Final Program

April 15, 2020 (Japan Standard Time)

9:00-10:30 **Special Invited Lecture 1**
Co-chairs: Takatsugu Ono (Kyushu Univ.), Hiroki Matsutani (Keio Univ.)

9:00-10:30 **Evolving Hardware Security Landscape in the AI Era**
Guru Prasadh Venkataramani (George Washington Univ., USA)

Abstract: Among the many emerging domains, Artificial Intelligence (AI) based applications have begun to permeate everywhere, and as such, ensuring their secure operation has become paramount. Meanwhile, hardware security is fast becoming a pressing problem with the growth of security attacks targeting processors and their peripherals. Therefore, hardware researchers have to better understand their designs and guard these critical AI applications against security flaws. In this talk, we will first visit the fundamental problems surrounding sensitive information leakage that dominate security research landscape and has the potential to hamper AI applications. More recently, timing channels have emerged as one of the most dangerous forms of information leakage leaving no physical evidence of an attack. We will explore the basic solutions needed to alleviate these timing channels that includes our early work to detect them on oft-used hardware structures. We will then go over some of our recent research findings that aims to find stronger indicators for timing channels. Next, we will see how more sophisticated adversaries may target inter-cache hardware mechanisms such as cache coherence protocols. Finally, we will wrap up the talk with some of my thoughts on how to design low-cost, hardware-software cooperative solutions to effectively defend against these attacks and how they could be customized for the AI domain.

10:30-11:00 **Break**

11:00-12:30 **Special Invited Lecture 2**
Co-chairs: Takatsugu Ono (Kyushu Univ.), Hiroki Matsutani (Keio Univ.)

11:00-12:30 **Using AI to Bridge the Gap Between AI Models and the Hardware of Today and Tomorrow**
Luis Ceze (Univ. of Washington, USA)

Abstract: There is an increasing need to bring machine learning to a wide diversity of hardware devices. Current frameworks rely on vendor-specific operator libraries and optimize for a narrow range of server-class GPUs. Deploying workloads to new platforms — such as mobile phones, embedded devices, and accelerators (e.g., FPGAs, ASICs) — requires significant manual effort. In this talk I will present our work on the TVM stack, which exposes graph-level and operator-level optimizations to provide performance portability to deep learning workloads across diverse hardware back-ends. TVM solves optimization challenges specific to deep learning, such as high-level operator fusion, mapping to arbitrary hardware primitives, and memory latency hiding. It also automates optimization of low-level programs to hardware characteristics by employing a novel, learning-based cost modeling method for rapid exploration of code optimizations. To address threat of changes in algorithms, models, operators, or numerical systems threaten to the viability of specialized hardware accelerators, we developed VTA, a programmable

deep learning architecture template tightly coupled to TVM. VTA achieves this flexibility via a parameterizable architecture, two-level ISA, and a JIT compiler. The TVM/VTA was incubated as an Apache Foundation project and is benefiting from a thriving community of developers. I will end the talk with ideas and possibilities for AI systems in a post-Moore's law world, including using hybrid molecular electronic systems for similarity search.

12:30-14:00 **Lunch Time Break**

14:00-15:40 **Poster Session**
Co-chairs: Koji Hashimoto (Fukuoka Univ.), Yuki Kobayashi (NEC)

Poster 1 **ReRAM Cell Reliability Variation Tolerated High-Speed Approximate Storage for Machine Learning**
Chihiro Matsui, Ken Takeuchi (Chuo Univ.)

Poster 2 **Xbyak_aarch64; JIT Assembler for Next Generation Supercomputer**
Kentaro Kawakami[1], Saitou Moriyuki[2], Kouji Kurihara[1], Naoto Fukumoto[1] ([1]Fujitsu Labs., [2]Fujitsu)

Poster 3 **Implementation of a Packet-Switching Router on a Multi-FPGA System**
Tomoki Shimizu, Kohei Ito, Kensuke Iizuka, Yugo Yamauchi, Kazuei Hironaka, Hideharu Amano (Keio Univ.)

Poster 4 **Implementation of an Application Utilized Multi-Switch on a Multi-FPGA System**
Kohei Ito[1], Kensuke Iizuka[1], Yugo Yamauchi[1], Kazuei Hironaka[1], Yao Hu[2], Michihiro Koibuchi[2], Hideharu Amano[1] ([1]Keio Univ., [2]National Institute of Informatics)

Poster 5 **Preliminary Performance Analysis of Distributed DNN Training with Relaxed Synchronization**
Koichi Shirahata, Amir Haderbache, Naoto Fukumoto (Fujitsu Labs.)

Poster 6 **Hardware design of AmoebaSat algorithm on FPGA for solving Boolean Satisfiability**
YingJie Yan, Hideharu Amano, Masashi Aono (Keio Univ.)

Poster 7 **Low-Latency Memory Packet Network Using Bypassing**
Yoshiya Shikama[1], Kawano Ryuta[1], Akram Ben Ahmed[1], Hiroki Matsutani[1], Michihiro Koibuchi[2], Hideharu Amano[1] ([1]Keio Univ., [2]National Institute of Informatics)

Poster 8 **Compiler Framework for Spatial Mapping CGRA using LLVM**
Ayaka Ohwada, Takuya Kojima, Hideharu Amano (Keio Univ.)

Poster 9 **Acceleration of Printed Circuit Board Analysis using CNN**
Eiji Ohta, Naoto Fukumoto, Yasumoto Tomita, Takuji Yamamoto (Fujitsu Labs.)

Poster 10 **SHA-256 Implementation on Coarse-Grained Reconfigurable Architecture**
Van Dai Phan, Thi Hong Tran, Yasuhiko Nakashima (Nara Institute of Science and Technology)

Poster 11 **A cache coherent protocol for Thru-Chip-Interface**
Tomohiro Teraoka, Takuya Kojima, Hideto Kayashima, Hideharu Amano (Keio Univ.)

April 16, 2020 (Japan Standard Time)

9:00-9:10 **Session I**

9:00-9:10 **Welcome and Opening Remarks**
Co-chairs: Yuki Kobayashi (NEC), Hiroki Matsutani (Keio Univ.)

Kunio Uchiyama	*Chair of the Organizing Committee*
Jose Renau	*Chair of IEEE/CS TCMM*
Tugumichi Shibata	*President of IEICE/ES*

9:10-10:00 **Session II**

9:10-10:00 **Keynote Presentation 1**
Co-chairs: Yasutaka Wada (Meisei Univ.), Yukinori Sato (Toyohashi Univ. of Technology)

Reconfigurable Cloud Scale AI
Aaron Smith (Microsoft, USA)

Abstract: TBA

10:00-10:10 **Break**

10:10-11:00 **Session III:**

10:10-11:00 **Keynote Presentation 2**
Co-chairs: Kunio Uchiyama (AIST), Yasushi Inoguchi (JAIST)

Disruptive Evolutions: Technology Challenges and Countermeasures
Shinichi Yoshioka (Renesas Electronics)

Abstract: Disruptive evolutions of digital transformation in industries and societies are invoking technology challenges to embedded processors and their solutions. After introducing key issues and criteria of those challenges in several market segments such as automotive, industry and infrastructure, countermeasure examples are to be explained showing how to address those issues clearing the criteria such as efficiencies (performance per power & cost), effective performance in real use cases, system robustness, easy-to-use/reuse SW and development environment. A real solution example with high-end processors and micro controllers for a mission critical application is introduced to summarize each technology explained in the presentation.

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

11:10-12:00 **Session IV:**

11:10-12:00 **Keynote Presentation 3**
Co-chairs: Tohru Ishihara (Nagoya Univ.), Yasuo Unekawa (Toshiba Electronic Devices & Storage)

How to Uplift the World with "Memory"
Kenichi Mori (Kioxia)

Abstract: NAND flash memory expanded its market and application and changed our lifestyle by reducing the cost per bit ($/GB). Many innovations were introduced to keep the cost trend and to improve the performance. Now, as the next step, 5G and AI are changing our society and advanced computing systems is required for such high level computing. Non-volatile memory is a key component to enable this paradigm shift. The challenges and opportunities of NAND flash and other emerging memories for next decades will be discussed.

12:00-13:00 **Lunch Time Break**

13:00-13:40 **Session V**

13:00-13:40 **Invited Presentation 1**
Co-chairs: Ryuichi. Sakamoto (Univ. of Tokyo), Atsutake Kosuge (Hitachi)

Virtualization for Non-volatile Memory Devices
Takahiro Hirofuchi (AIST)

Abstract: Non-volatile memory (NVM) technologies, being accessible in the same manner as DRAM, are considered indispensable for expanding main memory capacities. For example, Intel Optane DCPMM is a long-awaited product that drastically increases main memory capacities. However, a substantial performance gap exists between DRAM and NVM. This performance gap in main memory presents a new challenge to researchers; we need new system software technologies efficiently supporting emerging hybrid memory architecture. In this talk, I first present RAMinate, a hypervisor-based virtualization mechanism for hybrid memory systems, and a key technology to address the performance gap in main memory systems. It provides great flexibility in memory management and maximizes the performance of virtual machines (VMs) by dynamically optimizing memory mappings. Through experiments, we confirmed that even though a VM has only 1% of DRAM in its RAM, the performance degradation of the VM was drastically alleviated by memory mapping optimization. This talk will also cover performance emulation of memory devices, which is indispensable for system software studies targeting future memory devices. We developed a new NVM emulation mechanism that is not only light-weight but also aware of a read/write latency gap in NVM-based main memory. The emulator accurately emulates write-latencies of NVM-based main memory: in our experiments, it emulated the NVM write latencies in a range from 200 ns to 1000 ns with negligible errors from 0.2% to 1.1%. If the time for this talk remains, I will also introduce our ongoing research project that aims at massively energy-efficient memory subsystems by developing approximate computing techniques throughout whole the memory hierarchy. This is joint work with spintronics researchers. We are rethinking the design of memory cells, computer hardware and software from the scratch.

13:40-14:10 **Break**

14:10-15:25 **Session VI: Application specific processors and system**
Co-chairs: Yuichiro Shibata (Nagasaki Univ.), Yasutaka Wada (Meisei Univ.), Takumi Uezono (Hitachi)

14:10-14:35 **A RISC-V Processor with an Inter-Chiplet Wireless Communication Interface for Shape-Changeable Computers**
Junichiro Kadomoto, Hidetsugu Irie, Shuichi Sakai (Univ. of Tokyo)

14:35-15:00 **Space Responsive Multithreaded Processor (SRMTP) for Spacecraft Control**
Shota Nakabeppu, Yosuke Ide, Masahiko Takahashi, Yuta Tsukahara, Hiromi

Suzuki, Haruki Shishido, Nobuyuki Yamasaki (Keio Univ.)

15:00-15:25 **MMT-based Multi-channel Video Transmission System with Synchronous Processing Architecture**
Yasuhiro Mochida, Takahiro Yamaguchi, Ken Nakamura (NTT)

15:25-15:45 Break

15:45-16:35 Session VII: Cool Software
Co-chairs: Hiroyuki Takizawa (Tohoku Univ.), Shunsuke Sasaki (Toshiba Electoronic Devices & Storage), Atsutake Kosuge (Hitachi)

15:45-16:10 **User Insensible Sliding Firmware Update Technique for Flash-Area/Time-Cost Reduction toward Low-Power Embedded Software Replacement**
Jisu Kwon[1], Moon Gi Seok[2], Daejin Park[1] ([1]Kyungpook National Univ., Korea, [2]Nanyang Technological Univ., Singapore)

16:10-16:35 **XwattPilot: A Full-stack Cloud System Enabling Agile Development of Transprecision Software for Low-power SoCs**
Dionysios Diamantopoulos[1], Florian Scheidegger[1,2], Stefan Mach[2], Fabian Schuiki[2], Germain Haugou[2,3], Michael Schaffner[2], Frank K. Gürkaynak[2], Christoph Hagleitner[1], Cristiano Malossi[1], Luca Benini[2,4] ([1]IBM research, [2]Integrated Systems Lab., Switzerland, [3]GreenWaves Technologies, France, [4]Univ. of Bologna, Italy)

16:35-16:55 Break

16:55-17:45 Session VIII: Low Power Processors
Co-chairs: Kotaro Shimamura (Hitachi), Fumio Arakawa (Nagoya Univ.), Takumi Uezono (Hitachi)

16:55-17:20 **A 0.4-0.9V, 2.87pJ/cycle Near-Threshold ARM Cortex-M3 CPU with In-Situ Monitoring and Adaptive-Logic Scan**
Markus Hiienkari[1], Navneet Gupta[1], Jukka Teittinen[1], Jesse Simonsson[1], Matthew Turnquist[1], Jonas Eriksson[1], Risto Anttila[1], Ohto Myllynen[1,2], Hannu Rämäkkö[1], Sofia Mäkikyrö[1], Lauri Koskinen[1,2] ([1]Minima Processor, [2]Univ. of Turku, Finland)

17:20-17:45 **A 0.55V 6.3uW/MHz Arm Cortex-M4 MCU with Adaptive Reverse Body Bias and Single Rail SRAM**
Dennis Walter, Andre Scharfe, Alexander Oefelein, Florian Schraut, Heiner Bauer, Farkas Csazar, Robert Niebsch, Jörg Schreiter, Holger Eisenreich, Sebastian Höppner (Racyics GmbH, Germany)

April 17, 2020 (Japan Standard Time)

9:00-9:40 **Session IX**

9:00-9:40 **Invited Presentation 2**
Co-chairs: Yuki Kobayashi (NEC), Koyo Nitta (NTT), Hiroki Matsutani (Keio Univ.)

Intel Optane™ Data Center Persistent Memory - A True Breakthrough to Break the Traditional MemoryStorage Technologies Barriers
Jane Jianping Xu / Kaushik Balasubramanian (Intel, USA)

Abstract: Intel Optane™ DC persistent memory offers big and cost-effective persistent memory for all data center current and possibly future applications. With this new type of memory, a user will be able to accelerate data processing closer to CPUs with substantially lower latency in contrast with previously located on a NAND disk. The new Intel® Optane™ DC persistent memory redefines traditional architectures, offering a large and persistent memory tier at affordable cost. With breakthrough performance levels in memory intensive workloads, virtual machine density, and fast storage capacity Intel Optane™ DC persistent memory - combined 2nd gen Intel Xeon® Scalable processors - accelerates IT transformation to support the demands of the data era, with faster-than-ever-before analytics, cloud services, and next-generation communication services.

9:40-10:00 **Break**

10:00-10:50 **Session X: Memory Systems**
Co-chairs: Hajime Shimada (Nagoya Univ.), Ryohei Kobayashi (Univ. of Tsukuba), Yutaka Uematsu (Hitachi)

10:00-10:25 **Tileable Monolithic ReRAM Memory Design**
Meenatchi Jagasivamani[1], Candace Walden[1], Devesh Singh[1], Luyi Kang[1], Mehdi Asnaashari[2], Sylvain Dubois[2], Bruce Jacob[1], Donald Yeung[1] ([1]Univ. of Maryland, [2]Crossbar Inc., USA)

10:25-10:50 **Energy-efficient Design of an STT-RAM-based Hybrid Cache Architecture**
Masayuki Sato, Xue Hao, Kazuhiko Komatsu, Hiroaki Kobayashi (Tohoku Univ.)

10:50-11:00 **Break**

11:00-11:50 **Session XI: Keynote Presentation 4**
Co-chairs: Masato Suzuki (Socionext), Takuya Nakaike (IBM)

11:00-11:50 **An Extremely Quantized Deep Neural Network Accelerator for Edge Devices**
Hiroyuki Tokunaga (LeapMind)

Abstract: Deep learning greatly improved the performance in wide variety of research areas including image processing and audio processing. Although significant progress has been made in terms of accuracy, huge computation cost is still remaining as a big issue. Various ideas including *quantization* have been proposed to solve this issue. In terms of quantized neural network, *quantization* means a significant reduction of numerical precision of a computation, such as reducing from 32-bit floating point number to an 8-bit integer.

The ultimate quantization is making the weight and activation to just 1 bit. It is known that the neural network works even in such an extreme case, although the accuracy is slightly lowered. Techniques for reducing the numeric precision smaller than 8 bits is called extremely low bit quantization or ultra low bit quantization. In this problem setting, the computation of the inner product of the two vectors can be replaced by several bit manipulations. In this talk, we will give an overview of a new accelerator IP dedicated to extremely quantized neural network. Also, how to realize the extremely low bit quantization will be presented.

11:50-13:20 **Lunch Time Break**

13:20-14:35 **Session XII: Accelerators**
Co-chairs: Sugako Otani (Renesas Electronics), Salita Sombatsiri (NEC), Yutaka Uematsu (Hitachi)

13:20-13:45 **A Novel In-DRAM Accelerator Architecture for Binary Neural Network**
Haerang Choi[1,2], Yosep Lee[2], Jae-Joon Kim[3], Sungjoo Yoo[1] ([1]Seoul National Univ., [2]SK Hynix, [3]POSTECH, Korea)

13:45-14:10 **Non-Volatile Coarse Grained Reconfigurable Array Enabling Two-step Store Control for Energy Minimization**
Kimiyoshi Usami[1], Sosuke Akiba[1], Hideharu Amano[2], Takeharu Ikezoe[2], Keizo Hiraga[3], Kenta Suzuki[3], Yasuo Kanda[3] ([1]Shibaura Institute of Technology, [2]Keio Univ., [3]Sony Semiconductor Solutions)

14:10-14:35 **An Area-Efficient Implementation of Recurrent Neural Network Core for Unsupervised Anomaly Detection**
Takuya Sakuma, Hiroki Matsutani (Keio Univ.)

14:35-14:45 **Poster Award and Closing Remark**
Makoto Ikeda, Program Committee Co-chair (Univ. of Tokyo)