Final Program

April 18, 2018 *Main Hall(7th Floor), Yokohama Joho Bunka Center*
(Yokohama Media & Communications Center)

13:30-15:00 **Special Invited Lecture 1**
Chair: Tohru Ishihara (Kyoto Univ.)

13:30-15:00 **Energy-Efficient and Energy-Scalable Processing - Meeting the Varied Needs of the Internet of Things at Its Edge**
Massimo Alioto (National Univ. of Singapore, Singapore)

Abstract: The Internet of Things (IoT) is evolving as a complex ecosystem that enables ubiquitous sensing through the deployment of ultra-low cost miniaturized devices (the "IoT nodes" at its edge). As the communication-computation tradeoffs swings towards more computation to deal with power-hungry radios, energy efficient processing is now necessary for any smart IoT node. And it is often not sufficient, as occasional and vigorous performance boosts are required when performing on-chip data analytics and event evaluation. In other words, energy efficiency and scalability are both key properties of processing in IoT nodes. This talk addresses the fundamental challenges posed by IoT nodes, in terms of both energy efficiency and energy scalability. Operation at minimum energy is first discussed, describing the implications at the circuit and the architectural level. Techniques to scale down energy below the "natural" minimum energy point are discussed, leveraging recently proposed approaches at the circuit, microarchitectural and architectural level. The orthogonal design dimension of energy-quality scaling is then introduced as very promising way to keep scaling down the energy, even when performance is constrained. Overall, recent and on-going research on energy-efficient and energy-scalable processing for IoT shows that there is still substantial room for energy improvements. To ultimately make IoT nodes smarter, smaller and long-lived.

15:00-15:30 **Break**

15:30-17:00 **Special Invited Lecture 2**
Chair: Tohru Ishihara (Kyoto Univ.)

15:30-17:00 **High-Power-Efficiency Implementation of Neuromorphic Computing Systems with Memristors**
Yiran Chen (Duke Univ., USA)

Abstract: Inspired by the working mechanism of human brains, neuromorphic computing system (NCS) possesses a massively parallel architecture with closely coupled memory. NCS can be efficiently implemented by nonvolatile memories, e.g. memristor crossbar arrays, because of its analogy to matrix multiplication and high resistance resulting in low power consumption. However, memristor fabrication process cannot produce perfect devices: limited high/low resistance ratio and resistance level, varying resistance range and nonlinearity bring difficulties into hardware implementation. In this talk, we will start with spike and level versions of memristor based Neuromorphic chip prototypes using Integrate-and-Fire-Circuit and their applications in pattern recognitions, followed by the discussion on the challenges and our solutions on bridging the gap between software algorithm and hardware implementation. Both circuit design techniques and algorithm tailoring are included. (Author: Bonan Yan, Qing Yang, Chenchen Liu, Hai Li, and Yiran

Chen)

17:00-17:15 **Break**

17:15-18:05 **Keynote Presentation 1**

Co-chairs: Shinya Takamaeda-Yamazaki (Hokkaido Univ.), Masato Suzuki (Socionext)

17:15-18:05 **Tensor Processing Unit: A processor for neural network designed by Google**

Kaz Sato (Google, USA)

Abstract: Tensor Processing Unit (TPU) is a LSI designed by Google for neural network processing. TPU features a large-scale systolic array matrix unit that achieves outstanding performance-per-watt ratio. In this session we will learn how a minimalistic design philosophy and a tight focus on neural network inference use-cases enabled the high performance neural network accelerator chip.

April 19, 2018 *Main Hall(7th Floor), Yokohama Joho Bunka Center*
(Yokohama Media & Communications Center)

9:30-9:50 **Session I**

9:30-9:50 **Welcome and Opening Remarks**
Co-chairs: Yuki Kobayashi (NEC), Hiroki Matsutani (Keio Univ.)

Hideharu Amano	*Chair of the Organizing Committee*
Hironori Kasahara	*President of IEEE/CS*
Allen J. Baum	*Chair of IEEE/CS TCMM*
Hiroyuki Uenohara	*President of IEICE/ES*

9:50-10:40 **Session II**

9:50-10:40 **Keynote Presentation 2**
Co-chairs: Ryusuke Egawa (Tohoku Univ.), Yuki Kobayashi (NEC)

AMD EPYC$^{\text{TM}}$ Microprocessor Architecture
Jay Fleischman (AMD, USA)

Abstract: AMD will present the next-generation AMD EPYC™ microprocessor. This advanced processor is a Multi-Chip-Module (MCM) comprised of up to four System-on-a-Chip (SoC) die, codenamed “Zeppelin”. Each “Zeppelin” SoC contains eight high-performance AMD x86 cores, codenamed “Zen”, caches, memory controllers, IO controllers (such as PCIe and SATA), and integrated x86 southbridge chipset capabilities. All these functions are connected on the SoC and between multichip packages and multi-socket systems by AMD Infinity Fabric. Utilizing GLOBALFOUNDRIES’ 14nm LPP FinFET process technology, the four-die MCM EPYC™ microprocessor has over 19.2B transistors.

10:40-11:00 **Break**

11:00-11:50 **Session III:**

11:00-11:50 **Keynote Presentation 3**
Co-chairs: Megumi Ito (IBM), Takuya Nakaike (IBM)

Designing Deep Neural Network Accelerators with Analog Memory – A Device and Circuit Perspective
Pritish Narayanan (IBM Research - Almaden, USA)

Abstract: Deep Neural Networks (DNNs) have revolutionized the field of Artificial Intelligence in the last few years, demonstrating the capability to solve many challenging and meaningful machine learning tasks. However, training large neural net models using large amounts of data can often take days to weeks, even with today’s cutting-edge GPUs. Therefore, there is significant interest in the electrical engineering community to design and build new hardware systems that can accelerate these workloads and/or lower the energy consumption. One analog approach uses crossbar arrays of Non-Volatile Memory (NVM), wherein highly parallelized multiply-accumulate operations are performed at the location of the weight data. This is a non-Von Neumann architecture, which avoids expensive data transfers between the memory and the CPU, potentially achieving orders of magnitude performance/energy improvements. In this talk, I will present our group’s recent work towards achieving ‘computer science equivalent’ accuracies in such a system despite the presence of significant NVM non-idealities. I will also address circuit requirements, which tend to be significantly different from conventional memory design and discuss tradeoffs that influence area, effective performance and power.

11:50-12:00 **Break**

12:00-12:30 **Session IV: Poster Short Speeches**
Chair: Koji Hashimoto (Fukuoka Univ.)

Poster 1 **Evaluation of FPGA Routing Architecture with H-Tree Topology**
Theingi Myint, Motoki Amagasaki, Masahiro Iida, Toshinori Sueyoshi (Kumamoto Univ.)

Poster 2 **Flexible Automated Transistor sizing Tool for Scalable Logic Module Architecture**
Lamiae Haddacha, Motoki Amagasaki, Masahiro Iida, Toshinori Sueyoshi (Kumamoto Univ.)

Poster 3 **A Compact and Efficient Inference Technique for Deep Neural Networks on FPGAs**
Qian Zhao[1], Toki Matsumoto[1], Yukikazu Nakamoto[1], Shinya Honda[2], Kazutoshi Wakabayashi[3] ([1]Univ. of Hyogo, [2]Nagoya Univ., [3]NEC)

Poster 4 **LSTM Acceleration on a Multi-FPGA System**
Yugo Yamauchi, Hideharu Amano (Keio Univ.)

Poster 5 **Low-Power Freeze-Safe IoT-Driven Processor Development Using Large-Scale Heterogeneous Co-Simulation Method toward Communication-Centric Design Space Exploration**
Hyungyun Moon, Jeonghun Cho, Daejin Park (Kyungpook National Univ., Korea)

Poster 6 **Remote On-Demand Code Execution Framework using Code Memory Cloudification for Low-Power, Large-Scaled IoT Application**s
Dongkyu Lee, Jeonghun Cho, Daejin Park (Kyungpook National Univ., Korea)

Poster 7 **High-speed Hardware Implementation of 8-bit per Item Frequent Items Counter**
Katsumi Inoue[1], Trong-Thuc Hoang[2], Xuan-Thuan Nguyen[2], Hong-Thu Nguyen[2], Cong-Kha Pham[2] ([1]Advanced Original Technologies, [2]Univ. of Electro-Communications)

Poster 8 **A multi-FPGA accelerator for GoogLeNet**
Kensuke Iizuka, Kazusa Musha, Hideharu Amano (Keio Univ.)

Poster 9 **An Imprecise 4-2 Compressor Design with Low Error Rate**
Yu-Cheng Cheng, Yi-Fong Lin, Shao-Chi Liao, Tung-Chi Wu, Yen-Jen Chang (National Chung Hsing Univ., Taiwan)

Poster 10 **A micro-controller for MTJ-based Non-volatile Flip-flops for data verification**
Takeharu Ikezoe[1], Takuya Kojima[1], Hideharu Amano[1], Junya Akaike[2], Kimiyoshi Usami[2], Keizo Hiraga[3], Yusuke Shuto[3], Kojiro Yagami[3] ([1]Keio Univ., [2]Shibaura Institute of Tech., [3]Sony Semiconductor Solutions)

Poster 11 **An Ultra Low-power Automatic Body Bias Tuning Scheme Using SOTB technology**
Hayate Okuhara, Akram Ben Ahmed, Hideharu Amano (Keio Univ.)

Poster 12 **FPGA-based Accelerator for LZ77 with Parallel Duplication Check**
Seungdo Choi, Youngil Kim, Jongmin Chun, Yong Ho Song (Hanyang Univ., Korea)

Poster 13 **Low Power Dual Edge Triggered Flip Flop with a Switched Cross Coupled Inverter Chain and TGs**
Jaeheum Lee, Deokhwan Kim, Kyoungrok Cho (Chungbuk National Univ., Korea)

Poster 14 **Performance Evaluation of 3D-Stacked Processor under Temperature Constraints**
Naoya Niwa[1], Tomohiro Totoki[1], Hiroki Matsutani[1], Michihiro Koibuchi[2], Hideharu Amano[1] ([1]Keio Univ., [2]National Institute of Informatics)

Poster 15 **Evaluation System for Low-power LSIs using SOTB Technology towards Software-based Body Bias Control**
Kenji Oshiro, Shinsuke Hamada, Atsushi Koshiba, Mitaro Namiki (Tokyo Univ. of Agriculture and Technology)

Poster 16 **In-Situ Detector-Based AFS System on an FPGA**
Dam Minh Tung, Nguyen Van Toan, Jeong-Gun Lee (Hallym Univ., Korea)

Poster 17 **Performance Evaluation of Tsunami Simulation on FPGA by High-level Synthesis using OpenCL**
Fumiya Kono, Naohito Nakasato, Kensaku Hayashi, Alexander Vazhenin (Univ. of Aizu)

Poster 18 **ecTALK: Energy Efficient Coherent Transprecision Accelerators–The Bidirectional Long Short-Term Memory Neural Network Case**
Dionysios Diamantopoulos, Heiner Giefers, Christoph Hagleitner (IBM Research – Zurich, Switzerland)

Poster 19 **New Industrial Computing Platform (Concept Presentation)**
Shumpei Kawasaki[1], Kesami Hagiwara[2], Akira Tsukamoto ([1]SH Consulting, [2]Univ. of Electro-Communications)

12:30-13:50 Lunch Time Break

13:50-14:00 Poster Open: 7th floor poster show room

14:00-14:50 Session V

14:00-14:50 **Keynote Presentation 4**
Co-chairs: Yasutaka Wada (Meisei Univ.), Koyo Nitta (NTT)

Designing a Power and Energy Stack for Exascale Systems
Martin Schulz (Technische Univ. München, Germany)

Abstract: Both power and energy are major design constraints as we approach the exascale era. Hardware efforts alone will no longer be sufficient to tackle this problem; instead we need comprehensive software developments that go along with any hardware effort and that help manage these scarce resources. This will have a significant impact on all layers of the exascale software stack, starting from low-level measurement and control capabilities, interactions with runtimes and resource management systems, all the way to interfaces with applications. In this talk I will highlight several ongoing projects with partners in Japan, the US and Europe to create a comprehensive software stack that can tackle this challenge and help mitigate the impact we see in power and energy constraint systems. This will include work on reducing variability, active runtime control in parallel programs, OS-level resource management as well as suitable application interfaces. Combined, these efforts will lead us to a vertically integrated software stack that enables power- and energy-aware computing and can help deliver an exascale system in the coming years.

14:50-15:40 **Break (Poster Open: 7th floor poster show room)**

15:40-16:20 **Session VI: Design Methodologies**
Co-chairs: Yuichiro Shibata (Nagasaki Univ.), Kyoung-Rok Cho (Chungbuk National Univ.)

15:40-16:05 **Design Automation Methodology of a Critical Path Monitor for Adaptive Voltage Controls**
Ryosuke Kazami, Hayate Okuhara, Hideharu Amano (Keio Univ.)

16:05-16:20 **3D-Cool: Design and Development of Adaptive Thermal-Aware Three-Dimensional NoC-Based Multiprocessor Chip**
Vinod Pagracious[1], Ranjitha Dash[2], Ashok Kumar Turuk[2] ([1]American University in Dubai, UAE, [2]National Institute of Technology Rourkela, India)

16:20-16:35 **Break**

16:35-18:05 **Session VII: Panel Discussions**

Topics: “Challenges to the Scaling Limits: How Can We Achieve Sustainable Power-Performance Improvements?”
Organizer & Moderator: Koji Inoue (Kyushu Univ.)
Panelists: Takuya Araki (NEC)
Takumi Maruyama (Fujitsu)
Pritish Narayanan (IBM Research - Almaden)
Takashi Oshima (Hitachi)
Martin Schulz (Technische Univ. München)

April 20, 2018 *Main Hall(7th Floor), Yokohama Joho Bunka Center* *(Yokohama Media & Communications Center)*

9:30-10:20 **Session VIII**

9:30-10:20 **Keynote Presentation 5**
Co-chairs: Yukinori Sato (Toyohashi Univ. of Technology), Hiroki Matsutani (Keio Univ.)

Multiscale Dataflow ASICs – Easy, Fast, Low Cost
Oskar Mencer (Maxeler Technologies / Imperial College London, UK)

Abstract: AI algorithms are rapidly evolving while a lot of investment is being directed into custom ASICs for AI which can take years to design and build. Developing a single chip for AI is very difficult since we do not know which algorithms will be most popular for a particular task by the time the chip is finished, and in addition there are many tasks with a wide range of different optimal AI algorithms. We propose Multiscale Dataflow as a methodology and infrastructure to minimize the time and cost for making a new ASIC, and therefore allow for very fast and very efficient development of new AI chips immediately when new algorithms come out, or for specialist domains and challenges. Thus, our dataflow methodology allows us to adapt quickly and create AI chips for all future AI algorithms.

10:20-10:40 **Break (Poster Open: 7th floor poster show room)**

10:40-11:55 **Session IX: Neural Networks**
Co-chairs: Yuetsu Kodama (Riken), Sugako Otani (Renesas Electronics)

10:40-11:05 **XNORBIN: A 95 TOp/s/W Hardware Accelerator for Binary Convolutional Neural Networks**
Andrawes Al Bahou, Geethan Karunaratne, Renzo Andri, Lukas Cavigelli, Luca Benini (ETH Zurich, Switzerland)

11:05-11:30 **ecTALK: Energy Efficient Coherent Transprecision Accelerators - The Bidirectional Long Short-Term Memory Neural Network Case**
Dionysios Diamantopoulos, Heiner Giefers, Christoph Hagleitner (IBM Research - Zurich, Switzerland)

11:30-11:55 **EMAXVR: A Programmable Accelerator Employing Near ALU Utilization to DSA**
Takahiro Ichikura, Ryusuke Yamano, Yuma Kikutani, Renyuan Zhang, Yasuhiko Nakashima (Nara Institute of Science and Tech.)

11:55-13:20 **Lunch Time Break**

13:20-14:10 **Session X**

13:20-14:10 **Keynote Presentation 6**
Co-chairs: Hideharu Amano (Keio Univ.), Kunio Uchiyama (Hitachi)

Unlocking Hidden Performance: Examples from FPGA-Based Neural Nets
Ephrem Wu (Xilinx, USA)

Abstract: Reconfigurable numerical solutions often leave performance on the table, typically achieving only a third to a half of the potential throughput. Data movement between memory and compute in parallel algorithms presents a particularly difficult "feeding-the-beast" problem. It is possible to meet this challenge by mapping parallel algorithms to a minimalist hardware architecture, and by selecting numerical representations to reduce memory capacity, bandwidth, and energy. Drawing from our experience with a reconfigurable compute unit for neural networks, we present some principles to unlock latent FPGA performance. We believe that these principles are general enough to be applicable to other parallel numerical applications.

14:10-14:30 **Break (Poster Open: 7th floor poster show room)**

14:30-15:35 **Session XI: Signal Processing**
Co-chairs: Masanori Muroyama (Tohoku Univ.), Yasutaka Wada (Meisei Univ.)

14:30-14:55 **Data Selection and De-noising Based on Reliability for Long-Range and High-Pixel Resolution LiDAR**
Ken Tanabe, Hiroshi Kubota, Akihide Sai, Nobu Matsumoto (Toshiba)

14:55-15:20 **A Programmable Analog Calculation Unit for Vector Computations**
Noriyuki Uetake, Renyuan Zhang, Takashi Nakada, Yosuhiko Nakashima (Nara Institute of Science and Tech.)

15:20-15:35 **Subthreshold Logic for Low-Area and Energy Efficient True Random Number Generator**
Mathieu Coustans[1], Abdelkarim Cherkaoui[2], Laurent Fesquet[1,2], Christian Terrier[3], Stephanie Salgado[3], Thomas Eberhardt[3], Maher Kayal[1] ([1]Swiss Federal Institute of Tech., Switzerland, [2]Univ. Grenoble Alpes, France, [3]EM Microelectronic, Switzerland)

15:35-15:55 **Break**

15:55-17:40 **Session XII: Processor Architectures**
Co-chairs: Hajime Shimada (Nagoya Univ.), Kotaro Shimamura (Hitachi)

15:55-16:20 **An Energy-aware Set-level Refreshing Mechanism for eDRAM Last-Level Caches**
Masayuki Sato, Zehua Li, Ryusuke Egawa, Hiroaki Kobayashi (Tohoku Univ.)

16:20-16:45 **Power Performance Analysis of ARM Scalable Vector Extension**
Tetsuya Odajima, Yuetsu Kodama, Mitsuhisa Sato (Riken)

16:45-17:00 **A Two-stage-pipeline CPU of SH-2 Architecture Implemented on FPGA and SoC for IoT, Edge AI and Robotic Applications**
Kesami Hagiwara[1], Tomoichi Hayashi[2], Shumpei Kawasaki[3], Fumio Arakawa[5], Oleg Endo, Hayato Nomura[6], Akira Tsukamoto, Duong Nguyen[4], Binh Nguyen[4], Anh Tran[4], Hoan Hyunh[4], Ikuo Kudoh[3], Cong-Kha Pham[1] ([1]Univ. of Electro-Communications, [2]SH Consulting (at the time of development), [3]SH Consulting, [4]SH Consulting Viet Nam, [5]Nagoya Univ., [6]Univ. of Tokyo)

17:00-17:40 **<u>Invited Presentation</u>**

Designing the Next Billion Chips: How RISC-V is Revolutionizing Hardware
Yunsup Lee (SiFive, USA)

Abstract: Open source has revolutionized software. Now it's hardware's turn. In this talk, I present the chip design economics for today, introduce the free and open

RISC-V instruction set architecture, and talk about how RISC-V, open-source hardware, and SiFive are changing the chip design economics for the next billion chips that are being built for IoT, edge computing, machine learning, and artificial intelligence applications.

17:40-18:00 **Poster Award and Closing Remark**
Makoto Ikeda, Program Committee Co-chair (Univ. of Tokyo)