Dissemination Workshop

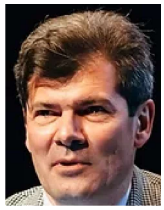
Monday 19 September 2022

Embedded Artificial Intelligence (EAI) – Devices, Systems, and Industrial Applications

(Room - Aula 1D) *Half day*

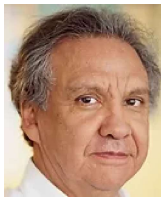
13:30-17:30

Chairs



Ovidiu Vermesan (SINTEF, NO)

Ovidiu Vermesan is Chef Scientist at SINTEF Digital, Oslo, where he is involved in applied research on future edge autonomous intelligent systems and edge AI, wireless sensing devices and networks, smart systems integration, microelectronics design of integrated systems (analogue and mixed signal), IIoT. He holds a PhD in Microelectronics and a Master of International Business (MIB). His applied research activities focus on advancing edge AI processing, embedded electronics, wireless and smart sensing technologies, and the convergence of these technologies and applying the developments to applications such as green mobility, energy, buildings, autonomous systems, electric connected, autonomous, and shared vehicles, and industrial manufacturing. He is currently the technical co-coordinator of the Artificial Intelligence for Digitising Industry (AI4DI) project.



Mario Diaz Nava (STMicroelectronics, FR)

Mario Diaz Nava has a Ph.D, and M.S. both in computer science, from Institut National Polytechnique de Grenoble, France, and B.S. in communications and electronics engineering from Instituto Politecnico National, Mexico. He has worked in STMicroelectronics since 1990. He currently has the position of ST Grenoble R&D Cooperative Programs Manager, and he has actively participated, for the last five years, in several H2020 IoT projects (ACTIVATE, IoF2020, Brain-IoT), working in key areas such as Security and Privacy, Smart Farming, IoT System modelling, and edge computing. He is currently leading the ANDANTE project devoted to developing neuromorphic ASICS for efficient AI/ML solutions at the edge. He has published more than 35 articles in these areas.

Björn Debaillie (imec, BE)

Björn Debaillie leads imec's collaborative R&D activities on cutting-edge IoT technologies in imec. As program manager, he is responsible for the operational management across programs and projects, and focusses on strategic collaborations and partnerships, innovation management, and public funding policies. As chief of staff, he is responsible for executive finance and operations management and transformations. Björn coordinates semiconductor-oriented public funded projects and seeds new initiatives on high-speed communications and neuromorphic sensing. He currently leads the 35M€ TEMPO project on neuromorphic hardware technologies, enabling low-power chips for computation-intensive AI applications (www.tempo-ecsel.eu). Björn holds patents and authored international papers published in various journals and conference proceedings. He also received several awards, was elected as IEEE Senior Member and is acting in a wide range of expert boards, technical program committees, and scientific/strategic think tanks.

Overview

The International on Embedded Artificial Intelligence (EAI) - Devices, Systems, and Industrial is part of the ESSCIRC ESSDERC 2022 European Solid-state Circuits and Devices Conference held in Milan, Italy, on 19

(SoC) accelerators and novel neuromorphic processing architectures.

This workshop brings together academics, researchers, and industry practitioners from various fields, including artificial intelligence, cyber-physical and embedded systems, microelectronics circuits and devices, edge computing, and autonomous integrated systems. Different views on todays and future embedded artificial intelligence are discussed the latest developments on devices, techniques, and industrial edge applications are presented. The workshop covers invited keynote talks and presentations with opportunities for researchers in the audience to interact with the speakers, discuss novel and exciting research, and establish new and fruitful collaborations in the field of embedded artificial intelligence.

The EAI workshop is co-organised by three large-scale ECSEL JU projects, AI4DI, ANDANTE, and TEMPO, to provide a platform to exchange knowledge and ideas among experts and professionals interested in advances in AI circuits and devices design, AI hardware architectures, industrial edge AI technologies, toolchains, and applications

13:30 - 13:40

European Embedded AI Ecosystem

Large-scale projects overview: ECSEL JU AI4DI, TEMPO, ANDANTE

Frank Badstuebner (Infineon Technologies, DE), Björn Debaillie (imec, BE), Mario Diaz Nava (STMicroelectronics, FR), Ovidiu Vermesan (SINTEF, NO)

13:40 - 14:00

Bio-inspired On-line Learning Circuits for Low-power Extreme-edge Spiking Neural Network Processing Systems

The presentation will address the challenges to develop novel brain-inspired processing methods, technologies and learning circuits for low-power extreme-edge Spiking Neural Network processing systems, and apply them to different application scenarios.

Speaker

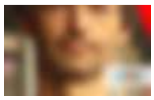
Giacomo Indiveri (University of Zurich, CH)

Giacomo Indiveri is a dual Professor at the Faculty of Science of the University of Zurich and at Department of Information Technology and Electrical Engineering of ETH Zurich, Switzerland. He is the director of the Institute of Neuroinformatics of the University of Zurich and ETH Zurich. 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. 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. His group uses these neuromorphic circuits to validate brain inspired computational paradigms in real-world scenarios, and to develop a new generation of fault-tolerant event-based neuromorphic computing 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.

14:00 - 14:15

Power Optimised Wafermap Classification for Semiconductor Process Monitoring

Today, the exploitation of AI solutions is very immersive and has wide applicability in almost all industrial fields. A key to profitability is the yield of the production and the quality of the product is above all. The control of wafer fabrication in the semiconductor industry is fundamental, especially due to the increasingly high daily



Paris, France, in 1994, where has been Assistant Professor in Computer Science until September 2004. From 1989 to 1996, F. Pétrot was one of the main contributors of the open-source Alliance VLSI CAD system, and from 1996 to 2004, he led a team focusing on the specification, simulation, and implementation of multiprocessor SoCs. He joined TIMA in September 2004, where he holds a professor position at Grenoble Institute of Technology, France. His research interests are in multiprocessor systems on chip architectures, including circuits and software aspects, and CAD tools for the design and evaluation of hardware/software systems.

14:15 - 14:30

Low-Power Analog In-memory Computing Neuromorphic Circuits

The neuromorphic circuits presented comprise synaptic weights and neurons including batch normalization, activation function, and offset cancelation circuits. These neuromorphic circuits comprise an effective 3.5 bits weight storage based on binary memory cells while the analog multiplication and addition operation is based on a voltage divider principle. To experimentally proof the working principle, three fully connected layers (50x20, 20x10 and 10x4) have been designed. The connection between layers is performed completely in the analog domain without ADCs and DACs in between. An inference state machine takes care of pipelining the layers for a proper operation during inference. The schematic and layouts of the neuromorphic circuits comprised in these layers have been automatically generated means a Fraunhofer IIS internal automation framework called UnilibPlus, a Python-based Cadence Virtuoso add-on. Simulation results of weight loading, transfer of input values, inference and read inference results via SPI interface show a correct operation of the designed ASIC achieving 9 nJ per inference with 5 us latency.

Speaker

Roland Müller (Fraunhofer IIS, DE)

Roland Müller obtained his bis B.Eng. at the OTH Regensburg in 2017 and his M.Sc. in 2019 at the FAU Erlangen, both in Electrical Engineering. In May 2019, he joined the Advanced Analog Circuits Group in the department of Integrated Circuits and Systems at Fraunhofer IIS, Erlangen (Germany), where he is working in the field of analog-mixed signal design of neural network accelerators with the main focus on the design automation of such circuits. Currently, he is working on his PhD degree with the preliminary thesis title Automatic synthesis, simulation and verification of analog-mixed-signal macro and base-level circuits for analog-mixed signal neural network accelerators with UnilibPlus. His main research interests include low power analog-mixed signal circuits, neuromorphic computing and electronic design automation.

14:30 - 14:45

Tools and Methodologies for Edge-AI Inference Accelerators

In the ANDANTE project several tools are developed to train and deploy either artificial neural networks, or spiking neural networks, in dedicated hardware accelerators. These tools provide hardware-aware training, automatic hardware generation, compilers, energy consumption estimation or simulation of the hardware implementations. The ANDANTE tools tackle the hardware/software co-design challenge when addressing neuromorphic hardware accelerators. A multidisciplinary approach combining neural network algorithm know-how, programming skills and integrated circuit design is necessary for developing software tools for neural network algorithms that consider the constrains of the hardware designs. The tools allow to optimize and verify the hardware design, reach the targeted KPIs, and reduce the time-to-market.

Speaker

Loreto Mateu (Fraunhofer IIS, DE)

Memory Devices with Lifelong Retention for Monolithic 3D-Inference Engine Applications

This presentation reports on back-end-of-line (BEOL) (process temperature<350°C) compatible vertically stacked indium gallium zinc oxide (IGZO) based multi-bit one-time programmable (OTP) ferroelectric (Fe) thin film transistor (TFT) memory devices with lifelong retention capability for the first time. Further, we have evaluated the performance of the IGZO-based OTP Fe-TFT as synaptic devices for a monolithic 3D (M3D) inference engine. The system-level simulation revealed inference accuracy of 97% for MNIST data in MLP neural network with a maximum accuracy loss of 1.5% over ten years without re-training. The proposed M3D inference engine also showed superior energy efficiency and cell area of 95.33 TOPS/W (binary) and 8F^2, respectively.

Speaker

Sourav De (Fraunhofer IPMS, DE)

Sourav De completed his Ph.D. from the Institute of Microelectronics, National Cheng Kung University. He is currently working as a scientist in Fraunhofer IPMS. His primary research interest involves emerging non-volatile memories, advanced logic technology, and non-conventional computing. Before joining Fraunhofer IPMS, he worked in Taiwan Semiconductor Research Institute to develop ferroelectric finFET devices.

15:00 - 15:30 *Break*

15:30 - 15:45

Addressing Physical and Functional Reverse Engineering Challenges for Advanced IoT Solutions with AI

Motivated by the threats of malicious modification and piracy arising from worldwide distributed supply chains, secured design flows, development tools and the physical and functional verification methods are paramount. The presentation gives an overview of a complete reverse engineering process for integrated circuits manufactured in technology nodes of 40nm and below. Building upon the presentation of individual reverse engineering process stages, this presentation connects analysis efforts and yields with their impact on hardware security. We evaluate the implementation of AI based methods enhancing the overall performance.

Speaker

Bernhard Lippmann (Infineon Technologies, DE)

Bernhard Lippmann received his diploma degree in Physics from the Technical University Munich (TUM), Germany in 1992. He worked with Hitachi Semiconductor Europe in Landshut, Type Engineering group from 1993 until 1998. Since 1999 he is with Infineon Technologies AG (Siemens), Digital Security Solutions Division in Munich. He has been involved in the project coordination of several public funded projects such as SYPASS, 2017 (www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/sypass), RESEC, 2019 (www.forschung-it-sicherheit-kommunikationssysteme.de/projekte/resec) and in ECSEL JU AI4DI project as supply chain leader for AI applications in semiconductor industry (https://ai4di.eu/).

15:45 - 16:00

Meeting the Latency and Energy Constraints on Timing-critical Edge-AI Systems

Smart devices, with AI capabilities, at the edge have demonstrated impressive application results. The current trend is to increase the image resolution and classification accuracy. However, computing object detection and classification tasks at the edge require both low latency and high-energy efficiency for these devices. In this talk, we will explore a novel architectural approach to overcome these limitations by using the attention mechanism of the human brain. The latter allows humans to selectively analyse the scene and thus limit the spent energy.

16:00 - 16:15

Low-power Neuromorphic Auditory Processing - Silicon to Applications

Event-driven computing on Neuromorphic hardware is an energy-efficient basis for real-time signal processing applications. Within ANDANTE, SynSense has pioneered new approaches for auditory processing applications using spiking neural networks. We will present our pipeline for building and deploying auditory processing applications to a new mixed-signal SNN ASIC, "Xylo".

Speaker

Dylan Muir (SynSense, CH)

Dylan Muir is the Vice President for Global Research Operations at SynSense. Dr. Muir is a specialist in architectures for neural computation. He has published extensively in computational and experimental neuroscience. At SynSense, he is responsible for the company research vision, and directing development of neural architectures for signal processing. Dr. Muir holds a Doctor of Science (PhD) from ETH Zürich, and undergraduate degrees (Honors) in Electronic Engineering and in Computer Science from QUT, Australia.

16:15 - 16:35

Embedded Artificial Intelligence

The presentation will give an overview on the impact of AI in semiconductor industry, as experienced by STMicroelectronics on two perspectives: the AI in semiconductor technology as enabler of Internet of Things (IoT), and the AI to manage the high level of complexity in semiconductor manufacturing.

Speaker

Sara Loi (STMicroelectronics, IT)

Sara Loi is a physicist (master's degree), expert in Optical Technologies, qualified expert in Radiation Protection and specialist in Health Physics. After a two-year experience in Optical Spectroscopy (Raman) for material characterization, she has been working for STMicroelectronics (Italy) since 2000. She started working in process technology R&D in photolithography area for Flash Memories, Embedded Memories and BCD Smart Power products. She has been assignee at IMEC (Interuniversity Research Center for Microelectronics, Belgium) for almost one year, involved in photolithography R&D of 45 nm technology node and beyond. She has worked for almost two years in the product division MEMS, Sensors and Analog, responsible of some programs on Gyroscope and Magnetic Sensor products. Since 2010, in the current position she has been responsible of R&D&I programs, funded in European FP7, ENIAC, ARTEMIS, H2020, Horizon Europe, ECSEL and KDT frameworks, or funded by national Ministry of Research (MIUR), national Ministry for Economic Development (MISE) or Lombardia Region. Sara has worked for three years as a Physicist at S. Raffaele Hospital (Milan) and National Institute of Tumors (Milan). She is author of several international publications and patents.

16:35 - 16:50

An Embedding Workflow for Tiny Neural Networks on ARM Cortex-M0(+) Cores

Neural networks are getting increasingly popular in always-on IoT edge devices for more precise and secure data analysis with less latency. However, due to the strict cost and power constraints, only the smallest microcontrollers could be used for such applications, making it difficult to use many of the available network frameworks. This presentation will introduce an end-to-end embedding workflow focused on tiny neural network deployment on ARM Cortex-M0(+) cores, covering all the steps, including network quantization, C code generation and performance verification. With a one-time development of a Python and C library following the method, one can easily embed different neural networks with very little manual effort afterward.

Architecting Edge AI Workflows for Predictive Maintenance in Industrial Applications

This presentation highlights on the importance of properly architecting the AI workflows for the design, development, and deployment of predictive maintenance applications in industrial applications at the edge. It is advocated that the selection of the frameworks/platforms employed for the task largely depend on the application, IIoT devices and their physical operating environments, rather than on the available knowledge and experience of the people, as suggested in similar literature. For this purpose, it was employed several existing AI frameworks and inference engines, that permit end-to-end solutions with various degree of automation and integration with Arm® Cortex®-M-based MCUs. The presentation addresses the verification and validation aspects of model design, development, and deployment of AI-based industrial applications and provides quantitative and qualitative insights that can ultimately allow the use of the right framework. The use case presented is a classification for predictive maintenance based on the vibration of generic rotating equipment (such as motors with vibration pumps, fans, compressors), common to many industrial applications, such as manufacturing.

Speaker

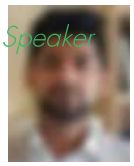
Ovidiu Vermesan (SINTEF, NO)

Ovidiu Vermesan is Chief Scientist at SINTEF Digital, Oslo, where he is involved in applied research on future edge autonomous intelligent systems and edge AI, wireless sensing devices and networks, smart systems integration, microelectronics design of integrated systems (analogue and mixed signal), IIoT. He holds a PhD in Microelectronics and a Master of International Business (MIB). His applied research activities focus on advancing edge AI processing, embedded electronics, wireless and smart sensing technologies, and the convergence of these technologies in different industrial sectors.

17:05 - 17:20

Food Ingredients Recognition Through Multi-label Learning

In this work, we present the need for an automated food assessment system and motivate the use Artificial System (AS) technology and digital cameras towards achieving this ambition. We concentrate on an important sub-problem, within automated food assessment, which is to accurately determine the key ingredients from dish images. This work describes a multi-label learning setting and evaluates various state-of-the-art deep learning models for recognizing the key ingredients in presented dish images.



Speaker

Ismail Rameez (Philips Research, NL)

Rameez Ismail is researcher at Royal Philips, whose research interests include Artificial Intelligence (AI) and High-Performance Computing (HPC). As a graduate, with master's in robotics and an Engineering Doctorate (EngD) in systems design, Rameez started his career with a brief stint (2016-2018) at NXP Semiconductors. Thereafter, in 2018, he joined Philips Research, as a scientist, with the ambition to accelerate the digital transformation of healthcare. Over the years, he has successfully applied advanced AI research to various healthcare use cases, including oral health, personal health & beauty and medical imaging.