AICAS 2021 Program Outline

June 6, 2021 (Sunday)	8:50 AM - 9:00 AM	**Opening**
	9:00 AM - 9:45 AM	**Keynote 01**
		Software-defined AI Chip – with the Emphasis on Architecture Innovation *Speaker: Shaojun Wei*
	9:50 AM - 10:35 AM	**Keynote 02**
		High-Performance and Energy-Efficient Circuit Technologies for Sub-5nm In-Memory/Near-Memory AI Accelerators *Speaker: Ram K. Krishnamurthy*
	10:40 AM - 11:15 AM	**Regular Session 01**
		Hardware for AI *Chairs: Jaydeep Kulkarni, Linghao Song*
	11:20 AM - 11:55 AM	**Regular Session 02**
		Neuro-inspired Circuits and Systems *Chairs: Jongsun Park, Biswas Dwaipayan*
	12:00 PM - 12:30 PM	**Tutorials**
June 7, 2021 (Monday)	9:00 AM - 9:45 AM	**Keynote 03**
		On-device AI for Augmented Reality (AR) Systems: Challenges and Opportunities *Speaker: Vikas Chandra*
	9:50 AM - 10:35 AM	**Regular Session 03**
		AI Accelerators *Chairs: Shimeng Yu, Bing Li*
	10:40 AM - 11:15 AM	**Special Session 01**
		Health Monitor & Autonomous Vehicle *Chairs: Kea-Tiong Tang, Tinoosh Mohsenin*
	11:20 AM - 11:55 AM	**Special Session 02**
		Design Automation & Compute-In-Memory *Chairs: Shimeng Yu, Debjyoti Bhattacharjee*
	12:00 PM - 12:30 PM	**Demo Session**
June 8, 2021 (Tuesday)	9:00 AM - 9:45 AM	**Keynote 04**
		Emerging AI Processing for Wireless Applications *Speaker: Young-Kai (Y.K.) Chen*
	9:50 AM - 10:35 AM	**Regular Session 04**
		Learning Models and Applications of Intelligent Systems *Chairs: John Paul Strachan, Hyungjun Kim*
	10:40 AM - 11:15 AM	**Special Session 03**
		Efficient Hardware Accelerator Design for ML: Architectural and Algorithmic Optimization *Chairs: Wei Zhang, Kea-Tiong Tang*
	11:20 AM - 11:55 AM	**Special Session 04**
		Micro AI and Low Power Autonomous Systems *Chair: Tinoosh Mohsenin*
June 9, 2021 (Wednesday)	9:00 AM - 9:45 AM	**Keynote 05**
		AI and Video Compression in the Era of Internet of Video Things *Speaker: Yen-Kuang Chen*
	9:50 AM - 10:35 AM	**Regular Session 05**
		Algorithm-Hardware Co-Design of ML Systems *Chairs: Somnath Paul, Jong-Hyeok Yoon*
	10:40 AM - 11:55 AM	**Special Session 05**
		Low Power Computer Vision Challenge *Chair: Yung-Hsiang Lu*
	12:00 PM - 12:30 PM	**Closing, Best Paper Award, AICAS 2022**

KEYNOTE 01

9:00am ~ 9:45am, Sunday June 6, 2021

Software-defined AI Chip - with the Emphasis on Architecture Innovation

Shaojun Wei

Tsinghua University, China

Abstract

Over the past decades, Application Specific Integrated Circuit (ASIC) is employed to meet specific system requirements. A wide variety and small volume are the advantages of ASIC. However, the diversity of applications contradicts the high investment in ASIC R&D while process technology going to 1Xnm. To realize an ASIC corresponding to the application in a low-cost way, a hardware scheme with the same topology as the C/C++ description should be the most direct implementation and the most efficient as well. Reconfigurable chip is both software and hardware programable. The hardware architecture and functions change dynamically in real-time with the change of software algorithm while ensuring flexibility. Thus it is also called a software-defined chip. The wide adaptability of the software-defined chip makes it a strong competitor to replace ASIC, FPGA, and general-purpose processors. Artificial intelligence (AI) is ubiquitous and AI chip has become a research hotspot in recent years. AI algorithms vary in different applications and will continue to evolve. AI services are migrating from cloud to edge nowadays. Performance demands and power consumption constraints require AI to deployed on an energy-efficient computing engine. Reconfigurable architecture is the ideal solution for intelligent computing since its programmability and dynamic reconfigurability of architecture can adapt algorithm evolution and diversity of applications and greatly improves energy efficiency as well. Dynamically reconfigurable technology brings the ability to bear the diversity and evolution of AI algorithms. Software-defined AI chips are expected to provide a new route for China's chip technology to get rid of imitation.

Biography

Dr. Shaojun Wei graduated from the Department of Radio & Electronics of Tsinghua University, China, in 1984 and received his Master degree in engineering. He received his Doctor degree in Applied Science from the Faculté Polytechnique de Mons (FPMs), Belgium, in 1991. Dr. Shaojun Wei is now the professor of Tsinghua University; Chief Scientist of the State Key Science and Technology Project; Member of the National Integrated Circuit Industry Development Advisory Committee; Vice President of China Semiconductor Industry Association (CSIA) and President of Fabless Chapter CSIA. Dr. Wei was the President & CEO of Datang Telecom Technology Co., Ltd. and the CTO of Datang Telecom Industry Group between 2001-2006. Dr. Wei has been working on VLSI design methodologies research and reconfigurable computing technology research. He has published more than 200 peer-reviewed papers and 6 monographs. He owns more than 130 patents including 18 US patents. Dr. Wei is the Fellow of Chinese Institute of Electronics (CIE) and the IEEE Fellow. Dr. Wei had won many awards including China National Second Award for Technology Invention (2015), China National Second Award for Technology Progress (2001), SIPO & WIPO Patent Golden Award (2003, 2015), First Award for Science and Technology of Ministry of Education (2014, 2019), China, First Award for Technology Invention of CIE (2012, 2017), EETimes China IC Design Achievement Award (2018), Aspencore Outstanding Contribution Award of the Year/Global Electronic Achievement Awards (2018) and SEMI Special Contribution Award (2019) and etc. He was selected to be the recipient of the 2020 IEEE CAS Industrial Pioneer Award.

KEYNOTE 02

9:50am ~ 10:35am, Sunday June 6, 2021

High-Performance and Energy-Efficient Circuit Technologies for Sub-5nm In-Memory/Near-Memory AI Accelerators

Ram K. Krishnamurthy

Intel Corporation, US

Abstract

This presentation will highlight some of the emerging challenges and opportunities for sub-5nm process machine learning and AI technologies in the rapidly evolving IoT industry. With Moore's law process technology scaling well into the nano-scale regime, future SoC platforms ranging from high performance cloud servers to ultra-low-power edge devices will demand advanced AI capabilities and energy-efficient deep neural networks. New and emerging IoT markets for autonomous vehicles, drones, and wearables require even higher performance at much lower cost while reducing energy consumption. Some of the prominent barriers to designing high performance and energy-efficient AI processors and SoCs in the sub-5nm technology nodes will be outlined. New paradigm shifts necessary for integrating special-purpose machine learning accelerators into next-generation SoCs will be explored. Emerging trends in SoC circuit design for machine learning and deep neural networks, specialized accelerators for in-memory and near-memory computing, reconfigurable multi-precision matrix multipliers, ultra-low-voltage logic, memory and clocking circuits, AI inference accelerators including binary neural networks and associated on-chip interconnect fabric circuits are described. Future brain-inspired neuromorphic computing circuit design challenges and technologies will also be reviewed. Specific chip design examples and case studies supported by silicon measurements and trade-offs will be discussed.

Biography

Ram K. Krishnamurthy received the B.E. degree in electrical engineering from the National Institute of Technology, Trichy, India, in 1993, the M.S. degree in electrical and computer engineering from the State University of New York, Buffalo, NY, USA, in 1994, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1997. He is currently a Senior Research Director and Senior Principal Engineer at Intel Labs, Hillsboro, OR, USA, where he heads the High Performance and Low Voltage Circuits Research Group. He has made circuit technology contributions to multiple generations of Intel's data center, client, FPGA, IoT, and AI products. Krishnamurthy has filed 320 patents and holds 180 issued patents. He has published 200 papers and four book chapters on high-performance and energy-efficient circuits. He serves as the Chair of the Semiconductor Research Corporation (SRC) Technical Advisory Board for the circuit design thrust. He served as the Technical Program Chair and the General Chair of the IEEE International Systems-on-Chip Conference and presently serves on the Conference's Steering Committee. He is an Adjunct Faculty with the Electrical and Computer Engineering Department, Oregon State University, Corvallis, OR, USA, where he taught advanced VLSI design. Krishnamurthy has received two Intel Achievement Awards for pioneering the first 64-bit Sparse-Tree ALU Technology and the first Advanced Encryption Standard hardware security accelerator on Intel products. He has received the IEEE International Solid State Circuits Conference Distinguished Technical Paper Award, IEEE European Solid State Circuits Conference Best Paper Award, Outstanding Industry Mentor Award from SRC, Intel awards for most patents filed and most patents issued, Intel Labs Gordon Moore Award, Alumni Recognition Award from Carnegie Mellon University, Distinguished Alumni Award from the State University of New York, MIT Technology Review's TR35 Innovator Award, and was recognized as a top ISSCC paper contributor. He has served as a Distinguished Lecturer of the IEEE Solid-State Circuits Society, a Guest Editor of the IEEE Journal of Solid State Circuits, an Associate Editor of the IEEE Transactions on VLSI Systems, and on the Technical Program Committees of ISSCC, CICC, and SOCC conferences. He is a Fellow of the IEEE and a Board Member of the Industry Advisory Board for the State University of New York.

KEYNOTE 03

9:00am ~ 9:45am, Monday June 7, 2021

On-device AI for Augmented Reality (AR) Systems: Challenges and Opportunities

Vikas Chandra

Facebook, US

Abstract

Augmented reality (AR) is a set of technologies that will fundamentally change the way we interact with our environment. It represents a merging of the physical and the digital worlds into a rich, context aware user interface delivered through a socially acceptable form factor such as eyeglasses. The majority of these novel experiences in AR systems will be powered by AI because of their superior ability to handle in-the-wild scenarios. In this presentation, we will discuss the challenges and opportunities in enabling AI-based algorithms in an energy-constrained AR system. One key AR use case is a personalized, proactive and context-aware Assistant that can understand the user's activity and their environment using audio-visual understanding. In this presentation, we will discuss the challenges and opportunities in both training and deployment of efficient audio-visual understanding on AR glasses. We show why it is imperative to co-design and co-optimize the AI algorithms and the hardware platforms jointly. We will present our early work to demonstrate the benefits and potential of such a co-design approach and discuss open research areas that are promising for the research community to explore.

Biography

Vikas Chandra leads the On-device AI applied research and engineering organization at Facebook Reality Labs responsible for developing real-time on-device computer vision, machine perception and speech/NLU technologies across AR/VR products. He received his Ph.D. in Computer Engineering from Carnegie Mellon University in 2004. He has held the positions of Visiting Scholar (2011 – 2014) and Visiting Faculty (2016 – 2017) at Stanford University. Dr. Chandra has authored 80+ research publications and is an inventor on 40+ US and international patents. He received the ACM-SIGDA Technical Leadership Award in 2009 and was invited to the 2017 National Academy of Engineering's Frontiers of Engineering Symposium. He is a senior member of IEEE. His research interests include all aspects of SW/HW co-design for efficient On-device AI.

KEYNOTE 04

9:00am ~ 9:45am, Tuesday June 8, 2021

Emerging AI Processing for Wireless Applications

Young-Kai (Y.K.) Chen

Defense Advanced Research Projects Agency（DARPA), US

Abstract

The emerging artificial intelligence and machine learning techniques promise new solutions and capabilities to unify the human world with the physical world and digital world together. This talk will illustrate a few challenges and solutions on utilizing artificial intelligence to enable next generation wireless communications.

Biography

Dr. Young-Kai (Y.K.) Chen, is a Program Manager with the Microsystems Technology Office at DARPA. Before joined DARPA in September 2017, he was a Senior Director at Nokia Bell Labs supporting research groups to develop high speed electronics and optoelectronics for advanced wireless and optic fiber communication networks. Dr. Chen and his teams had contributed to first commercial integrated DFB-EAM devices, silicon-based frontend ICs for 3G/4G/5G millimeter-wave backhaul radios, and launched the first commercial 100Gbps optical transponders in 2011. His team also developed silicon photonics ICs for microwave photonics and 100G data links.

Dr. Chen received his Ph.D. degree from Cornell University. He was an Adjunct Professor at Columbia University, National Taiwan University and National Chiao-Tung University. Dr. Chen is a Fellow of Bell Labs, IEEE and OSA, a member of the Academia Sinica and National Academy of Engineering, and a recipient of IEEE David Sarnoff Award and Edison Patent Award.

KEYNOTE 05

9:00am ~ 9:45am, Wednesday June 9, 2021

AI and Video Compression in the Era of Internet of Video Things

Yen-Kuang Chen

Alibaba Group, US

Abstract

We are at the very beginning of the era of Internet of Video Things (IoVT), where many cameras collect a huge amount of visual data to be analyzed. IoVT will become more important as the numbers of cameras and applications are growing exponentially in the coming years. Humans cannot process all the videos and it is critical to use artificial intelligence (AI) to process the data. Many challenges arise fulfilling the era of IoVT, e.g., accuracy, energy efficiency, processing speed. This talk will discuss the following question: Can we design efficient IoVT systems by co-optimizing video compression and computer vision algorithms?

Biography

Yen-Kuang Chen received his Ph.D. degree from Princeton University and is a Senior Director and Chief Scientist of Computing Technology, Computing Technology Lab, DAMO Academy, Alibaba. His research areas span from emerging applications that can utilize the true potential of multimedia and Internet of Things (IoT) to computer architecture that can embrace emerging applications. He has 80+ US patents and 100+ technical publications. He is one of the key contributors to Supplemental Streaming SIMD Extension 3 and Advanced Vector Extension in Intel microprocessors. He is recognized as an IEEE Fellow for his contributions to algorithm-architecture co-design for multimedia signal processing.

TUTORIALS

Design of CMOS Annealing Processor for Solving Combinatorial Optimization Problems

Bongjin Kim

University of California, Santa Barbara, US

Abstract

Annealing processors based on the convergence property of the Ising model offer an attractive means for solving non-deterministic polynomial-time hard (NP-hard) combinatorial optimization problems. Quantum annealing processors exploiting quantum tunneling effect implemented >1K spins using 100K+ Josephson junctions. However, the practical application of quantum annealers is limited due to the extreme operating conditions (e.g., extremely low operating temperature, 15mK) and the associated huge power consumption (e.g., 25kW) as well as high development and operation cost. Alternatively, low-power annealing processors based on simulated annealing have been developed using low-cost CMOS process. This tutorial introduces fundamentals and recent progresses in the design of CMOS annealing processors.

Biography

Bongjin Kim received his PhD degree from the University of Minnesota in 2014. After his PhD, he worked on design techniques and methodologies for communication circuits and microarchitectures at Rambus and Stanford University as a senior staff and a postdoctoral research fellow. After working as an assistant professor at Nanyang Technological University in Singapore for three years (from 2017 to 2020), he joined the Department of Electrical and Computer Engineering at UC Santa Barbara. His research team develops innovative integrated circuits and system solutions using traditional CMOS logic and emerging technologies to solve challenging problems in fundamental science and accelerate computations and communications. Target applications include, but not limited to, artificial intelligence, machine learning, robotics, and alternative computing. He received the Doctoral Dissertation Fellowship Award at the University of Minnesota and the ISLPED International Low Power Design Contest Award. His research works have been published in peer-reviewed conferences and journals, including the International Solid-State Circuits Conference (ISSCC), VLSI Symposium, Custom Integrated Circuits Conference (CICC), and Journal of Solid-State Circuits (JSSC). He has served on the technical program committee for Design Automation Conference (DAC) and the IEEE Solid-State Circuits Letter (SSC-L) editorial review board.

Computing with Oscillating Neural Networks

Siegfried Karg

IBM Research Zurich, Switzerland

Abstract

Machine-learning applications are penetrating virtually every aspect of modern society, e.g. natural-language processing for virtual personal assistants, image recognition for social media, mobility data for traffic prediction. Nonetheless, already fairly simple tasks in human everyday life such as recognizing patterns from a stream of images brings even the most powerful computer chip at its limits. The underlying von-Neumann architecture with its physically separated processor and memory appears as a major bottleneck for faster and more power-efficient inference machines. In-memory or in-sensor computation has been proposed to reduce data traffic and latency. This offers the opportunity to move specific tasks into the analog domain and process data inherently parallel and power efficient. However, these neuromorphic computing approaches come with substantial challenges for system design, mixed-signal electronic circuits and physical modelling.

Development of materials with non-linear properties such as memristive oxides or metal-insulator transition materials offer new routes to brain-inspired data processing. As an example, we will show a novel and alternative neuromorphic computing paradigm based on oscillating neural networks (ONN). Area and power efficient relaxation oscillators based on the metal-insulator phase-transition material VO2 act as 'neurons'. Tunable memristors are envisioned as the 'synapses' of the network. Inspired by neural oscillations or brain waves, in ONN the information is encoded in the phase of coupled oscillators.
We will demonstrate the neuro-computational computing power of these phase-modulated ONN by means of experiments and simulations. The scope of the project ranges from material and fabrication aspects of VO2 layers, over device fabrication and characterization and modelling to circuit-level ONN implementations and simulation of ML applications. Key questions on topics such as influence of device and process variability, prospects of the ONN architecture for compute performance and power consumption will be addressed.

Biography
Dr. Siegfried Karg is a Research Staff Member at IBM Research – Zurich and a Senior Member of IEEE. He worked on the physics and materials science of organic and polymer devices (OLEDs, OFETs and electrochemical cells). He has conducted research on memory applications such as resistive oxide RAM and on capacitorless eDRAM based on III-V semiconductor transistors. Moreover, he investigated the one-dimensional electronic properties of InAs nanostructures. His current research fields include brain-inspired computing applications exploiting oscillatory neural networks (with electronic oscillators based on the metal-insulator transition of VO2). S. Karg has authored about 100 scientific publications and holds more than 30 patents.

Neuromorphic Computing – Opportunities and Challenges

Abhronil Sengupta

The Pennsylvania State University, US

Abstract
While research in designing brain-inspired algorithms have attained a stage where such Artificial Intelligence platforms are being able to outperform humans at several cognitive tasks, an often-unnoticed cost is the huge computational expenses required for running these algorithms in hardware. Bridging the computational efficiency gap necessitates the exploration of devices, circuits and algorithms that provide a better match to the computational primitives of biological processing – neurons and synapses, and which require a significant rethinking of traditional von-Neumann based computing. The tutorial will discuss innovations in emerging post-CMOS technologies that are revealing immense possibilities of implementing a plethora of neural and synaptic functionalities by single device structures that can be operated at very low terminal voltages. The tutorial will also discuss recent innovations at the algorithm front on exploring event-driven Spiking Neural Networks for large-scale machine learning tasks. Such neuromorphic systems can potentially provide significantly lower computational overhead in contrast to standard deep learning platforms, especially in sparse, event-driven application domains with temporal information processing. I will conclude the presentation by providing the prospects of enabling end-to-end cognitive intelligence across the computing stack that combines knowledge from devices and circuits to machine learning and computational neuroscience.

Biography
Dr. Abhronil Sengupta is an Assistant Professor in the School of Electrical Engineering and Computer Science at Penn State University. He is also affiliated with the Department of Materials Science and Engineering and the Materials Research Institute (MRI). Dr. Sengupta received the PhD degree in Electrical and Computer Engineering from Purdue University in 2018 and the B.E. degree from Jadavpur University, India in 2013. He worked as a DAAD (German Academic Exchange Service) Fellow at the University of

Hamburg, Germany in 2012, and as a graduate research intern at Circuit Research Labs, Intel Labs in 2016 and Facebook Reality Labs in 2017. Dr. Sengupta has published over 60 articles in referred journals and conferences and holds 4 granted/pending US patents. He serves on the IEEE Circuits and Systems Society Technical Committee on Neural Systems and Applications, Editorial Board of Neuromorphic Computing and Engineering, Frontiers in Neuroscience journals and the Technical Program Committee of several international conferences like DAC, ICCAD, ISQED and GLSVLSI. He has been awarded the IEEE Circuits and Systems Society (CASS) Outstanding Young Author Award (2019), IEEE SiPS Best Paper Award (2018), Schmidt Science Fellows Award nominee (2017), Bilsland Dissertation Fellowship (2017), CSPIN Student Presenter Award (2015), Birck Fellowship (2013), the DAAD WISE Fellowship (2012). His work on neuromorphic computing has been highlighted in media by MIT Technology Review, US Department of Defense, American Institute of Physics among others. Dr. Sengupta is a member of the IEEE Electron Devices Society (EDS) and Circuits and Systems (CAS) Society, the Association for Computing Machinery (ACM) and the American Physical Society (APS).

Artificial Intelligence & Robotics Design challenges and perspectives

Théophile Gonos

A.I.Mergence, France

Abstract

Robotics and Artificial Intelligence are part of a technological revolution in nowadays human society. But many constrains are raised by the specific needs of robotics in hardware – motors, batteries, sensors – as well as in software, framework compatibilities for instance. Robotic systems development also needs a multiple-skill team and a good communication between different parts of the field: mechanics, electronics, and computer science.

Moreover, despite the recent development, power consumption and computational capabilities remain a major bottleneck of robotics advances. Legal aspects also need to be addressed such as responsibility of robots or more recently the European legislation such as RGPD.

On the contrary, for the past several years, the development of smartphones and IoT devices have increased the computational capabilities and drastically reduced the power consumption of embedded chips. More specifically, a current line of research focuses on designing bio-inspired low-energy AI chips.

In this talk we address the main challenges that a company needs to cope with to build robotic and embedded systems. Throughout the talk, we will use the company's robot E4 as a study case, designed to ensure building protection. We will also briefly address cultural and social threats for creating such devices and speak about market perspectives.

Biography

Théophile Gonos received a PhD degree from the School of "Informatics" of the University of Edinburgh (UK) in 2011, for his work on bioinspired adaptive sensing. He also received an MSc degree in Robotics and Intelligent Systems from UPMC (Paris-Sorbonne) and a BSc degree in Computer Science from Denis Diderot University, Paris. His fields of expertise in Artificial Intelligence are on multi-agent systems, modular robotics, as well as bioinspired systems and neural networks. He founded the company A.I.Mergence in 2015, which is developing an autonomous safety robot named E4, aiming to ensure indoor security.

REGULAR SESSION 01

10:40am ~ 11:15am, Sunday June 6, 2021

Hardware for AI

Chairs: ***Jaydeep Kulkarni - UT Austin, Linghao Song - UCLA***

1 **Exploiting Memristors for Neuromorphic Reinforcement Learning**
Cong Shi, Jing Lu, Ying Wang, Ping Li, Min Tian

2 **MLFlash-CIM: Embedded Multi-Level NOR-Flash Cell Based Computing in Memory Architecture for Edge AI Devices**
Sitao Zeng, Yuxin Zhang, Zhiguo Zhu, Zhaolong Qin, Chunmeng Dou, Xin Si, Qiang Li

3 **Single RRAM Cell-Based In-Memory Accelerator Architecture for Binary Neural Networks**
Hyunmyung Oh, Hyungjun Kim, Nameun Kang, Yulhwa Kim, Jihoon Park, Jae-Joon Kim

4 **End-to-End 100-TOPS/W Inference with Analog In-Memory Computing: Are We There Yet?**
Gianmarco Ottavi, Geethan Karunaratne, Francesco Conti, Irem Boybat, Luca Benini, Davide Rossi

5 **Neural Network Acceleration and Voice Recognition with a Flash-Based In-Memory Computing SoC**
Liang Zhao, Shifan Gao, Shengbo Zhang, Xiang Qiu, Fan Yang, Jie Li, Zezhi Chen, Yi Zhao

6 **An Efficient and Fast Softmax Hardware Architecture (EFSHA) for Deep Neural Networks**
Muhammad Awais Hussain, Tsung-Han Tsai

7 **Efficient Digital Implementation of n-Mode Tensor-Matrix Multiplication**
Christian Gianoglio, Edoardo Ragusa, Rodolfo Zunino, Paolo Gastaldo

8 **Hardware-Algorithm Co-Design Enabling Efficient Event-Based Object Detection**
Brian Crafton, Andrew Paredes, Evan Gebhardt, Arijit Raychowdhury

9 **Open Discussion**

REGULAR SESSION 02

11:20am ~ 11:55am, Sunday June 6, 2021

Neuro-inspired Circuits and Systems

Chairs: ***Jongsun Park - Korea University, Biswas Dwaipayan - IMEC***

1 **A Memristor Model with Concise Window Function for Spiking Brain-Inspired Computation**
Jiawei Xu, Deyu Wang, Feng Li, Lianhao Zhang, Dimitrios Stathis, Yu Yang, Yi Jin, Anders Lansner, Ahmed Hemani, Zhuo ...

2 **Hardware Approximation of Exponential Decay for Spiking Neural Networks**
Sherif Eissa, Sander Stuijk, Henk Corporaal

3 **FL-HDC: Hyperdimensional Computing Design for the Application of Federated Learning**
Cheng-Yen Hsieh, Yu-Chuan Chuang, An-Yeu Wu

4 **Event-Driven Continuous-Time Feature Extraction for Ultra Low-Power Audio Keyword Spotting**
Soufiane Mourrane, Benoit Larras, Andreia Cathelin, Antoine Frappé

5 **Online Detection of Vibration Anomalies Using Balanced Spiking Neural Networks**
Nik Dennler, Germain Haessig, Matteo Cartiglia, Giacomo Indiveri

6 **Performance of Crossbar Based Long Short Term Memory with Aging Memristors**
Aswani A R, Rohan Kumar, Jai Tripathi, Alex James

7 **An Ultra-Low Latency Multicast Router for Large-Scale Multi-Chip Neuromorphic Processing**
Chen Ding, Yuxiang Huan, Hao Jia, Yulong Yan, Fanxi Yang, Zhuo Zou, Lirong Zheng

8 **Graph-Based Spatio-Temporal Backpropagation for Training Spiking Neural Networks**
Yulong Yan, Haoming Chu, Xin Chen, Yi Jin, Yuxiang Huan, Lirong Zheng, Zhuo Zou

9 **Characterization of Drain Current Variations in FeFETs for PIM-Based DNN Accelerators**
Nathan Miller, Zheng Wang, Saurabh Dash, Asif Khan, Saibal Mukhopadhyay

10 **Open Discussion**

REGULAR SESSION 03

9:50am ~ 10:35am, Monday June 7, 2021

AI Accelerators

Chairs: ***Shimeng Yu - GaTech, Bing Li - Capitcal Normal University, China***

1 **Multiple-Precision Floating-Point Dot Product Unit for Efficient Convolution Computation**
Kai Li, Wei Mao, Xinang Xie, Quan Cheng, Huan Xie, Zhenjiang Dong, Hao Yu

2 **EILE: Efficient Incremental Learning on the Edge**
Xi Chen, Chang Gao, Tobi Delbruck, Shih-Chii Liu

3 **An 8.62 µW Processor for Autism Spectrum Disorder Classification Using Shallow Neural Network**
Abdul Rehman Aslam, Nauman Hafeez, Hadi Heidari, Muhammad Awais Bin Altaf

4 **Quantized Fully Convolution Neural Network for HW Implementation of Human Posture Recognition**
Alessandro Russo, Gian Domenico Licciardo, Luigi Di Benedetto, Alfredo Rubino, Rosalba Liguori, Alessandro Naddeo, Ni...

5 **TCBNN: Error-Correctable Ternary-Coded Binarized Neural Network**
Cheng-Di Tsai, Tingyu Chen, Hsiao-Wen Fu, Tsung-Chu Huang

6 **Efficient FPGA Implementation of Approximate Singular Value Decomposition Based on Shallow Neural Networks**
Hamoud Younes, Ali Ibrahim, Mostafa Rizk, Maurizio Valle

7 **Efficient Zero-Activation-Skipping for On-Chip Low-Energy CNN Acceleration**
Min Liu, Yifan He, Hailong Jiao

8 **An Energy-Efficient Quad-Camera Visual System for Autonomous Machines on FPGA Platform**
Zishen Wan, Yuyang Zhang, Arijit Raychowdhury, Bo Yu, Yanjun Zhang, Shaoshan Liu

9 **Energy Efficient Computing with Heterogeneous DNN Accelerators**
Md Shazzad Hossain, Ioannis Savidis

10 **Flexible-Width Bit-Level Compressor for Convolutional Neural Network**
Junhan Zhu, Xiaoliang Chen, Li Du, Haoran Geng, Yichuan Bai, Yuandong Li, Yuan Du, Zhongfeng Wang

11 **Open Discussion**

SPECIAL SESSION 01

10:40am ~ 11:15am, Monday June 7, 2021

Health Monitor & Autonomous Vehicle

Chairs: ***Kea-Tiong Tang - National Tsing Hua University, Tinoosh Mohsenin - UBMC***

Part I: Health Monitor

1 **Self-Aware Anomaly-Detection for Epilepsy Monitoring on Low-Power Wearable Electrocardiographic Devices**
Farnaz Forooghifar, Amin Aminifar, Tomas Teijeiro, Amir Aminifar, Jesper Jeppesen, Sandor Beniczky, David Atienza

2 **Towards Smart and Efficient Health Monitoring Using Edge-Enabled Situational-Awareness**
Sina Shahhosseini, Anil Kanduri, Milad Asgari Mehrabadi, Emad Kasaeyan Naeini, Dongjoo Seo, Sung-Soo Lim, Amir Mohamm...

3 **Demography-Aware COVID-19 Confinement with Game Theory**
Sreenitha Kasarapu, Rakibul Hassan, Setareh Rafatirad, Houman Homayoun, Sai Manoj Pudukotai Dinakarrao

4 **CoughNet: A Flexible Low Power CNN-LSTM Processor for Cough Sound Detection**
Hasib-Al Rashid, Arnab Neelim Mazumder, Utteja Panchakshara Kallakuri Niyogi, Tinoosh Mohsenin

Part II: Autonomous Vehicle

5 **An 176.3 GOPs Object Detection CNN Accelerator Emulated in a 28nm CMOS Technology**
Ying-Cheng Lu, Ching-Wen Chen, Ching-Chun Pu, Yang-Tung Lin, Jyun-Kai Jhan, Shu-Ping Liang, Wei-Lun Tseng, Chi-Shi Ch...

6 **A Bio Inspired Motion Detection Circuit Model for the Computation of Optical Flow: the Spatial-Temporal - Filtering Reichardt Model**
Hsin-Yu Wu, Wei-Tse Kao, Harrison Hao-Yu Ku, Cheng-Te Wang, Chih-Cheng Hsieh, Ren-Shuo Liu, Kea-Tiong Tang, Chung-Chu...

7 **iAMEC, an Intelligent Autonomous Mover for Navigation in Indoor People Rich Environments**
Yin-Tsung Hwang, Kuan-Hung Chen, Chih-Peng Fan, Yeong-Kang Lai, Chung-Bin Wu, Hsiao-Ping Tsai, Wei-Liang Lin, Kuang-H...

8 **An AI AUV Enabling Vision-Based Diver-Following and Obstacle Avoidance with 3D-Modeling Dataset**
Yu-Cheng Chou, Hsin-Hung Chen, Chau-Chang Wang, Hui-Min Chou, Chua-Chin Wang

9 **Open Discussion**

SPECIAL SESSION 02

11:20am ~ 11:55am, Monday June 7, 2021

Design Automation & Compute-In-Memory

Chairs: ***Shimeng Yu - GaTech, Debjyoti Bhattacharjee - IMEC***

1 **A Flexible and Fast PyTorch Toolkit for Simulating Training and Inference on Analog Crossbar Arrays**
Malte Rasch, Diego Moreda, Tayfun Gokmen, Manuel Le Gallo-Bourdeau, Fabio Carta, Cindy Goldberg, Kaoutar El Maghraoui...

2 **Design Tools for Resistive Crossbar Based Machine Learning Accelerators**
Indranil Chakraborty, Sourjya Roy, Shrihari Sridharan, Mustafa Ali, Aayush Ankit, Shubham Jain, Anand Raghunathan

3 **MNSIM-TIME: Performance Modeling Framework for Training-In-Memory Architectures**
Kaizhong Qiu, Zhenhua Zhu, Yi Cai, Hanbo Sun, Yu Wang, Huazhong Yang

4 **NeuroSim Validation with 40nm RRAM Compute-in-Memory Macro**
Anni Lu, Xiaochen Peng, Wantong Li, Hongwu Jiang, Shimeng Yu

5 **Trends in Analog and Digital Intensive Compute-in-SRAM Designs**
Rishabh Sehgal, Jaydeep Kulkarni

6 **Compute-in-RRAM with Limited On-Chip Resources**
Anni Lu, Xiaochen Peng, Shimeng Yu

7 **MRAM-Based BER Resilient Quantized Edge-AI Networks for Harsh Industrial Conditions**
Vivek Parmar, Manan Suri, Kazutaka Yamane, Taeyoung Lee, Nyuk Leong Chung, Vinayak Bharat Naik

8 **uPIM: Performance-Aware Online Learning Capable Processing-in-Memory**
Sathwika Bavikadi, Purab Ranjan Sutradhar, Amlan Ganguly, Sai Manoj Pudukotai Dinakarrao

9 **Open Discussion**

DEMO SESSION

12:00pm ~ 12:30pm, Monday June 7, 2021

Chair: ***Hai Li - Duke***

1 **Energy-Efficient Intelligent EPTS Device Using Novel DCNN-Based Dynamic Sensor Activation**
Hyunsung Kim, Jaehee Kim, Young-Seok Kim, Mijung Kim, Youngjoo Lee

2 **An Ultra-Low-Power Real-Time Hand-Gesture Recognition System for Edge Applications**
Yuncheng Lu, Zehao Li, Tony Tae Hyoung Kim

3 **Real-Time Language Recognition Using Hyperdimensional Computing on Phase-Change Memory Array**
Geethan Karunaratne, Abbas Rahimi, Manuel Le Gallo-Bourdeau, Giovanni Cherubini, Abu Sebastian

4 **A Real-Time Face Recognition System by Efficient Hardware-Software Co-Design on FPGA Socs**
Hao Wang, Shan Cao, Shugong Xu

5 **Live Demo: An 176.3 GOPs Object Detection CNN Accelerator Emulated in a 28nm CMOS Technology**
Ying-Cheng Lu, Ching-Wen Chen, Ching-Chun Pu, Yang-Tung Lin, Jyun-Kai Jhan, Shu-Ping Liang, Wei-Lun Tseng, Chi-Shi Ch..

6 **Open Discussion**

REGULAR SESSION 04

9:50am ~ 10:35am, Tuesday June 8, 2021

Learning Models and Applications of Intelligent Systems

Chairs: ***John Paul Strachan - HPE Lab, Hyungjun Kim - POSTECH***

1 **Unbalanced Bit-Slicing Scheme for Accurate Memristor-Based Neural Network Architecture**
Sumit Diware, Anteneh Gebregiorgis, Rajiv Joshi, Said Hamdioui, Rajendra Bishnoi

2 **Integer Quadratic Integrate-and-Fire (IQIF): A Neuron Model for Digital Neuromorphic Systems**
Wen-Chieh Wu, Chen-Fu Yeh, Alexander James White, Cheng-Te Wang, Zuo-Wei Yeh, Chih-Cheng Hsieh, Ren-Shuo Liu, Kea-Tio...

3 **Federated Regularization Learning: An Accurate and Safe Method for Federated Learning**
Tianqi Su, Meiqi Wang, Zhongfeng Wang

4 **A Novel Multi-Scale Dilated 3D CNN for Epileptic Seizure Prediction**
Ziyu Wang, Jie Yang, Mohamad Sawan

5 **Contention-Aware Adaptive Model Selection for Machine Vision in Embedded Systems**
Basar Kutukcu, Sabur Baidya, Anand Raghunathan, Sujit Dey

6 **A Two-Layer LSTM Deep Learning Model for Epileptic Seizure Prediction**
Shiva Maleki Varnosfaderani, Rihat Rahman, Nabil J. Sarhan, Levin Kuhlmann, Eishi Asano, Aimee Luat, Mohammad Alhawar...

7 **TempDiff: Temporal Difference-Based Feature Map-Level Sparsity Induction in CNNs with <4% Memory Overhead**
Udari De Alwis, Massimo Alioto

8 **ECG-TCN: Wearable Cardiac Arrhythmia Detection with a Temporal Convolutional Network**
Thorir Mar Ingolfsson, Xiaying Wang, Michael Hersche, Alessio Burrello, Lukas Cavigelli, Luca Benini

9 **Smart Refrigerator Inventory Management Using Convolutional Neural Networks**
Taeho Lee, Shin-Woo Kang, Tae-Hyun Kim, Jin-Sung Kim, Hyuk-Jae Lee

10 **On-Chip Pixel Reconstruction Using Simple CNN for Sparsely Read CMOS Image Sensor**
Wilfred Kisku, Amandeep Kaur, Deepak Mishra

11 **Open Discussion**

SPECIAL SESSION 03

10:40am ~ 11:15am, Tuesday June 8, 2021

Efficient Hardware Accelerator Design for ML: Architectural and Algorithmic Optimization

Chairs: ***Wei Zhang - Hong Kong University of Science and Technology, Kea-Tiong Tang - National Tsing Hua University***

1 **HPPU: An Energy-Efficient Sparse DNN Training Processor with Hybrid Weight Pruning**
Yang Wang, Yubin Qin, Leibo Liu, Shaojun Wei, Shouyi Yin

2 **Improving System Latency of AI Accelerator with On-Chip Pipelined Activation Preprocessing and Multi-Mode Batch Inference**
Wenxuan Chen, Zheng Wang, Ming Lei, Bo Dong, Zhuo Wang, Yongkui Yang, Chao Chen, Weiyu Guo, Chen Liang, Qian Zhang, W...

3 **Exploiting Weight Statistics for Compressed Neural Network Implementation on Hardware**
Prachi Kashikar, Sharad Sinha, Ajeet Kumar Verma

4 **FPGA-Accelerated Agent-Based Simulation for COVID-19**
Lei Fu, Ce Guo, Wayne Luk

5 **Design Optimization for ADMM-Based SVM Training Processor for Edge Computing**
Shuo-An Huang, Yi-Yen Hsieh, Chia-Hsiang Yang

6 **A Quality-Oriented Reconfigurable Convolution Engine Using Cross-Shaped Sparse Kernels for Highly-Parallel CNN Acceleration**
Chi-Wen Weng, Chao-Tsung Huang

7 **Tile-Based Architecture Exploration for Convolutional Accelerators in Deep Neural Networks**
Yang-Tsai Chen, Yu-Xiang Yen, Chun-Tse Chen, Tzu-Yu Chen, Chih-Tsun Huang, Jing-Jia Liou, Juin-Ming Lu

8 **An Energy-Efficient Hardware Accelerator for Hierarchical Deep Reinforcement Learning**
Aidin Shiri, Bharat Prakash, Arnab Neelim Mazumder, Nicholas R. Waytowich, Tim Oates, Tinoosh Mohsenin

9 **Open Discussion**

SPECIAL SESSION 04

11:20am ~ 11:55am, Tuesday June 8, 2021

Micro AI and Low Power Autonomous Systems

Chair: ***Tinoosh Mohsenin - UBMC***

1 **Energy-Efficient Deep Reinforcement Learning Accelerator Designs for Mobile Autonomous Systems**
Juhyoung Lee, Changhyeon Kim, Donghyeon Han, Sangyeob Kim, Sangjin Kim, Hoi-Jun Yoo

2 **Automated Tuning of End-to-End Neural Flight Controllers for Autonomous Nano-Drones**
Vlad Niculescu, Lorenzo Lamberti, Daniele Palossi, Luca Benini

3 **Efficient FPGA Implementation of a Convolutional Neural Network for Radar Signal Processing**
Jingchi Zhang, Yihao Huang, Huanrui Yang, Michael Martinez, Granger Hickman, Jeffrey Krolik, Hai Li

4 **Software/Hardware Co-Design for Multi-Modal Multi-Task Learning in Autonomous Systems**
Cong Callie Hao, Deming Chen

5 **Evaluation of Machine Learning-Based Detection Against Side-Channel Attacks on Autonomous Vehicle**
Han Wang, Soheil Salehi, Hossein Sayadi, Avesta Sasan, Tinoosh Mohsenin, Sai Manoj P D, Setareh Rafatirad, Houman Hom...

6 **An Efficient and Low-Power MLP Accelerator Architecture Supporting Structured Pruning, Sparse Activations and Asymmetric Quantization for Edge Computing**
Wei-Chen Lin, Ya-Ju Chang, Juinn-Dar Huang

7 **LPE: Logarithm Posit Processing Element for Energy-Efficient Edge-Device Training**
Yang Wang, Dazheng Deng, Leibo Liu, Shaojun Wei, Shouyi Yin

8 **Open Discussion**

REGULAR SESSION 05

9:50am ~ 10:35am, Wednesday June 9, 2021

Algorithm-Hardware Co-Design of ML Systems

Chairs: ***Somnath Paul - Intel, Jong-Hyeok Yoon - DGIST***

1. **Quantization Strategy for Pareto-Optimally Low-Cost and Accurate CNN**
   Kengo Nakata, Daisuke Miyashita, Asuka Maki, Fumihiko Tachibana, Sasaki Shinichi, Jun Deguchi, Ryuichi Fujimoto
2. **Dynamically-Biased Fixed-Point LSTM for Time Series Processing in AIoT Edge Device**
   Jinhai Hu, Wang Ling Goh, Yuan Gao
3. **iELAS: An ELAS-Based Energy-Efficient Accelerator for Real-Time Stereo Matching on FPGA Platform**
   Tian Gao, Zishen Wan, Yuyang Zhang, Bo Yu, Yanjun Zhang, Shaoshan Liu, Arijit Raychowdhury
4. **LOMA: Fast Auto-Scheduling on DNN Accelerators Through Loop-Order-Based Memory Allocation**
   Arne Symons, Linyan Mei, Marian Verhelst
5. **Analyzing the Energy-Latency-Area-Accuracy Trade-Off Across Contemporary Neural Networks**
   Vikram Jain, Linyan Mei, Marian Verhelst
6. **Integer-Only Approximated MFCC for Ultra-Low Power Audio NN Processing on Multi-Core MCUs**
   Marco Fariselli, Manuele Rusci, Joel Cambonie, Eric Flamand
7. **Adaptable Approximation Based on Bit Decomposition for Deep Neural Network Accelerators**
   Taha Soliman, Cecilia De La Parra, Andre Guntoro, Norbert Wehn
8. **Two-Phase Scheme for Trimming QTMT CU Partition Using Multi-Branch Convolutional Neural Networks**
   Pin-Chieh Fu, Chia-Cheng Yen, Nien-Chen Yang, Jia-Shung Wang
9. **Ensemble of Pruned Networks for Reliable Classifiers**
   Zhen Gao, Han Zhang, Xiaohui Wei, Jiajun Xiao, Shulin Zeng, Guangjun Ge, Yu Wang, Pedro Reviriego
10. **Memory Efficient Invertible Neural Networks for Class-Incremental Learning**
    Guillaume Hocquet, Olivier Bichler, Damien Querlioz
11. **Open Discussion**

SEPCIAL SESSION 05

10:40am ~ 11:55am, Wednesday June 9, 2021

Low Power Computer Vision Challenge

Chair: ***Yung-Hsiang Lu - Purdue University***

1 **Welcome**
George Thiruvathukal

2 **Listen to Look: Efficient Visual Recognition with Audio**
Ruohan Gao

3 **2020 Low-Power Computer Vision Challenge Summary**
Xiao Hu

4 **Preprocessing for energy efficient scene text recognition**
SeongHwan Jeong

5 **E2VTS: Energy-Efficient Video Text Spotting from Unmanned Aerial Vehicles**
Zhenyu Hu

6 **Open Discussion**