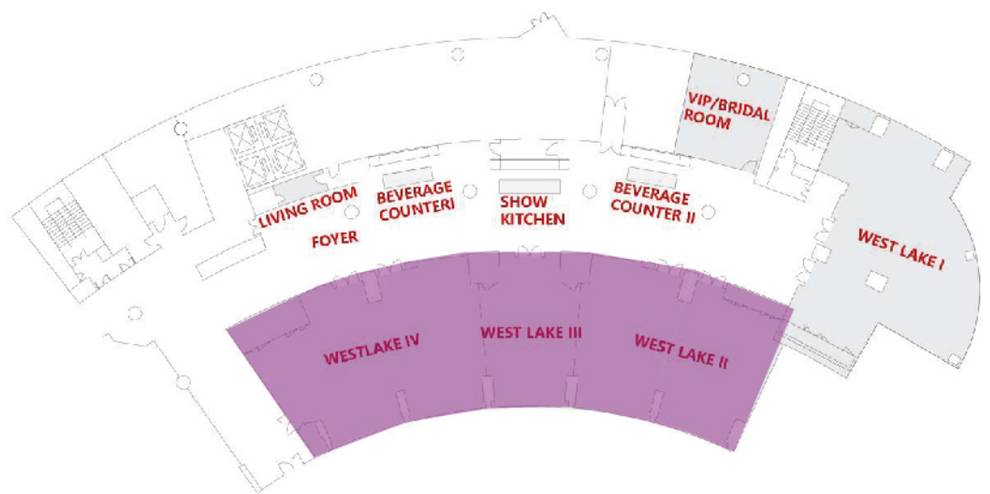
Conference Information

Venue

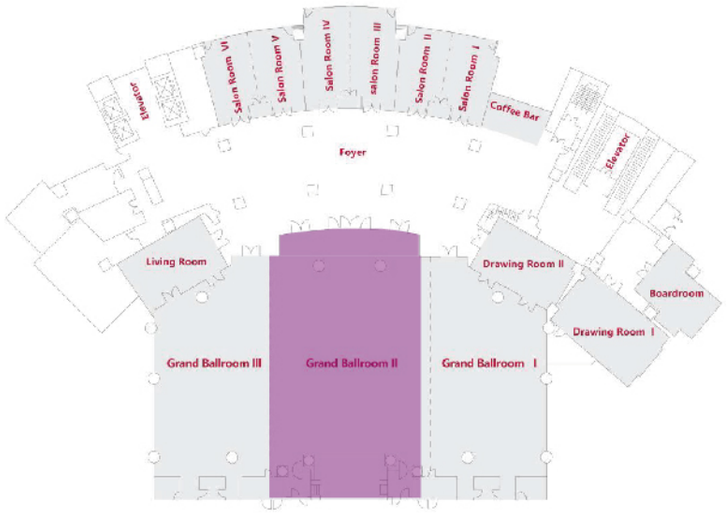
Grand Hyatt Hangzhou
Address: 28 Hu Bin Road, Hangzhou, China, 310006
Tel:+86 571 8779 1234
Website: https://www.hyatt.com/en-US/hotel/china/grand-hyatt-hangzhou/hangz

Venue Map

2F Meeting Room



B2 Banquet Grand Ballroom



Lunch

Lunch will be provided to all the on-site attendees.

Date	Monday, June 12, 2023	Tuesday, June 13, 2023
Time	12:00-13:30	12:20-13:30
Venue	Grand Café (Buffet) (1F)	

Please make sure that you have to the lunch coupon in your conference bag. Please be noted that you have to keep the coupon well and show it to the restaurants when having lunch.

Banquet

Banquet will be provided to all the on-site attendees on the evening of June 12.

Date	Monday, June 12, 2023
Time	18:30-20:30
Venue	Grand Hyatt Lobby（B2）

Please make sure that you have to show the Banquet coupon in your conference bag. Please be noted that you have to keep the coupon well and show it to the staff when having Banquet.

Coffee Break

Coffee and Cookies will be served at the following hours.

Date	Sunday, June 11,2023		Monday, June 12, 2023		Tuesday, June 13, 2023	
Time	10:00-10:30	17:00-18:00	9:45-10:00	16:00-16:15	9:45-10:00	15:45-16:00
Venue	Grand Hyatt Lobby (2F)					

Sponsors

Diamond

T-HEAD

Gold

Alibaba Cloud arm

Silver

无锡国家“芯火”双创基地(平台)
WUXI NATIONAL XINHUO INNOVATION BASE(PLATFORM)

Bronze

PiMCHIP 芯动科技 FLYSLICE SynSense 时识科技 KEYSIGHT TECHNOLOGIES Unlocking Measurement Insights for 75 Years
DFZK 中科曙光 Sugon 佰欧生物 BAIOU BIOLOGICAL

Support

IEEE CAS 西湖大学 WESTLAKE UNIVERSITY Cen BRAIN NEURO Tech 杭州

T-HEAD

Company Profile

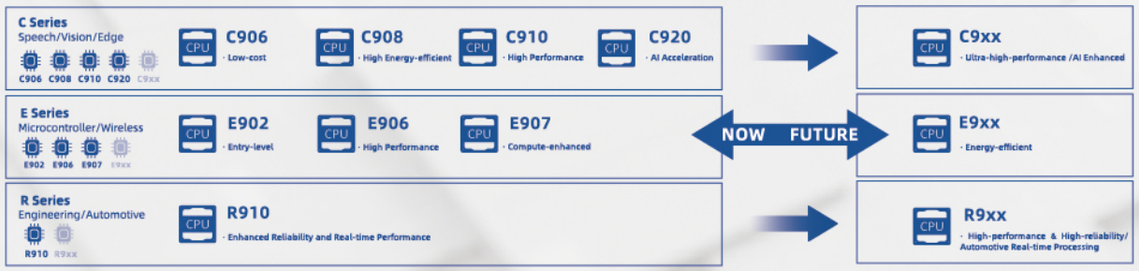
Established in September 2018, T-Head Semiconductor Co., Ltd. is a wholly-owned semiconductor business entity of Alibaba Group. T-head's products range from data center chips, IoT chips to CPU IPs covering both cloud and edge application scenarios. T-Head is dedicated to drive innovation with technology and to embrace digital intelligence future with the chip power.

Server CPU Yitian 710

Yitian 710 is the first data center CPU developed by T-Head. It's an industry-leading cloud-native processor with robust performance, excellent energy efficiency, and high bandwidth. At present, Yitian 710 has been applied in multiple business scenarios within Alibaba Group and also been serving a number of Internet technology companies through Alibaba cloud.

Xuantie CPU IP

Xuantie CPU IP is the cornerstone of intelligent, secure and edge-cloud-integrated chip architecture, providing the computing core for the digital age. Xuantie has adhered to the path of self-development of core technologies ever since its birth. Its new series of products adopt the open source RISC-V architecture and are widely used in various fields such as computing vision, data storage, industrial IoT, network communication, smart home, and biometrics. Up to now, total licensed chips shipped has exceeded 3 billion pieces.



https://www.t-head.cn

arm

Arm Technology is Defining the Future of Computing

From the sensor to the smartphone and the supercomputer.
The future is built on Arm.

关注 Arm 社区
了解更多资讯