Goto Session:

- 1.1 Opening Session: Plenary, Awards Ceremony & Keynote Addresses
- UB01 Session 1
  - 2.2 Energy Efficient Neural Networks
  - 2.3 High-Level Synthesis
  - 2.4 Model Checking
  - 2.5 GPU and GPU-based heterogeneous system management
  - 2.6 Circuit Locking and Camouflaging
  - 2.7 Special Session: Spintronics based New Computing Paradigms and Applications
  - 2.8 Enabling ICT Innovations for European SMEs
- UB02 Session 2
  - 3.1 Executive Session: Design Automation for Quantum Computing
  - 3.2 Approximate and Near-Threshold Computing
  - 3.3 Optimization Techniques for MPSoCs
  - 3.4 Optimizing Computing with Neuromorphic Architectures and Accelerators
  - 3.5 Memory Reliability
  - 3.6 Real-time Multiprocessing
  - 3.8 Innovative Products for Autonomous Driving (part 1)
- UB03 Session 3
  - IP1 Interactive Presentations
    - 4.1 Executive Session: Exact Synthesis and SAT
    - 4.2 Domain Specific Design Methodologies
    - 4.3 System Modelling for Simulation and Optimisation
    - 4.4 Overcoming the Limitations of Worst-Case IC Design
    - 4.5 Test: innovative infrastructures and ATPG techniques
    - 4.6 Special Session: Securing Power-constrained System-on-Chips: Challenges and Opportunities
    - 4.7 Adaptive Reliable Computing Using Memristive and Reconfigurable Hardware
    - 4.8 Components for Secure IoT Systems
- UB04 Session 4
  - Exhibition Reception
  - 5.1 Special Day Session on Future and Emerging Technologies: Challenges for the Design of Microfluidic Devices: EDA for your Lab-on-a-Chip
  - 5.2 Smart Energy and Automotive Systems
  - 5.3 Heterogeneous multi-level caching
  - 5.4 Special Session: Lightweight Security for Resources-Constrained Internet-of-Things Applications
  - 5.5 Emerging Technologies for Future Computing
  - 5.6 Reliability improvement and evaluation techniques
  - 5.7 Software-centric techniques for embedded systems
  - IP2 Interactive Presentations
- UB05 Session 5
  - 6.1 Special Day Session on Future and Emerging Technologies: Transistors for Digital NanoSystems: The Road Ahead
  - 6.2 Memory Security
  - 6.3 Advances in AMS/RF Design & Test Automation and Beyond
  - 6.4 Modeling, Control and Scheduling for Cyber-Physical Systems
  - 6.5 Special Session: Three Years of Low-Power Image Recognition Challenge
  - 6.6 Special Session: Three Years of Low-Power Image Recognition Challenge
  - 6.8 Innovative Products for Autonomous Driving (part 2)
- UB06 Session 6
  - 7.0 LUNCH TIME KEYNOTE SESSION: From Inverse Design to Implementation of Robust and Efficient Photonics for Computing
- UB07 Session 7
  - 7.1 Special Day Session on Future and Emerging Technologies: Theoretical and practical aspects of verification of quantum computers
  - 7.2 Run-time power estimation and optimization
  - 7.3 Advances in Logic Synthesis and Technology Mapping
  - 7.4 DRAM and NVMs
  - 7.5 Reliability Modeling and Mitigation
  - 7.6 Special Session: Next Generation Processors and Architectures for Deep Learning
  - 7.7 Rigorous design, analysis, and monitoring of dependable embedded systems
  - 7.8 22FDX - the superior technology for IoT, RF, Automotive and Mobility: Advanced Design Methodologies for Ultra-low Power Solutions
  - IP3 Interactive Presentations
- UB08 Session 8
  - 8.1 Special Day Session on Future and Emerging Technologies: NanoSystems: Connecting Devices, Architectures, and Applications
  - 8.2 EU Projects: Novel Technologies, Predictable Architectures and Worst-Case Execution Times
  - 8.3 Real-time intelligent methods for energy-efficient approaches in CNN and biomedical applications
  - 8.4 Efficient and reliable memory and computing architectures
1.1 Opening Session: Plenary, Awards Ceremony & Keynote Addresses

Date: Tuesday, March 20, 2018
Time: 08:30 - 10:30
Location / Room: Großer Saal

Chair: Jan Madsen, DATE 2018 General Chair, DTU, DK, Contact Jan Madsen
Co-Chair: Ayse Coskun, DATE 2018 Programme Chair, Boston University, US, Contact Ayse Coskun

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30</td>
<td>1.1.1</td>
<td>WELCOME ADDRESSES</td>
<td>Jan Madsen¹ and Ayse Coskun²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>¹DTU, DK; ²Boston University, US</td>
</tr>
<tr>
<td>08:45</td>
<td>1.1.2</td>
<td>PRESENTATION OF AWARDS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2018 EDAA Achievement Award</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EDAA Outstanding Dissertations Award 2017</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DATE Fellow Award</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IEEE Fellow Award</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IEEE CEDA and CS TTTC Outstanding Service Contribution Award 2017</td>
<td></td>
</tr>
<tr>
<td>09:15</td>
<td>1.1.3</td>
<td>KEYNOTE ADDRESS: THE RESPONSIBILITY SENSITIVE SAFETY (RSS) FORMAL MODEL TOWARD SAFETY GUARANTEES FOR AUTONOMOUS VEHICLES</td>
<td>Amnon Shashua, Intel Corporation, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Amnon Shashua, Intel Corporation, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: In recent years, car makers and tech companies are racing toward self-driving cars. A critical component in getting society acceptance to the technology is to find a way to guarantee safety. The prevailing common wisdom is a data-driven empirical approach for safety validation where the more mileage driven the better the maturity of the system must be. I will describe a model in which the sources of errors due to Planning (the actions and decisions for negotiating motion in traffic) can be fended out from the data driven approach through a formal model of the common sense behind human judgment of what it means to cause an accident and how to define actions that will guarantee that the AV will never cause an accident due to Planning. The model creates a clear distinction of what can be certified by regulators and what should be left to the judgment of AV manufacturers. The RSS model also puts in context the conversation of &quot;ethical dilemmas&quot; by providing a formal framework for the discussion.</td>
<td></td>
</tr>
</tbody>
</table>
1.1.4 KEYNOTE ADDRESS: PROGRAMMING LIVING CELLS: DESIGN AUTOMATION TO MAP CIRCUITS TO DNA

Speaker: Christopher Voigt, MIT, US

Abstract
Platforms are being established to facilitate large genetic engineering projects. A desired cellular function is divided into systems that can be developed independently and then combined. Genetic sensors allow cells to receive environmental and cell state information. Sensory information is integrated by genetic circuits, which control the conditions and timing of a response. The circuit outputs are connected to actuators that control what the cell is doing, from building molecules to moving and communicating. Design automation tools from the electronics industry are applied to map a circuit design to a DNA sequence. Collectively, this enables a wide range of applications, for example cells that communicate to build a material, navigate the human body to treat a disease, or protect plants by responding to the environment.

Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

UB01 Session 1

Date: Tuesday, March 20, 2018
Time: 10:30 - 12:30
Location / Room: Booth 1, Exhibition Area

<table>
<thead>
<tr>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>UB01.1</td>
<td>ARCHON: AN ARCHITECTURE-OPEN RESOURCE-DRIVEN CROSS-LAYER MODELLING FRAMEWORK</td>
<td>Fei Xia¹, Ashur Rafeev¹, Mohammed Al-Hayannis², Alexei Iliasov², Rishad Shafik¹, Alexander Romanovsky¹ and Alex Yakovlev²&lt;br&gt;¹Newcastle University, GB; ²Newcastle University, UK and University of Technology and HCED, IQ</td>
</tr>
<tr>
<td></td>
<td>Abstract</td>
<td>This demonstration showcases a modelling method for large complex computing systems focusing on many-core types and concentrating on the cross-layer aspects. The resource-driven models aim to help system designers reason about, analyse, and ultimately design such systems across all conventional computing and communication layers, from application, operating system, down to the finest hardware details. The framework and tool support the notion of selective abstraction and are suitable for studying such non-functional properties such as performance, reliability and energy consumption.</td>
</tr>
<tr>
<td></td>
<td>More information ...</td>
<td></td>
</tr>
</tbody>
</table>

UB01.2 TOPOLINANO & MAGCAD: A DESIGN AND SIMULATION FRAMEWORK FOR THE EXPLORATION OF EMERGING TECHNOLOGIES

Authors: Umberto Garlando and Fabrizio Riente, Politecnico di Torino, IT

Abstract
We developed a design framework that enables the exploration and analysis of emerging beyond-CMOS technologies. It is composed of two powerful tools: ToPoNaNano and MagCAD. Different technologies are supported, and new ones could be added thanks to their modular structure. ToPoNaNano starts from a VHDL description of a circuit and performs the place&route following the technological constraints. The resulting circuit can be simulated both at logical or physical level. MagCAD is a layout editor where the user can design custom circuits, by placing basic elements of the selected technology. The tool can extract a VHDL netlist based on compact models of placed elements derived from experiments or physical simulations. Circuits can be verified with standard VHDL simulators. The design workflow will be demonstrated at the U-booth to show how these tools could be a valuable help in the studying and development of emerging technologies and to obtain feedbacks from the scientific community.

More information ...
UB01.3 ADVANCED SIMULATION OF QUANTUM COMPUTATIONS

Authors:
Zulehner Alwin and Robert Wille, Johannes Kepler University Linz, AT

Abstract
Quantum computation is a promising emerging technology which allows for substantial speed-ups compared to classical computation. Since physical realizations of quantum computers are in their infancy, most research in this domain still relies on simulations on classical machines. This causes an exponential overhead which current simulators try to tackle with straightforward array-based representations and massive hardware power. There also exist solutions based on decision diagrams (graph-based approaches) that try to tackle the complexity by exploiting redundancies in quantum states and operations. However, they did not get established since they yield speedups only for certain benchmarks. Here, we demonstrate a new graph-based simulation approach which clearly outperforms state-of-the-art simulators. By this, users can efficiently execute quantum algorithms even if the respective quantum computers are not broadly available yet.

More information ...

UB01.4 OTPG: SPECIFICATION-BASED CONSTRUCTION OF ONLINE TPGS FOR MICROPROCESSORS

Authors:
Mikhail Chupikko, Alexander Kamkin and Andrei Tatamkov, ISP RAS, RU

Abstract
This work presents an approach to construction of online test program generators (TPGs). The approach is intended to use specifications of ISA presented in MML/mmSL specification languages. They are processed by a meta-generator to obtain their binary representations with meta information and a test generation core compatible with the target microprocessor. The test generation core is loaded as a binary image into the target microprocessor's memory (for experiments we're using QEMU for MIPS) and produces test cases to be processed (incl. results checking) by an executor. It should be noticed that the meta-generator and the executor are not obligatory run at the same microprocessor (especially, if it is highly incomplete). The final goal of the project is to propose a method of obtaining online TPGs for a wide range of ISAs, and to develop a mature tool implementing this method.

More information ...

UB01.5 ABSYNTH: A COMPREHENSIVE APPROACH TO FRONT TO BACK ANALOG BLOCK DESIGN AUTOMATION

Authors:
Abhaya Chandra Kammara S.,1 Sidney Pontes-Filho2 and Andreas König2
1TU Kaiserslautern, DE; 2 University of Kaiserslautern, DE

Abstract
ABSYNTH was first presented in CEBIT 2014 where complete, practical circuit sizing approaches have been shown using meta-heuristics on trusted simulators. This tool was then proven by its use in design of several cells in a research project. Here, we present the extension to our nested optimization approach that creates a symmetric and well matched layout in every step for every instance in the population of the swarm, that is extracted in our flow to provide feedback to the cost function impacting on the population update for more viable and robust circuits. The layout optimization presented in this DEMO works with Cadence Layout design tools. Our initial focus is motivated by Industry 4.0, IoT, on cells for signal conditioning electronics with reconfigurability and Self-X features.[1] Abhaya C. Kammara, L.Palanichamy, and A. König, "Multi-Objective optimization and visualization for analog automation", Complex. Intell. Syst, Springer, DOI 10.1007/s40747-016-0027-3, 2016

More information ...

UB01.6 WARE: WEARABLE ELECTRONICS DIRECTIONAL AUGMENTED REALITY

Authors:
Gabriele Miorandi, Walter Vendraminetto, Federico Fraccaro, Davide Quaglia and Gianluca Benedetti
University of Verona, IT; REDLab Srl, IT; Wagoo LLC, IT; Wagoo Italia srls, IT

Abstract
Augmented Reality (AR) currently require large form factors, weight, cost and frequent recharging cycles that reduce usability. Connectivity, image processing, localization, and direction evaluation lead to high power usage and cost. A multi-antenna system, patented by the industrial partner, enables a new generation of smart eye-wear that elegantly requires less hardware, connectivity, and power to provide AR functionalities. They will allow users to directionally locate nearby radio emitting sources that highlight objects of interest (e.g., people or retail items) by using existing standards like Bluetooth Low Energy, Apple's iBeacon and Google’s Eddystone. This booth will report the current level of research addressed by the Computer Science Department of University of Verona, Wagoo LLC, and Wagoo Italia srls. In the presented demo, different objects emit an "I am here" signal and a prototype of the smart glasses shows the information related to the observed object.

More information ...

UB01.7 IDEEA: DESIGN SPACE EXPLORATION FOR FUNCTIONAL-LEVEL APPROXIMATION

Authors:
Marcello Traiola1, Mario BarbareschP, Marcello Traiola2 and Alberto Bossi3
1LIRMM, FR; 2DIEETI - University of Naples Federico II, IT; 3LIRMM - University of Montpellier / CNRS, FR

Abstract
Approximate Computing (AxC) aims at enabling the production of computing systems which can support the rising performance demands and can improve the energy efficiency. AxC exploits the gap between the level of accuracy required by the users and the precision provided by the computing system, for achieving diverse optimizations. Various AxC techniques have been proposed so far for several applications and, unfortunately, existing approaches are application specific and a general and systematic methodology to automatically define approximate algorithms is still an open challenge. In this work we introduce a methodology which makes use of mutation techniques to obtain approximate versions of a given application described as a C/C++ code. We designed and implemented IDEEA, an automatic tool exploiting (i) a source-to-source manipulation technique and (ii) an Evolutionary search engine, in order to search for the best functional approximation version of the given C/C++ code.

More information ...

UB01.8 IIP GENERATORS TO EASE ANALOG IC DESIGN

Authors:
Benjamin Prautsch, Uwe Eichler and Torsten Reich, Fraunhofer Institute for Integrated Circuits IIS/EAS, DE

Abstract
Semiconductor technology has shown significant progress over the last decades. Digital EDA (electronic design automation) allowed that this progress could be converted to high-performance digital ICs. Analog components are part of Systems-on-Chip (SoC) too, but analog EDA lags far behind. Therefore, a lot of effort was spent to automate analog IC design. Major reasons are constraint-based layout-aware optimization tools using predefined layout templates or pure automation as well as analog generators containing expert knowledge. While optimization is a holistic top-down approach, generators allow parameterized and fast bottom-up generation of critical schematic and layout parts, pre-planned by experienced designers. With IIP Generators, we follow three use cases to ease analog design: 1) design on higher hierarchy levels, 2) development of hierarchical high-level IIPs, and 3) automated design porting due to highly technology-independent blocks down to 22nm.

More information ...

UB01.9 CIJTAG: CONCURRENT IJTAG DEMONSTRATOR

Author:
Kienz-Baath René, Hämmerläppdtad University of Applied Sciences, DE

Abstract
The flexibility of on-chip instrument access enabled by IEEE 1687 (IJTAG) has shown tremendous improvements in modern industrial designs. Due to a constantly increasing spectrum of tasks performed through 1687 networks such as performing test operations during production test, on-line test operations as well as operating health monitors the test requirements in modern designs increase dramatically with respect to test performance, responsiveness and low power. These requirements have a major impact on the design of such test infrastructures. In complex designs with large test infrastructures it might be challenging to comply with the large spectrum of requirements. Concurrent IJTAG is novel partitioning concept to a reconfigurable test infrastructure in order to enable an independent operation of different sections of the test infrastructure. The proposed demonstrator shows the first FPGA-based implementation of concurrent IJTAG test infrastructures.

More information ...
Panelists:
- Challenges to be overcome

More information ...

2.1 Executive Panel: How Electronics May Change Our Lives, and the World

Date: Tuesday, March 20, 2018
Time: 11:30 - 13:00
Location / Room: Saal 2

Chair:
Antun Domic, Synopsys, US, Contact Antun Domic

Innovation runs strong in our industry. A 17 qubits Quantum Computer (QC) has been shipped to Deft University researchers; an initiative to make cloud QC commercially available for businesses and research has been announced; if, and once available QC may change the landscape of finance, imaging diagnostics, pharmacology, meteorology and, of course, security all the way. After decades of domination by general purpose CPU and GPU, innovation is disrupting computing architectures: Massively parallel Tensor Processing Units (TPU) have demonstrated that a computer can learn from past experience, and then beat a 9 dan human professional Go player, or classify zillions of images with unprecedented accuracy and speed. Autonomous “Things” in which a wide breadth of sensors feed a processor with huge amounts of data, that are analyzed in order to make decisions that are then sent to actuators with minimal if any human supervision are emerging; Advanced Driver-Assistance Systems (ADAS) are the top of the iceberg, exceeding SAE level 3 requirements, ADAS have been adopted by scores of automakers, and all the top 10 automakers have already announced plans toward SAE level 4, and 5 availability before the end of this decade. Finally, computers are digital, but the world is analog; scores of sensors and actuators are the eyes, the ears, the nose, and the arms of the most advanced processors, and the most advanced applications could not exist without them. Today, sensors are designed at the established technology nodes, and often manufactured using electro-mechanical processes which make their integration with their host processors challenging. Is this going to continue, or will they be submerged by digital, and eventually be designed, integrated, and manufactured using the very same emerging technology nodes? Electronics may truly change our lives, and the world, if we will be able to expand the scope of QC beyond cryptography, and the scope of AI beyond image recognition, if ADAS technology will not be overwhelmed by legal. IF “More than Moore” will join forces with “More of Moore”, IF... This Executive Panel, moderated by Dr. Antun Domic, Synopsys CTO, gathers world experts to discuss the many opportunities that lie ahead, and the challenges to be overcome

Panelists:
- Loic Lietar, Greenwaves Technologies, FR
- Martin Roesler, Microsoft, US
- Horst Symanzik, Bosch Sensortec, DE
- Olivier Temam, Google, US
- Martin Duncan, STMicroelectronics, IT
Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

2.2 Energy Efficient Neural Networks

Date: Tuesday, March 20, 2018
Time: 11:30 - 13:00
Location / Room: Kont. 6

Chair: Hai (Helen) Li, Duke University, US, Contact Hai (Helen) Li

Co-Chair: Muhammad Shafique, Vienna University of Technology (TU Wien), AT, Contact Muhammad Shafique

This session focuses on energy efficient neural network architectures. The first paper proposes a methodology that enables aggressive voltage scaling of accelerator weight memories to improve the energy-efficiency of DNN accelerators. The second paper introduces methods to optimize the memory usage in DNN training. The third paper presents HyperPower, that enables efficient Bayesian optimization and random search in the context of power- and memory-constrained hyper-parameter optimization for NNs running on a given hardware platform. Finally, the last paper presents a new sparse matrix format to maximize the inference speed of the LSTM accelerator. The session also includes 2 IP papers ReCom and SparseNN, which both focus on energy efficiency of neural networks.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:30</td>
<td>2.2.1</td>
<td>MATIC: LEARNING AROUND ERRORS FOR EFFICIENT LOW-VOLTAGE NEURAL NETWORK ACCELERATORS</td>
<td>Sung Kim, University of Washington, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>As a result of the increasing demand for deep neural network (DNN)-based services, efforts to develop dedicated hardware accelerators for DNNs are growing rapidly. However, while accelerators with high performance and efficiency on convolutional deep neural networks (Conv-DNNs) have been developed, less progress has been made with regards to fully-connected DNNs (FC-DNNs). In this paper, we propose MATIC (Memory Adaptive Training with In-situ Canaries), a methodology that enables aggressive voltage scaling of accelerator weight memories to improve the energy-efficiency of DNN accelerators. To enable accurate operation with voltage overscaling, MATIC combines the characteristics of destructive SRAM reads with the error resilience of neural networks in a memory-adaptive training process. Furthermore, PVT-related voltage margins are eliminated using bit-cells from synaptic weights as in-situ canaries to track runtime environmental variation. Demonstrated on a low-power DNN accelerator that we fabricate in 65 nm CMOS, MATIC enables up to 60-80 mV of voltage overscaling (3.3x total energy reduction versus the nominal voltage), or 18.6x application error reduction.</td>
</tr>
<tr>
<td>12:00</td>
<td>2.2.2</td>
<td>MAXIMIZING SYSTEM PERFORMANCE BY BALANCING COMPUTATION LOADS IN LSTM ACCELERATORS</td>
<td>Junki Park, POSTECH, KR</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>The LSTM is a popular neural network model for modeling or analyzing the time-varying data. The main operation of LSTM is a matrix-vector multiplication and it becomes sparse (spMV) due to the widely-accepted weight pruning in deep learning. This paper presents a new sparse matrix format, named CBSR, to maximize the inference speed of the LSTM accelerator. In the CBSR format, speed-up is achieved by balancing out the computation loads over PEs. Along with the new format, we present a simple network transformation to completely remove the hardware overhead incurred when using the CBSR format. Also, the detailed analysis on the impact of network size or the number of PEs is performed, which lacks in the prior work. The simulation results show 16<del>38% improvement in the system performance compared to the well-known CSC/CSR format. The power analysis is also performed in 65nm CMOS technology to show 9</del>22% energy savings.</td>
</tr>
</tbody>
</table>
MODNN: MEMORY OPTIMAL DNN TRAINING ON GPUs

**Authors:**
Xiaoming Chen, Institute of Computing Technology, Chinese Academy of Sciences, CN
Xiaobo Shoun Hu, University of Notre Dame, US

**Abstract**
Deep Neural Networks (DNNs) play a key role in prevailing machine learning applications. Resistive random-access memory (ReRAM) is capable of both computation and storage, contributing to the acceleration on DNNs process in memory. Besides, DNNs have a significant amount of zero weights, which provides a possibility to reduce computation cost by skipping infeasible calculations on zero weights. However, the irregular distribution of zero weights in DNNs makes it difficult for resistive accelerators to take advantage of the sparsity, because resistive accelerators have a high reliance on regular matrix-vector multiplication in ReRAM. In this work, we propose ReCom, the first resistive accelerator to support sparse DNN processing. ReCom is an efficient resistive accelerator for compressed deep neural networks, where DNN weights are structurally compressed to eliminate zero parameters and become more friendly to computation in ReRAM, and zero DNN activations are also considered at the same time. Two technologies, Structurally-compressed Weight Oriented Fetching (SWOF) and In-layer Pipeline for Memory and Computation (IPMC), are particularly proposed to efficiently process the compressed DNNs in ReRAM. In our evaluation, ReCom can achieve 3.37x speedup and 2.41x energy efficiency compared to a state-of-the-art resistive accelerator.

Download Paper (PDF; Only available from the DATE venue WiFi)

MODNN: MEMORY OPTIMAL DNN TRAINING ON GPUs

**Authors:**
Xiaoming Chen, Institute of Computing Technology, Chinese Academy of Sciences, CN
Xiaobo Shoun Hu, University of Notre Dame, US

**Abstract**
Deep Neural Networks (DNNs) play a key role in prevailing machine learning applications. Resistive random-access memory (ReRAM) is capable of both computation and storage, contributing to the acceleration on DNNs process in memory. Besides, DNNs have a significant amount of zero weights, which provides a possibility to reduce computation cost by skipping infeasible calculations on zero weights. However, the irregular distribution of zero weights in DNNs makes it difficult for resistive accelerators to take advantage of the sparsity, because resistive accelerators have a high reliance on regular matrix-vector multiplication in ReRAM. In this work, we propose ReCom, the first resistive accelerator to support sparse DNN processing. ReCom is an efficient resistive accelerator for compressed deep neural networks, where DNN weights are structurally compressed to eliminate zero parameters and become more friendly to computation in ReRAM, and zero DNN activations are also considered at the same time. Two technologies, Structurally-compressed Weight Oriented Fetching (SWOF) and In-layer Pipeline for Memory and Computation (IPMC), are particularly proposed to efficiently process the compressed DNNs in ReRAM. In our evaluation, ReCom can achieve 3.37x speedup and 2.41x energy efficiency compared to a state-of-the-art resistive accelerator.

Download Paper (PDF; Only available from the DATE venue WiFi)

MODNN: MEMORY OPTIMAL DNN TRAINING ON GPUs

**Authors:**
Xiaoming Chen, Institute of Computing Technology, Chinese Academy of Sciences, CN
Xiaobo Shoun Hu, University of Notre Dame, US

**Abstract**
Deep Neural Networks (DNNs) play a key role in prevailing machine learning applications. Resistive random-access memory (ReRAM) is capable of both computation and storage, contributing to the acceleration on DNNs process in memory. Besides, DNNs have a significant amount of zero weights, which provides a possibility to reduce computation cost by skipping infeasible calculations on zero weights. However, the irregular distribution of zero weights in DNNs makes it difficult for resistive accelerators to take advantage of the sparsity, because resistive accelerators have a high reliance on regular matrix-vector multiplication in ReRAM. In this work, we propose ReCom, the first resistive accelerator to support sparse DNN processing. ReCom is an efficient resistive accelerator for compressed deep neural networks, where DNN weights are structurally compressed to eliminate zero parameters and become more friendly to computation in ReRAM, and zero DNN activations are also considered at the same time. Two technologies, Structurally-compressed Weight Oriented Fetching (SWOF) and In-layer Pipeline for Memory and Computation (IPMC), are particularly proposed to efficiently process the compressed DNNs in ReRAM. In our evaluation, ReCom can achieve 3.37x speedup and 2.41x energy efficiency compared to a state-of-the-art resistive accelerator.

Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

2.3 High-Level Synthesis

Date: Tuesday, March 20, 2018
Time: 11:30 - 13:00
Location / Room: Kont. 1

Chair:
Selma Saidi, Hamburg University of Technology, DE, Contact Selma Saidi

Co-Chair:
Daniel Ziener, University of Twente, NL, Contact Daniel Ziener

This session addresses high-level synthesis for easing the development of application-specific designs. First, user-guided optimizations for high-level synthesis based on innovative resource prediction using CNNs will be discussed for an area-reduction advisor. The second paper proposes a look-ahead scheduling scheme to minimize the area for functional units. The final talk presents a direct HLS synthesis path of software-customizable floating-point cores that does not rely on external libraries or floating-point code generators.

11:30 2.3.1 SENSEI: AN AREA-REDUCTION ADVISOR FOR FPGA HIGH-LEVEL SYNTHESIS

Speaker:
Hsuan Hsiao, University of Toronto, CA

Authors:
Hsuan Hsiao and Jason H. Anderson, University of Toronto, CA

Abstract
High-level synthesis (HLS) provides an easy-to-use abstraction for designing hardware circuits. However, standard datatypes in high-level languages are overprovisioned for typical applications, incurring extra area since the underlying FPGA hardware can support arbitrary bitwidths. This area inefficiency can be overcome by enabling the use of arbitrary-width datatypes at the source code level. However, this requires that HLS users spend time and effort on examining all program variables and quantifying their area impact, which can be intractable especially with large, complex programs and time-consuming synthesis. We propose Sensei, an advisor that predicts the post-synthesis area savings brought about by reducing bitwidth and presents users with a ranking of program variables and their area impact. Equipped with a convolutional neural network (CNN)-based predictor, Sensei achieves high area prediction accuracy and enables rapid exploration of area-saving opportunities.

Download Paper (PDF; Only available from the DATE venue WiFi)

12:00 2.3.2 A FAST AND EFFECTIVE LOOKAHEAD AND FRACTIONAL SEARCH BASED SCHEDULING ALGORITHM FOR HIGH-LEVEL SYNTHESIS

Speaker:
Shantanu Dutt, University of Illinois at Chicago, US

Authors:
Shantanu Dutt and Ouwen Shi, University of Illinois at Chicago, US

Abstract
We present a latency-constrained iterative list scheduling type algorithm, FALLS, to minimize the total number of functional units (FUs) allocated, and thus the total area, in high-level synthesis designs. The algorithm incorporates a novel lookahead technique to selectively schedule available operations by allocating the needed FUs earlier or reserving available FUs for scheduling more timing-urgent operations later, such that no additional FU is needed and higher FU utilization is obtained. Further, a fractional search framework is developed to iteratively estimate the number of FUs of each function type required in the final design based on the current scheduling solution and FU utilization, and reiterate the lookahead-based list scheduling with the new FU allocation estimate to further increase FU utilization. Extensive experiments conducted over several DFGs and a wide range of latency constraints demonstrate that FALLS is much more effective than other approximate state-of-the-art algorithms in both number of FUs and total FU area, and has a much smaller runtime. Results also show that FALLS has only an average 5.5% optimality gap compared to an optimal integer linear programming (ILP) formulation, but is 278k times faster. FALLS also performs much better in architectural (FU + mux/demux + register) area, interconnect congestion and number of interconnects than approximate algorithms, and is at most 4.0% worse in them than the ILP method.

Download Paper (PDF; Only available from the DATE venue WiFi)
2.4 Model Checking

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:00</td>
<td>IP1-3</td>
<td>ACCLIB: ACCELERATORS AS LIBRARIES</td>
<td>Jacob Stevens, Purdue University, US</td>
</tr>
</tbody>
</table>

Abstract

In this work, we propose ACCLIB, a design framework that allows software developers to utilize existing libraries of pre-designed hardware accelerators with specific capabilities for accelerating software functions. The key contributions of ACCLIB are:

1. Parameterization: ACCLIB allows the specification of the functional behavior of the hardware accelerator, including operations, data types, and interface definitions.
2. Formal Verification: ACCLIB uses formal verification techniques to ensure the correctness of the accelerator's implementation.
3. Hardware Abstraction: ACCLIB abstracts away the low-level details of the hardware accelerator, allowing software developers to focus on higher-level abstractions.
4. Offloading: ACCLIB provides a mechanism for offloading computation to the hardware accelerator, improving performance and efficiency.

ACCLIB is implemented using a formal verification tool and integrates with existing design flows, enabling software developers to specify and verify hardware accelerators within their existing development processes.

Download Paper (PDF; Only available from the DATE venue WiFi)

Coffee Break 13:00 - 14:30

End of session

Lunch in Großer Saal and Saal 1
<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:30</td>
<td>2.4.1</td>
<td>EFFICIENT VERIFICATION OF MULTI-PROPERTY DESIGNS (THE BENEFIT OF WRONG ASSUMPTIONS)</td>
<td>Eugene Goldberg, Diffblue, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Authors:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Eugene Goldberg(^1), Matthias Gudemann(^2), Daniel Kroening(^2) and Rajdeep Mukherjee(^4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1)Diffblue, US; (^2)Diffblue, DE; (^3)Diffblue, GB; (^4)Oxford University, GB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>We consider the problem of efficiently checking a set of safety properties P(_1),...,P(_k) of one design. We introduce a new approach called Ja-verification, where Ja stands for “Just-Assume” (as opposed to “assume-guarantee”). In this approach, when proving a property P(_j) one assumes that every property P(_i) for j (\neq i) holds. The process of proving properties either results in showing that P(_j),...,P(_k) hold without any assumptions or finding a “debugging set” of properties. The latter identifies a subset of failed properties that are the first to break. The design behaviors that cause the properties in the debugging set to fail must be fixed first. Importantly, in our approach, there is no need to prove the assumptions used. We describe the theory behind our approach and report experimental results that demonstrate substantial gains in performance, especially in the cases where a small debugging set exists.</td>
</tr>
<tr>
<td>12:00</td>
<td>2.4.2</td>
<td>COMBINING PDR AND REVERSE PDR FOR HARDWARE MODEL CHECKING</td>
<td>Tobias Seufert, University of Freiburg, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Authors:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tobias Seufert and Christoph Scholl, University Freiburg, DE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>In the last few years IC3 resp. PDR attracted a lot of attention as a SAT-based hardware verification approach without needing to unroll the transition relation as in Bounded Model Checking (BMC). Motivated by different strengths of forward and backward traversal already observed in BDD-based model checking and by an exponential complexity gap between original PDR and its reverted counterpart “Reverse PDR” (which starts its analysis with the initial states instead of the unsafe states as in the original PDR), we take a closer look at Reverse PDR and we present a combined forward/backward version of PDR that inherits the advantages of both original and Reverse PDR. Our experimental results on benchmarks from the Hardware Model Checking Competition demonstrate clear benefits of the combined approach.</td>
</tr>
<tr>
<td>12:30</td>
<td>2.4.3</td>
<td>SYMBOLIC QUICK ERROR DETECTION USING SYMBOLIC INITIAL STATE FOR PRE-SILICON VERIFICATION</td>
<td>Mohammad Rahmani Fadíathe, University of Kaiserslautern, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Authors:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mohammad Rahmani Fadíathe(^1), Joakim Urdahl(^1), Snivivasà Shashanik Nuthakkî(^2), Subhasish Mitra(^2), Dominik Stoffel(^1) and Wolfgang Kunz(^1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1)University of Kaiserslautern, DE; (^2)Stanford University, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>Driven by the demand for highly customizable processor cores for IoT and related applications, there is a renewed interest in effective but low-cost techniques for verifying systems-on-chip (SoCs). This paper revisits the problem of processor verification and presents a radically different approach when compared to the state of the art. The proposed approach is highly automated and leverages recent progress in the field of post-silicon validation by the method of Quick Error Detection (QED) and Symbolic Quick Error Detection (SQED). In this paper, we modify SQED by incorporating a symbolic initial state in its BMC-based analysis and generalize the approach into the SQ2QED method. As a first advantage, SQ2QED can separate logic bugs from electrical bugs in QED-based post-silicon validation. Secondly, it also makes a strong contribution to pre-silicon verification by proving that the execution of each instruction is independent of its context in the program. The manual efforts for the proposed approach are orders of magnitude smaller than for conventional property checking. Our experimental results demonstrate the potential of SQ2QED using the Aquarius open-source processor example.</td>
</tr>
<tr>
<td>12:45</td>
<td>2.4.4</td>
<td>VERIFICATION OF TREE-BASED HIERARCHICAL READ-COPY UPDATE IN THE LINUX KERNEL</td>
<td>Paul McKenney, IBM Linux Technology Center, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Authors:</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lihao Liang(^1), Paul McKenney(^2), Daniel Kroening(^3) and Tom Melham(^3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(^1)University of Oxford, GB; (^2)IBM Linux Technology Center, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>Read-Copy Update (RCU) is a scalable, high-performance Linux-kernel synchronization mechanism that runs low-overhead readers concurrently with updaters. Production-quality RCU implementations are decidedly non-trivial and their stringent validation is mandatory. This suggests use of formal verification. Previous formal verification efforts for RCU either focus on simple implementations or use modeling languages. In this paper, we construct a model directly from the source code of Tree RCU in the Linux kernel, and use the CBMC program analyzer to verify its safety and liveness properties. To the best of our knowledge, this is the first verification of a significant part of RCU’s source code—an important step towards integration of formal verification into the Linux kernel’s regression test suite.</td>
</tr>
</tbody>
</table>
Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

2.5 GPU and GPU-based heterogeneous system management

Date: Tuesday, March 20, 2018
Time: 11:30 - 13:00
Location / Room: Konf. 3

Chair:
Andrea Morongiu, Università di Bologna, IT, Contact Andrea Morongiu

Co-Chair:
Carles Hernandez, BSC, ES, Contact Carles Hernández

GPUs are at the heart of several modern heterogeneous systems, where the common communication paradigm between the CPU and the GPU is shared memory. The papers in this session propose novel techniques to deal with i) efficient shared memory management and ii) GPU multiprocessor scheduling in presence of process variation. The first paper focuses on GPU-based heterogeneous systems with a shared last-level cache (SLLC) and proposes a novel metric that combines CPU/GPU miss count and "hit utility" to devise an effective cache-way partitioning. The second paper proposes a technique to mitigate the negative effects of cache and memory controller sharing in GPUs running multiple workloads. The third paper discusses a HW technique to mitigate the effects of hardware variability (e.g., process variations (PVs) and negative bias temperature instability (NBTI)) in GPU Streaming Processors (SPs).

2.5.1 HVSM: HARDWARE-VARIABILITY AWARE STREAMING PROCESSORS' MANAGEMENT POLICY IN GPUS

Speaker:
Jingweijia Tan, Jilin University, CN
Authors:
Jingweijia Tan1 and Kaige Yan2
1Jilin University, CN; 2College of Communication Engineering, Jilin University, CN

Abstract
GPUs are widely used in general-purpose high performance computing field due to their highly parallel architecture. In recent years, a new era with nanometer scale integrated circuit manufacture process has come, as a consequence, GPU's computation capability gets even stronger. However, as process technology scales down, hardware variability, e.g., process variations (PVs) and negative bias temperature instability (NBTI), has a higher impact on the chip quality. The parallelism of GPU desires high consistency of hardware units on chip, otherwise, the worst unit will inevitably become the bottleneck. So the hardware variability becomes a pressing concern to further improve GPUs' performance and lifetime, not only in integrated circuit fabrication, but more in GPU architecture design. Streaming Processors (SPs) are the key units in GPUs, which perform most of parallel computing operations. Therefore, in this work, we focus on mitigating the impact of hardware variability in GPU SPs. We first model and analyze SPs performance variations under hardware variability. Then, we observe that both PV and NBTI have large impact on SPs performance. We further observe unbalanced SP utilization, e.g., some SPs are idle while others are active, during program execution. Leveraging both observations, we propose a Hardware Variability-aware SPs Management policy (HVSM), which dynamically prioritizes the fast SPs, regroups SPs in a two-level granularity and dispatches computation in appropriate SPs. Our experimental results show HVSM effectively reduces the impact of hardware variability, which can translate to 28% performance improvement or 14.4% lifetime extension for a GPU chip.

Download Paper (PDF; Only available from the DATE venue WiFi)
**2.5.2 THROUGHPUT OPTIMIZATION AND RESOURCE ALLOCATION ON GPUs UNDER MULTI-APPLICATION EXECUTION**

**Speaker:**
Irakis Anagnostopoulos, Southern Illinois University Carbondale, US

**Authors:**
Srinivas Reddy Punyala, Theodoros Marinakis, Arash Komae and Irakis Anagnostopoulos, Southern Illinois University Carbondale, US

**Abstract**
Platform heterogeneity prevails as a solution to the throughput and computational challenges imposed by parallel applications and technology scaling. Specifically, Graphics Processing Units (GPUs) are based on the Single Instruction Multiple Thread (SIMT) paradigm and they can offer tremendous speed-up for parallel applications. However, GPUs were designed to execute a single application at a time. In case of simultaneous multi-application execution, due to the GPUs' massive multi-threading paradigm, applications compete against each other using destructively the shared resources (caches and memory controllers) resulting in significant throughput degradation. In this paper, a methodology for minimizing interference in shared resources and providing efficient concurrent execution of multiple applications on GPUs is presented. Particularly, the proposed methodology (i) performs application classification; (ii) analyzes the per-class interference; (iii) finds the best matching between classes; and (iv) employs an efficient resource allocation. Experimental results showed that the proposed approach increases the throughput of the system for two concurrent applications by an average of 36% compared to the default execution and 10% compared to an exhaustive profile-based optimization technique.

Download Paper (PDF; Only available from the DATE venue WiFi)

**2.5.3 SET VARIATION-AWARE SHARED LAST-LEVEL CACHE MANAGEMENT FOR CPU-GPU HETEROGENEOUS ARCHITECTURE**

**Speaker:**
Xin Li, Shandong University, CN

**Authors:**
Zhaoqiang Li, Lei Ju, Hongjun Dai, Xin Li, Mengying Zhao and Zhiping Jia, Shandong University, CN

**Abstract**
Heterogeneous CPU-GPU multicore processors on chip (HMPSoC) becomes a popular architecture choice for high performance embedded systems, where shared last-level cache (LLC) management becomes a critical design consideration. We observe that within a sampling period, CPU and GPU may have distinct access behaviors over various LLC sets. In this work, we propose a light-weighted and fine-grained cache management policy to cope with the GPU-CPU access behavior variation among cache sets. In particular, CPU and GPU requests are prioritized disparately in each LLC set during cache block insertion and promotion, based on the per-core utility behaviors and a per-set CPU-GPU miss counter. Experimental results show that our LLC management scheme outperforms the two state-of-the-art schemes TAP-RRIP and LSIP by 12.6% and 10.01%, respectively.

Download Paper (PDF; Only available from the DATE venue WiFi)

**13:00 HPXA: A HIGHLY PARALLEL XML PARSER**

**Speaker:**
Smruti Sarangi, IIT Delhi, IN

**Authors:**
Isaar Ahmad, Sanjog Patil and Smruti R. Sarangi, IIT Delhi, IN

**Abstract**
State of the art XML parsing approaches read an XML file byte by byte, and use complex finite state machines to process each byte. In this paper, we propose a new parser, HPXA, which reads and processes 16 bytes at a time. We designed most of the components ab initio, to ensure that they can process multiple XML tokens in parallel. We propose two basic elements - a sparse 1D array compactor, and a hardware unit called LTMAdder that takes its decisions based on adding the rows of a lower triangular matrix. We demonstrate that we are able to process 16 bytes in parallel with very few pipeline stalls for a suite of widely used XML benchmarks. Moreover, for a 28nm technology node, we can process XML data at 106 Gbps, which is roughly 6.5X faster than competing prior work.

Download Paper (PDF; Only available from the DATE venue WiFi)

**2.6 Circuit Locking and Camouflaging**

**Date:**
Tuesday, March 20, 2018

**Time:**
11:30 - 13:00

**Location / Room:** Konf. 4

**Chair:**
Debdeep Mukhopadhyay, IIT Kharagpur, IN, Contact Debdeep Debdeep Mukhopadhyay
Intellectual property piracy, counterfeiting and reverse-engineering are serious threats for the supply chain in advanced microelectronics. This session presents novel approaches to protect circuits against these threats. The techniques deploy nanotechnology and novel timing scheme to obtain efficient protections.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:30</td>
<td>2.6.1</td>
<td>CYCLIC LOCKING AND MEMRISTOR-BASED OBfuscATION AGAINst CYCSAT AND INSIDE FOUNDRY ATTACKS</td>
<td>Hai Zhou, Northwestern University, US; Amin Rezaei, Yuanqi Shen, Shuyu Kong, Jie Gu and Hai Zhou, Northwestern University, US</td>
</tr>
<tr>
<td>12:00</td>
<td>2.6.2</td>
<td>TIMINGCAMOUFLAGE: IMPROVING CIRCUIT SECURITY AGAINst COUNTERFEITING BY UNCONVENTIONAL TIMING</td>
<td>Li Zhang, Technical University of Munich, DE; Grace Li Zhang¹, Bing Li², Bei Yu³, David Z. Pan⁴ and Ulf Schlichtmann⁵</td>
</tr>
<tr>
<td>12:30</td>
<td>2.6.3</td>
<td>ADVANCING HARDWARE SECURITY USING POLYMORPHIC AND STOCHASTIC SPIN-HALL EFFECT DEVICES</td>
<td>Satwik Patnaik, New York University, AE; Satwik Patnaik¹, Nikhil Rangarajan², Johann Knechtel³, Ozgur Sinanoglu⁴ and Shaloo Rakheja²</td>
</tr>
</tbody>
</table>

**Abstracts**

1. **CYCLIC LOCKING AND MEMRISTOR-BASED OBfuscATION AGAINst CYCSAT AND INSIDE FOUNDRY ATTACKS**
   - Speaker: Hai Zhou, Northwestern University, US
   - Abstract: The high cost of IC design has made chip protection one of the first priorities of the semiconductor industry. Although there is a common impression that combinational circuits must be designed without any cycles, circuits with cycles can be combinational as well. Such cyclic circuits can be used to reliably lock ICs. Moreover, since memristor is compatible with CMOS structure, it is possible to efficiently obfuscate cyclic circuits using polymorphic memristor-CMOS gates. In this case, the layouts of the circuits with different functionalities look exactly identical, making it impossible even for an inside foundry attacker to distinguish the defined functionality of an IC by looking at its layout. In this paper, we propose a comprehensive chip protection method based on cyclic locking and polymorphic memristor-CMOS obfuscation. The robustness against state-of-the-art key-pruning attacks is demonstrated and the overhead of the polymorphic gates is investigated.

2. **TIMINGCAMOUFLAGE: IMPROVING CIRCUIT SECURITY AGAINst COUNTERFEITING BY UNCONVENTIONAL TIMING**
   - Speaker: Li Zhang, Technical University of Munich, DE
   - Abstract: With recent advances in reverse engineering, attackers can reconstruct a netlist to counterfeit chips by opening the die and scanning all layers of original chips. This relatively easy counterfeiting is made possible by the use of the standard simple clocking scheme where all combinational blocks function within one clock period. In this paper, we propose a method to invalidate the assumption that a netlist completely represents the function of a circuit. With the help of wave-pipelining paths, this method forces attackers to capture delay information from manufactured chips, which is a very challenging task because we also introduce false paths. Experimental results confirm that wave-pipelining paths and false paths can be constructed in benchmark circuits successfully with only a negligible cost, while the potential attack techniques can be thwarted.

3. **ADVANCING HARDWARE SECURITY USING POLYMORPHIC AND STOCHASTIC SPIN-HALL EFFECT DEVICES**
   - Speaker: Satwik Patnaik, New York University, AE
   - Abstract: Protecting intellectual property (IP) in electronic circuits has become a serious challenge in recent years. Logic locking/encryption and layout camouflaging are two prominent techniques for IP protection. Most existing approaches, however, particularly those focused on CMOS integration, incur excessive design overheads resulting from their need for additional circuit structures or device-level modifications. This work leverages the innate polymorphism of an emerging spin-based device, called the giant spin-Hall effect (GSHE) switch, to simultaneously enable locking and camouflaging within a single instance. Using the GSHE switch, we propose a powerful primitive that enables cloaking all the 16 Boolean functions possible for two inputs. We conduct a comprehensive study using state-of-the-art Boolean satisfiability (SAT) attacks to demonstrate the superior resilience of the proposed primitive in comparison to several others in the literature. While we tailor the primitive for deterministic computation, it can readily support stochastic computation; we argue that stochastic behavior can break most, if not all, existing SAT attacks. Finally, we discuss the resilience of the primitive against various side-channel attacks as well as invasive monitoring at runtime, which are arguably even more concerning threats than SAT attacks.
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018

- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

2.7 Special Session: Spintronics based New Computing Paradigms and Applications

Date: Tuesday, March 20, 2018
Time: 11:30 - 13:00
Location / Room: Kont. 5

Chair:
Zhao Weisheng, Beihang University, CN, Contact Weisheng Zhao

Co-Chair:
Tahoori Mehdi, Karlsruhe Institute of Technology, DE, Contact Mehdi Tahoori

Recent technology nodes, the well-known “moore’s law” tends to slow down. Indeed, the continuously decreasing size of the CMOS transistors and operating frequencies result in serious power consumption, heat dissipation and reliability issues. Among the solutions investigated to overcome these limitations, the use of emerging nano-devices mixed (or not) with CMOS circuits is often referred as the « More than Moore » concept. In particular, logic circuits based on non-volatile memories can be an efficient solution to reduce the power, to improve the reliability and can offer new paradigms for computing. We are convinced that this research field has become a hot topic for the DATE community. The aim of this session is to bring together the worldwide leading experts (from respectively USA, Belgium, China and Germany) related to this hot topic to share the most recent results and discuss the future challenges. Different computing paradigms will be involved in this special session benefiting from interesting nature of spintronics devices. The invited speakers will talk about devices, design and compact modeling aspects, and applications, permitting a full development platform from devices to circuit & systems based on spintronics.

Time Label Presentation Title Authors
11:30 2.7.1 MAIN MEMORY ORGANIZATION TRADE-OFFS WITH DRAM AND STT-MRAM OPTIONS BASED ON EXTENDED GEMS/NVMAIN SIMULATION FRAMEWORK
Manu Komalan, IMEC, BE
Manu Komalan1, Oh Hyung Rock1, Matthias Hartmann1, Sushil Saikhande1, Christian Tenllado2, Jose Ignacio Gomez2, Gouri Sankar Kar1, Amaud Fumemont1, Francy Cattao1, Sophie Senn2, David Novo1, Abdoulaye Gamatie3, Lionel Torres2
IMEC, BE; 1University Complutense de Madrid (UCM), ES; 2LIRMM, FR; 3French National Centre for Scientific Research (CNRS), FR; 4CNRS LIRMM / University of Montpellier, FR; 5University of Montpellier, FR

Abstract
Current main memory organizations in embedded and mobile application systems are DRAM dominated. The everincreasing gap between today’s processor and memory speeds makes the DRAM subsystem design a major aspect of computer system design. However, the limitations to DRAM scaling and other associated challenges like refresh provides some undesired trade-offs between performance, energy and area to be made by architecture designers. Several emerging NVM options are being explored to at least partly remedy this but today it is very hard to assess the viability of these proposals because the simulations are not fully based on realistic assumptions on the NVM memory technologies and on the system architecture level. In this paper, we propose to use realistic, calibrated STTMRAM models and a well calibrated cross-layer simulation and exploration framework, named SEAT, to better consider technologies aspects and architecture constraints. We will focus on general purpose/mobile SoC multi-core architectures. We will highlight results for a number of relevant benchmarks, representatives of numerous applications based on actual system architecture.

Download Paper (PDF; Only available from the DATE venue WiFi)
Many cognitive algorithms such as neural networks cannot be efficiently executed by von Neumann architectures, the performance of which is constrained by the memory wall between microprocessor and memory hierarchy. Hence, researchers started to investigate new computing paradigms such as neuromorphic computing that can adapt their structure to the topology of the algorithms and accelerate their executions. New computing units have been also invented to support this effort by leveraging emerging nanodevices. In this work, we will discuss the opportunity of implementing neuromorphic computing systems with spintronic devices. We will also provide insights on how spintronic devices fit into different part of neuromorphic computing systems. Approaches to optimize the circuits are also discussed.

Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

2.8 Enabling ICT Innovations for European SMEs

Date: Tuesday, March 20, 2018
Time: 11:30 - 13:00
Location / Room: Exhibition Theatre

Organisers:
Rainer Leupers, RWTH Aachen, DE, Contact Rainer Leupers
Bernd Janson, ZENIT GmbH, DE, Contact Bernd Janson

Moderator:
Luca Fanucci, University of Pisa, IT, Contact Luca Fanucci

Technology-driven business in Europe fails more often compared to other regions such as the USA and China. Turning results into products is a challenge which starts at the very beginning of an idea for a new technology and obliges researchers to cooperate with business experts and investors. The European Commission started its Smart Anything Everywhere (SAE) Initiative to foster transfer from research to business in the areas of Cyber Physical Systems (CPS) and via the instrument of Digital Innovation Hubs (DIH).

Two SAE initiative projects will be presented in the workshop session, supported by two speakers from industry presenting their products consisting of programming technology SLX (Silexica) and software for automated driving (BASELABS). The workshop session will introduce the SAE approach and their individual funding schemes for European university-industry cooperation. The technology transfer concept focuses on direct cooperation between universities and SMEs supported by open innovation networks and other stakeholders like investors. The session speakers will demonstrate in a pragmatic way and by use of concrete examples how technology transfer can be initiated and implemented in practice and to overcome the associated pitfalls and use the innovation opportunities. The mix of presentations ensures that both academic and industrial viewpoints and concerns are adequately addressed. The workshop session will hence be of interest to a large audience. Amongst others, the goal is to motivate more stakeholders to engage in international technology transfer.

During the session, SAE representatives will share their experiences and insights as researcher, founder, entrepreneur, investor or consultant.

Time Label Presentation Title Authors
11:30 2.8.1 PRESENTATION OF TETRAMAX
Speaker: Rainer Leupers, RWTH Aachen, DE
Abstract
TETRAMAX as part of the SAE Initiative started in 2017 and is funded by Horizon 2020. The project supports application experiments between academia and industry (SMEs) related to Internet of Things (IoT) technologies and focusing on customized low energy computing (CLEC).
The project builds on three major activity lines:
(1) Stimulating, organizing, co-funding, and evaluating different types of cross-border Application Experiments, providing "EU added value" via innovative CLEC technologies to first-time users and broad markets in European ICT-related industries.
(2) Building and leveraging a new European CLEC competence center network, offering technology brokerage, one-stop shop assistance and CLEC training to SMEs and mid-caps, and with a clear evolution path towards new regional digital innovation hubs where needed, and
(3) Paving the way towards self-sustainability based on pragmatic and customized long-term business plans
The project impact will be measured based on 50+ performance indicators. The immediate ambition of TETRAMAX within its duration is to support 50+ industry clients and 3rd parties in the entire EU with innovative technologies, leading to an estimated revenue increase of 25 Mio. € based on 50+ new or improved CLEC-based products, 10+ entirely new businesses/SMEs initiated, as well as 30+ new permanent jobs and significant cost and energy savings in product manufacturing. Moreover, in the long term, TETRAMAX will be the trailblazer towards a reinforced, profitable, and sustainable ecosystem infrastructure, providing CLEC competence, services and a continuous innovation stream at European scale, yet with strong regional presence as favoured by SMEs.
2.8.2 THE FED4SAE PROJECT, ACCELERATING EUROPEAN CPS SOLUTIONS TO MARKET

Speaker: Isabelle Dor, COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES, FR

Abstract
Fed4SAE - accelerating European CPS solutions to market - will boost digitization of European industry by strengthening companies' competitiveness in the CPS market under the Smart Anything Everywhere Initiative. The H2020 project aims at creating a pan-European network of Digital Innovation Hubs (DIH) by leveraging existing regional technology or business ecosystems across complete value chains and multiple competencies. The network within Fed4SAE will enable start-ups, SMEs and mid-cap companies in all sectors to build and create new digital products and services. The project mission also includes innovation management and links these companies to suppliers and investors in order to create innovative CPS solutions and accelerate their development and industrialization. The Fed4SAE project will fund industrial projects by means of the cascade funding process set by the European Commission and through an adapted, fast-response and agile approach to attract innovative companies. Three open calls have taken/will take place in 2017 and 2018 in order to support the best projects in accordance with several criteria: innovative solution in terms of technical expertise, mature solution with a good technology readiness level, impacting solution thanks to efficient innovation management. Companies will be able to use industrial CPS programmable platforms in combination with expertise and know-how from the R&D Advanced Platform according to the application domains. The ultimate goal of each project within Fed4SAE is to provide a complete solution combining hardware and software, available to be tested in the market environment prior to large deployment in the targeted market - this deployment will be supported to enhance the innovation management.

2.8.3 OPEN INNOVATION BUSINESS BASED ON EFFICIENT NETWORKING

Speaker: Bernd Janson, ZENIT GmbH, DE

Abstract
Open innovation is based on strong networks between academia and industry. ICT developments depend greatly on open innovation due to short innovation cycles and strong competition. To build an open innovation network which operates in a regional, national and international context was the idea behind the Enterprise Europe Network which started in 2008. The overall aim is to support the competitiveness and innovation capabilities of SMEs in Europe. Today, the Enterprise Europe Network is the largest innovation network in the world. It addresses every need in the whole value chain of the innovation process - from idea to product. Bernd Janson will explain how 600 partners and over 6000 consultants worldwide work together to improve the performance of SMEs in Europe. He also explains the Network's role within Tetramax.

2.8.4 THE SILEXICA MULTICORE SOLUTION

Speaker: Juan Eusse, Silexica GmbH, DE

Abstract
Silexica was founded in 2014 as a spin-off from RWTH Aachen University and expanded rapidly to complete an $8 million “Series A” round of financing in November 2016. The following year saw Silexica open offices in Silicon Valley and Japan and win multicore solution projects with companies including Denso, Ricoh and Fujitsu. Silexica continues to work with RWTH Aachen on research projects to develop and enhance SLX, Silexica’s programming technology SLX uses state-of-the-art compiler technology and full heterogeneity awareness to support software professionals in the most challenging projects. Juan Eusse’s presentation will talk about the transfer of technology from university to a limited company, forming a strong relationship between both parties and learning from each other. He will provide an insight into SLX with real industrial examples of how the company and technologies have since developed.

2.8.5 TRANSFERRING RESEARCH RESULTS TO SAFETY-RELEVANT PRODUCTS: CASE STUDY ON AUTOMATED DRIVING SOFTWARE

Speaker: Robert Schubert, BASELABS GmbH, DE

Abstract
While the transfer from research to market exploitation is already a challenge in itself, additional tasks arise when it comes to products that are used for safety-critical applications. This case study will focus on the example of BASELABS, a software company which focuses on automated driving. The presentation will highlight the path from research via pre-development towards safety-certified software products and the different business models related to each stage. The objective is to provide insights and best practices helping researchers and entrepreneurs with similar challenges.

2.8.6 INTENTA GMBH

Speaker: Basel Fardi, Intenta GmbH, DE

Abstract
Intenta is on the cutting edge of research and development in the fields of image processing, data fusion, and object/person recognition and detection. Intenta’s focus is the development and marketing of product lines based on smart sensor technologies. Founded in 2011, the company has experienced continual growth to over 160 employees in 2017.
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018

- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

UB02 Session 2

Date: Tuesday, March 20, 2018
Time: 12:30 - 15:00
Location / Room: Booth 1, Exhibition Area

<table>
<thead>
<tr>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>UB02.1</td>
<td>ARCHON: AN ARCHITECTURE-OPEN RESOURCE-DRIVEN CROSS-LAYER MODELLING FRAMEWORK</td>
<td>Fei Xia1, Ashur Rafiev1, Mohammed Al-Hayanni1, Alexei Iliasov1, Rishad Shafik1, Alexander Romanovsky1 and Alex Yakovlev1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1Newcastle University, GB; 2Newcastle University, UK and University of Technology and HCED, IQ</td>
</tr>
<tr>
<td></td>
<td>Abstract</td>
<td>This demonstration showcases a modeling method for large complex computing systems focusing on many-core types and concentrating on the cross-layer aspects. The resource-driven models aim to help system designers reason about, analyze, and ultimately design such systems across all conventional computing and communication layers, from application, operating system, down to the finest hardware details. The framework and tool support the notion of selective abstraction and are suitable for studying such non-functional properties such as performance, reliability and energy consumption.</td>
</tr>
</tbody>
</table>

UB02.2 GENERATING FULL-CUSTOM SCHEMATICS IN A MIXED-SIGNAL TOP-DOWN DESIGN FLOW

Authors: Tobias Markus1, Markus Mueller2 and Ulrich Bruening1
1University of Heidelberg, DE; 2Extoll GmbH, DE

Abstract: Design time is one of the precious assets in the cycle of hardware design. The top down methodology has been used in digital designs very successfully and now we also apply it for analog and mixed signal designs. Generating most of the structures automatically saves time and avoids errors. A Top Down Design Flow for Mixed Signal Designs is used which generates the schematic structure from the system RNM representation. Since the structural verilog part of the system level design will automatically generate the schematic structure it is only the functional part which is missing and has to be implemented by the analog designer. Some often used blocks can be used as an entry point to partially generate parts of the design in the schematic and furthermore even parts of the layout. We will demonstrate this design method with an example project.

UB02.3 OSC MULTICORE STENCIL PROCESSOR: ONE INSTRUCTION-SET COMPUTER-BASED MULTICORE PROCESSOR FOR STENCIL COMPUTING

Authors: Kaoru Saso, Jing Yuan Zhao and Yuko Hara-Azumi, School of Engineering, Tokyo Institute of Technology, JP

Abstract: Subtract and Branch on NEGative with 4 operands (SUBNEG4) is one of One Instruction-Set Computers that execute only one type of instruction. Thanks to its simplicity, SUBNEG4 has only 1/20th circuit area and 1/10x power consumption against MIPS processor. As SUBNEG4 is Turing-complete, it is suitable for parallel computing by multiple cores, while keeping its low-power feature. Our on-going project is seeking for effective use and deployment of SUBNEG4 cores on embedded systems. Our booth will demonstrate the significant speed-up by a SUBNEG4-based many-core processor against a conventional processor, for stencil computing. Our 64-core processor efficiently handles 2D von-Neumann neighborhood stencils, e.g., wave simulation by Verlet integration and 2D Jacobi iteration, to compute 64 points simultaneously. We show that small many-core processors can be realized even with such large number of cores while achieving good speed-up for heaving computation.
Abstract

While digital design automation is highly developed, analog design automation still remains behind the demands. Previous circuit synthesis approaches, which are usually based on optimization algorithms, do not satisfy industrial requirements. A promising alternative is given by procedural approaches (also known as “generators”); They (a) emulate experts’ decisions, thus (b) make expert knowledge re-usable and (c) can consider all relevant aspects and constraints implicitly. Nowadays, generators are successfully applied in analog layout (PCBs, PCells). We aim at an entire design flow completely based on procedural automation techniques. This flow will consist of procedures for the generation of schematics and layouts for every typical analog circuit class, such as amplifier, bandgap, filter a.s.o. In our presentation we give an overview on such a design flow and we show an approach for capturing an analog circuit designer’s strategy as an executable “expert design plan”.

More information ...
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

3.1 Executive Session: Design Automation for Quantum Computing

Date: Tuesday, March 20, 2018
Time: 14:30 - 16:00
Location / Room: Saal 2

Chair:
Charbon Edoardo, TU Delft / EPFL, NL, Contact Charbon Edoardo

Co-Chair:
Große Daniel, University of Bremen, DE, Contact Daniel Große

Recent developments in quantum hardware indicate that systems featuring more than 50 physical qubits are within reach. At this scale, classical simulation will no longer be feasible and there is a possibility that such quantum devices may outperform even classical supercomputers at certain tasks. With the rapid growth of qubit numbers and coherence times comes the increasingly difficult challenge of quantum program compilation. This entails the translation of a high-level description of a quantum algorithm to hardware-specific low-level operations which can be carried out by the quantum device. Some parts of the calculation may still be performed manually due to the lack of efficient methods. This, in turn, may lead to a design gap, which will prevent the programming of a quantum computer. In this session, we discuss the challenges in fully-automatic quantum compilation. We motivate directions for future research to tackle these challenges. Yet, with the algorithms and approaches that exist today, we demonstrate how to automatically perform the quantum programming flow from algorithm to a physical quantum computer for a simple algorithmic benchmark, namely the hidden shift problem. We present and use two tool flows which invoke RevKit. One which is based on ProjectQ and which targets the IBM Quantum Experience or a local simulator, and one which is based on Microsoft's quantum programming language Q#.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30</td>
<td>3.1.1</td>
<td>QUANTUM ALGORITHMS: THE QUEST FOR SCALABLE PROGRAMMING, SYNTHESIS, AND TEST</td>
<td>Martin Roetteler, Microsoft, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>tbd</td>
</tr>
<tr>
<td>15:00</td>
<td>3.1.2</td>
<td>PROJECTQ: A SOFTWARE FRAMEWORK FOR PROGRAMMING QUANTUM COMPUTERS</td>
<td>Thomas Haener, ETHZ, CH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>tbd</td>
</tr>
<tr>
<td>15:30</td>
<td>3.1.3</td>
<td>REVKIT: AUTOMATIC COMPILATION AND DESIGN SPACE EXPLORATION FOR QUANTUM PROGRAMS</td>
<td>Mathias Soeken, Integrated System Laboratory – EPFL, CH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>tbd</td>
</tr>
</tbody>
</table>
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

3.2 Approximate and Near-Threshold Computing

Date: Tuesday, March 20, 2018
Time: 14:30 - 16:00
Location / Room: Konf. 6

Chair:
Semeen Rehman, Vienna University of Technology (TU Wien), AT, Contact Semeen Rehman

Co-Chair:
Saibal Mukhopadhyay, Georgia Tech., US, Contact Saibal Mukhopadhyay

This session focuses on approximate and near-threshold computing. The first paper proposes a novel dynamic virtual machine (VM) allocation method, while guaranteeing quality of service (QoS) requirements. The second paper introduces and presents an adaptive simulation methodology in which neurons in the region of interest (ROI) follow highly accurate biological models while the other neurons follow computation-friendly models. Finally the last paper shows an approximate computing technique to perform approximate computing with memory, avoiding redundant computation when encountering similar input patterns. The session also includes one IP paper on approximate big data computing.

3.2.1 ENERGY PROPORTIONALITY IN NEAR-THRESHOLD COMPUTING SERVERS AND CLOUD DATA CENTERS: CONSOLIDATING OR NOT?

Speaker:
Ali Pahlevan, Embedded Systems Lab (ESL), EPFL, CH

Authors:
Ali Pahlevan1, Yasir Mahmood Qureshi2, Marina Zapater1, Andrea Bartolini2, Davide Rossi2, Luca Benini2 and David Atienza1

1Embedded Systems Lab (ESL), EPFL, CH; 2Integrated System Laboratory ETH, Zurich, CH; 3Energy Efficient Embedded Systems (EEES) Lab – DEI, University of Bologna, IT

Abstract
Cloud Computing aims to efficiently tackle the increasing demand of computing resources, and its popularity has led to a dramatic increase in the number of computing servers and data centers worldwide. However, as effect of post-Dennard scaling, computing servers have become power-limited, and new system-level approaches must be used to improve their energy efficiency. This paper first presents an accurate power modelling characterization for a new server architecture based on the FD-SOI process technology for near-threshold computing (NTC). Then, we explore the existing energy vs. performance trade-offs when virtualized applications with different CPU utilization and memory footprint characteristics are executed. Finally, based on this analysis, we propose a novel dynamic virtual machine (VM) allocation method that exploits the knowledge of VM's characteristics together with our accurate server power model for next-generation NTC-based data centers, while guaranteeing quality of service (QoS) requirements. Our results demonstrate the inefficiency of current workload consolidation techniques for new NTC-based data center designs, and how our proposed method provides up to 45% energy savings when compared to state-of-the-art consolidation-based approaches.

Download Paper (PDF; Only available from the DATE venue WiFi)
### 3.2.2 Lookup Table Allocation for Approximate Computing with Memory Under Quality Constraints

**Speaker:**
Yun Long, Georgia Institute of Technology, US  
Yun Long, Xueyuan She and Saibal Mukhopadhyay, Georgia Institute of Technology, US

**Abstract**

Due to the large-scale and biophysically plausible nature of the brain, biophysical neural network (BNN) modeling requires solving multiple terms, coupled and non-linear differential equations, making simulation computationally complex and memory intensive. This paper presents an adaptive simulation methodology in which neurons in the region of interest (ROI) follow high biological accurate models while the other neurons follow computation friendly models. To enable ROI based approximation, we propose a generic template based computing algorithm which unifies the data structure and computing flow for various neuron models. We implement the algorithms on CPU, GPU and embedded platforms, showing a 11x speedup with insignificant loss of biological details in the region of interest.

**Download Paper (PDF; Only available from the DATE venue WiFi)**

### 3.2.3 Accelerating Biophysical Neural Network Simulation with Region of Interest Based Approximation

**Speaker:**
Seyed Morteza Nabavinejad1, Xin Zhan2, Reza Azimi2, Maziar Goudarzi1 and Sherief Reda2  
1Sharif University of Technology, IR; 2Brown University, US

**Abstract**

To limit the peak power consumption of a cluster, a centralized power capping system typically assigns power caps to the individual servers, which are then enforced using local capping controllers. Consequently, the performance and throughput of the servers are affected, and the runtime of jobs is extended as a result. We observe that servers in big data processing clusters often execute big data applications that have different tolerance for approximate results. To mitigate the impact of power capping, we propose a new power-capping aware resource manager for Approximate Big data processing (CAB) that takes into consideration the minimum Quality-of-Result (QoR) of the jobs. We use industry standard feedback power capping controllers to enforce a power cap quickly, while, simultaneously modifying the resource allocations to various jobs based on their progress rate, target minimum QoR, and the power cap such that the impact of capping on runtime is minimized. Based on the applied cap and the progress rates of jobs, CAB dynamically allocates the computing resources (i.e., number of cores and memory) to the jobs to mitigate the impact of capping on the final time. We implement CAB in Hadoop-2.7.3 and evaluate its improvement over other methods on a state-of-the-art 28-core Xeon server. We demonstrate that CAB minimizes the impact of power capping on runtime by up to 39.4% while meeting the minimum QoR constraints.

**Download Paper (PDF; Only available from the DATE venue WiFi)**
This session presents innovative techniques for optimizing several aspects in multi-processor/core system design. The first paper proposes an effective and efficient design space exploration technique for designing domain-specific platforms. The second paper proposes a novel task allocation and scheduling scheme to maximize soft-error reliability while satisfying lifetime reliability constraints for soft-real-time MPSoCs. Third paper proposes a virtual resource manager that monitors the access behavior and predicts the node-to-node interconnect performance.

15:00  3.3.2  VARIATION-AWARE TASK ALLOCATION AND SCHEDULING FOR IMPROVING RELIABILITY OF REAL-TIME MPSoCs

Speaker: Junlong Zhou, Nanjing University of Science and Technology, CN
Authors: Junlong Zhou1, Tongquan Wei2, Mingsong Chen3, Xiaobo Sharon Hu2, Yue Ma2, Xiaoyan Zang2 and Jianming Yan4
1Nanjing University of Science and Technology, CN; 2East China Normal University, CN; 3University of Notre Dame, US; 4Meitu.com Corporation, CN
Abstract
Both soft-error reliability (SER) due to transient faults and lifetime reliability (LTR) due to permanent faults are key concerns in real-time MPSoCs. Existing works have investigated related problems, however, most of them only focus on one of the two reliability concerns. A few efforts do consider both types of reliability together, but ignore the impacts of hardware- and application-level variations on reliability, thus are not applicable to state-of-the-art MPSoCs under variations. In this paper, we focus on increasing SER without sacrificing LTR since transient faults occur much more frequently than permanent faults. Specifically, we propose a novel task allocation and scheduling scheme to maximize SER while satisfying a LTR constraint for soft real-time MPSoCs. Considering that SER is the objective while LTR is a constraint in our problem, and LTR is highly related to core temperature profiles, we dedicate to investigating the effects of variations in core soft-error rate, task vulnerability to soft errors, and task execution time on SER. To the best of our knowledge, our work is the first attempt that jointly handles the two reliability issues as well as taking into account the effects of variations on reliability. Experimental results show that our scheme improves the SER by up to 66% as compared to a number of representative existing approaches while meeting the same LTR constraint.

Download Paper (PDF; Only available from the DATE venue WiFi)

16:00  3.3.3  EXACT MULTI-OBJECTIVE DESIGN SPACE EXPLORATION USING ASPMT

Speaker: Kai Neubauer, University of Rostock, DE
Authors: Kai Neubauer1, Philipp Wanko2, Torsten Schaub3 and Christian Haubelt1
1University of Rostock, DE; 2University of Potsdam, DE
Abstract
An efficient Design Space Exploration (DSE) is imperative for the design of modern, highly complex embedded systems in order to steer the development towards optimal design points. The early evaluation of design decisions at system-level abstraction layer helps to find promising regions for subsequent development steps in lower abstraction levels by diminishing the complexity of the search problem. In recent works, symbolic techniques, especially Answer Set Programming (ASP) modulo Theories (ASPMt), have been shown to find feasible solutions of highly complex system-level synthesis problems with non-linear constraints very efficiently. In this paper, we present a novel approach to a holistic system-level DSE based on ASPmT. To this end, we include additional background theories that concurrently guarantee compliance with hard constraints and perform the simultaneous optimization of several design objectives. We implement and compare our approach with a state-of-the-art preference handling framework for ASP. Experimental results indicate that our proposed method produces better solutions with respect to both diversity and convergence to the true Pareto front.

Download Paper (PDF; Only available from the DATE venue WiFi)
3.4 Optimizing Computing with Neuromorphic Architectures and Accelerators

Date: Tuesday, March 20, 2018
Time: 14:30 - 15:30
Location / Room: Konf. 2
Chair:
Dimitrios Soudris, NTUA, GR, Contact Dimitrios Soudris
Co-Chair:
Ioana Vatajelu, University of Grenoble–Alpes, TIMA Laboratory, FR, Contact Elena Ioana Vatajelu

Creating performance and power efficient acceleration techniques is a major challenge. In this session, various approaches are presented toward this direction for neural network applications and GPUs. A wide range of optimization techniques are discussed, including application-level optimizations, system-level solutions, matrix optimizations, and accuracy vs. computations trade-offs.

3.4.1 STRUCTURE OPTIMIZATIONS OF NEUROMORPHIC COMPUTING ARCHITECTURES FOR DEEP NEURAL NETWORKS

Speaker:
Heechun Park, SNUCAD, KR
Authors:
Heechun Park and Taewhan Kim, Seoul National University, KR
Abstract
This work addresses a new structure optimization of neuromorphic computing architectures. This enables to speed up the DNN (deep neural network) computation twice as fast as, theoretically, that of the existing architectures. Precisely, we propose a new structural technique of mixing both of the dendritic and axonal based neuromorphic cores in a way to totally eliminate the inherent non-zero waiting time between cores in the DNN implementation. In addition, in conjunction with the new architecture we propose a technique of maximally utilizing computation units so that the resource overhead of total computation units can be minimized. We have provided a set of experimental data to demonstrate the effectiveness (i.e., speed and area) of our proposed architectural optimizations: ~2x speedup with no accuracy penalty on the neuromorphic computation or improved accuracy with no additional computation time.
Download Paper (PDF; Only available from the DATE venue WiFi)

3.4.2 CCR: A CONCISE CONVOLUTION RULE FOR SPARSE NEURAL NETWORK ACCELERATORS

Speaker:
Jiajun Li, Institute of Computing Technology, Chinese Academy of Sciences, CN
Authors:
Jiajun Li, Guihai Yan, Wenyen Lu, Shuhao Jiang, Shijun Gong, Jingya Wu and Xiaowei Li, Institute of Computing Technology, Chinese Academy of Sciences, CN
Abstract
Convolutional Neural networks (CNNs) have achieved great success in a broad range of applications. As CNN-based methods are often both computation and memory intensive, sparse CNNs have emerged as an effective solution to reduce the amount of computation and memory accesses while maintaining the high accuracy. However, dense CNN accelerators can hardly benefit from the reduction of computations and memory accesses due to the lack of support for irregular and sparse models. This paper proposes a concise convolution rule (CCR) to diminish the gap between sparse CNNs and dense CNN accelerators. CCR transforms a sparse convolution into multiple effective and intuitive ones. The ineffective convolutions in which either the neurons or synapses are all zeroes do not contribute to the final results and the computations and memory accesses can be eliminated. The effective convolutions in which both the neurons and synapses are dense can be easily mapped to the existing dense CNN accelerators. Unlike prior approaches which trade complexity for flexibility, CCR advocates a novel approach to reaping the benefits from the reduction of computation and memory accesses as well as the acceleration of the existing dense architectures without intrusive PE modifications. As a case study, we implemented a sparse CNN accelerator, SparseK, following the rationale of CCR. The experiments show that SparseK achieved a speedup of ~2.9 ms/s on VGG16 compared to a comparably provisioned dense architecture. Compared with state-of-the-art sparse accelerators, SparseK can improve the performance and energy efficiency by 1.8x and 1.5x, respectively.
Download Paper (PDF; Only available from the DATE venue WiFi)
HIPE: HMC INSTRUCTION PREDICATION EXTENSION APPLIED ON DATABASE PROCESSING

Speaker:
Diego Tomé, Centrum Wiskunde & Informatica (CWI), BR

Authors:
Diego Gomes Tomé1, Paulo Cesar Santos2, Luigi Carro2, Eduardo Cunha de Almeida3 and Marco Antonio Zanata Alves3
1Federal University of Paraná, BR; 2UFRGS, BR; 3UFPR, BR

Abstract
The recent Hybrid Memory Cube (HMC) is a smart memory which includes functional units inside one logic layer of the 3D stacked memory design. In order to execute instructions inside the Hybrid Memory Cube (HMC), the processor needs to send instructions to be executed near data, keeping most of the pipeline complexity inside the processor. Thus, control-flow and data-flow dependencies are all managed inside the processor, in such way that only update instructions are supported by the HMC. In order to solve data-flow dependencies inside the memory, previous work proposed HMC Instruction Vector Extensions (HIVE), which embeds a high number of functional units with an interlock register bank. In this work, we propose HMC Instruction Prediction Extensions (HIPE), that supports predicated execution inside the memory, in order to transform control-flow dependencies into data-flow dependencies. Our mechanism focuses on removing the high latency iteration between the processor and the smart memory during the execution of branches that depends on data processed inside the memory. In this paper, we evaluate a balanced design of HIVE comparing to x86 and HMC executions. After we show the HIPE mechanism results when executing a database workload, which is a strong candidate to use smart memories. We show interesting trade-offs of performance when comparing our mechanism to previous work.

Download Paper (PDF; Only available from the DATE venue WiFi)

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

3.5 Memory Reliability

Date: Tuesday, March 20, 2018
Time: 14:30 - 16:00
Location / Room: Konf. 3

Chair:
Jose Pineda, NXP, NL, Contact Jose Pineda

Co-Chair:
Mehdi Tahoori, Karlsruhe Institute of Technology, DE, Contact Mehdi Tahoori

This session discusses reliability issues for different on-chip and off-chip memory technologies. The first paper uses important sampling to reduce the number of Monte Carlo simulations to obtain failure rates for advanced SRAM memories. The second paper performs degradation analysis for FinFET memories. The third paper discusses reliability issues for solid state memories.
INVESTIGATING POWER OUTAGE EFFECTS ON RELIABILITY OF SOLID-STATE DRIVES

Speaker:
Hossein Asadi, Sharif University of Technology, IR

Authors:
Saba Ahmadian, Farhad Taheri, Mehrshad Lotfi, Maryam Karimi and Hossein Asadi, Sharif University of Technology, IR

Abstract
Solid-State Drives (SSDs) are recently employed in enterprise servers and high-end storage systems in order to enhance performance of storage subsystem. Although employing high speed SSDs in the storage subsystems can significantly improve system performance, it comes with significant reliability threat for write operations upon power failures. In this paper, we present a comprehensive analysis investigating the impact of workload dependent parameters on the reliability of SSDs under power failure for variety of SSDs (from top manufacturers). To this end, we first develop a platform to perform two important features required for study: a) a realistic fault injection into the SSD in the computing systems and b) data loss detection mechanism on the SSD upon power failure. In the proposed physical fault injection platform, SSDs experience a real discharge phase of Power Supply Unit (PSU) that occurs during power failure in data centers which was neglected in previous studies. The impact of workload dependent parameters such as workload Working Set Size (WSS), request size, request type, access pattern, and sequence of accesses on the failure of SSDs is carefully studied in the presence of realistic power failures. Experimental results over thousands number of fault injections show that data loss occurs even after completion of the request (up to 700ms) where the failure rate is influenced by the type, size, access pattern, and sequence of IO accesses while other parameters such as workload WSS has no impact on the failure of SSDs.

Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

### 3.6 Real-time Multiprocessing

**Date:** Tuesday, March 20, 2018  
**Time:** 14:30 - 16:00  
**Location / Room:** Konf. 4

**Chair:**  
Jian-Jia Chen, TU Dortmund, DE, Contact Jian-Jia Chen

**Co-Chair:**  
Rolf Ernst, TU Braunschweig, DE, Contact Rolf Ernst

The session details on various aspects of real-time multiprocessors, where special focus is put on workload-aware scheduling. Network-on-Chips, security and synchronization constraints. The first paper improves the overall schedulability by strategically arranging the workload among processors. The second paper reduces the pessimism in the analysis of NoC. The third paper considers security-related workloads whilst maintaining feasibility of schedules. The fourth paper presents an implementation of SDF graphs by means of OS-synchronization primitives.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30</td>
<td>3.6.1</td>
<td>WORKLOAD-AWARE HARMONIC PARTITIONED SCHEDULING FOR PROBABILISTIC REAL-TIME SYSTEMS</td>
<td>Jiankang Ren, Dalian University of Technology, CN</td>
</tr>
</tbody>
</table>

**Authors:**  
Jiankang Ren, Ran Bi, Xiaoyan Su, Qian Liu, Guowei Wu and Guozhen Tan, Dalian University of Technology, CN

**Abstract**  
Multiprocessor platforms, widely adopted to realize real-time systems nowadays, bring the probabilistic characteristic to such systems because of the performance variations of complex chips. In this paper, we present a harmonic partitioned scheduling scheme with workload awareness for periodic probabilistic real-time tasks on multiprocessors under the fixed-priority preemptive scheduling policy. The key idea of this research is to improve the overall schedulability by strategically arranging the workload among processors based on the exploration of the harmonic relationship among probabilistic real-time tasks. In particular, we define a harmonic index to quantify the harmony among probabilistic real-time tasks. This index can be obtained via the harmonic period transformation and probabilistic cumulative worst case utilization calculation of these tasks. The proposed scheduling scheme first sorts tasks with respect to the workload, then packs them to processors one by one aiming at minimizing the increase of harmonic index caused by the task assignment. Experiments with randomly generated task sets show significant performance improvement of our proposed approach over the existing harmonic partitioned scheduling algorithm for probabilistic real-time systems.  
Download Paper (PDF; Only available from the DATE venue WiFi)

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:00</td>
<td>3.6.2</td>
<td>BUFFER-AWARE BOUNDS TO MULTI-POINT PROGRESSIVE BLOCKING IN PRIORITY-PREEMPTIVE NOCs</td>
<td>Leandro Indrusiak, University of York, GB</td>
</tr>
</tbody>
</table>

**Authors:**  
Leandro Indrusiak¹, Alan Burns² and Borislav Nikolic²  
¹University of York, GB; ²Clister/INESC TEC, ISEP, IPP, PT

**Abstract**  
This paper aims to reduce the pessimism of the analysis of the multi-point progressive blocking (MPB) problem in real-time priority-preemptive wormhole networks-on-chip. It shows that the amount of buffering on each network node can influence the worst-case interference that packets can suffer along their routes, and it proposes a novel analytical model that can quantify such interference as a function of the buffer size. It shows that, perhaps counter-intuitively, smaller buffers can result in lower upper-bounds on interference and thus improved schedulability. Didactic examples and large-scale experiments provide evidence of the strength of the proposed approach.  
Download Paper (PDF; Only available from the DATE venue WiFi)
A DESIGN-SPACE EXPLORATION FOR ALLOCATING SECURITY TASKS IN MULTICORE REAL-TIME SYSTEMS

Speaker: Monowar Hasan, University of Illinois, BD

Authors: Monowar Hasan¹, Sibin Mohan¹, Rodolfo Pellizzoni² and Rakesh Bobba³
¹University of Illinois at Urbana-Champaign, US; ²University of Waterloo, CA; ³Oregon State University, US

Abstract
The increased capabilities of modern real-time systems (RTS) introduce more security threats. Recently, frameworks that integrate security tasks without perturbing the real-time tasks have been proposed, but they only target single core systems. However, modern RTS are migrating towards multicore platforms. This makes the problem of integrating security mechanisms more complex, as designers now have multiple choices for where to allocate the security tasks. In this paper, we propose Hydra, a design space exploration algorithm that finds an allocation of security tasks into existing (viz., legacy) multicore RTS using the concept of opportunistic execution. Hydra allows security tasks to operate with existing real-time tasks without perturbing system parameters or normal execution patterns, while still meeting the desired monitoring frequency for intrusion detection. Our evaluation using a representative real-time control system (along with synthetic tasksets for a broader design space exploration) illustrates the efficacy of the proposed mechanism.

Download Paper (PDF; Only available from the DATE venue WiFi)

DESIGN AND ANALYSIS OF SEMAPHORE PRECEDENCE CONSTRAINTS: A MODEL-BASED APPROACH FOR DETERMINISTIC COMMUNICATIONS

Speaker: Yassine Ouhammou, LIAS / ENSMA & University of Poitiers, FR

Authors: Thanh-Dat Nguyen¹, Yassine OUHAMMOU¹, Emmanuel GROLLEAU¹, Julien Forge², Claire Pagetti³ and Pascal RICHARD¹
¹LIAS/ENSMA, FR; ²ULTRA/University of Lille 1, FR; ³ONERA / DTIM, FR

Abstract
Architecture Analysis and Design Language (AADL) is a standard in avionics system design. However, the communication patterns provided by AADL are not sufficient to the current context of Real-Time Embedded System (RTES) in which some multi-periodic communication patterns may occur. We propose an extension of a precedence model between tasks of different periods (multiperiodic communication). This relies on the Semaphore Precedence Constraint (SPC) model that is inspired from the concept of Semaphore, and more specifically on the m−n producer/consumer paradigm. We reinforce the SPC semantics by allowing cycles in the precedence graph. We also present another viewpoint on the periodicity of tasks system using SPC based on a graph apart from the encoding technique presented in the SPC seminal work. An implementation of SPC in AADL and its associated analysis tool are also provided to study the temporal behaviour of systems using SPC.

Download Paper (PDF; Only available from the DATE venue WiFi)

ONE-WAY SHARED MEMORY

Speaker and Author: Martin Schoeberl, Technical University of Denmark, DK

Abstract
Standard multicore processors use the shared main memory via the on-chip caches for communication between cores. However, this form of communication has two limitations: (1) it is hardly time predictable and therefore not a good solution for real-time systems and (2) this shared memory is a bottleneck in the system. This paper presents a communication architecture for time-predictable multicore systems where core-local memories are distributed on the chip. A network-on-chip constantly copies data from a sender core-local memory to a receiver core-local memory. As this copying is performed in one direction we call this architecture a one-way shared memory. With the use of time-division multiplexing for the memory accesses and the network-on-chip routers we achieve a time-predictable solution where the communication latency and bandwidth can be bounded. An example architecture for a 3x3 core processor and 32-bit wide links and memory ports provides a cumulative bandwidth of 29 bytes per clock cycle. Furthermore, the evaluation shows that this architecture, due to its simplicity, is small compared to other network-on-chip solutions.

Download Paper (PDF; Only available from the DATE venue WiFi)

Coffee Break in Exhibition Area

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
● Coffee Break 10:30 - 11:30
● Lunch Break 13:00 - 14:30
● Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
● Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
● Coffee Break 10:00 - 11:00
● Lunch Break 12:30 - 14:30
● Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
● Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
● Coffee Break 10:00 - 11:00
● Lunch Break 12:30 - 14:00
● Keynote Lecture in “Saal 2” 13:20 - 13:50
● Coffee Break 15:30 - 16:00

3.8 Innovative Products for Autonomous Driving (part 1)
Organiser:
Hans-Jürgen Brand, IDT/ZMDI, DE, Contact Hans-Jürgen Brand

The workshop on Innovative Products for Autonomous Driving includes 2 sessions (part 2: session 6.8). This session will highlight how to design functional safety products, how 5G will enable connected cars and foundry solutions for manufacturing chips for autonomous driving.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30</td>
<td>3.8.1</td>
<td>MICROELECTRONICS-DRIVEN INNOVATION IN MOBILITY</td>
<td>Christian Wolf¹ and Hans-Jürgen Brand² &lt;br&gt;¹IDT Europe GmbH, DE; ²IDT/ZMDI, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>The 2nd car and mobility revolution is predominantly enabled by innovative semiconductor products. The new megatrends in mobility such as vehicle electrification, vehicle connectivity and autonomous driving are creating a diversifying demand for new automotive semiconductors. The presentation will show the major trends in this area and also focus on approaches how Functional Safety - a key requirement for automotive semiconductors - can be handled in the design process of those products.</td>
</tr>
<tr>
<td>15:00</td>
<td>3.8.2</td>
<td>5G CONNECTED CARS</td>
<td>Stanislav Mudriievskyi and Vincent Latzko, Technical University Dresden, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>While autonomous driving already promises more comfort and safety, connected driving makes it possible to use the new strategies to improve the safety of road traffic, significantly reduce CO2 emissions and increase traffic efficiency. Additional 5G networking possibilities will remove the fundamental limitation of today's autonomous approaches that are used for controlling the vehicle by means of the onboard installed sensors only. It will be possible to use the information gained by the sensors of all neighbor vehicles as well as the environment or the existing infrastructure (e. g. surveillance cameras at crossroads, highways, geolocal weather sensors, etc.). All these can be virtually merged in the network, resulting in better decision-making.</td>
</tr>
<tr>
<td>15:30</td>
<td>3.8.3</td>
<td>FOUNDRY SOLUTIONS FOR AUTONOMOUS DRIVING</td>
<td>Alexander Muffler, X-Fab Semiconductor Foundries AG, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>ICs developed for autonomous driving do not only have to be designed according respective ASIL levels. They also need to be based and manufactured on highly reliable semiconductor processes. X-FAB is the leading analog/mixed-signal and MEMS foundry group manufacturing silicon wafers for the automotive market with a track record of more than 25 years as autonomous foundry. Already at process development, the clear focus is on highly reliable semiconductor processes. This also applies to IPs such as Flash, EEPROM, digital IPs like RAM and ROM, all for high temperature application up to 175 °C junction temperature. The PDKs - which are the interface between IC designers and the silicon - are developed by X-FAB with the target to give chip designers all necessary tools on hand to enable first-time-right mixed-signal IC development for harsh environments. X-FAB does not only provide simulation models which behave as close as possible to the real silicon, but also tools for lifetime calculation based on user defined mission profiles. X-FAB also develops NWIPs explicitly targeting automotive needs. These IP cores include for example error correction and detection modes and are especially suited for autonomous driving cars.</td>
</tr>
</tbody>
</table>

16:00 End of session

Coffee Break in Exhibition Area

Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

**Lunch Breaks (Großer Saal + Saal 1)**

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

**Tuesday, March 20, 2018**
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

**Wednesday, March 21, 2018**
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17.00

**Thursday, March 22, 2018**
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

UB03 Session 3

**Date:** Tuesday, March 20, 2018  
**Time:** 15:00 - 17:30  
**Location / Room:** Booth 1, Exhibition Area
## UB03.1 TOPOLOGY & MAGCAD: A DESIGN AND SIMULATION FRAMEWORK FOR THE EXPLORATION OF EMERGING TECHNOLOGIES

**Authors:** Umberto Garlando and Fabrizio Riente, Politecnico di Torino, IT

**Abstract**

We developed a design framework that enables the exploration and analysis of emerging beyond-CMOS technologies. It is composed of two powerful tools: ToPolNano and MagCAD. Different technologies are supported, and new ones could be added thanks to their modular structure. ToPolNano starts from a VHDL description of a circuit and performs the placed route following the technological constraints. The resulting circuit can be simulated both at logical or physical level. MagCAD is a layout editor where the user can design custom circuits, by placing basic elements of the selected technology. The tool can extract a VHDL netlist based on compact models of placed elements derived from experiments or physical simulations. Circuits can be verified with standard VHDL simulators. The design workflow will be demonstrated at the U-booth to show how those tools could be a valuable help in the studying and development of emerging technologies and to obtain feedbacks from the scientific community.

[More information...](#)

## UB03.2 GENERATING FULL-CUSTOM SCHEMATICS IN A MIXED-SIGNAL TOP-DOWN DESIGN FLOW

**Authors:** Tobias Markus¹, Markus Mueller² and Ulrich Brunner³

¹University of Heidelberg, DE; ²Exxol GmbH, DE

**Abstract**

Design time is one of the precious assets in the cycle of hardware design. The top down methodology has been used in digital designs very successfully and now we also apply it for analog and mixed signal designs. Generating most of the structures automatically saves time and avoids errors. A Top Down Design Flow for Mixed Signal Designs is used which generates the schematic structure from the system RNM representation. Since the structural verilog part of the system level design will automatically generate the schematic structure it is only the functional part which is missing and has to be implemented by the analog designer. Some often used blocks can be used as an entry point to partially generate parts of the design in the schematic and furthermore even parts of the layout. We will demonstrate this design method with an example project.

[More information...](#)

## UB03.3 DISGUIsing THE INTERconnectS: EFFICIENT PROTECTION OF DESIGN IP

**Authors:** Johann Knechtel¹, Satwik Patnak², Mohammed Ashraf³ and Ozgur Sinanoglu³

¹NYU Abu Dhabi, AE; ²New York University, US; ³New York University Abu Dhabi, AE

**Abstract**

Ensuring the trustworthiness and security of electronics has become an urgent challenge in recent years. Among various concerns, the protection of design intellectual property (IP) is to be addressed, due to outsourcing trends for the manufacturing supply chain and malicious end-user. In other words, adversaries either residing in the off-shore fab or in the field may want to obtain and pirate the design IP. As classical design tools do not consider such threats, there is clearly a need for security-aware EDA techniques. Here we present novel but proven techniques for efficient protection of design IP, embedded in an industrial-level design flow using Cadence Innovus. The key idea in our work is that disguising the interconnects is supremely suitable for protecting design IP, while inducing only little additional cost and providing strong resilience. We share our customized libraries with the community, and we demonstrate our design flow and its security measures.

[More information...](#)

## UB03.4 HARDENING THE HARDWARE: A REVERSE-ENGINEERING RESILIENT SECURE CHIP

**Authors:** Ashraaj Sangupura¹, Muhammad Yasin², Mohammed Nabeel³, Mohammed Ashraf³, Jayavijayan Rajendran⁴ and Ozgur Sinanoglu³

¹New York University, AE; ²New York University, US; ³New York University Abu Dhabi, AE; ⁴Texas A&M, US

**Abstract**

With the globalization of integrated circuit (IC) supply chain, the semiconductor industry is facing a number of threats, such as Intellectual Property (IP) piracy, hardware Trojans, and counterfeiting. To defend against such attacks at the hardware level, logic locking was proposed as a promising countermeasure. Yet, several recent attacks have completely undermined its security by successfully retrieving the secret key. Here, we present stripped-functionality logic locking (SFLL), which resists all existing attacks by hiding a part of the functionality in the form of a secret key. We leverage security-aware synthesis to develop a computer-aided design (CAD) framework that meets the desired security criterion at a minimal cost of 5%, 0.5%, and 8% for power, performance, and area, respectively. Moreover, we taped out a chip, the first such prototype of its kind, by applying our technique on an industry-level processor, namely, ARM Cortex-M0 microcontroller in 65nm technology.

[More information...](#)

## UB03.5 RECONFIGURABLE SELF-TIMED DATAFLOW ACCELERATOR

**Authors:** Daniil Sokolov, Alessandro de Gennaro and Andrey Makhov, Newcastle University, GB

**Abstract**

Many applications require reconfigurable pipelines to handle incoming data items differently depending on their values or the operating mode. Currently, reconfigurable synchronous pipelines are the mainstream of dataflow accelerators. However, there are certain advantages to be gained from self-timed dataflow processing, e.g. robustness to unstable power supply, data-dependent performance, etc. To become attractive for industry, reconfigurable asynchronous pipelines need a formal behavioural and design automation. This demo will present a design flow for the specification, verification and synthesis of reconfigurable self-timed pipelines using Dataflow Structure formalism in Workcraft (https://workcraft.org/). As a case study we will use an asynchronous accelerator for Ordinal Pattern Encoding (OPE) with reconfigurable pipeline depth. We will exhibit the resultant OPE chip fabricated in TSMC 65nm to show the benefits of reconfigurability and asynchrony for dataflow processing.

[More information...](#)

## UB03.6 SPANNER: SELF-REPAIRING SPIKING NEURAL NETWORK CONTROLLER FOR AN AUTONOMOUS ROBOT

**Authors:** Alan Milani¹, Anju Johnson¹, James Hilder¹, Andy Tyrrell¹, Jon Timmis¹, Junxiu Liu², Shivan Kasim², Jim Harkin² and Liam McDaid²

¹University of York, GB; ²Ulster University, GB

**Abstract**

The human brain is remarkably resilient, and is able to self-repair following injury or a stroke. In contrast, electronic systems typically exhibit limited self-repair capabilities, and cannot recover from faults. We demonstrate a bio-inspired approach to self-repair that allows an autonomous robot to recover from faults in its artificial brain. Astrocytes are support cells in the human brain that interact with neurons to regulate synaptic activity. We have modelled this interaction to create a spiking neural network that can self-repair when synapses between neurons are damaged, by strengthening redundant pathways. We demonstrate a robot platform controlled by a self-repairing spiking neural network that is implemented on an FPGA. We demonstrate that injecting faults into the synapses of the network initially causes the robot to behave erratically, but that the neural controller is able to automatically repair itself, thus allowing the robot to resume normal function.

[More information...](#)
Deep Neural Networks (DNNs) play a key role in prevailing machine learning applications. Resistive random-access memory (ReRAM) is capable of both computation and storage, contributing to the acceleration on DNNs process in memory. Besides, DNNs have a significant amount of zero weights, which provides a possibility to reduce computation cost by sparsity, because resistive accelerators have a high reliance on regular matrix-vector multiplication in ReRAM. In this work, we propose ReCom, the first resistive accelerator to support sparse DNN processing. ReCom is an efficient resistive accelerator for compressed deep neural networks, where DNN weights are structurally compressed to eliminate zero parameters and become more friendly to computation in ReRAM, and zero DNN activations are also considered at the same time. Two technologies, Structurally-compressed Weight and multiversioning and code transformations in the context of runtime autotuning. mARGOt is an autotuner that allows application adaptation to changing conditions and goals. Clava is a source-to-source compiler to transform C/C++ programs, including code instrumentation and integration with components such as mARGOt. We will demonstrate how to use Clava to integrate the mARGOt autotuner in an example application, and several mARGOt functionalities exposed through a Clava API.

More information ...

CLAVA-MARGOT: CLAVA + MARGOT = C/C++ TO C/C++ COMPILER AND RUNTIME AUTOTUNING FRAMEWORK

Authors:
João Bispo1, Davide Gadioli2, Pedro Pinto1, Emanuele VitaliP, Hamid Arabnejad1, Gianluca Palermo2, Cristina Silvano2, Jorge G. Barbosa1 and João M. P Cardoso1
1Porto University, PT; 2Politecnico di Milano (POLIMI), IT

Abstract
Current computing platforms consist of heterogeneous architectures. To efficiently target those platforms, compilers can be extend with code transformations and insertion of code to interface to runtime autotuning schemes, which tune application parameters according to: the actual execution, target architecture, and workload. We present an approach consisting of a C/C++ source-to-source compiler (Clava) and an autotuner (mARGOt). They are part of the toolflow of the FET-HPC ANTAREX project and allow parallelization, multiversioning and code transformations in the context of runtime autotuning. mARGOt is an autotuner that allows application adaptation to changing conditions and goals. Clava is a source-to-source compiler to transform C/C++ programs, including code instrumentation and integration with components such as mARGOt. We will demonstrate how to use Clava to integrate the mARGOt autotuner in an example application, and several mARGOt functionalities exposed through a Clava API.

More information ...

CLAVA-MARGOT: CLAVA + MARGOT = C/C++ TO C/C++ COMPILER AND RUNTIME AUTOTUNING FRAMEWORK

Authors:
João Bispo1, Davide Gadioli2, Pedro Pinto1, Emanuele VitaliP, Hamid Arabnejad1, Gianluca Palermo2, Cristina Silvano2, Jorge G. Barbosa1 and João M. P Cardoso1
1Porto University, PT; 2Politecnico di Milano (POLIMI), IT

Abstract
Current computing platforms consist of heterogeneous architectures. To efficiently target those platforms, compilers can be extend with code transformations and insertion of code to interface to runtime autotuning schemes, which tune application parameters according to: the actual execution, target architecture, and workload. We present an approach consisting of a C/C++ source-to-source compiler (Clava) and an autotuner (mARGOt). They are part of the toolflow of the FET-HPC ANTAREX project and allow parallelization, multiversioning and code transformations in the context of runtime autotuning. mARGOt is an autotuner that allows application adaptation to changing conditions and goals. Clava is a source-to-source compiler to transform C/C++ programs, including code instrumentation and integration with components such as mARGOt. We will demonstrate how to use Clava to integrate the mARGOt autotuner in an example application, and several mARGOt functionalities exposed through a Clava API.

More information ...

T-CREST: THE OPEN-SOURCE REAL-TIME MULTICORE PROCESSOR

Authors:
Martin Schoebeli, Luca Pezzarossa and Jens Sparse, Technical University of Denmark, DK

Abstract
Future real-time systems, such as advanced control systems or real-time image recognition, need more powerful processors, but still a system where the worst-case execution time (WCET) can be statically predicted. Multicore processors are one answer to the need for more processing power. However, it is still an open research question how to best organize and implement time-predictable communication between processing cores. T-CREST is an open-source multicore processor for research on time-predictable computer architecture. It consists of several Patmos processors connected by various time-predictable communication structures: access to shared off-chip, access to shared on-chip memory, and the Agil network-on-chip for fast inter-processor communication. T-CREST is supported by open-source development tools, such as compilation and WCET analysis. To best of our knowledge, T-CREST is the only fully open-source architecture for research on future real-time multicore architectures.

More information ...

FPGA-BASED HARDWARE ACCELERATOR FOR DRUG DISCOVERY

Authors:
Ghaith Tarawneh, Alessandro de Gennaro, Georgy Lukyanov and Andrey Mokrov, Newcastle University, GB

Abstract
We present an FPGA-based hardware accelerator for drug discovery, developed during the EPSRC programme grant POETS (EP/N031768/1) in partnership with e-Therapeutics, an Oxford based drug discovery company. e-Therapeutics is pioneering a novel form of drug discovery based on analyzing protein interactome networks (https://www.youtube.com/watch?v=wQFp7muuggA). This approach can discover suitable drug candidates much more efficiently compared to wet lab testing but requires considerable computing power, particularly because commodity computers are generally inefficient at analyzing large-scale networks. The presented accelerator, consisting of an FPGA board with a silicon-mapped protein interactome plus accompanying software frameworks and tools, can deliver a 1000x speed up in this application compared to software running on commodity computers. We will showcase demos in which we run in-silico analysis of protein interactomes to test drug effects and visualize the results in real-time.

More information ...

CIJTAG: CONCURRENT IJTAG DEMONSTRATOR

Author:
Krenz-Baath René, Hamm-Lippstadt University of Applied Sciences, DE

Abstract
The flexibility of on-chip instrument access enabled by IEEE 1687 (IJTAG) has shown tremendous improvements in modern industrial designs. Due to a constantly increasing spectrum of tasks performed through 1687 networks such as performing test operations during production test, on-line test operations as well as operating health monitors the test requirements in modern designs increase dramatically with respect to test performance, responsiveness and low power. These requirements have a major impact on the design of such test infrastructures. In complex designs with large test infrastructures it might be challenging to comply with the large spectrum of requirements. Concurrent IJTAG is a novel partitioning concept to a reconfigurable test infrastructure in order to enable an independent operation of different sections of the test infrastructure. The proposed demonstrator shows the first FPGA-based implementation of concurrent IJTAG test infrastructures.

More information ...

IP1 Interactive Presentations

Date: Tuesday, March 20, 2018
Time: 16:00 - 16:30
Location / Room: Conference Level, Foyer

Interactive Presentations run simultaneously during a 30-minute slot. Additionally, each IP paper is briefly introduced in a one-minute presentation in a corresponding regular session.
SPARSEN: AN ENERGY-EFFICIENT NEURAL NETWORK ACCELERATOR EXPLOITING INPUT AND OUTPUT SPARSITY

Speaker: Jingyang Zhu, Hong Kong University of Science and Technology, HK
Authors: Jingyang Zhu, Jibing Jiang, Xiu Chen and Chi-Ying Tsui, Hong Kong University of Science and Technology, HK
Abstract: The large computational complexity poses a challenge to the hardware design. In this work, we leverage the intrinsic activation sparsity of DNN to substantially reduce the execution cycles and the energy consumption. An end-to-end training algorithm is proposed to develop a lightweight (less than 5% overhead) run-time predictor for the output activation sparsity on the fly. Furthermore, an energy-efficient hardware architecture, SPARSEN, is proposed to exploit both the input and output sparsity. SPARSEN is scalable architecture with distributed memories and processing elements connected through a dedicated on-chip network. Compared with the state-of-the-art accelerators which only exploit the input sparsity, SPARSEN can achieve a 10%-70% improvement in throughput and a power reduction of around 50%.

ACCLIB: ACCELERATORS AS LIBRARIES

Speaker: Jacob R. Stevens, Purdue University, US
Authors: Jacob Stevens1, Yue Du2, Vivek Kozhikkot2 and Anand Raghunathan1
1Purdue University, US; 2IBM, US; T*nel Corporation, US
Abstract: Accelerator-based computing, which has been a mainstay of System-on-Chips (SoC) is of growing interest to a wider range of computing systems. However, the significant design effort required to develop a computational target for acceleration, design a hardware accelerator, verify the correctness of the accelerator, integrate the accelerator into the system, and write applications to use the accelerator, is a major bottleneck to the widespread adoption of accelerator-based computing. The classical approach to this problem is based on top-down methodologies such as automatic HW/SW partitioning and high-level synthesis (HLS). While HLS has advanced significantly and is seeing increased adoption, it does not leverage the ability of experienced human designers to craft highly optimized RTL. Our approach is to develop a framework that allows software developers to utilize existing libraries of pre-designed hardware accelerators automatically with no prior knowledge of the function of the accelerators, with minimal knowledge of hardware design, and with minimal design effort. To accomplish this, ACCLIB uses formal verification techniques to match a target software function with a functionally equivalent accelerator from a library of accelerators. It also generates the required HW/SW interfaces as well as the code necessary to offload the computation to the accelerator. We validate ACCLIB by applying it to accelerate six different applications using a library of hardware accelerators in just over one hour per application, demonstrating that the proposed approach has the potential to lower the barrier to adoption of accelerator-based computing.

HPXA: A HIGHLY PARALLEL XML PARSER

Speaker: Smruti Sarangi, IIT Delhi, IN
Authors: Israr Ahmad, Sanjog Patil and Smruti R. Sarangi, IIT Delhi, IN
Abstract: We present HPXA, which reads and processes 16 bytes at a time. We designed our component to process XML data at 106 Gbps, which is roughly 6.5X faster than competing prior work.

QOR-AWARE POWER CAPping FOR APPROXIMATE BIG DATA PROCESSING

Speaker: Sherief Reda, Brown University, US
Authors: Seyed Morteza Nabavinejad1, Xin Zhan2, Reza Azimi, Maziar Goudarzi1 and Sherief Reda2
1Sharif University of Technology, IR; 2Brown University, US
Abstract: We propose a new parser, HPXA, which reads and processes 16 bytes at a time. We designed our component to process XML data at 106 Gbps, which is roughly 6.5X faster than competing prior work.

EXACT MULTI-OBJECTIVE DESIGN SPACE EXPLORATION USING ASPM

Speaker: Kai Neubauer, University of Rostock, DE
Authors: Kai Neubauer1, Philipp Wanka2, Torsten Schaub3 and Christian Haubelt1
1University of Rostock, DE; 2University of Potsdam, DE
Abstract: Exact Design Space Exploration (DSE) is imperative for the design of modern, highly complex embedded systems in order to steer the development towards optimal design points. The early evaluation of design decisions at system-level abstraction layer helps to find promising regions for subsequent development steps in lower abstraction levels by diminishing the complexity of the search problem. In recent works, symbolic techniques, especially Answer Set Programming (ASP) modulo Theories (ASPmT), have been shown to find feasible solutions of highly complex system-level synthesis problems with non-linear constraints very efficiently. In this paper, we present a novel approach to a holistic system-level DSE based on ASPmT. To this end, we include additional background theories that concurrently guarantee compliance with hard constraints and perform the simultaneous optimization of several design objectives. We implement and compare our approach with a state-of-the-art preference handling framework for ASP. Experimental results indicate that our proposed method produces better solutions with respect to both diversity and convergence to the true Pareto front.
HIPE: HMC INSTRUCTION PREDICATION EXTENSION APPLIED ON DATABASE PROCESSING

Speaker: Diego Tomé, Centrum Wiskunde & Informatica (CWI), BR
Authors: Diego Gomes Tomé1, Paulo Cesar Santos2, Luigi Camp2, Eduardo Cunha de Almeida2 and Marco Antonio Zanata Alves2
1Federal University of Paraná, BR; 2UFPR, BR

Abstract
The recent Hybrid Memory Cube (HMC) is a smart memory which includes functional units inside one logic layer of the 3D stacked memory design. In order to execute instructions inside the Hybrid Memory Cube (HM), the processor needs to send instructions to be executed near data, keeping most of the pipeline complexity inside the processor. Thus, control-flow and data-flow dependences are all managed inside the processor, in such way that only update instructions are supported by the HMC. In order to solve data-flow dependences inside the memory, previous work proposed HMC Instruction Vector Extensions (HIVE), which embeds a high number of functional units with an interblock register bank. In this work, we propose HMC Instruction Prediction Extensions (HIPE), that supports predicated execution inside the memory, in order to transform control-flow dependences into data-flow dependences. Our mechanism focuses on removing the high latency iteration between the processor and the smart memory during the execution of branches that depends on data processed inside the memory. In this paper, we evaluate a balanced design of HIVE comparing it with HMC executions. After we show the HIPE mechanism results when executing a database workload, which is a strong candidate to use smart memories. We show interesting trade-offs of performance when comparing our mechanism to previous work.

Download Paper (PDF; Only available from the DATE venue WiFi)

PARAMETRIC FAILURE MODELING AND YIELD ANALYSIS FOR STT-MRAM

Speaker: Sarath Mohanachandran Nair, Karlsruhe Institute of Technology, DE
Authors: Sarath Mohanachandran Nair, Rajendra Bishnoi and Mehdi Tahoori, Karlsruhe Institute of Technology, DE

Abstract
The emerging Spin Transfer Torque Magnetic Random Access Memory (STT-MRAM) is a promising candidate to replace conventional on-chip memory technologies due to its advantages such as non-volatility, high density, scalability and unlimited endurance. However, as the technology scales, yield loss due to extreme parametric variations is becoming a major challenge for STT-MRAM because of its higher sensitivity to process variations as compared to CMOS memories. In addition, the parametric variations in STT-MRAM exacerbate its stochastic switching behavior, leading to both test time fails and reliability failures in the field. Since an STT-MRAM memory array consists of both CMOS and magnetic components, it is important to consider variations in both these components to obtain the failures at the system level. In this work, we model the parametric failures of STT-MRAM at the system level considering the correlation among bit-cells as well as the impact of peripheral components. The proposed approach provides realistic fault distribution maps and equip the designer to investigate the efficacy of different combinations of defect tolerance techniques for an effective design-for-yield exploration.

Download Paper (PDF; Only available from the DATE venue WiFi)

AN EFFICIENT RESOURCE-OPTIMIZED LEARNING PREFETCHER FOR SOLID STATE DRIVES

Speaker: Rui Xu, University of Science and Technology of China, CN
Authors: Rui Xu, Xi Jin, Linfeng Tao, Shuazhi Guo, Zikun Xiang and Teng Tian, Strongly-Coupled Quantum Matter Physics, Chinese Academy of Sciences, School of Physical Sciences, University of Science and Technology of China, Hefei, Anhui, China, CN

Abstract
In recent years, solid-state drives (SSDs) have been widely deployed in modern storage systems. To increase the performance of SSDs, prefetchers for SSDs have been designed both at operating system (OS) layer and flash translation layer (FTL). Prefetchers in FTL have many advantages like OS-independence, easy-using, and compatibility. However, due to the limitation of computing capabilities and memory resources, existing prefetchers in FTL merely employ simple sequential prefetching which may incur high penalty cost for I/O access stream with complex patterns. In this paper, an efficient learning prefetcher implemented in FTL is proposed. Considering the resource limitation of SSDs, a learning algorithm based on Markov chains is employed and optimized so that high hit ratio and low penalty cost can be achieved even for complex access patterns. To validate our design, a simulator with the prefetcher is designed and implemented based on Flashsim. The TPC-H benchmark and an application launch trace are tested on the simulator. According to experimental results of the TPC-H benchmark, more than 90% of memory cost can be saved in comparison with a previous design at OS layer. The hit ratio can be increased by 24.1% and the number of times of misprefetching can be reduced by 95.8% in comparison with the simple sequential prefetching strategy.

Download Paper (PDF; Only available from the DATE venue WiFi)

BRIDGING DISCRETE AND CONTINUOUS TIME MODELS WITH ATOMS

Speaker: George Ungureanu, KTH Royal Institute of Technology, SE
Authors: George Ungureanu1, José E. G. de Medeiros2 and Ingo Sander3
1KTH Royal Institute of Technology, SE; 2University of Brasilia, BR

Abstract
Recent trends in replacing traditionally digital components with analog counterparts in order to overcome physical limitations have led to an increasing need for rigorous modeling and simulation of hybrid systems. Combining the two domains under the same set of semantics is not straightforward and often leads to chaotic and non-deterministic behavior due to the lack of a common understanding of aspects concerning time. We propose an algebra of primitive interactions between continuous and discrete aspects of systems which enables their description within two orthogonal layers of computation. We show its benefits from the perspective of modeling and simulation, through the example of an RC oscillator modeled in a formal framework implementing this algebra.

Download Paper (PDF; Only available from the DATE venue WiFi)
IP1-12

**Abstract**
Virtual platform (VP) technology is an established enabler of embedded system design. However, the sheer number of CPU models and VPs in modern multi-core systems forms a performance bottleneck. Hybrid simulation addresses this issue by executing parts of the embedded software stack on the host. Although the approach is significantly faster, hybridization can not cope with higher software layers, e.g., Operating Systems (OSs). Thus, this paper presents the OS-aware Host EXTension (OXEH) framework to accelerate VPs while expanding the applicability of hybridization. OXEH is evaluated on various system layers, yielding speedups between 2.99x-21.14x with specific benchmarks.

**Authors**
Róbert Lajos Bücs1, Maximilian Frick2, Rainer Leupers3, Gerd Ascheid1, Stephan Tobies1 and Andreas Hoffmann2
1Institute for Communication Technologies and Embedded Systems, RWTH Aachen University, DE; 2Synopsys GmbH, DE

**Speaker**
Róbert Lajos Bücs

**Download Paper (PDF; Only available from the DATE venue WiFi)**

---

IP1-13

**Abstract**
An effective full-system virtual prototype is critical for early-stage systems design exploration. Generally, however, traditional acceleration approaches of virtual prototypes cannot accurately analyze system performance and model non-deterministic inter-component interactions due to the unpredictability of simulation progress. In this paper, we propose an effective virtualization-assisted approach for modeling and performance analysis. First, we develop a deterministic synchronization process that manages the interactions affecting the data dependency in chronological order to model inter-component interactions consistently. Next, we create accurate timing and bus contention models based on runtime operation statistics for analyzing system performance. We implement the proposed virtualization-assisted approach on an off-the-shelf System-on-Chip (SoC) board to demonstrate the effectiveness of our idea. The experimental results show that the proposed approach runs 12-77 times faster than a commercial virtual prototyping tool and performance estimation is only 3-6% apart from real systems.

**Authors**
Hein-I Wu, National Tsing Hua University, Department of Computer Science, Hsinchu, Taiwan, TW

**Speaker**
Hein-I Wu

**Download Paper (PDF; Only available from the DATE venue WiFi)**

---

IP1-14

**Abstract**
Adaptive voltage scaling (AVS) has been used widely to compensate for process, voltage, and temperature variations as well as for power optimization of integrated circuits. The current industrial state-of-the-art AVS approaches using Process Monitor Boxes (PMBs) have shown several limitations such as huge characterization effort, which makes these approaches very expensive, and a low accuracy that results in extra margins, which consequently lead to yield loss and performance limitations. To overcome those limitations, in this paper we propose an alternative solution using transition fault test patterns, which is able to eliminate the need for PMBs, while improving the accuracy of voltage estimation. The paper shows, using simulation of ISCAS'99 benchmarks with 28nm FD-SOI library, that AVS using transition fault testing (TF-based AVS) results in an error as low as 5.33%. The paper also shows that the PMB approach can only account for 85% of the uncertainty in voltage measurements, which results in power waste, while the TF-based approach can account for 99% of that uncertainty.

**Authors**
Mahroo Zandrashimi1, TU Delft, NL

**Speaker**
Mahroo Zandrashimi

**Download Paper (PDF; Only available from the DATE venue WiFi)**

---

IP1-15

**Abstract**
DRAM technology is scaling aggressively that results in high leakage power, worse data retention time behavior, and large process variations. Due to these process variations, vendors provide large guard bands on various DRAM currents and timing specifications that are over pessimistic. Detailed knowledge on the DRAM retention behavior and currents for the average case allow to improve memory system performance and energy efficiency of specific applications by moving away from worst case behavior. In this paper, we present an advanced measurement platform to investigate off-the-shelf DDR4 DRAMs' retention behavior, and to precisely measure various DRAM currents (IDDs and IPPs) at a wide range of operating temperatures. Error Checking and Correction (ECC) schemes are popular in correcting randomly scattered single bit errors. Since retention failures also occur randomly, ECCs can be used to improve DRAM retention behavior. Therefore, for the first time, we show the influence of ECC on the retention behavior of recent DDR4 DRAMs, and how it varies across various DRAM architectures considering detailed structure of the DRAM (true-cell devices / mixed-cell devices).

**Authors**
Deepak M. Mathew, University of Kaiserslautern, DE

**Speaker**
Deepak M. Mathew

**Download Paper (PDF; Only available from the DATE venue WiFi)**

---

IP1-16

**Abstract**
Emerging nanotechnologies such as ambipolar carbon nanotube field effect transistors (CNTFETs) and silicon nanowire FETs (SiNFETs) provide ambipolar devices allowing the design of more complex logic primitives than those found in today's typical CMOS libraries. When switching, such devices show a behavior not seen in simpler CMOS and FinFET cells, making unsuitable the existing delay fault testing approaches. We provide a Boolean model of switching ambipolar devices to support delay fault testing of logic cells based on such devices both in Boolean and Pseudo-Boolean satisfiability engines.

**Authors**
Marcello Dalpasso1, Davide Bertozzi2 and Michele Favalli2
1DEI - UNIV. of Padova, IT; 2DE - Univ. of Ferrara, IT

**Speaker**
Davide Bertozzi

**Download Paper (PDF; Only available from the DATE venue WiFi)**
4.1 Executive Session: Exact Synthesis and SAT

Date: Tuesday, March 20, 2018
Time: 17:00 - 18:30
Location / Room: Saal 2

Chair: Patrick Vuillod, Synopsys, FR, Contact Patrick Vuillod
Co-Chair: Amaru Luca, Synopsys, US, Contact Luca Amaru

Exact synthesis and SAT-based methods open new opportunities in design automation flows, where attaining the best possible logic implementation is key. This executive session covers recent advances on these two topics, which are tightly related, from both academic and industrial standpoints. The first paper shows how to find optimal circuit, on small number of variables, using SAT-solvers. The frontiers of circuits achievable by this method are discussed, together with known open problems. The second paper presents recent advancements on exact synthesis, with focus on implicit enumeration methods. Improvements on the SAT-formulation are delineated, which enable complex constraints to be considered while solving exact synthesis. The third paper introduces a redundancy removal engine based on SAT. Its integration in a commercial EDA tool is described, detailing challenges and opportunities arising in an industrial synthesis environment.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:00</td>
<td>4.1.1</td>
<td>IMPROVING CIRCUIT SIZE UPPER BOUNDS USING SAT-SOLVERS</td>
<td>Alexander Kulikov, Steklov Mathematical Institute at St. Petersburg, RU</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker and Author:</td>
<td>Alexander Kulikov</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>Boolean circuits is arguably the most natural model for computing Boolean functions. Despite intensive research, for many functions, we still do not know what optimal circuits look like. In this paper, we discuss how SAT-solvers can be used for constructing optimal circuits for functions on moderate number of variables. We first discuss why this problem is important and then indicate the current frontiers: what can and cannot be found by state-of-the-art SAT-solvers, and for what functions we are interested in finding efficient circuits.</td>
</tr>
<tr>
<td></td>
<td>17:30</td>
<td>4.1.2 PRACTICAL EXACT SYNTHESIS</td>
<td>Winston Haaswijk, EPFL, CH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Winston Haaswijk</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>Winston Haaswijk, EPFL, CH, Eleonora Testa, Alan Mishchenko, Luca G. Amaru, Robert K. Brayton, Giovanni De Micheli, EPFL, CH, University of California, Berkeley, US, Synopsys Inc., US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>In this paper, we discuss recent advances in exact synthesis, considering both their efficient implementation and various applications in which they can be employed. We emphasize on solving exact synthesis through Boolean satisfiability (SAT) encodings. Different SAT encodings for exact synthesis are compared, and examined the applications to multi-level logic synthesis, in both area and depth optimization. Another application of SAT based exact synthesis is optimization under many constraints. These constraints can, e.g., be a fixed fanout or delay constraints. Finally, we end our discussion by proposing directions for future research in exact synthesis.</td>
</tr>
<tr>
<td>18:00</td>
<td>4.1.3</td>
<td>SAT-BASED REDUNDANCY REMOVAL</td>
<td>Kishanu Debnath, Synopsys, IN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker and Author:</td>
<td>Kishanu Debnath</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>Logic optimization is an integral part of digital circuit design. It reduces design area and power consumption, and quite often improves circuit delay as well. Redundancy removal is a key step in logic optimization, in which redundant connections in the circuit are determined and replaced by constant values 0 or 1. The resulting circuit is simplified, resulting in area and power savings. In this paper, we describe a redundancy removal approach for combinational circuits based on a combination of logic simulation and SAT. We show that this approach can handle large industrial strength designs in a reasonable amount of CPU time.</td>
</tr>
</tbody>
</table>

4.2 Domain Specific Design Methodologies

Date: Tuesday, March 20, 2018
Time: 17:00 - 18:30
Location / Room: Konf. 6

Chair: Frédéric Pétrot, Grenoble Institute of Technology, FR, Contact Frédéric Pétrot
Co-Chair: Lars Bauer, Kaiserslautern Institute of Technology, DE, Contact Lars Bauer

In the quest for high efficiency, design methodologies specialize to particular domains. At first, a case study for approximate computing in the field of biometric security is presented. The second talk proposes a framework that uses a genetic algorithm to find an optimal mapping of artificial neural networks onto GPU + multicore systems. Finally, a method is presented that controls by
introduces compile-time analysis to improve parallel SystemC simulation. Workloads from big-data applications. The second paper minimizes the energy consumption by reducing accesses to off-chip memory for convolutional neural networks (CNNs). The third paper demonstrated in the areas of machine learning, image processing, and computer vision. In this paper we make the case for a new direction for approximate computing in the field of biometric security with a comprehensive case study of iris scanning. We devise an end-to-end flow from an input camera to the final list encoding that produces sufficiently accurate final results despite relying on intermediate approximate computational steps. Unlike previous methods which evaluated approximate computing techniques on individual algorithms, our flow consists of a complex SW/HW pipeline of four major algorithms that eventually compute the list encoding from input live camera feeds. In our flow, we identify overall eight approximation knobs at both the algorithmic and hardware levels to trade-off accuracy with runtime. To identify the optimal values for these knobs, we devise a novel design space exploration technique based on reinforcement learning with a recurrent neural network agent. Finally, we fully implement and test our proposed methodologies using both benchmark dataset images and live images from a camera using an FPGA-based SoC. We show that we are able to reduce the runtime of the system by 48% on top of an already HW accelerated design, while meeting industry-standard accuracy requirements for iris scanning systems.

**AN EFFICIENT RESOURCE-OPTIMIZED LEARNING PREFETCHER FOR SOLID STATE DRIVES**

Speaker: Rui Xu, University of Science and Technology of China, China

Authors: Rui Xu, Xi Jin, Linteng Tao, Shuaizhi Guo, Zikun Xiang and Teng Tian, Strongly-Coupled Quantum Matter Physics, Chinese Academy of Sciences, School of Physical Sciences, University of Science and Technology of China, HeFei, Anhui, China, CN

Abstract: In recent years, solid state drives (SSDs) have been widely deployed in modern storage systems. To increase the performance of SSDs, prefetchers for SSDs have been designed both at operating system (OS) layer and flash translation layer (FTL). Prefetchers in FTL have many advantages like OS-independence, easy-using, and compatibility. However, due to the limitation of computing capabilities and memory resources, existing prefetchers in FTL merely employ simple sequential prefetching which may incur high penalty cost for I/O access stream with complex patterns. In this paper, an efficient learning prefetcher implemented in FTL is proposed. Considering the resource limitation of SSDs, a learning algorithm based on Markov chains is employed and optimized so that high hit ratio and low penalty cost can be achieved even for complex access patterns. To validate our design, a simulator with the prefetcher is designed and implemented based on Flashsim. The TPC-H benchmark and an application launch trace are tested on the simulator. According to experimental results of the TPC-H benchmark, more than 90% of memory cost can be saved in comparison with a previous design at OS layer. The hit ratio can be increased by 24.1% and the number of times of misprefetching can be reduced by 95.8% in comparison with the simple sequential prefetching strategy.

**FLASH READ DISTURB MANAGEMENT USING ADAPTIVE CELL BIT-DENSITY WITH IN-PLACE REPROGRAMMING**

Speaker: Tai-Chou Wu, Yu-Ping Ma and Li-Pin Chang, National Chiao-Tung University, TW

Authors: Tai-Chou Wu, Yu-Ping Ma and Li-Pin Chang, National Chiao-Tung University, TW

Abstract: Read disturbance is a circuit-level noise induced by flash read operations. Read refreshing employs data migration to prevent read disturbance from corrupting useful data. However, it costs frequent block erasure under read-intensive workloads. Inspired by software-controlled cell bit-density, we propose to reserve selected threshold voltage levels as guard levels to extend the tolerance of read disturbance. Blocks with guard levels have a low cell bit-density, but they can store frequently read data without frequent read refreshing. We further propose to convert a high-density block into a low-density one using in-place reprogramming to reduce the need for data migration. Our approach reduced the number of blocks erased due to read refreshing by up to 85% and the average read response time by up to 22%.

**HTF-MPR: A HETEROGENEOUS TENSORFLOW MAPPER TARGETING PERFORMANCE USING GENETIC ALGORITHMS AND GRADIENT BOOSTING REGRESSORS**

Speaker: Nader Bagherzadeh, University of California, Irvine, US

Authors: Ahmad Albaqami, Maryam S. Hosseini and Nader Bagherzadeh, University of California, Irvine, US

Abstract: TensorFlow is a library developed by Google to implement Artificial Neural Networks using computational dataflow graphs. The neural network has many iterations during training. A distributed, parallel environment is ideal to speedup learning. Parallelism requires proper mapping of devices to Tensorflow operations. We developed HTF-MPR framework for that reason. HTF-MPR utilizes a genetic algorithm approach to search for the best mapping that outperforms the default Tensorflow mapper. By using Gradient Boosting Regressors to create a fitness predictive model, the search space is expanded which increases the chances of finding a solution mapping. Our results on well-known neural network benchmarks, such as ALEXNET, MNIST softmax classifier, and VGG-16, show an overall speedup in the training stage by 1.18, 3.33, and 1.13, respectively.

**HTF-MPR: A HETEROGENEOUS TENSORFLOW MAPPER TARGETING PERFORMANCE USING GENETIC ALGORITHMS AND GRADIENT BOOSTING REGRESSORS**

Speaker: Soheil Hashemi, Hokchhay Tann, Francesco Buttafuoco and Sherief Reda, Brown University, US

Authors: Soheil Hashemi, Hokchhay Tann, Francesco Buttafuoco and Sherief Reda, Brown University, US

Abstract: TensorFlow is a library developed by Google to implement Artificial Neural Networks using computational dataflow graphs. The neural network has many iterations during training. A distributed, parallel environment is ideal to speedup learning. Parallelism requires proper mapping of devices to Tensorflow operations. We developed HTF-MPR framework for that reason. HTF-MPR utilizes a genetic algorithm approach to search for the best mapping that outperforms the default Tensorflow mapper. By using Gradient Boosting Regressors to create a fitness predictive model, the search space is expanded which increases the chances of finding a solution mapping. Our results on well-known neural network benchmarks, such as ALEXNET, MNIST softmax classifier, and VGG-16, show an overall speedup in the training stage by 1.18, 3.33, and 1.13, respectively.

**IP1-10-363**

**AN EFFICIENT RESOURCE-OPTIMIZED LEARNING PREFETCHER FOR SOLID STATE DRIVES**

Speaker: Rui Xu, University of Science and Technology of China, China

Authors: Rui Xu, Xi Jin, Linteng Tao, Shuaizhi Guo, Zikun Xiang and Teng Tian, Strongly-Coupled Quantum Matter Physics, Chinese Academy of Sciences, School of Physical Sciences, University of Science and Technology of China, HeFei, Anhui, China, CN

Abstract: In recent years, solid state drives (SSDs) have been widely deployed in modern storage systems. To increase the performance of SSDs, prefetchers for SSDs have been designed both at operating system (OS) layer and flash translation layer (FTL). Prefetchers in FTL have many advantages like OS-independence, easy-using, and compatibility. However, due to the limitation of computing capabilities and memory resources, existing prefetchers in FTL merely employ simple sequential prefetching which may incur high penalty cost for I/O access stream with complex patterns. In this paper, an efficient learning prefetcher implemented in FTL is proposed. Considering the resource limitation of SSDs, a learning algorithm based on Markov chains is employed and optimized so that high hit ratio and low penalty cost can be achieved even for complex access patterns. To validate our design, a simulator with the prefetcher is designed and implemented based on Flashsim. The TPC-H benchmark and an application launch trace are tested on the simulator. According to experimental results of the TPC-H benchmark, more than 90% of memory cost can be saved in comparison with a previous design at OS layer. The hit ratio can be increased by 24.1% and the number of times of misprefetching can be reduced by 95.8% in comparison with the simple sequential prefetching strategy.

**4.3 System Modelling for Simulation and Optimisation**

**End of session**

**Exhibition Reception** in Exhibition Area

The Exhibition Reception will take place on Tuesday in the exhibition area, where free drinks for all conference delegates and exhibition visitors will be offered. All exhibitors are welcome to also provide drinks and snacks for the attendees.
18:00 4.3.1  CAMP: ACCURATE MODELING OF CORE AND MEMORY LOCALITY FOR PROXY GENERATION OF BIG-DATA APPLICATIONS
Speaker: Andreas Gerstlauer, University of Texas at Austin, US
Authors: Reena Panda, Xinmin Zheng, Andreas Gerstlauer and Lizy John, The University of Texas at Austin, US
Abstract
Fast and accurate design-space exploration is a critical requirement for enabling future hardware designs. However, big data applications are often complex targets to evaluate on early performance models (e.g., simulators or RTL models) owing to their complex software-stacks, significantly long run times, system dependencies and the limited speed of performance models. To overcome the challenges in benchmarking complex big data applications, in this paper, we propose a proxy generation methodology, CAMP that can generate miniature proxy benchmarks, which are representative of the performance of big data applications and yet converge to results quickly without needing any complex software stack support. Prior system-level proxy generation techniques model core locally features in detail, but abstract out memory locally modeling using simple stride-based models, which results in poor cloning accuracy for most applications. CAMP accurately models both core-performance and memory locality, along with modeling the feedback loop between the two. CAMP replicates core performance by modeling the dependencies between instructions, instruction types, control-flow behavior, etc. CAMP also adds a memory locally profiling approach that captures spatial and temporal locality of applications. Finally, we propose a novel proxy replay methodology that integrates the core and memory locality models to create accurate system-level proxy benchmarks. We demonstrate that CAMP proxies can mimic the original application's performance behavior and that they can capture the performance feedback loop well. For a variety of real-world big-data applications, we show that CAMP achieves an average cloning accuracy of 89%. We believe this is a new capability that can facilitate for overall system (core and memory subsystem) design exploration.
Download Paper (PDF; Only available from the DATE venue WiFi)

18:30 4.3.2  SMARTSHUTTLE: OPTIMIZING OFF-CHIP MEMORY ACCESSES FOR DEEP LEARNING ACCELERATORS
Speaker: Guihui Yan, Institute of Computing Technology, Chinese Academy of Sciences, CN
Authors: Jiajun Li, Guihui Yan, Wenyuan Lu, Shuhao Jiang, Shijun Gong, Jingya Wu and Xiaowei Li, Institute of Computing Technology, Chinese Academy of Sciences, CN
Abstract
Convolutional Neural Network (CNN) accelerators are rapidly growing in popularity as a promising solution for deep learning based applications. Though optimizations on computation have been intensively studied, the energy efficiency of such accelerators remains limited by off-chip memory accesses since their energy cost is magnitudes higher than other operations. Minimizing off-chip memory access volume, therefore, is the key to higher energy efficiency. However, there exists a dilemma of minimizing the access of which data types. We observed that sticking to minimizing the access of one data type cannot fit the varying shapes of convolutional layers in CNNs. To overcome this problem, this paper proposed a adaptive layer partitioning and scheduling scheme, called SmartShuttle, which can adaptively switch among the specific data reuse oriented scheduling schemes and the corresponding layer partitioning schemes to dynamically match different shapes of convolutional layers. Specifically, SmartShuttle takes both data reusability and sparsity into account since they have significant impact on the memory access volume. The experimental results show that SmartShuttle achieves a performance at 434.8 multiply and accumulations (MACs)/DRAM access for VGG-16, and 526.3 MACs/DRAM access for AlexNet, which outperforms the state-of-the-art approach (Eyeriss) by 52.2% and 52.6%, respectively.
Download Paper (PDF; Only available from the DATE venue WiFi)

18:00 4.3.3  PORT CALL PATH SENSITIVE CONFLICT ANALYSIS FOR INSTANCE-AWARE PARALLEL SYSTEMC SIMULATION
Speaker: Tim Schmidt, Student, US
Authors: Tim Schmidt, Zhongqi Cheng and Rainer Doemer, University of California, Irvine, US
Abstract
Many SystemC approaches expect a thread safe and conflict free model from the designer. Alternatively, an advanced compiler can identify and avoid possible parallel access conflicts. While manual conflict resolution can theoretically be more precise, it is impractical for real-world applications because of the inherent complexities. Here automatic compiler-based analysis is preferred which provides conservative conflict avoidance with minimal false positives. This paper introduces a novel compiler technique called port call path analysis that greatly reduces the amount of false positive conflicts resulting in significantly increased simulation speed. Experimental results show that the new analysis reduces the amount of false conflicts by up to 98% and, on a 4-core processor, speeds up the simulation up to 3x for a NoC particle simulator and 3.5x for a bitcoin miner SystemC model.
Download Paper (PDF; Only available from the DATE venue WiFi)

18:30 IP-1
11, 142  BRIDGING DISCRETE AND CONTINUOUS TIME MODELS WITH ATOMS
Speaker: George Ungureanu, KTH Royal Institute of Technology, SE
Authors: George Ungureanu1, José E. G. de Medeiros2 and Ingo Sander1
1KTH Royal Institute of Technology, SE; 2University of Brasilia, BR
Abstract
Recent trends in replacing traditionally digital components with analog counterparts in order to overcome physical limitations have led to an increasing need for rigorous modeling and simulation of hybrid systems. Combining the two domains under the same set of semantic is not straightforward and often leads to chaotic and non-deterministic behavior due to the lack of a common understanding of aspects concerning time. We propose an algebra of primitive interactions between continuous and discrete aspects of systems which enables their description within two orthogonal layers of computation. We show its benefits from the perspective of modeling and simulation, through the example of an RC oscillator modeled in a formal framework implementing this algebra.
Download Paper (PDF; Only available from the DATE venue WiFi)

18:31 IP-1
12, 436  OHEX: OS-AWARE HYBRIDIZATION TECHNIQUES FOR ACCELERATING MPSOC FULL-SYSTEM SIMULATION
Speaker: Robert Lajos Bücs, Institute for Communication Technologies and Embedded Systems, RWTH Aachen University, DE
Authors: Robert Lajos Bücs1, Maximilian Fricke2, Rainer Leupers1, Gerd Asche1, Stephan Tobias2 and Andreas Hoffmann2
1Institute for Communication Technologies and Embedded Systems, RWTH Aachen University, DE; 2Synopsys GmbH, DE
Abstract
Virtual platform (VP) technology is an established enabler of embedded system design. However, the sheer number of CPU models in modern multi-core VPs forms a performance bottleneck. Hybrid simulation addresses this issue by executing parts of the embedded software stack on the host. Although the approach is significantly faster, hybridization can not cope with higher software layers, e.g., Operating Systems (OS). Thus, this paper presents the OS-aware Host EXTension (OHEX) framework to accelerate VPs while expanding the applicability of hybridization. OHEX is evaluated on various system layers, yielding speedups between 2.99x-21.14x with specific benchmarks.
Download Paper (PDF; Only available from the DATE venue WiFi)
### 4.4 Overcoming the Limitations of Worst-Case IC Design

**Date:** Tuesday, March 20, 2018  
**Time:** 17:00 - 18:30  
**Location / Room:** Konf. 2

**Co-Chair:**  
Vasilis Pavlidis, University of Manchester, GB.  
Contact: Vasilis Pavlidis

The session illustrates novel approaches to lower the high voltage and timing guard-bands affecting the performance of computing systems. The first talk introduces a methodology to increase the resiliency towards timing errors at ultra-low-voltages. Then, the placement of the timing monitor infrastructure is investigated in the second talk. The illustration of a mechanism to reliably tune the core. The key contribution is to perform auto-tuning of the coefficients of the feedback loop of the IVR based on the performance of the digital cores. Simulations using an h-f IVR Simulink model and digital logic in 45nm CMOS process shows that the proposed performance driven auto-tuning demonstrates potential for up to 12% increase in system performance under inductance and threshold variation.

**TRIDENT: A COMPREHENSIVE TIMING ERROR RESILIENT TECHNIQUE AGAINST CHOKE POINTS AT NTC**  
**Speaker:** Aatreyi Bal, Utah State University, US  
**Authors:** Aatreyi Bal, Sanghamitra Roy and Koushik Chakraborty, Utah State University, US  
**Abstract**  
Near Threshold Computing (NTC) systems have been inherently plagued with heightened process variation (PV) sensitivity. Choke points are an intriguing manifestation of this PV sensitivity. In this paper, we explore the probability of minimum timing violations, caused by choke points, in an NTC system and, their non-trivial impacts on the system reliability. We show that conventional timing error mitigation techniques are inefficient in tackling choke point induced minimum timing violations. Consequently, we propose a comprehensive error mitigation technique, Trident, to tackle choke points, at NTC. Trident offers a 1.37× performance improvement and a 1.1× energy efficiency gain over Razor at NTC, with minimal overheads.

Download Paper (PDF; Only available from the DATE venue WiFi)

**BAYESIAN THEORY BASED SWITCHING PROBABILITY CALCULATION METHOD OF CRITICAL TIMING PATH FOR ON-CHIP TIMING SLACK MONITORING**  
**Speaker:** Byung Su Kim, Samsung Electronics, Foundry, KR  
**Authors:** Byung Su Kim and Joon-Sung Yang.  
1Samsung Electronics, KR, 2Sungkyunkwan University, KR  
**Abstract**  
Accurate in-situ monitoring is urgently required for an adaptive performance control system and post silicon validation. For accurate in-situ monitoring, a direct probing method is presented in which monitors directly measure a path delay from real critical timing paths. However, we may not be able to predict when the timing slack monitors would activate since the activation depends on a design structure and input patterns. If a timing slack monitor is rarely activated by timing critical paths, the observability from this monitor would be low and the monitor possibly can be discarded. For this reason, we propose a novel timing slack monitoring methodology based on switching probability of timing critical paths. Switching probability and correlation on critical timing paths are formulated, and the proposed method finds a list of critical path endpoints for the timing slack monitor insertion under given power and area constraints. Experimental results with ISCAS'89 circuits show that, compared to the method which places monitors for all worst critical paths, 16.67 ~ 97.2% of timing slack monitors are removed and 32.56 ~ 96.88% of dynamic power reduction from the monitors is achieved by the proposed method.

Download Paper (PDF; Only available from the DATE venue WiFi)

**PERFORMANCE BASED TUNING OF AN INDUCTIVE INTEGRATED VOLTAGE REGULATOR DRIVING A DIGITAL CORE AGAINST PROCESS AND PASSIVE VARIATIONS**  
**Speaker:** Venkata Chaitanya Krishna Chekuri, Georgia Institute of Technology, US  
**Authors:** Venkata Chaitanya Krishna Chekuri, Monodeep Kar, Arvind Singh and Sibaib Mukhopadhyay, Georgia Institute of Technology, US  
**Abstract**  
This paper presents an auto-tuning method for fully integrated voltage regulators (IVRs) driving digital cores against variations in passive as well as process/temperature of the core. The key contribution is to perform auto-tuning of the coefficients of the feedback loop of the IVR based on the performance of the digital core. Simulations using a high-frequency IVR Simulink model and digital logic in 45nm CMOS process shows that the proposed performance driven auto-tuning demonstrates potential for up to 12% increase in system performance under inductance and threshold variation.

Download Paper (PDF; Only available from the DATE venue WiFi)
**4.5 Test: innovative infrastructures and ATPG techniques**

**Date:** Tuesday, March 20, 2018  
**Time:** 17:00 - 18:00  
**Location / Room:** Konf. 3

**Chair:**  
Deepak Mathew, University of Kaiserslautern, DE

**Co-Chair:**  
Lukasz Rybak, Mentor Graphics Poland, PL

The session addresses hot challenges for 2.5D and 3D integration and asynchronous circuits, and introduces solutions for improving ATPG efficiency.

### 4.5.1 Pre-Assembly Testing of Interconnects in Embedded Multi-Die Interconnect Bridge (EMIB) Dies

**Speaker:** Krishnendu Chakrabarty, Duke University, US  
**Authors:** Sudipta Mondal and Krishnendu Chakrabarty, Duke University, US

**Abstract**  
The embedded multi-die interconnect bridge (EMIB) is an advanced packaging technology for 2.5D integration. This paper presents a bridge test architecture based on the proposed IEEE Std. P1838. The proposed test method enables access to interconnects at a pre-assembly stage by pairing the interconnects using metal shorts and probing on coarse-pitch C4 bumps. It can efficiently detect resistive-open and resistive-short defects in the bridge interconnects and micro-bumps. Simulation results are presented to evaluate the range of defects that can be detected by the proposed method.

Download Paper (PDF; Only available from the DATE venue WiFi)

### 4.5.2 On the Reuse of Timing Resilient Architecture for Testing Path Delay Faults in Critical Paths

**Speaker:** Luciano Ost, University of Leicester, GB  
**Authors:** Felipe Kuentzer, Leonardo Juracy and Alexandre Amory, PUCRS University, BR

**Abstract**  
Energy efficiency has become one of the most common and important demands for contemporary applications, increasing the desire for chips that operate near the threshold voltage levels, which unfortunately worsens the effects of process, voltage, and temperature (PVT) variability. An alternative solution to cope with PVT variations are the timing resilient architectures, such as the synchronous Razor family and the asynchronous Blade template, that rely on error-detection logic (EDL) to detect and recover from timing violations. On one hand, the use of timing resilient architectures makes the path delay testing more challenging because it is not a matter of simple pass or fails the test. On the other hand, we show that timing resilient architectures, such as Blade, present opportunities to design low-cost online delay testing of the critical paths. Results show the area overhead and fault coverage using functional testing on a 32-bit MIPS CPU and a crypto core.

Download Paper (PDF; Only available from the DATE venue WiFi)
CHARACTERIZATION OF POSSIBLY DETECTED FAULTS BY ACCURATELY COMPUTING THEIR DETECTION PROBABILITY

Speaker:
Jan Burchard, Mentor, a Siemens Business, DE

Authors:
Jan Burchard¹, Dominik Erb² and Bernd Becker²

¹University of Freiburg, DE; ²Rhinecon Technologies, DE

Abstract
With ever more complex and larger VLSI devices and higher and higher reliability requirements, high quality test with a large fault and defect coverage is becoming even more relevant. At the same time, when unspecified or unknown input values (X-values) have to be considered in a pattern, commercial ATPG tools are sometimes not capable of determining whether a fault can be tested - but there is at least a chance to detect the fault, as \(0/X\) or \(1/X\) could be propagated to at least one output. Consequently, these faults are considered to be possibly detected and often counted towards the overall fault coverage with a weighting factor. However, as the actual probability to detect these faults with the considered test pattern is not taken into account, this could lead to an over- or underestimation of their real fault coverage, falsifying the test results. We introduce a \(\delta\delta\delta\varepsilon\) based characterization algorithm for this class of faults. This new algorithm is, for the first time, able to accurately compute the detection probability for faults marked as possibly detected by state-of-the-art commercial tools. Our experimental results for the largest ITC'99 benchmarks as well as larger industrial circuits show that our algorithm can accurately determine the detection probability for most of the possibly detected faults and also identify faults that are completely untestable or found with a probability of 100% irrespective of the assignment of the inputs with an X-value. Furthermore, they show that the detection probability is circuit dependent and consequently should not just be estimated by a simple weighting factor but requires a more in-depth evaluation. Otherwise, there is a high risk that the achieved results could clearly be to optimistic or pessimistic with regard to the real fault coverage.

Download Paper (PDF; Only available from the DATE venue WiFi)
This session discusses reliability analysis and enhancement of memristive computing, addressing the non-linear behavior and the development of a logic synthesis flow for defect tolerance. The

Alessandro Cilardo, University of Naples Federico II, IT,
Co-Chair:
Walter Weber, NAMLAB, DE,
Location / Room:
Time:
Date:

4.7 Adaptive Reliable Computing Using Memristive and Reconfigurable Hardware

Date: Tuesday, March 20, 2018
Time: 17:00 - 18:30
Location / Room: Konf. 5
Chair:
Walter Weber, NAMLAB, DE, Contact Walter Weber
Co-Chair:
Alessandro Cilardo, University of Naples Federico II, IT, Contact Alessandro Cilardo

This session discusses reliability analysis and enhancement of memristive computing, addressing the non-linear behavior and the development of a logic synthesis flow for defect tolerance. The session also focuses on adapting the precision of the heterogenous hardware to fit application requirements.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>18:05</td>
<td>4.6.4</td>
<td>ENERGY-SECURE SWARM POWER MANAGEMENT</td>
</tr>
</tbody>
</table>

Speaker: Pradip Bose, IBM Corporation, US
Authors: Augusto Vega, Alper Buyuktosunoglu and Pradip Bose, IBM T. J. Watson Research, US

Abstract: We present a visionary concept of a distributed (or decentralized) power/thermal control mechanism that applies the bio-inspired artificial intelligence paradigm of swarm intelligence. The target use case is a future many-core processor. Preliminary results based on a swarm simulator are presented. The talk then addresses the challenge of making such power control systems secure against energy attacks - i.e. maliciously launched virus programs that are designed to disrupt the power control mechanism and cause performance shortfalls or even physical damage from over-heating.

Download Paper (PDF; Only available from the DATE venue WiFi)
Abstract
Emerging metal oxide resistive switching random access memory (RRAM) device and RRAM crossbar have shown great potential in computing matrix-vector multiplication. However, due to the non-linear distribution of resistance levels in RRAM devices, state-of-the-art multi-bit RRAM cannot accomplish the multi-bit computing task accurately. In this paper, we propose fault-tolerant schemes to rescue RRAM-based computation with non-linear resistance levels. We classify the resistance level distributions in RRAM into three types, and the corresponding models are proposed to analyze the computation characteristics. We propose two theoretical conditions for the resistance levels to determine if an RRAM device can support multi-bit matrix computation. For the linear model, the least squares method is used to reduce the computing error. When the resistance distribution obeys the proposed power model, a logarithmic operation is used to decode the multiplication results and accomplish accuracy computing. For exponential model, since the device cannot complete the multi-bit matrix-vector multiplication from hardware level, we propose online and offline quantization methods to make the neural computing algorithms friendly to RRAM device. Simulation results show that the root-mean-square error improves around 4% with the linear model and more than 99% with the power model. After quantization, the accuracy of ResNet-18 using RRAM with exponential resistance levels can be improved to the same accuracy with ideal linear RRAM devices.

Download Paper (PDF; Only available from the DATE venue WiFi)
In this Exhibition Workshop leading suppliers from the microelectronics industry present their newest technical solutions for designing and securing the IoT systems of the upcoming digital age. By Jürgen Haase, edacentrum, DE.

Organiser:
Location / Room: Exhibition Theatre
Time: 17:00 - 18:30

4.8 Components for Secure IoT Systems

Time Label Presentation Title Authors
17:30 IP2-2, 199 A CO-DESIGN METHODOLOGY FOR SCALABLE QUANTUM PROCESSORS AND THEIR CLASSICAL ELECTRONIC INTERFACE Speaker: Jeroen van Dijk, Delft University of Technology, NL Authors: Jeroen van Dijk1, Andrei Vladimirescu2, Masoud Babaei1, Eduard Chardon1 and Fabio Sebastian1
1Delft University of Technology, NL; 2University of California, Berkeley, US Abstract A quantum computer fundamentally comprises a quantum processor and a classical controller. The classical electronic controller is used to correct and manipulate the qubits, the core components of a quantum processor. To enable quantum computers scalable to millions of qubits, as required in practical applications, the simultaneous optimization of both the classical electronic and quantum systems is needed. In this paper, a co-design methodology is proposed for obtaining an optimized qubit performance while considering practical trade-offs in the control circuits, such as power consumption, complexity, and cost. The SPINE (SPIN Emulator) toolset is introduced for the co-design and co-optimization of electronic/quantum systems. It comprises a circuit simulator enhanced with a Verilog-A modeling the quantum behavior of single-electron spin qubits. Design examples show the effectiveness of the proposed methodology in the optimization, design and verification of a whole electronic/quantum system.
Download Paper (PDF; Only available from the DATE venue WiFi)

18:31 IP2-3, 757 APPROXIMATE QUATERNARY ADDITION WITH THE FAST CARRY CHAINS OF FPGAS Speaker: Philip Bisk, University of California, Riverside, US Authors: Sina Boroumand1, Hadi P. Afshar2 and Philip Bisk3 1University of Tehran, IR; 2Qualcomm Research, US; 3University of California, Riverside, US Abstract A heuristic is presented to efficiently synthesize approximate adder trees on Altera and Xilinx FPGAs using their carry chains. The mapper constructs approximate adder trees using an approximate quaternary adder as the fundamental building block. The approximate adder trees are smaller than exact adder trees, allowing more operators to fit into a fixed-area device, trading off arithmetic accuracy for higher throughput.
Download Paper (PDF; Only available from the DATE venue WiFi)

18:32 IP2-4, 424 NN COMPACTOR: MINIMIZING MEMORY AND LOGIC RESOURCES FOR SMALL NEURAL NETWORKS Speaker: Seongmin Hong, Hongik University, KR Authors: Seongmin Hong1, Inho Lee1 and Yongjun Park2 1Hongik University, KR; 2Hanyang University, KR Abstract Special neural accelerators are an appealing hardware platform for machine learning systems because they provide both high performance and energy efficiency. Although various neural accelerators have recently been introduced, they are difficult to adapt to embedded platforms because current neural accelerators require high memory capacity and bandwidth for the fast preparation of synaptic weights. Embedded platforms are often unable to meet these memory requirements because of their limited resources. In FPGA-based IoT (internet of things) systems, the problem becomes even worse since computation units generated from logic blocks cannot be fully utilized due to the small size of block memory. In order to overcome this problem, we propose a novel dual-track quantization technique to reduce synaptic weight width based on the magnitude of the value while minimizing accuracy loss. In this value-adaptive technique, large and small value weights are quantized differently. In this paper, we present a fully automatic framework called NN Compactor that generates a compact neural accelerator by minimizing the memory requirements of synaptic weights through dual-track quantization and minimizing the logic requirements of PUs with minimum recognition accuracy loss. For the three widely used datasets of MNIST, CNAE-9, and Forest, experimental results demonstrate that our compact neural accelerator achieves an average performance improvement of 6.4x over a baseline embedded system using minimal resources with minimal accuracy loss.
Download Paper (PDF; Only available from the DATE venue WiFi)

18:30 End of session
Exhibition Reception in Exhibition Area
The Exhibition Reception will take place on Tuesday in the exhibition area, where free drinks for all conference delegates and exhibition visitors will be offered. All exhibitors are welcome to also provide drinks and snacks for the attendees.

4.8 Components for Secure IoT Systems

Date: Tuesday, March 20, 2018
Time: 17:00 - 18:30
Location / Room: Exhibition Theatre
Organiser:
Jürgen Haase, edacentrum, DE. Contact Jürgen Haase
In this Exhibition Workshop leading suppliers from the microelectronics industry present their newest technical solutions for designing and securing the IoT systems of the upcoming digital age. By elaborating on their technical approaches and the experience made during the design and in the field, they will provide attendees with valuable advice for the challenges in their own job

Time Label Presentation Title Authors
17:00 4.8.1 SECURING THE INTERNET OF THINGS WITH TI SIMPLELINK PLATFORM Speaker: Roger Monk, Texas Instruments Europe, FR Abstract With billions of IoT devices getting connected to the Internet, it is ever more important to make these devices are as secure and robust as possible. These devices should be protected from running malicious software and it is critical that the sensitive user data that these device handle is kept secret. These security requirements add significant responsibility to the system-on-chip solutions at their heart. The SimpleLink Wi-Fi family of devices have been instrumental in enabling small, power-optimised IoT devices leveraging existing Wi-Fi infrastructure. The first generation CC3100/CC3200 provides secure network socket connectivity to enable secure data connection to remote servers and services. The next generation of this family, CC3120/CC3220, significantly extends these capabilities by not only offering the latest network security ciphers, but also an advanced security platform to protect system assets for the entire life cycle of the product, offering customers further confidence and protection. This presentation aims to detail the challenges of security in today’s IoT products and how the architecture of this latest generation of embedded Wi-Fi platform has been designed to efficiently address the technical challenges presented and how these advanced and differentiated security capabilities can be exposed and enabled for all users via a ‘simple’ SimpleLink API.
DEVELOPMENT OF A NEAR-THRESHOLD DIGITAL CELL LIBRARY AND A DESIGN FLOW FOR IOT SENSOR SYSTEMS

Speaker: Jörg Dobiasi, X-FAB, DE

Abstract
Optimized digital standard cells which efficiently operate in the near-threshold voltage (NTV) region are one basic enabler for the next generation of smart sensor systems especially of IoT. While a significant reduction of both dynamic power and leakage power is necessary to meet the power requirements of such systems, a reasonable performance still needs to be supported, to enable on-chip pre-processing and analysis of the sensor data.

The presentation will provide an overview about the development of a near-threshold digital library implemented in X-FAB's 0.18 µm Silicon-on-Insulator technology carried out in the framework of the BMBF-funded project RoMuLiS [1]. A digital ultra-low-power logic library was developed based on the standard CMOS technique, which operates in NTV region with 700 ... 800mV operating voltage at -20 °C ... 85 °C. Additionally supporting cells like level shifters and an NTV I/O pad cell library with ESD protection have been developed. The cells have been implemented with a NTV test chip, using a power-aware digital implementation flow. The test chip has been manufactured and characterized. The test results prove the function of the NTV logic cells in the specified voltage and temperature range and demonstrate the feasibility and possibilities of the development of further NTV logic cells.

FULL CUSTOM MEMS DESIGN: NEW METHODS FOR THE ANALYSIS OF PARASITIC ELECTROSTATIC EFFECTS

Speaker: Axel Hadd, Robert Bosch GmbH, DE

Abstract
Microelectromechanical systems (MEMS) are widely used in IoT devices. Due to the lack of sophisticated component libraries for MEMS, highly optimized MEMS sensors are currently designed using a polygon-driven design flow. The advantage of this design flow is its accurate mechanical simulation, but it lacks of methods for analyzing the parasitic electrostatic effects arising from the electric coupling between (stationary) wiring and the mechanical structures. For a robust and secure MEMS design, it is necessary to analyze, to optimize and finally to include these parasitics into the MEMS-ASIC co-simulation.

The presentation will provide an overview about the development of new methods for the analysis of parasitic electrostatic effects by a 3D field-solver carried out in the framework of the BMBF-funded project RoMuLiS. The developed methods include a rule based structure recognition algorithm, which allows the identification of meaningful MEMS sensor parts out of the plain graphical polygon representation of the MEMS layout. The mapping of the extracted RC-values to the recognized elements of the MEMS sensor enables a detailed analysis and optimization of actual MEMS sensors. This method is upgraded by a feature that enables the parasitics arising from in-plane, sensor-structure motion to be extracted quasi-dynamically.

END OF SESSION

18:30

End of session
Exhibition Reception in Exhibition Area

The Exhibition Reception will take place on Tuesday in the exhibition area, where free drinks for all conference delegates and exhibition visitors will be offered. All exhibitors are welcome to also provide drinks and snacks for the attendees.

EWCMS: AN EMBEDDED WALK-CYCLE MONITORING SYSTEM USING BODY AREA COMMUNICATION AND SECURE LOW-POWER DYNAMIC SIGNALING

Authors: Shahzad Muzaffar1, Markus Mueller2 and Ibrahim (Abe) M. Elfadel2
1Masdar Institute, Khalifa University of Science and Technology, AE; 2Khalifa University of Science and Technology, AE

Abstract
The demo presents a novel ultra-low power, embedded, and wearable walk-cycle monitoring system with applications in areas such as healthcare, robotics, sports medicine, physical therapy, prosthesis, and animal sports. Customized shoes with sensors continuously measure the forces, and an electronic digital assistant is used to analyze the acquired measurements in real-time by employing an IMU free and self-synchronizing method in order to estimate weight and study motion patterns. To achieve ultra-low power operation, the human body is used as a communication medium between the sensors and the digital assistant. The single-channel behavior of the human body is accommodated with a novel, simple yet robust single wire signaling technique, Pulsed-Index Communication (PIC), that significantly reduces the system footprint and overall power consumption as it eliminates the need for clock and data recovery. The system prototype has been rigorously and successfully tested.

DEVELOPMENT OF A NEAR-THRESHOLD DIGITAL CELL LIBRARY AND A DESIGN FLOW FOR IOT SENSOR SYSTEMS

Authors: Tobias Markus1, Markus Mueller2 and Ulrich Bruening3
1University of Heidelberg, DE; 2Extol GmbH, DE

Abstract
Power optimized digital standard cell libraries are a "must have" when it comes to the implementation of power efficient smart sensor systems for IoT applications. In the framework of the BMBF-funded project RoMuLiS, the Fritz Huettinger Chair of Microelectronics of the University of Freiburg and X-FAB jointly develop near- and sub-threshold digital standard cell libraries for both ultra low-voltage and ultra low-power applications.

18:30

End of session
Exhibition Reception in Exhibition Area

The Exhibition Reception will take place on Tuesday in the exhibition area, where free drinks for all conference delegates and exhibition visitors will be offered. All exhibitors are welcome to also provide drinks and snacks for the attendees.
**UB04.3 PRIME: PLATFORM- AND APPLICATION-AGNOSTIC RUN-TIME POWER MANAGEMENT OF HETEROGENEOUS EMBEDDED SYSTEMS**

**Authors:** Domenico Balsamo, Graeme M. Bragg, Charles Leech and Geoff V. Merrett, University of Southampton, GB

**Abstract**

Increasing energy efficiency and reliability at runtime is a key challenge of heterogeneous many-core systems. We demonstrate how contributions from the PRIME project integrate to enable application- and platform-agnostic runtime management that respects application performance targets. We consider opportunities to enable runtime management across the system stack and we enable cross-layer interactions to trade-off power and reliability with performance and accuracy. We consider a system as three distinct layers, with abstracted communication between them, which enables the direct comparison of different approaches, without requiring specific application or platform knowledge. Application-agnostic runtime management is demonstrated with a selection of runtime managers from PRIME, including linear regression modelling and predictive thermal management, operating across multiple applications. Platform-independent runtime management is demonstrated using two heterogeneous platforms.

**More information ...**

**UB04.4 OTPG: SPECIFICATION-BASED CONSTRUCTION OF ONLINE TPGS FOR MICROPROCESSORS**

**Authors:** Mikhail Chupilko, Alexander Kamkin and Andrei Tatamkov, ISP RAS, RU

**Abstract**

This work presents an approach to construction of online test program generators (TPGs). The approach is intended to use specifications of ISA presented in nML/mmuSL specification languages. They are processed by a meta-generator to obtain their binary representations supplied with meta information and a test generation core compatible with the target microprocessor. The test generation core is loaded as a binary image into the target microprocessor’s memory (for experiments we’re using QEMU for MIPS) and produces test cases to be processed (incl. results checking) by an executor. It should be noticed that the meta-generator and the executor are not obligatory run at the same microprocessor (especially, if it is highly incomplete). The final goal of the project is to propose a method of obtaining online TPGs for a wide range of ISAs, and to develop a mature tool implementing this method.

**More information ...**

**UB04.7 T-CREST: THE OPEN-SOURCE REAL-TIME MULTICORE PROCESSOR**

**Authors:** Martin Schoberl, Luca Pezzarossa and Jens Sparse, Technical University of Denmark, DK

**Abstract**

Future real-time systems, such as advanced control systems or real-time image recognition, need more powerful processors, but still a system where the worst-case execution time (WCET) can be statically predicted. Multicore processors are one answer to the need for more processing power. However, it is still an open research question how to best organize and implement time-predictable communication between processing cores. T-CREST is an open-source multicore processor for research on time-predictable computer architecture. It consists of several Patmos processors connected by various time-predictable communication structures: access to shared off-chip, access to shared on-chip memory, and the Argo network-on-chip for fast inter-processor communication. T-CREST is supported by open-source development tools, such as compilation and WCET analysis. To best of our knowledge, T-CREST is the only fully open-source architecture for research on future real-time multicore architectures.

**More information ...**

**UB04.10 EXPERIENCE-BASED AUTOMATION OF ANALOG IC DESIGN**

**Authors:** Florian Leber and Juergen Scheible, Reutlingen University, DE

**Abstract**

While digital design automation is highly developed, analog design automation still remains behind the demands. Previous circuit synthesis approaches, which are usually based on optimization algorithms, do not satisfy industrial requirements. A promising alternative is given by procedural approaches (also known as "generators"). They (a) emulate experts' decisions, thus (b) make expert knowledge re-usable and (c) can consider all relevant aspects and constraints implicitly. Nowadays, generators are successfully applied in analog layout (Pcells, Pycells). We aim at an entire design flow completely based on procedural automation techniques. This flow will consist of procedures for the generation of schematics and layouts for every typical analog circuit class, such as amplifier, bandgap, filter a.s.o. In our presentation we give an overview on such a design flow and we show an approach for capturing an analog circuit designer's strategy as an executable "expert design plan".

**More information ...**

**End of session**

---

**Exhibition-Reception Exhibition Reception**

**Date:** Tuesday, March 20, 2018  
**Time:** 18:30 - 19:30  
**Location / Room:** Exhibition Area

The Exhibition Reception will take place on Tuesday in the exhibition area, where free drinks for all conference delegates and exhibition visitors will be offered. All exhibitors are welcome to also provide drinks and snacks for the attendees.   

**Time** | **Label** | **Presentation Title** | **Authors**
--- | --- | --- | ---
19:30 | End of session | 

---

**5.1 Special Day Session on Future and Emerging Technologies: Challenges for the Design of Microfluidic Devices: EDA for your Lab-on-a-Chip**

**Date:** Wednesday, March 21, 2018  
**Time:** 08:30 - 10:00  
**Location / Room:** Saal 2

**Chair:** Chakrabarty Krishnendu, Duke University, US, Contact Krishnendu Chakrabarty

This session introduces experts from design automation to the field of microfluidic devices. Those devices, often also referred to as labs-on-chip, allow for conducting biological, chemical, and/or medical experiments with fluids on a nano- or even picolitre scale automatically on miniaturized devices. By this, they revolutionized point-of-care diagnostics, chemo-fluidic logic, and more. The speakers in this session will introduce those areas and show how microfluidic devices help here. At the same time, they will cover how design automation can actually advance the further development of this emerging technology and how this can help to broaden the scope of applications for it.

**Time** | **Label** | **Presentation Title** | **Authors**
--- | --- | --- | ---
5.1.1 POINT-OF-CARE DIAGNOSTICS 2.0: STANDARDS, DESIGN AUTOMATION, AND CONSUMER ELECTRONICS FOR THE NEXT GENERATION OF DIAGNOSTIC DEVICES

Author:
Emmanuel Delamarche, IBM Research, CH

Abstract
Diagnoses are ubiquitous in healthcare because they support prevention, diagnosis and treatment of diseases. Specifically, point-of-care diagnostics are particularly attractive for identifying diseases near patients, quickly, and in many settings and scenarios. One of our contributions to the field of microfluidics is the development of capillary-driven microfluidic chips for highly miniaturized immunoassays. In this presentation, I will review how to program capillary flow and encode specific functions to form a microfluidic element that can easily be assembled into self-powered devices for immunoassays, reaching unprecedented levels of precision for manipulating samples and reagents. I will also reflect on the fragmented approaches that our community has in developing microfluidic-based diagnostics, which is exacerbated by the fragmented nature of the in vitro diagnostic market. Standards, design automation, consumer electronic components, and smartphones may play a critical role in helping to rationalize our development and utilization of the next generation of point-of-care diagnostic devices.

5.2 Smart Energy and Automotive Systems

Date: Wednesday, March 21, 2018
Time: 08:30 - 10:00
Location / Room: Kont. 6
This session presents the latest advancements in battery and photovoltaic system management and optimization, as well as novel approaches towards efficient environmental mapping for autonomous driving and cloud-connected vehicles.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30</td>
<td>5.2.1</td>
<td>SOH-AWARE ACTIVE CELL BALANCING STRATEGY FOR HIGH POWER BATTERY PACKS</td>
<td>Alma Proebstl, Technical University of Munich, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Alma Proebstl, Technical University of Munich, DE</td>
<td>Authors: Alma Proebstl, Sangyoung Park, Swaminathan Narayanaswamy, Sebastian Steinhorst and Samaaj Chakraborty</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Short drive range due to limited battery capacity and high battery degradation costs persist to be the main deterrents to the wide adoption of Electric Vehicles (EVs). High power battery packs consisting of a large number of battery cells require extensive management, such as State of Charge (SOC) balancing and thermal management, in order to keep the operating conditions within a safe range. In this paper, we propose a novel State of Health (SOH)-aware active cell balancing technique, which is capable of extending the cycle life of the whole battery pack. In contrast to the state-of-the-art active cell balancing techniques, the proposed technique allows cells to have different SOC values such that aging is mitigated when an EV trip does not require the full capacity. Based on the observation that prefiring cells with higher SOH over cells with lower SOH extends cycle life, the technique identifies the charge transfers between cells that would benefit the most. We find that with our proposed scheme, aging could be mitigated by up to 23.5% over passive cell balancing and 17.6% over active SOC cell balancing.</td>
<td></td>
</tr>
</tbody>
</table>

| 09:00  | 5.2.2 | GIS-BASED OPTIMAL PHOTOVOLTAIC PANEL FLOORPLANNING FOR RESIDENTIAL INSTALLATIONS | Sara Vino, Politecnico di Torino, IT                                   |
|        |       | Speaker: Sara Vino, Politecnico di Torino, IT               | Authors: Sara Vino, Lorenzo Bottacchi, Edoardo Patti, Andrea Acquaviva, Ennio Maci and Massimo Poncino, Politecnico di Torino, IT |
|        |       | Abstract: Shading is a crucial issue for the placement of PV installations, as it heavily impacts power production and the corresponding return on investment. Nonetheless, residential rooftop installations still rely on rule-of-thumb criteria and on gross estimates of the shading patterns, while more optimized approaches focus solely on the identification of suitable surfaces (e.g., roofs) in a larger geographic area (e.g., city or district). This work addresses the challenge of identifying an optimal (with respect to the overall energy production) placement of PV panels on a roof. The novel aspect of the proposed solution lies in the possibility of having a sparse, irregular placement of individual modules so as to better exploit the variance of solar data. The latter are represented in terms of the distribution of irradiance and temperature values over the roof, as elaborated from historical traces and Geographical Information System (GIS) data. Experimental results will prove the effectiveness of the algorithm through three real world case studies, and that the generated optimal solutions allow to increase power production by up to 28% with respect to rule-of-thumb solutions. |

| 09:30  | 5.2.3 | CELL-BASED UPDATE ALGORITHM FOR OCCUPANCY GRID MAPS AND HYBRID MAP FOR ADAS ON EMBEDDED GPUS | Jörg Fickenscher, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), DE |
|        |       | Speaker: Jörg Fickenscher, Jens Schlumberger, Frank Hannig, Mohamed Essayed Bouzoua and Jürgen Teich | Authors: Jörg Fickenscher, Jens Schlumberger, Frank Hannig, Mohamed Essayed Bouzoua and Jürgen Teich |
|        |       | Abstract: Advanced Driver Assistance Systems (ADASs), such as autonomous driving, require the continuous computation and update of detailed environment maps. Today’s standard processors in automotive Electronic Control Units (ECUs) struggle to provide enough computing power for those tasks. Here, new architectures, like Graphics Processing Units (GPUs), might be a promising accelerator candidate for ECUs. Current algorithms have to be adapted to these new architectures when possible, or new algorithms have to be designed to take advantage of these architectures. In this paper, we propose a novel parallel update algorithm, called cell-based update algorithm for occupancy grid maps, which exploits the highly parallel architecture of GPUs and overcomes the shortcomings of previous implementations based on the Bresenham algorithm on such architectures. A second contribution is a new hybrid map, which takes the advantages of the classic occupancy grid map and reduces the computational effort of those. All algorithms are parallelized and implemented on a discrete GPU as well as on an embedded GPU (Nvidia Tegra K1 Jetson board). Compared with the state-of-the-art Bresenham algorithm as used in the case of occupancy grid maps, our parallelized cell-based update algorithm and our proposed hybrid map approach achieve speedups of up to 2.5 and 4.5, respectively. |

| 10:00  | IP2-5 | IMPROVING FAST CHARGING EFFICIENCY OF RECONFIGURABLE BATTERY PACKS | Alexander Lamprecht, TUM CREATE, SG                                   |
|        |       | Speaker: Alexander Lamprecht | Authors: Alexander Lamprecht, Swaminathan Narayanaswamy and Sebastian Steinhorst |
|        |       | Abstract: Recently, reconfigurable battery packs that can dynamically modify the electrical connection topology of their individual cells are gaining importance. While several circuit architectures and management algorithms are proposed in the literature, the electrical characteristics of the reconfiguration circuit architectures are not sufficiently studied so far. In this paper, we derive a detailed analytical model for a state-of-the-art reconfiguration architecture capturing the losses introduced by the parasitic resistances of the circuit components. For the first time, we propose a novel fast charging strategy using the reconfiguration architecture that significantly reduces the power losses in comparison to conventional battery packs. Moreover, using the analytical model, we highlight the challenges faced by existing reconfiguration architectures using state-of-the-art components and we derive the specifications for the switches which are essential for improving the energy efficiency of such reconfigurable battery packs. |

Download Paper (PDF; Only available from the DATE venue WiFi)
Cloud-Assisted Control of Ground Vehicles Using Adaptive Computation Offloading Techniques

Speaker:
Soheil Samii, General Motors R&D, Warren, MI 48090, US

Authors:
Arun Adiththan1, Ramesh S2 and Soheil Samii2
1City University of New York, US; 2General Motors R&D, US

Abstract
The existing approaches to design efficient safety-critical control applications are constrained by limited in-vehicle sensing and computational capabilities. In the context of automated driving, we argue that there is a need to leverage resources "out-of-the-vehicle" to meet the sensing and powerful processing requirements of sophisticated algorithms (e.g., deep neural networks). To realize the need, a suitable computation offloading technique that meets the vehicle safety and stability requirements, even in the presence of unreliable communication network, has to be identified. In this work, we propose an adaptive offloading technique for control computations into the cloud. The proposed approach considers both current network conditions and control application requirements to determine the feasibility of leveraging remote computation and storage resources. As a case study, we describe a cloud-based path following controller application that leverages crowdsensed data for path planning.

Download Paper (PDF; Only available from the DATE venue WiFi)
1. WALL: A WRITEBACK-AWARE LLC MANAGEMENT FOR PCM-BASED MAIN MEMORY SYSTEMS
   
   **Abstract**
   In this paper, we propose WALL, a novel writeback-aware LLC management scheme to reduce the number of LLC writebacks and consequently improve performance, energy efficiency, and lifetime of a PCM-based main memory system. First, we investigate the writeback behavior of LLC sets and show that writebacks are not uniformly distributed among sets; some sets observe much higher writeback rates than others. We then propose a writeback-aware set-balancing mechanism, which employs the underutilized LLC sets with few writebacks as an auxiliary storage for storing the evicted dirty lines of sets with frequent writebacks. We also propose a simple and effective writeback-aware replacement policy to avoid the eviction of the writeback blocks that are highly reused after being evicted from the cache. Our experimental results show that WALL achieves an average of 26.6% reduction in the total number of LLC writebacks, compared to the baseline scheme, which uses the LRU replacement policy. As a result, WALL can reduce the memory energy consumption by 19.2% and enhance PCM lifetime by 1.25x, on average, on an 8-core system with a 4GB PCM main memory, running memory intensive applications.

2. DESIGN AND INTEGRATION OF HIERARCHICAL-PLACEMENT MULTI-LEVEL CACHES FOR REAL-TIME SYSTEMS
   
   **Abstract**
   Enabling timing analysis in the presence of caches has been pursued by the real-time embedded systems (RTES) community for years due to cache's huge potential to reduce software's worst-case execution time (WCET). However, caches heavily complicate timing analysis due to hard-to-predict access patterns, with few works dealing with time analysability of multi-level cache hierarchies. For measurement-based timing analysis (MBTA) techniques - widely used in domains such as avionics, automotive, and rail - we propose several cache hierarchies amenable to MBTA. We focus on a probabilistic variant of MBTA (or MBPTA) that requires caches with time-randomized behavior whose execution time variability can be captured in the measurements taken during system's test runs. For this type of caches, we explore and propose different multi-level cache setups. From those, we choose a cost-effective cache hierarchy that we implement and integrate in a 4-core LEON3 RTL processor model and prototype in a FPGA. Our results show that our proposed setup implemented in RTL results in better (reduced) WCET estimates with similar implementation cost and no impact on average performance w.r.t. other MBPTA-amenable setups.

3. LARS: LOGICALLY ADAPTABLE RETENTION TIME STT-RAM CACHE FOR EMBEDDED SYSTEMS
   
   **Abstract**
   STT-RAMs have been studied as a promising alternative to SRAMs in embedded systems' caches and main memories. STT-RAMs are attractive due to their low leakage power and high density; STT-RAMs, however, also have drawbacks of long write latency and high dynamic write energy. A popular solution to this drawback relaxes the retention time to lower both write latency and energy, and uses a dynamic refresh scheme that refreshes data blocks to prevent them from prematurely expiring. However, the refreshes can incur overheads, thus limiting optimization potential. In addition, this solution only provides a single retention time, and cannot adapt to applications' variable retention time requirements. In this paper, we propose LARS (Logically Adaptable Retention Time STT-RAM) cache as a viable alternative for reducing the write energy and latency. LARS cache comprises of multiple STT-RAM units with different retention times, with only one unit on at a given time. LARS dynamically determines which STT-RAM unit to power on during runtime, based on executing applications' needs. Our experiments show that LARS cache is low-overhead, and can reduce the average energy and latency by 35.8% and 13.2%, respectively, as compared to the dynamic refresh scheme.

4. FUSIONCACHE: USING LLC TAGS FOR DRAM CACHE
   
   **Abstract**
   DRAM caches have been shown to be an effective way to utilize the bandwidth and capacity of 3D stacked DRAM. Although they can capture the spatial and temporal data locality of applications, their access latency is still substantially higher than conventional on-chip SRAM caches. Moreover, their tag access latency and storage overheads are excessive. Storing tags for a large DRAM cache in DRAM is impractical as it would occupy a significant fraction of the processor chip. Storing them in the DRAM itself incurs high access overheads. Attempting to cache the DRAM tags on the processor adds a constant delay to the access time. In this paper, we introduce FusionCache, a DRAM cache that offers more efficient tag accesses by fusing DRAM cache tags with the tags of the on-chip Last Level Cache (LLC). We observe that, in an inclusive cache model where the DRAM cachelines are multiples of on-chip SRAM cachelines, LLC tags could be re-purposed to access a large part of the DRAM cache contents. Then, accessing DRAM cache tags incurs zero additional latency in the common case.
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

5.4 Special Session: Lightweight Security for Resources-Constrained Internet-of-Things Applications

Date: Wednesday, March 21, 2018
Time: 08:30 - 10:00
Location / Room: Konf. 2

Chair:
Halak Basel, Southampton University, GB, Contact Basel Halak

Co-Chair:
Jin Yier, University of Florida, US, Contact Yier Jin

This special session includes four papers: the first paper addresses the first question, it presents a lightweight cryptographic primitive based on physical unclonable functions, the second and third papers tackle the second and the third questions. They present two security protocols, for authentication and attestation respectively, which are specifically developed for resources-constrained IoT platforms. The fourth paper addresses the last challenge, it presents a solution which exploits existing on-chip hardware structure to detect abnormal and suspicious behaviours of an embedded system.

Time | Label | Presentation Title
--- | --- | ---
08:30 | 5.4.1 | COST EFFICIENT DESIGN OF MODELLING ATTACKS-RESISTANT PHYSICAL UNCLONABLE FUNCTIONS

Speaker:
Basel Halak, Southampton University, GB

Authors:
Mohd Syafiq Mispan1, Haibo Su1, Mark Zwolinski2 and Basel Halak3
1Electronics and Computer Science Department, Southampton University, GB; 2University of Southampton, GB; 3Southampton University, GB

Abstract
Physical Unclonable Functions (PUFs) exploit the intrinsic manufacturing process variations to generate a unique signature for each silicon chip; this technology allows building lightweight cryptographic primitive suitable for resource-constrained devices. However, the vast majority of existing PUF design is susceptible to modeling attacks using machine learning techniques, this means it is possible for an adversary to build a mathematical clone of the PUF that have the same challenge/response behavior of the device. Existing approaches to solve this problem include the use of hash functions, which can be prohibitively expensive and render PUF technology as the suitable candidate for lightweight security. This work presents a challenge permutation and substitution techniques which are both area and energy efficient. We implemented two examples of the proposed solution in 65-nm CMOS technology, the first using a delay-based structure design (an Arbiter-PUF), and the second using sub-threshold current design (two-choose-one PUF or TCO-PUF). The resiliency of both architectures against modeling attacks is tested using an artificial neural network machine learning algorithm. The experiment results show that it is possible to reduce the predictability of PUFs to less than 70% and a fractional area and power costs compared to existing hash function approaches.

Download Paper (PDF; Only available from the DATE venue WiFi)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully
Lunch Breaks (Großer Saal + Saal 1)

Coffee Breaks in the Exhibition Area

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully
registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
☐ Coffee Break 10:30 - 11:30
☐ Lunch Break 13:00 - 14:30
☐ Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
☐ Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
☐ Coffee Break 10:00 - 11:00
☐ Lunch Break 12:30 - 14:30
☐ Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
☐ Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
☐ Coffee Break 10:00 - 11:00
☐ Lunch Break 12:30 - 14:00
☐ Keynote Lecture in “Saal 2” 13:20 - 13:50
☐ Coffee Break 15:30 - 16:00
5.5 Emerging Technologies for Future Computing

Date: Wednesday, March 21, 2018
Time: 08:30 - 10:00
Location / Room: Kont. 3

Chair:
Aida Todri-Sanial, CNRS, FR, Contact Aida Todri-Sanial

Co-Chair:
Mariagrazia Graziano, Politecnico di Torino, IT, Contact Mariagrazia Graziano

A wide overview of emerging technologies to enable novel computing paradigms. The session covers topics from carbon nanotube thin film transistors for flexible electronics, novel 3D interconnects using inductive coupling links, physical design of quantum cellular automata, and improving reliability of quantum logic cell implementation.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 08:30  | 5.5.1   | COMPACT MODELING OF CARBON NANOTUBE THIN FILM TRANSISTORS FOR FLEXIBLE CIRCUIT DESIGN | Author: Leilai Shao, University of California, Santa Barbara, US  
Authors: Leilai Shao, Tsung-Ching Huang, Ting Li, Zhenan Bao, Raymond Beausoleil, and Tim Cheng  
1University of California Santa Barbara, US; 2Hewlett Packard Labs, US; 3Stanford University, US; 4HPE Labs, US; 5HKUST, HK  
Abstract  
Carbon nanotube thin film transistor (CNT-TFT) is a promising candidate for flexible electronics, because of its high carrier mobility and great mechanical flexibility. An accurate and trustworthy device model for CNT-TFTs, however, is still missing. In this paper, we present a SPICE-compatible compact model for CNT-TFT circuit simulation and validate the proposed model based on fabricated CNT-TFTs and Pseudo-CMOS circuits. The proposed CNT-TFT model enables circuit designers to explore design space by adjusting device parameters, supply voltages and transistor sizes to optimize the noise margin (NM) and power-delay product (PDP), which are the key metrics for larger scale CNT-TFT circuits. We further propose a design framework to effectively optimize the NM and PDP to facilitate greater automation of flexible circuit design based on CNT-TFTs.  
Download Paper (PDF; Only available from the DATE venue WiFi) |
| 09:00  | 5.5.2   | A HIGH-SPEED DESIGN METHODOLOGY FOR INDUCTIVE COUPLING LINKS IN 3D-ICs              | Author: Benjamin Fletcher, University of Southampton, GB  
Authors: Benjamin Fletcher, Shidhartha Das, and Terence Mak  
1University of Southampton, GB; 2ARM Ltd., GB  
Abstract  
Inductive coupling links (ICLs) are gaining traction as an alternative to through silicon vias (TSVs) for 3D integration, promising high-bandwidth connectivity without the inflated fabrication costs associated with TSV-enabled processes. For power-efficient ICL design, optimisation of the utilized physical inductor geometries is essential, however typically necessitates the use of finite element analysis (FEA) in addition to manual parameter fitting, a process that can take several hours even for a single geometry. As a result, the generation of optimised inductor designs poses a significant challenge. In this paper, we address this challenge, presenting a CAD-tool for Optimisation of Inductive coupling Links for 3D-ICs (COIL-3D). COIL-3D uses a rapid solver based upon semi-empirical expressions to quickly and accurately characterise a given link, in conjunction with a high-speed refined optimisation flow to find optimal inductor geometries for use in ICL-based 3D-ICs. The proposed solver achieves an average accuracy within 9.1% of commercial FEA software tools, and the proposed optimisation flow reduces the search time by 26 orders of magnitude. This work unlocks new potential for power-efficient 3D integration using inductive coupling links.  
Download Paper (PDF; Only available from the DATE venue WiFi) |
| 09:30  | 5.5.3   | AN EXACT METHOD FOR DESIGN EXPLORATION OF QUANTUM-DOT CELLULAR AUTOMATA           | Author: Marcel Walter, University of Bremen, DE  
Authors: Marcel Walter, Robert Wille, Daniel Grosse, Frank Sill Torres, and Rolf Drechsler  
1University of Bremen; 2Johannes Kepler University Linz, AT; 3University of Bremen/DFKI GmbH, DE; 4Federal University of Minas Gerais, BR  
Abstract  
Quantum-dot Cellular Automata (QCA) are an emerging computation technology in which basic states are represented by nanosize particles and logic operations are conducted through corresponding effects such as Coulomb interaction. This allows to overcome physical boundaries of conventional solutions such as CMOS and, hence, constitutes a promising direction for future computing devices. Despite these promises, however, the development of (automatic) design methods for QCAs is still in its infancy. In fact, QCA circuits are mainly designed manually thus far and only few heuristics are available. This frequently leads to unsatisfactory results and generally makes it hard to evaluate the quality of respective QCA designs. In this work, we propose an exact solution for the design of QCA circuits that can be configured e.g. to generate circuits that satisfy certain design objectives and/or physical constraints. For the first time, this allows for design exploration of QCA circuits. Experimental evaluations and case studies demonstrate the benefit of proposed method.  
Download Paper (PDF; Only available from the DATE venue WiFi) |
| 09:45  | 5.5.4   | ACCURATE MARGIN CALCULATION FOR SINGLE FLUX QUANTUM LOGIC CELLS                  | Author: Soheil Nazar Shahsavani, Bo Zhang and Massoud Pedram, University of Southern California, US  
Authors: Soheil Nazar Shahsavani, Bo Zhang, and Massoud Pedram, University of Southern California, US  
Abstract  
This paper presents a novel method for accurate margin calculation of single flux quantum (SFQ) logic cells in a superconducting electronic circuit. The proposed method can be utilized as a figure of merit to estimate the robustness of a logic cell without the need for expensive Monte-Carlo simulations. This is achieved through efficient state-space exploration of all parameters in the cell structure. Using proposed approach, distinct parameter dispersion (DPD) based yield of SFQ cells increases by 55% on average, compared with state-of-the-art techniques.  
Download Paper (PDF; Only available from the DATE venue WiFi) |
Coffee Breaks in Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Coffee Break in Exhibition Area

Coffee Break 10:30 - 11:30
Coffee Break 16:00 - 17:00
Coffee Break 15:30 - 16:00
Coffee Break 10:00 - 11:00
Coffee Break 12:30 - 14:30
Coffee Break 13:00 - 14:30
Coffee Break 16:00 - 17:00
Coffee Break 10:00 - 11:00
Coffee Break 12:30 - 14:30
Coffee Break 10:00 - 11:00
Coffee Break 16:00 - 17:00
Coffee Break 15:30 - 16:00
Coffee Break 10:00 - 11:00
Coffee Break 12:30 - 14:30
Coffee Break 13:00 - 14:30
Coffee Break 10:00 - 11:00
Coffee Break 12:30 - 14:30
Coffee Break 15:30 - 16:00
5.6 Reliability improvement and evaluation techniques

Date: Wednesday, March 21, 2018
Time: 08:30 - 10:00
Location / Room: Kont. 4

Chair:
Stefano Di Carlo, Politecnico di Torino, IT, Contact Stefano Di Carlo

Co-Chair:
Vasileios Tenentes, University of Southampton, GB, Contact Vasileios Tenentes

This session introduces reliability improvement approaches using dynamic recovery, redundant multithreading, aging mitigation and optimization of metastability effects, spanning from the system to the circuit layer. Also, cross-layer resilience evaluation via fault injection for complex microprocessors is presented.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30</td>
<td>5.6.1</td>
<td>IMPROVING RELIABILITY FOR REAL-TIME SYSTEMS THROUGH DYNAMIC RECOVERY</td>
<td>Yue Ma, University of Notre Dame, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Authors: Yue Ma(^1), Tam Chantam(^2), Robert P. Dick(^2) and Xiaobo Sharon Hu(^3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors: University of Notre Dame, US; University of Michigan, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract:</td>
<td>Technology scaling has increased concerns about transient faults due to soft errors and permanent faults due to lifetime wear processes. Although researchers have investigated related problems, they have either considered only one of the two reliability concerns or presented simple recovery allocation algorithms that cannot effectively use available time slack to improve soft-error reliability. This paper introduces a framework for improving soft-error reliability while satisfying lifetime reliability and real-time constraints. We present a dynamic recovery allocation technique that guarantees to recover any failed task if the remaining slack is adequate. Based on this technique, we propose two scheduling algorithms for task sets with different characteristics to improve system-level soft-error reliability. Lifetime reliability requirements are satisfied by reducing core frequencies for appropriate tasks, thereby reducing wear due to temperature and thermal cycling. Simulation results show that the proposed framework reduces the probability of failure by at least 8% and 73% on average compared to existing approaches.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
<tr>
<td>09:00</td>
<td>5.6.2</td>
<td>OPTIMAL METASTABILITY-CONTAINING SORTING NETWORKS</td>
<td>Johannes Bund, Saarland University, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Authors: Johannes Bund(^1), Christoph Lenzen(^2) and Moti Medina(^2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors: Saarland University, Saarland Informatics Campus, DE; Max Planck Institute for Informatics, Saarland Informatics Campus, DE; From 1/10/2017 in The Department of Electrical and Computer Engineering Ben-Gurion University, IL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract:</td>
<td>When setup/hold times of bistable elements are violated, they may become metastable, i.e., enter a transient state that is neither digital 0 nor 1 [Marino 81]. In general, metastability cannot be avoided, a problem that manifests whenever taking discrete measurements of analog values. Metastability of the output then reflects uncertainty as to whether a measurement should be rounded up or down to the next possible integral measurement outcome. Surprisingly, Lenzen &amp; Medina (ASYNC 2016) showed that metastability can be contained, i.e., measurement values can be correctly sorted without resolving metastability first. However, both their work and the state of the art by Bund et al. (DATE 2017) leave open whether such a solution can be as small and fast as standard sorting networks. We show that this is indeed possible, by providing a circuit that sorts Gray code inputs (possibly containing a metastable bit) and has asymptotically optimal depth and size. Concretely, for 10-channel sorting networks and 16-bit wide inputs, we improve by 48.46% in delay and by 71.58% in area over Bund et al. Our simulations indicate that straightforward transistor-level optimization is likely to result in performance on par with standard (non-containing) solutions.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
<tr>
<td>09:30</td>
<td>5.6.3</td>
<td>MAUI: MAKING AGING USEFUL, INTENTIONALLY</td>
<td>Shou-Chun Li, Department of Computer Science, National Chiao Tung University, TW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Authors: Kai-Chiang Wu(^1), Tien-Hung Tseng(^2) and Shou-Chun Li(^1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors: National Chiao Tung University, TW; National Chiao Tung University, Taiwan, TW</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract:</td>
<td>Device aging, which causes significant loss on circuit performance and lifetime, has been a primary factor in reliability degradation of nanoscale designs. In this paper, we propose to take advantage of aging-induced clock skews (i.e., make them useful for aging tolerance) by manipulating these time-varying skews to compensate for the performance degradation of logic networks. The goal is to assign achievable reasonable aging-induced clock skews in a circuit, such that overall performance degradation due to aging can be minimized, that is, the lifespan can be maximized. On average, 25% aging tolerance can be achieved with insignificant design overhead.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
<tr>
<td>09:45</td>
<td>5.6.4</td>
<td>EXPERT: EFFECTIVE AND FLEXIBLE ERROR PROTECTION BY REDUNDANT MULTITHREADING</td>
<td>HeeSo So, Yonsei University, KR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Authors: HeeSo So(^1), Moslem Didehban(^2), Yohan Ko(^1), Aviral Shrivastava(^2) and Hyungwoong Lee(^1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors: Yonsei University, KR; Arizona State University, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract:</td>
<td>Resiliency is a first-order design concern in modern microprocessor design. Compiler-level Redundant MultiThreading (RMT) schemes are promising because of their capability to detect the manifestation of hardware transient and permanent faults. In this work, we propose EXPERT, a compiler-level RMT scheme which can detect the manifestation of hardware faults in all hardware components. EXPERT transformation generates a checker thread for program main execution thread. These redundant threads execute simultaneously on two physically different cores of a multi-core processor. They perform mostly same computations, however, after each memory write operation committed by the main thread, the checker thread loads back the written data from the memory and checks it against its own locally computed values. If they match, execution continues. Otherwise, the error flag will be raised. Our processor-wide statistical transient and permanent fault injection experiments show that EXPERT error coverage is ~65 better than the state-of-the-art scheme.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
</tbody>
</table>
### 5.7 Software-centric techniques for embedded systems

**Date:** June 21, 2018  
**Time:** 08:30 - 10:00  
**Location / Room:** Kont. 5

**Chair:**  
Marc Geilen, Eindhoven University of Technology, NL;  
Contact Marc Geilen

**Co-Chair:**  
Daniel Ziener, University of Twente, NL;  
Contact Daniel Ziener

Modern heterogeneous architectures pose new challenges for energy-efficient embedded realizations. The talks in this session address these challenges using software techniques, such as approximate computing, task scheduling, and memory and power-management.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00</td>
<td>IP2-11, 579</td>
<td>PRECISE EVALUATION OF THE FAULT SENSITIVITY OF OOO SUPERSCALAR PROCESSORS</td>
<td>Daniel Mueller-Gritschneider¹, Martin Dittrich¹, Josef Weinzierl¹, Eric Cheng², Subhasish Mitra² and Ulf Schlichtmann¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Martin Dittrich, Technical University of Munich, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>Daniel Mueller-Gritschneider¹, Martin Dittrich¹, Josef Weinzierl¹, Eric Cheng², Subhasish Mitra² and Ulf Schlichtmann¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>ETISS is an instruction set simulator (ISS) for Virtual Prototypes (VPs) modeled with SystemC/TLM. In this paper, we propose the extension ETISS-ML, which enables a multi-level simulation that switches between ISS-level and register transfer level (RTL) to accurately evaluate the impact of soft errors in the pipeline of a RISC processor. ETISS-ML achieves close-to-RTL accurate fault injection simulation results with close-to-ISS simulation performance with a speed up gain up to 100× compared to RTL. For this, we propose an approach to dynamically determine the length of the RTL simulation period. The high simulation performance of ETISS-ML enables an ultra-efficient and accurate evaluation of cross-layer resiliency techniques for embedded applications, which requires running a large number of fault injections for long simulation scenarios. This is demonstrated on a case study of a Microcontroller Unit (MCU) executing a control algorithm for adaptive cruise control. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
</tbody>
</table>

### 10:01 Coffee Break in Exhibition Area

Coffee Break in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

**Tuesday, March 20, 2018**

- Coffee Break 10:30 - 11:30  
- Lunch Break 13:00 - 14:30  
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20  
- Coffee Break 16:00 - 17:00

**Wednesday, March 21, 2018**

- Coffee Break 10:00 - 11:00  
- Lunch Break 12:30 - 14:30  
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20  
- Coffee Break 16:00 - 17:00

**Thursday, March 22, 2018**

- Coffee Break 10:00 - 11:00  
- Lunch Break 12:30 - 14:00  
- Keynote Lecture in “Saal 2” 13:20 - 13:50  
- Coffee Break 15:30 - 16:00

### 5.7 Software-centric techniques for embedded systems

**Date:** Wednesday, March 21, 2018  
**Time:** 08:30 - 10:00  
**Location / Room:** Kont. 5

**Chair:**  
Marc Geilen, Eindhoven University of Technology, NL;  
Contact Marc Geilen

**Co-Chair:**  
Daniel Ziener, University of Twente, NL;  
Contact Daniel Ziener

Modern heterogeneous architectures pose new challenges for energy-efficient embedded realizations. The talks in this session address these challenges using software techniques, such as approximate computing, task scheduling, and memory and power-management.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00</td>
<td>IP2-11, 579</td>
<td>PRECISE EVALUATION OF THE FAULT SENSITIVITY OF OOO SUPERSCALAR PROCESSORS</td>
<td>Daniel Mueller-Gritschneider¹, Martin Dittrich¹, Josef Weinzierl¹, Eric Cheng², Subhasish Mitra² and Ulf Schlichtmann¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Martin Dittrich, Technical University of Munich, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>Daniel Mueller-Gritschneider¹, Martin Dittrich¹, Josef Weinzierl¹, Eric Cheng², Subhasish Mitra² and Ulf Schlichtmann¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>ETISS is an instruction set simulator (ISS) for Virtual Prototypes (VPs) modeled with SystemC/TLM. In this paper, we propose the extension ETISS-ML, which enables a multi-level simulation that switches between ISS-level and register transfer level (RTL) to accurately evaluate the impact of soft errors in the pipeline of a RISC processor. ETISS-ML achieves close-to-RTL accurate fault injection simulation results with close-to-ISS simulation performance with a speed up gain up to 100× compared to RTL. For this, we propose an approach to dynamically determine the length of the RTL simulation period. The high simulation performance of ETISS-ML enables an ultra-efficient and accurate evaluation of cross-layer resiliency techniques for embedded applications, which requires running a large number of fault injections for long simulation scenarios. This is demonstrated on a case study of a Microcontroller Unit (MCU) executing a control algorithm for adaptive cruise control. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
<tr>
<td>Time</td>
<td>Label</td>
<td>Presentation Title</td>
<td>Authors</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>-------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>08:30</td>
<td>5.7.1</td>
<td>HEPREM: ENABLING PREDICTABLE GPU EXECUTION ON HETEROGENEOUS SOC</td>
<td>Björn Forsberg, ETH Zürich, CH</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Speaker:</strong> Björn Forsberg, ETH Zürich, CH; Luca Benini(^2) and Andrea Marongiu(^3)</td>
<td><strong>Authors:</strong> Björn Forsberg(^1), Luca Benini(^2) and Andrea Marongiu(^3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong> Heterogeneous systems-on-a-chip are increasingly embracing shared memory designs, in which a single DRAM is used for both the main CPU and an integrated GPU. This architectural paradigm reduces the overheads associated with data movements and simplifies programmability. However, the deployment of real-time workloads on such architectures is troublesome, as memory contention significantly increases execution time of tasks and the pessimism in worst-case execution time (WCET) estimates. The Predictable Execution Model (PREM) separates memory and computation phases in real-time codes, then arbitrates memory phases from different tasks such that only one core at a time can access the DRAM. This paper revisits the original PREM proposal in the context of heterogeneous SoCs, proposing a compiler-based approach to make GPU codes PREM-compliant. Starting from high-level specifications of computation offloading, suitable program regions are selected and separated into memory and compute phases. Our experimental results show that the proposed technique is able to reduce the sensitivity of GPU kernels to memory interference to near zero, and achieves up to a 20× reduction in the measured WCET.</td>
<td></td>
</tr>
<tr>
<td>09:00</td>
<td>5.7.2</td>
<td>CIRCUIT CARVING: A METHODOLOGY FOR THE DESIGN OF APPROXIMATE HARDWARE</td>
<td>Itaria Scarabottolo, USI Lugano, CH</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Speaker:</strong> Itaria Scarabottolo, USI Lugano, CH; Giovanni Ansaboni and Laura Pozzi, USI Lugano, CH</td>
<td><strong>Authors:</strong> Itaria Scarabottolo, Giovanni Ansaboni and Laura Pozzi, USI Lugano, CH</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong> Systems-on-Chip (SoCs) commonly couple low-power processors and dedicated hardware accelerators, which allow the execution of high-workload and/or timing-critical applications while relying on constrained resources. The functions performed by accelerators are often robust with respect to approximations that, when implemented in HW, can lead to circuits with tangibly lower area and power consumption. Research in approximate computing aims at developing effective strategies to explore the ensuing correctness/efficiency trade-offs. In this context, we address the challenge of approximate circuit design in an innovative way, called here Circuit Carving, which consists in identifying the maximum portion of an exact circuit that can be discarded from it, or carved out, to derive an inexact version not exceeding an error threshold. We achieve this goal by proposing an algorithm based on binary tree exploration, bounded by conditions extracted from the circuit topology. Our approach can be applied to any combinational circuit, without a-priori knowledge of its functionality. The proposed algorithm allows back-tracking in order to never be trapped in local minima, and identifies the exact influence of each circuit gate on the output correctness, resulting in inexact circuits with higher efficiency and accuracy with respect to state-of-the-art greedy strategies.</td>
<td></td>
</tr>
<tr>
<td>09:15</td>
<td>5.7.3</td>
<td>ICNN: AN ITERATIVE IMPLEMENTATION OF CONVOLUTIONAL NEURAL NETWORKS TO ENABLE ENERGY AND COMPUTATIONAL COMPLEXITY AWARE DYNAMIC APPROXIMATION</td>
<td>Avesta Sasan, George Mason University, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Speaker:</strong> Avesta Sasan, George Mason University, US; Khatayoun Neshatpour, Farnaz Behnia, Houman Homayoun and Avesta Sasan, George Mason University, US</td>
<td><strong>Authors:</strong> Avesta Sasan, George Mason University, US; Khatayoun Neshatpour, Farnaz Behnia, Houman Homayoun and Avesta Sasan, George Mason University, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong> With Convolutional Neural Networks (CNN) becoming more of a commodity in the computer vision field, many have attempted to improve CNN in a bid to achieve better accuracy to a point that CNN accuracies have surpassed that of human's capabilities. However, with deeper networks, the number of computations and consequently the power needed per classification has grown considerably. In this paper, we propose iterative CNN (ICNN) by reformulating the CNN from a single feed-forward network to a series of sequentially executed smaller networks. Each smaller network processes a small set of sub-sampled input features and enhances the accuracy of the classification. Upon reaching an acceptable classification confidence, further possessing of smaller networks is discarded. The proposed network architecture allows the CNN function to be dynamically approximated by creating the possibility of early termination and performing the classification with far fewer operations compared to a conventional CNN. Initial results show that this iterative approach competes with the original larger networks in terms of accuracy while incurring far less computational complexity by detecting many images in early iterations.</td>
<td></td>
</tr>
<tr>
<td>09:30</td>
<td>5.7.4</td>
<td>TASK SCHEDULING FOR MANY-CORES WITH S-NUCA CACHES</td>
<td>Anuj Pathania and Joerg Henkel, Karlsruhe Institute of Technology, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Speaker:</strong> Anuj Pathania and Joerg Henkel, Karlsruhe Institute of Technology, DE; Katayoun Neshatpour, Farnaz Behnia, Houman Homayoun and Avesta Sasan, George Mason University, US</td>
<td><strong>Authors:</strong> Anuj Pathania and Joerg Henkel, Karlsruhe Institute of Technology, DE; Katayoun Neshatpour, Farnaz Behnia, Houman Homayoun and Avesta Sasan, George Mason University, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong> A many-core processor may comprise a large number of processing cores on a single chip. The many-core's last-level shared cache can potentially be physically distributed alongside the cores in the form of cache banks connected through a Network on Chip (NoC). Static Non-Uniform Cache Access (S-NUCA) memory address mapping policy provides a scalable mechanism for providing the cores quick access to the entire last-level cache. By design, S-NUCA introduces a unique topology-based performance heterogeneity and we introduce a scheduler that can exploit it. The proposed scheduler improves performance of the many-core by 9.93% in comparison to a state-of-the-art generic many-core scheduler with minimal run-time overheads.</td>
<td></td>
</tr>
<tr>
<td>09:45</td>
<td>5.7.5</td>
<td>KVSSD: CLOSE INTEGRATION OF LSM TREES AND FLASH TRANSLATION LAYER FOR WRITE-EFFICIENT KV STORE</td>
<td>Sung-Ming Wu, National Chiao-Tung University, TW</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Speaker:</strong> Sung-Ming Wu, National Chiao-Tung University, TW; Kai-Heang Lin and LiPin Chang, National Chiao-Tung University, TW</td>
<td><strong>Authors:</strong> Sung-Ming Wu, Kai-Heang Lin and LiPin Chang, National Chiao-Tung University, TW</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong> Log-Structured Merge (LSM) trees are a write-optimized data structure for lightweight, high-performance Key-Value (KV) store. Solid State Disks (SSDs) provide acceleration of KV operations on LSM trees. However, this hierarchical design involves multiple software layers, including the LSM tree, host file system, and Flash Translation Layer (FTL), causing cascading write amplifications. We propose KVSSD, a close integration of LSM trees and the FTL, to manage write amplifications from different layers. KVSSD exploits the FTL mapping mechanism to implement copy-free compaction of LSM trees, and it enables direct data allocation in flash memory for efficient garbage collection. In our experiments, compared to the hierarchical design, our KVSSD reduced the write amplification by 88% and improved the throughput by 347%.</td>
<td></td>
</tr>
</tbody>
</table>

Download Paper (PDF; Only available from the DATE venue WiFi)
STREAMFTL: STREAM-LEVEL ADDRESS TRANSLATION SCHEME FOR MEMORY CONSTRAINED FLASH STORAGE
Speaker: Dongkun Shin, Sungkyunkwan University, KR
Authors: Hyukjong Kim, Kyuhwa Han and Dongkun Shin, Sungkyunkwan University, KR
Abstract
Although much research efforts have been devoted to reducing the size of address mapping table which consumes DRAM space in solid state drives (SSDs), most SSDs still use page-level mapping for high performance in their firmware called flash translation layer (FTL). In this paper, we propose a novel FTL scheme, called StreamFTL. In order to reduce the size of the mapping table in SSDs, StreamFTL maintains a mapping entry for each stream, which consists of several logical pages written at contiguous physical pages. Unlike extent, which is used by previous FTL schemes, the logical pages in a stream do not need to be contiguous. We show that StreamFTL can reduce the size of the mapping table by up to 90% compared to page-level mapping scheme.

ONLINE CONCURRENT WORKLOAD CLASSIFICATION FOR MULTI-CORE ENERGY MANAGEMENT
Speaker: Karunakar Reddy Basireddy, University of Southampton, GB
Authors: Karunakar Reddy Basireddy1, Amit Kumar Singh2, Geoff V. Merrett1 and Bashir M. Al-Hashimi1
1University of Southampton, GB; 2University of Essex, GB
Abstract
Modern embedded multi-core processors are organized as clusters of cores, where all cores in each cluster operate at a common Voltage-frequency (V-f). Such processors often need to execute applications concurrently, exhibiting varying and mixed workloads (e.g. compute- and memory-intensive) depending on the instruction mix and resource sharing. Runtime adaptation is key to achieving energy savings without trading-off application performance with such workload variabilities. In this paper, we propose an online energy management technique that performs concurrent workload classification using the metric Memory Reads Per Instruction (MRPI) and pro-actively selects an appropriate V-f setting through workload prediction. Subsequently, it monitors the workload prediction error and performance loss, quantified by Instructions Per Second (IPS) at runtime and adjusts the chosen V-f to compensate. We validate the proposed technique on an Odroid-XU3 with various combinations of benchmark applications. Results show an improvement in energy efficiency of up to 69% compared to existing approaches.

Coffee Break in Exhibition Area

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

IP2 Interactive Presentations
Date: Wednesday, March 21, 2018
Time: 10:00 - 10:30
Location / Room: Conference Level, Foyer

Interactive Presentations run simultaneously during a 30-minute slot. Additionally, each IP paper is briefly introduced in a one-minute presentation in a corresponding regular session.
IN-GROWTH TEST FOR MONOLITHIC 3D INTEGRATED SRAM
Speaker: Yixun Zhang, Shanghai Jiao Tong University, CN
Authors: Pu Pang1, Yixun Zhang1, Tianjian Li1, Sung Kyu Lim2, Quan Chen3, Xiaoyao Liang3 and Li Jiang1
1Shanghai Jiao Tong University, CN; 2Georgia Tech, US
Abstract
Monolithic three-dimensional integration (3D-I) directly fabricates tiers of integrated circuits upon each other and provides millions of vertical interconnections with interlayer vias (ILVs). It thus brings higher integration density and communication capability compared with three-dimensional stacked integration (3D-Si). However, the Known-Good-Die problem haunting 3D-Si—a faulty tier causes the failure of the entire stack—also occurs in 3D-I. Lack of efficient test methodologies such as the pre-bond testing in 3D-Si, 3D-I may have a more significant yield drop and thus its cost may be unacceptable for mainstream adoption. This paper introduces a novel in-growth test method for 3D-I SRAM. We propose a novel Design-for-Test (DfT) methodology to enable the proposed in-growth test on cell-level partitioned incomplete SRAM cells. We also build a statistical model of cost and discover a prospective judgment to determine whether or not to stop the fabrication, in order to prevent from raising the cost of fabricating more tiers upon the imperable tiers. We find that a "sweet point" exists in the judgement, which can minimize the overall cost. Experimental results show the effectiveness of our proposed test methodology.
Download Paper (PDF; Only available from the DATE venue WiFi)

A CO-DESIGN METHODOLOGY FOR SCALABLE QUANTUM PROCESSORS AND THEIR CLASSICAL ELECTRONIC INTERFACE
Speaker: Jeroen van Dijk, Delft University of Technology, NL
Authors: Jeroen van Dijk1, Andrei Vladimirescu2, Masoud Babaie1, Edoardo Charbon1 and Fabio Sebastiani1
1Delft University of Technology, NL; 2University of California, Berkeley, US
Abstract
A quantum computer fundamentally comprises a quantum processor and a classical controller. The classical electronic controller is used to correct and manipulate the qubits, the core components of a quantum processor. To enable quantum computers scalable to millions of qubits, as required in practical applications, the simultaneous optimization of both the classical electronic and quantum systems is needed. In this paper, a co-design methodology is proposed for obtaining an optimized qubit performance while considering practical trade-offs in the control circuits, such as power consumption, complexity, and cost. The SPINE (SPIN Emulator) toolset is introduced for the co-design and co-optimization of electronic/quantum systems. It comprises a circuit simulator enhanced with a Verilog-A model emulating the quantum behavior of single-electron spin qubits. Design examples show the effectiveness of the proposed methodology in the optimization, design and verification of a whole electronic/quantum system.
Download Paper (PDF; Only available from the DATE venue WiFi)

APPROXIMATE QUATERNARY ADDITION WITH THE FAST CARRY CHAINS OF FPGAS
Speaker: Philip Brisk, University of California, Riverside, US
Authors: Sina Boroumand1, Hadi P. Ahn2 and Philip Brisk2
1University of Tehran, IR; 2Qualcomm Research, US; 3University of California, Riverside, US
Abstract
A heuristic is presented to efficiently synthesize approximate adder trees on Altera and Xilinx FPGAs using their carry chains. The mapper constructs approximate adder trees using an approximate quaternary adder as the fundamental building block. The approximate adder trees are smaller than exact adder trees, allowing more operators to fit into a fixed-area device, trading off arithmetic accuracy for higher throughput.
Download Paper (PDF; Only available from the DATE venue WiFi)

NN COMPACTOR: MINIMIZING MEMORY AND LOGIC RESOURCES FOR SMALL NEURAL NETWORKS
Speaker: Seongmin Hong, Hongik University, KR
Authors: Seongmin Hong1, Inho Lee1 and Yongjun Park2
1Hongik University, KR; 2Hanyang University, KR
Abstract
Special neural accelerators are an appealing hardware platform for machine learning systems because they provide both high performance and energy efficiency. Although various neural accelerators have recently been introduced, they are difficult to adapt to embedded platforms because current neural accelerators require high memory capacity and bandwidth for the fast preparation of synaptic weights. Embedded platforms are often unable to meet these memory requirements because of their limited resources. In FPGA-based IoT (Internet of things) systems, the problem becomes even worse since computation units generated from logic blocks cannot be fully utilized due to the small size of block memory. In order to overcome this problem, we propose a novel dual-track quantization technique to reduce synaptic weight width based on the magnitude of the value while minimizing accuracy loss. In this value-adaptive technique, large and small value weights are quantized differently. In this paper, we present a fully automatic framework called NN Compactor that generates a compact neural accelerator by minimizing the memory requirements of synaptic weights through dual-track quantization and minimizing the logic requirements of PUs with minimum recognition accuracy loss. For the three widely used datasets of MNIST, CNAE-9, and Forest, experimental results demonstrate that our compact neural accelerator achieves an average performance improvement of 6.4x over a baseline embedded system using minimal resources with minimal accuracy loss.
Download Paper (PDF; Only available from the DATE venue WiFi)

IMPROVING FAST CHARGING EFFICIENCY OF RECONFIGURABLE BATTERY PACKS
Speaker: Alexander Lamprecht, TUM CREATE, SG
Authors: Alexander Lamprecht1, Swaminathan Narayanaswamy2 and Sebastian Steinhorst2
1TUM CREATE, SG; 2Technical University of Munich, DE
Abstract
Recently, reconfigurable battery packs that can dynamically modify the electrical connection topology of their individual cells are gaining importance. While several circuit architectures and management algorithms are proposed in the literature, the electrical characteristics of the reconfiguration circuit architectures are not sufficiently studied so far. In this paper, we derive a detailed analytical model for a state-of-the-art reconfiguration architecture capturing the losses introduced by the parasitic resistances of the circuit components. For the first time, we propose a novel fast charging strategy using the reconfiguration architecture that significantly reduces the power losses in comparison to conventional battery packs. Moreover, using the analytical model, we highlight the challenges faced by existing reconfiguration architectures using state-of-the-art components and we derive the specifications for the switches which are essential for improving the energy efficiency of such reconfigurable battery packs.
Download Paper (PDF; Only available from the DATE venue WiFi)
CLOUD-ASSISTED CONTROL OF GROUND VEHICLES USING ADAPTIVE COMPUTATION OFFLOADING TECHNIQUES

Speaker:
Schehl Sami, General Motors R&D, Warren, MI 48090, US

Authors:
Shubham Rai1, Ansh Rupani2, Dennis Walter1, Michael Raitza1, André Heinzig1, Christian Mays1, Walter Weber4 and Akash Kumar1
1Technische Universität Dresden, DE; 2Technische Universität Dresden, DE; 3Birla Institute of Technology and Science Pilani, Hyderabad Campus, IN; 4NanMLab GmbH, DE; 5NanMLab gGmbH and CfAED, DE

Abstract
The existing approaches to design efficient safety-critical control applications is constrained by limited in-vehicle sensing and computational capabilities. In the context of automated driving, we argue that there is a need to leverage resources “out-of-the-vehicle” to meet the sensing and powerful processing requirements of sophisticated algorithms (e.g., deep neural networks). To realize the need, a suitable computation offloading technique that meets the vehicle safety and stability requirements, even in the presence of unreliable communication network, has to be identified. In this work, we propose an adaptive offloading technique for control computations into the cloud. The proposed approach considers both current network conditions and control application requirements to determine the feasibility of leveraging remote computation and storage resources. As a case study, we describe a cloud-based path following controller application that leverages crowdsensed data for path planning.
Abstract

Modern embedded multi-core processors are organized as clusters of cores, where all cores in each cluster operate at a common Voltage-frequency (V-f). Such processors often need to execute applications concurrently, exhibiting varying and mixed workloads (e.g. compute- and memory-intensive) depending on the instruction mix and resource sharing. Runtime adaptation is key to achieving energy savings without trading-off application performance with such workload variabilities. In this paper, we propose an online energy management technique that performs concurrent workload classification using the metric Memory Reads Per Instruction (MRPI) and pro-actively selects an appropriate V-f setting through workload prediction. Subsequently, it monitors the workload prediction error and performance loss, quantified by Instructions Per Second (IPS) at runtime and adjusts the chosen V-f to compensate. We validate the proposed technique on an Odroid-XU3 with various combinations of benchmark applications. Results show an improvement in energy efficiency of up to 69% compared to existing approaches.

Download Paper (PDF; Only available from the DATE venue WiFi)
Logic encryption is a hardware security technique that uses extra key inputs to prevent unauthorized use of a circuit. With the discovery of the SAT-based attack, new encryption techniques such as SARLock and Anti-SAT are proposed, and further combined with traditional logic encryption techniques, to guarantee both high error rates and resilience to the SAT-based attack. In this paper, the SAT-based bit-flipping attack is presented. It first separates the two groups of keys via SAT-based bit-flippings, and then attacks the traditional encryption and the SAT-resilient encryption, by conventional SAT-based attack and by-passing attack, respectively. The experimental results show that the bit-flipping attack successfully returns a circuit with the correct functionality and significantly reduces the execution time compared with other advanced attacks.

Download Paper (PDF; Only available from the DATE venue WiFi)

AMS VERIFICATION METHODOLOGY REGARDING SUPPLY MODULATION IN RF SOCs INDUCED BY DIGITAL STANDARD CELLS

Speaker: Fabian Speicher, RWTH Aachen University, DE
Authors: Fabian Speicher, Jonas Meier, Soheil Aghaie, Ralf Wunderlich and Stefan Heinen, RWTH Aachen University, DE

Abstract
Nanoscale CMOS enables and forces the use of digital-centric RF architectures, where timing resolution is traded for analog resolution. Simultaneously, digital circuits act as aggressors endangering the performance of the time continuous digital and analog parts. The switching activities of logic cells result in power supply variations which lead to jitter in the digital signal paths and causes interference coupling to the analog paths, appearing as e.g. phase noise, crosstalk, unwanted frequency conversion, etc. Since today's commonly used AMS simulation methods are limited to register-transfer level (RTL) models for the digital domain, the electrical behavior caused by digital switching is not considered. Here, a method for modeling logic cells with regard to power supply noise is presented using the available characterization data of a standard cell library. It covers the influence of switching on the supply voltage as well as influences of supply variations on the digital path delay and their feedthrough to blocks of the RF domain. A fast event-driven simulation of an entire AMS system regarding the mentioned aspects is enabled. The method is demonstrated on a digital-centric transmitter to detect the effects on system level.

Download Paper (PDF; Only available from the DATE venue WiFi)
VIRTUAL PROTOTYPE MAKANI: ANALYZING THE USAGE OF POWER MANAGEMENT TECHNIQUES AND EXTRA-FUNCTIONAL PROPERTIES BY USING VIRTUAL PROTOTYPING

Author: Sören Schreiner, OFFIS – Institute for Information Technology, DE

Abstract
My PhD work consists of analyzing the correct usage of power management techniques, as well as the analysis of extra-functional properties, including power and timing properties, in MPSoCs. Especially in safety-critical environments the power management gets safety-critical too, since it is able to influence the overall system behavior. To demonstrate my methodologies a mixed-critical multi-rocket system and its corresponding virtual prototype is used. The multi-rocket system's avionics is served by a Xilinx Zynq 7000 MPSoC. The hardware architecture includes ARM and MicroBlaze cores, a NoC for communication and peripherals. The MPSoC processes the flight algorithms with triple modular redundancy and a mission-critical video processing task. The virtual prototype consists of a virtual platform and an environmental model. The virtual platform is equipped with my measuring tool libraries to generate traces of the observed power management techniques and the extra-functional properties.

More information ...

SPANNER: SELF-REPAIRING SPIKING NEURAL NETWORK CONTROLLER FOR AN AUTONOMOUS ROBOT

Authors: Alan Milani1, Anju Johnson1, James Hilder1, David Halliday1, Andy Tyrrell1, Jon Timmis1, Junxiu Liu2, Shvan Karim1,2 and Jim Harkin1,2
1University of York, GB; 2Ulster University, GB

Abstract
The human brain is remarkably resilient, and is able to self-repair following injury or a stroke. In contrast, electronic systems typically exhibit limited self-repair capabilities, and cannot recover from faults. We demonstrate a bio-inspired approach to self-repair that allows an autonomous robot to recover from faults in its artificial brain. Astrocytes are support cells in the human brain that interact with neurons to regulate synaptic activity. We have modelled this interaction to create a spiking neural network that can self-repair when synapses between neurons are damaged, by strengthening redundant pathways. We demonstrate a robot platform controlled by a self-repairing spiking neural network that is implemented on an FPGA. We demonstrate that injecting faults into the synapses of the network initially causes the robot to behave erratically, but that the neural controller is able to automatically repair itself, thus allowing the robot to resume normal function.

More information ...

EVOAPPROX: LIBRARIES AND GENERATORS OF APPROXIMATE CIRCUITS

Authors: Lukas Sekanina, Zdeněk Vašíček and Vojtěch Mrazek, Brno University of Technology, CZ

Abstract
Our contribution deals with a fully automated functional approximation methodology for combinational digital circuits. We present open libraries of approximate circuits and tools performing desired approximations. Our approach uses a multi-objective genetic programming-based method to automatically design approximate k-bit adders and multipliers (k = 8, 12, 16, 32). All circuits can be downloaded from [1] at the level of a source code (C, Verilog, and Matlab). Several error metrics are pre-calculated and formal guarantees are given in terms of these errors. By means of an interactive web interface the user can easily find the best trade-off between the error and electrical parameters provided for 45/90/180 nm technology process. We will also demonstrate the circuit design flow developed. References: [1] http://www.fkt.vutbr.cz/research/groups/ehw/approx/

More information ...

REPABIT: AUTOMATED DESIGN OF RELOCATABLE PARTIAL BITSTREAMS

Authors: Jens Rettkowski1 and Diana Göhringer2
1Ruhr-University Bochum, DE; 2Technische Universität Dresden, DE

Abstract
Dynamic partial reconfiguration of FPGAs enables the replacement of hardware modules at runtime without disturbing remaining hardware modules. The standard vendor tools generate a specific partial bitstream for each reconfigurable region. Relocation generates a partial bitstream in such a way, that it can be moved to different regions. Hence, the number and the time to generate bitstreams is reduced. In this work, RepaBit is presented that automates the generation of relocatable partial bitstreams for Xilinx Vivado. TCL scripts check the compatibility of resource footprints and arrange identical partition pins in all regions for the connection of relocatable modules with the remaining design. Feedthrough routes are avoided using the isolation design flow from Xilinx. The results show an overhead of LUTs by 0.7% and a frequency reduction of only 1.5%. Nevertheless, RepaBit simplifies the design and reduces the design time as well as the needed memory for storing the partial bitstreams.

More information ...

ULITAG: CONCURRENT IJTAG DEMONSTRATOR

Author: Krenz-Baath René, Hamm-Lippstadt University of Applied Sciences, DE

Abstract
The flexibility of on-chip instrument access enabled by IEEE 1687 (ULITAG) has shown tremendous improvements in modern industrial designs. Due to a constantly increasing spectrum of tasks performed through 1687 networks such as performing test operations during production test, on-line test operations as well as operating health monitors the test requirements in modern designs increase dramatically with respect to test performance, responsiveness and low power. These requirements have a major impact on the design of such test infrastructures. In complex designs with large test infrastructures it might be challenging to comply with the large spectrum of requirements. Concurrent ULITAG is novel partitioning concept to a reconﬁgurable test infrastructure in order to enable an independent operation of different sections of the test infrastructure. The proposed demonstrator shows the first FPGA-based implementation of concurrent ULITAG test infrastructures.

More information ...

EXPERIENCE-BASED AUTOMATION OF ANALOG IC DESIGN

Authors: Florian Leber and Juergen Scheible, Reutlingen University, DE

Abstract
While digital design automation is highly developed, analog design automation still remains behind the demands. Previous circuit synthesis approaches, which are usually based on optimization algorithms, do not satisfy operational requirements. A promising alternative is given by procedural approaches (also known as "generators"). They (a) emulate experts' decisions, thus (b) make expert knowledge reusable and (c) consider all relevant aspects and constraints implicitly. Nowadays, generators are successfully applied in analog layout (Poells, Pycells). We aim at an entire design flow completely based on procedural automation techniques. This flow will consist of procedures for the generation of schematics and layouts for every typical analog circuit class, such as amplifier, bandgap, filter a.s.o. In our presentation we give an overview on such a design flow and we show an approach for capturing an analog circuit designer's strategy as an executable "expert design plan".

More information ...

End of session
Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

6.1 Special Day Session on Future and Emerging Technologies: Transistors for Digital NanoSystems: The Road Ahead

Date: Wednesday, March 21, 2018
Time: 11:00 - 12:30
Location / Room: Saal 2

Chair:
Aitken Rob, ARM, US, Contact Robert Aitken

This session presents energy-efficient digital design approaches using new transistor ideas and their experimental demonstrations. Examples include negative capacitance-based gate control, carbon nanotube-based channels, and polarity-control by design.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:00</td>
<td>6.1.1</td>
<td>NEGATIVE CAPACITANCE TRANSISTORS</td>
<td>Michael Hoffmann, NaMLab gGmbH, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A transistor looking from the gate essentially acts as a series combination of two</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>capacitors: the gate oxide capacitor and the channel capacitor. When the gate</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>oxide is an appropriate ferroelectric, this series combination can stabilize the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ferroelectric material at a state of negative capacitance. At this state, the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>total capacitance of the series combination is enhanced, leading to more charge</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>at the channel at the same voltage. This boost of charge, in turn, leads to larger</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>current at the same voltage. In fact, this boost makes it possible to reduce</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>supply voltage of transistors below the traditional Boltzmann limit — often</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>termed as the Boltzmann tyranny. In the recent years, many groups around the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>world, both in academy and in the industry, have demonstrated the fundamental</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>effect and the Negative Capacitance Transistors. In this work, we shall describe</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>the physical origin of the negative capacitance effect and our current</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>understanding of the recent experimental work. We shall also discuss potential</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ways to optimize devices that could lead to significant improvement in energy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>efficiency for next generation computers.</td>
<td></td>
</tr>
<tr>
<td>11:30</td>
<td>6.1.2</td>
<td>CARBON NANOTUBE FILM-BASED CMOS AND OPTOELECTRONIC DEVICES AND INTEGRATED SYSTEMS</td>
<td>Lian-Mao Peng, Peking University, CN</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Carbon nanotube (CNT)-based electronics has been considered one of the most</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>promising candidates to replace Si complementary metal-oxide-semiconductor (CMOS)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>technology, which will soon meet its performance limit. Prototype device studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>on individual CNTs revealed that CNT based devices have the potential to outperform</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Si CMOS technology in both performance and power consumption, especially at sub-10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>nm technology nodes, which are close to the theoretical limits; and various</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>optoelectronic device such as light-emitting diodes, photodetectors and photovoltaic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(PV) cells have been demonstrated. In this talk, I will discuss the use of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>randomly oriented CNT film to build CNT CMOS and optoelectronic devices, and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>show that the performance of CNT film devices and systems can be dramatically</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>improved by optimizing the material purity, device structure and fabrication</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>processes, thus yielding CNT devices with outstanding performance comparable to</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>that of Si CMOS and ICs working in the Grhio regime, and integrated electronic and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>optoelectronic systems for communications between nanoelectronic circuits using</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CNT devices.</td>
<td></td>
</tr>
</tbody>
</table>
Towards high-performance polarity-controllable FETs with 2D materials

Speaker:
Pierre-Emmanuel Gaillardon, University of Utah, US

Authors:
Giovanni V. Resta1, Jorge Romero Gonzalez2, Yashwanth Balaji3, Tarun Kumar Agarwal3, Dennis Lin3, Francy Catthoor3, Iuliana P. Radu3, Giovanni De Micheli1 and Pierre-Emmanuel Gaillardon4

1Integrated System Laboratory – EPFL, CH; 2Laboratory of NanoIntegrated Systems (LNIS), Department of Electrical and Computer Engineering, University of Utah, US; 3IMEC, BE; 4University of Utah, US

Abstract
As scaling of conventional silicon-based electronics is reaching its ultimate limit, two-dimensional semiconducting materials of the transition-metal-dichalcogenides family, such as MoS2 and WSe2, are considered as viable candidates for next-generation electronic devices. Fully relying on electrostatic doping, polarity-controllable devices, that use additional gate terminals to modulate the Schottky barriers at source and drain, can strongly take advantages of 2D materials to achieve high on/off ratio and low leakage floor. Here, we provide an overview of the latest advances in 2D material processes and growth. Then, we report on the experimental demonstration of polarity-controllable devices fabricated on 2D-WSe2 and study the scaling trends of such devices using ballistic self-consistent quantum simulations. Finally, we discuss the circuit-level opportunities of such technology.

Download Paper (PDF; Only available from the DATE venue WiFi)

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

6.2 Memory Security

Date: Wednesday, March 21, 2018
Time: 11:00 - 12:30
Location / Room: Kont. 6

Chair:
Francesco Regazzoni, ALaRI USI, CH, Contact Francesco Regazzoni

Co-Chair:
Todd Austin, University of Michigan, US, Contact Todd Austin

Papers in this session address the problem of dealing with secure memory architectures and cover the whole memory hierarchy from cache to storage. Different levels of the hierarchy need different protection mechanisms, such as: protecting information from attacks on untrusted clouds and ensuring integrity of the main memory used by the CPU. Finally, the last paper identifies cache attacks and vulnerabilities on MPSoCs using Networks on Chips.

Time Label Presentation Title Authors
<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 11:00 | 6.2.1 | DYNAMIC SKEWED TREE FOR FAST MEMORY INTEGRITY VERIFICATION | Saru Vign, Nanyang Technological University, SG  
Authors: Saru Vign, Jiang Guijuan and Lam Siew Kei, Nanyang Technological University, SG  
Abstract  
Memory authentication techniques often employ an integrity tree as a countermeasure against replay, spoofing and splicing attacks. However, the balanced memory integrity trees used in existing approaches lead to excessive memory access overheads for runtime verification. In this paper, we propose a framework to dynamically construct a customized integrity tree based on the data access patterns to reduce the overhead of runtime verification. The proposed framework can adapt the memory integrity tree structure at runtime such that the nodes that correspond to frequently accessed data are placed closer to the root. We have validated the effectiveness of our approach on the Altera NIOS II processor with an external DRAM. Experimental results based on applications from widely used CHStone and SNU Real-Time benchmarks demonstrate that the proposed approach can lead to an average performance gain of 30% compared to the conventional means of using balanced memory integrity trees. In addition, to preserve data confidentiality, we implemented the encryption/decryption operations using custom instructions on the NIOS II processor to notably reduce the overall overhead of memory security. Download Paper (PDF; Only available from the DATE venue WiFi) |
| 11:30 | 6.2.2 | EARTHQUAKE - A NOC-BASED OPTIMIZED DIFFERENTIAL CACHE-COLLISION ATTACK FOR MPSOCs | Cezar Rodolfo W. Reinbrecht, UFRGS, BR  
Authors: Cezar Rodolfo Wedig Reinbrecht,1 Bruno Endres Forlin, Andreas Zankl, and Alvin Glova  
Abstract  
Multi-Processor Systems-on-Chips (MPSoCs) are a platform for a wide variety of applications and use-cases. The high on-chip connectivity, the programming flexibility, and the reuse of IPs, however, also introduce security concerns. Problems arise when applications with different trust and protection levels share resources of the MPSoC, such as processing units, cache memories, and the Network-on-Chip (NoC) communication structure. If a program gets compromised, an adversary can observe the use of these resources and infer (potentially secret) information from other applications. In this work, we explore the cache-based attack by Bogdanov et al., which infers the cache activity of a target program through timing measurements and exploits collisions that occur when the same cache location is accessed for different program inputs. We implement this differential cache-collision attack on the MPSoC Glass and introduce an optimized variant of it, the Earthquake Attack, which leverages the NoC-based communication to increase attack efficiency. Our results show that Earthquake performs well under different cache line and MPSoC configurations, illustrating that cache-collision attacks are considerable threats on MPSoCs. Download Paper (PDF; Only available from the DATE venue WiFi) |
| 12:00 | 6.2.3 | A FAST AND RESOURCE EFFICIENT FPGA IMPLEMENTATION OF SECRET SHARING FOR STORAGE APPLICATIONS | Jakob Stangl, Austrian Institute of Technology (AIT), AT  
Authors: Jakob Stangl, Thomas Lorünser, and Sai Dinaikarao  
Abstract  
Outsourcing data into the cloud gives wide benefits and opportunities to customers. Besides these advantages, new challenges such as confidentiality and accessibility have to be addressed. One approach to overcome these challenges is by applying secret sharing in a distributed storage setting, known as cloud of clouds approach. For this purpose we present a new hardware architecture of a wide parametrizable secret sharing core. Performance metrics for various applied bit-widths of secret words are given, which are crucial for benefits of higher level protocols in the cloud of clouds approach. Additionally, a complete system which is able to operate in a network environment is presented. The achieved throughputs are in the order of Gbit/s. It is significantly faster than similar comparable hardware architectures and orders of magnitude higher than software implementations. Download Paper (PDF; Only available from the DATE venue WiFi) |
| 12:30 | IP2-15, 958 | AIM: FAST AND ENERGY-EFFICIENT AES IN-MEMORY IMPLEMENTATION FOR EMERGING NON-VOLATILE MAIN MEMORY | Jingtong Hu, University of Pittsburgh, US  
Authors: Mimi Xie,1 Shuangchen Li,2 Amin Rezaei,3 Hai Zhou, Northwood University, US  
Abstract  
Non-volatile main-memory-based systems pose an opportunity for an attacker to ready access sensitive information on the memory because of its long retention time. While real-time memory encryption with dedicated AES engine can address this vulnerability, it incurs extra performance and energy overheads. As an alternative, we propose an AES in-memory implementation, AIM, to encrypt the whole part of the memory only when it is necessary. We leverage the benefits offered by the in-memory computing architecture to address the challenges of the bandwidth intensive encryption application. We take advantage of NVM’s intrinsic logic operation capability to implement the AES task. Embracing the massive parallelism inside the memory, AIM outperforms existing mechanisms with higher throughput yet lower energy consumption. Compared with state-of-the-art AES engine running at 2.1GHz, AIM can speed up the encryption process by 80 times for a 1GB NVM. Download Paper (PDF; Only available from the DATE venue WiFi) |
| 13:30 | IP2-16, 748 | SAT-BASED BIT-FLIPPING ATTACK ON LOGIC ENCRYPTIONS | Yuangang Wang,1 Andreas Zankl, and Johanna Sepulveda  
Authors: Yuangang Wang,1 Andreas Zankl, and Johanna Sepulveda  
Abstract  
Logic encryption is a hardware security technique that uses extra key inputs to prevent unauthorized use of a circuit. With the discovery of the SAT-based attack, new encryption techniques such as SARLock and Anti-SAT are proposed, and further combined with traditional logic encryption techniques, to guarantee both high error rates and resilience to the SAT-based attack. In this paper, the SAT-based bit-flipping attack is presented. It first separates the two groups of keys via SAT-based bit-flippings, and then attacks the traditional encryption and the SAT-resistant encryption, by conventional SAT-based attack and by-passing attack, respectively. The experimental results show that the bit-flipping attack successfully returns a circuit with the correct functionality and significantly reduces the execution time compared to other advanced attacks. Download Paper (PDF; Only available from the DATE venue WiFi) |
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (Lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

6.3 Advances in AMS/RF Design & Test Automation and Beyond

This session brings together new design and test automation developments for AMS/RF systems and beyond. Papers in the session cover a wide range of exciting topics from circuit optimization to design tools and verification. The topics include innovative combination of principal component analysis and evolutionary computation applied to analog/RF IC optimization; hybrid automation approach for SAR ADC design aimed at IoT applications; and design space exploration for wireless systems. Interactive papers discuss AMS circuit testbenches, modeling and simulation of systems that combine continuous and discrete time components, and AMS verification.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:00</td>
<td>6.3.1</td>
<td>ENHANCED ANALOG AND RF IC SIZING METHODOLOGY USING PCA AND NSGA-II OPTIMIZATION KERNEL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Nuno Lourenco, Instituto de Telecomunicações, PT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors: Tiago Pessoa, Nuno Lourenco, Ricardo Martins, Ricardo Povea and Nuno Horta, Instituto de Telecomunicações, Instituto Superior Técnico – Universidade de Lisboa, PT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: State-of-the-art design of analog and radio frequency integrated circuits is often accomplished using sizing optimization. In this paper, an innovative combination of principal component analysis (PCA) and evolutionary computation is used to increase the optimizer’s efficiency. The adopted NSGA-II optimization kernel is improved by applying the genetic operators of mutation and crossover on a transformed design-space, obtained from the latest set of solutions (the parents) by PCA. By applying crossover and mutation on variables that are projections of the principal components, the optimization moves more effectively, finding solutions with better performances, in the same amount of time, than the standard NSGA-II optimization kernel. The proposed method was validated in the optimization of two widely used analog circuits, an amplifier and a voltage controlled oscillator, reaching wider solutions sets, and in some cases, solutions sets that can be almost 3 times better in terms of hypervolume.</td>
</tr>
<tr>
<td>11:30</td>
<td>6.3.2</td>
<td>A SYSTEMIC-BASED SIMULATOR FOR DESIGN SPACE EXPLORATION OF SMART WIRELESS SYSTEMS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Gabriele Morandi, University of Verona, IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors: Gabriele Morandi, Francesco Stefanni, Federico Fraccaroli and Davide Quaglia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1University of Verona, IT; 2REDaLab s.r.l., IT; 3Wagoo Italia s.r.l.s., IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Smart wireless techniques are at the core of many today’s telecommunication and networked embedded systems where performance are enhanced by intertwining radio frequency (RF) and digital aspects. Therefore their design requires to focus on both domains. Traditional approaches for their simulation rely either on different domain-specific tools or on analog-mixed-signal modeling languages. In the former case, the simulation of the whole platform in the same session is not possible while in the latter case, simulation performance are limited by the computationally most intensive domain (usually RF). We present an extension of the SystemC Network Simulation Library that allows to simulate antenna details and node position together with digital hardware and software. The validation on a real wearable system shows that the proposed simulation approach achieves a good trade-off between accuracy and speed thus allowing fast exploration of various configurations in the early phase of the design flow without recurring to the expensive and time-consuming creation of physical prototypes.</td>
</tr>
</tbody>
</table>

Download Paper (PDF; Only available from the DATE venue WiFi)
An algebra for modeling continuous time systems

Speaker: José Medeiros, University of Brasília, BR

Authors: José E. G. de Medeiros¹, George Ungureanu² and Ingo Sander²

¹University of Brasília, BR; ²KTH Royal Institute of Technology, SE

Abstract

Advancements on analog integrated design have led to new possibilities for complex systems combining both continuous and discrete time modules on a single chip. However, this also increases the complexity of design flow and need for automation in order to address the challenges between the two domains, as the interactions between them should be better understood. We believe that a common language for describing continuous and discrete time compositions is beneficial for such a goal and a step towards it is to gain insight and describe more fundamental building blocks. In this work we present an algebra based on the General Purpose Analog Computer, a theoretical model of computation recently updated as a continuous time equivalent of the Turing Machine.

Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

6.4 Modeling, Control and Scheduling for Cyber-Physical Systems

Date: Wednesday, March 21, 2018
Time: 11:00 - 12:30
Location / Room: Konf. 2

Chair: Shiyan Hu, Michigan Tech., US, Contact Shiyan Hu
Co-Chair: Franco Fummi, University of Verona, IT, Contact Franco Fummi

The fast advancement of cyber-physical systems has been presenting significant design challenges. The papers in this session address these CPS design challenges across layers of control, communication, computation and embedded microarchitecture. They include methodologies for modeling and integrating heterogeneous models to build CPS virtual platforms, routing and scheduling messages with control stability consideration for networked CPS, designing feedback control of EtherCAT networks for reliability enhancement, and scheduling tasks with consideration of cache to maximize control performance.

### 6.4.1 Automatic Integration of Cycle-Accurate Descriptions with Continuous-Time Models for Cyber-Physical Virtual Platforms

**Speaker:** Rouhollah Mahfouzi, Linköping University, SE
**Authors:**
- Rouhollah Mahfouzi2, Amir Aminifar2, Soheil Samii3, Ahmed Rezine1, Petru Ele1 and Zebo Peng3
- 1Linköping University, SE; 2Swiss Federal Institute of Technology in Lausanne (EPFL), CH; 3General Motors Research & Development, US

**Abstract**

Development of cyber-physical systems’ control algorithms usually relies on architecture-agnostic abstract models, often leading to ineffective implementations. This paper presents a technique to automatically integrate cycle-accurate models of digital HW components with continuous-time physical models. It proposes a solution to the semantic gap between the involved models of computation. Furthermore, model generation and integration for both Simulink-based proprietary environment and FMI-based portable standard are presented. The aim of such techniques is to produce cyber-physical virtual platforms: a powerful tool to refine control algorithms up to their SW implementations on the actual HW platform.

Download Paper (PDF; Only available from the DATE venue WiFi)

### 6.4.2 Stability-Aware Integrated Routing and Scheduling for Control Applications in Ethernet Networks

**Speaker:** Rouhollah Mahfouzi, Linköping University, SE
**Authors:**
- Rouhollah Mahfouzi1, Amir Aminifar2, Soheil Samii3, Ahmed Rezine1, Petru Ele1 and Zebo Peng3
- 1Linköping University, SE; 2Swiss Federal Institute of Technology in Lausanne (EPFL), CH; 3General Motors Research & Development, US

**Abstract**

Real-time communication over Ethernet is becoming important in various application areas of cyber-physical systems such as industrial automation and control, avionics, and automotive networking. Since such applications are typically time-critical, Ethernet technology has been enhanced to support time-driven communication through the IEEE 802.1 TSN standards. The performance and stability of control applications is strongly impacted by the timing of the network communication. Thus, in order to guarantee stability requirements, when synthesizing the communication schedule and routing, it is needed to consider the degree to which control applications can tolerate message delays and jitters. In this paper, we jointly solve the message scheduling and routing problem for networked cyber-physical systems based on the time-triggered Ethernet TSN standards. Moreover, we consider this communication synthesis problem in the context of control applications and guarantee their worst-case stability, taking explicitly into consideration the impact of communication delay and jitter on control quality. Considering the inherent complexity of the network communication synthesis problem, we also propose new heuristics to improve synthesis efficiency without any major loss of quality. Experiments demonstrate the effectiveness of the proposed solutions.

Download Paper (PDF; Only available from the DATE venue WiFi)
12:00 6.4.3 FEEDBACK CONTROL OF REAL-TIME ETHERCAT NETWORKS FOR RELIABILITY ENHANCEMENT IN CPS
Speaker: Tongquan Wei, East China Normal University, CN
Authors: Lijing Li¹, Peijin Cong², Junlong Zhuo¹, Tongquan Wei¹, Mingsong Chen¹ and Xiaobo Sharan Hu³
¹East China Normal University, CN; ²Nanjing University of Science and Technology, CN; ³Université de Notre Dame, US
Abstract: EtherCAT has become one of the leading real-time Ethernet solutions for networked industrial systems where a reliable communication infrastructure is needed due to highly error-prone environments. However, existing work on EtherCAT mainly focuses on clock synchronization and timeliness improvement. The reliability of EtherCAT-based networked systems has largely been ignored. In this paper, we present a PID-based feedback control scheme that aims at enhancing reliability of networked systems under timing and system resource constraints. Instead of automatic repeat request method (ARQ), a forward error control technique is introduced to achieve the required system reliability at a lower deadline miss rate of messages. The PID-based feedback control scheme can also improve the stability of a system in terms of deadline miss rate in the presence of bursty errors. Simulation results show that the proposed scheme can achieve reliability enhancement of up to 79% compared to benchmarking methods.
Download Paper (PDF; Only available from the DATE venue WiFi)

12:15 6.4.4 CACHE-AWARE TASK SCHEDULING FOR MAXIMIZING CONTROL PERFORMANCE
Speaker: Wanli Chang, Singapore Institute of Technology, SG
Authors: Wanli Chang¹, Debayan Roy², Xiaobo Sharan Hu³ and Samarjit Chakravarty²
¹Singapore Institute of Technology, SG; ²Technical University of Munich, DE; ³University of Notre Dame, US
Abstract: Embedded control applications are widely implemented on small, low-cost and resource-constrained microcontrollers, e.g., in the automotive domain. Conventionally, control algorithms are designed using model-based approaches, without considering the details of the implementation platform. This leads to inefficient utilization of the resources.
With the emergence of the cyber-physical system (CPS)-oriented thinking, there has lately been a strong interest in co-design of control algorithms and their implementation platforms. Some recent efforts have shown that a schedule on multiple applications with more on-chip cache reuse is able to improve the control performance. However, it has not been studied how the control performance can be maximized for a given schedule and how an optimal schedule can be computed. In this work, we propose a two-stage framework to compute the schedule maximizing the overall control performance of all the applications. First, a holistic controller design taking all the sampling periods and sensing-to-actuation delays in a schedule into account is presented, aiming to maximize the overall control performance. Second, a hybrid search algorithm for discrete decision space is reported to efficiently compute an optimal schedule. Experimental results on a case study with multiple automotive applications show that a significant improvement of 10-20% in control performance can be achieved by the proposed cache-aware scheduling approach.
Download Paper (PDF; Only available from the DATE venue WiFi)

12:30 135 TTW: A TIME-TRIGGERED WIRELESS DESIGN FOR CPS
Speaker: Romain Jacob, ETH Zurich, CH
Authors: Romain Jacob¹, Leong Zhang², Marco Zimmerling³, Jan Beutel¹, Samarjit Chakravarty² and Lothar Thiele¹
¹ETH Zurich, CH; ²Technical University of Munich, DE; ³Technische Universität Dresden, DE
Abstract: Wired fieldbuses have long been proven effective in supporting Cyber-Physical Systems (CPS). However, various domains are now striving for wireless solutions due to ease of deployment or novel functionality requiring the ability to support mobile devices. Low-power wireless protocols have been proposed in response to this need, but requirements of a large class of CPS applications can still not be satisfied. We thus propose Time-Triggered Wireless (TTW), a distributed low-power wireless system design that minimizes communication energy consumption and offers end-to-end timing predictability, runtime adaptability, reliability, and low latency. Evaluation shows a 2x reduction in communication latency and 33-40% lower radio-on time compared with DRP, the closest related work, validating the suitability of TTW for new exciting wireless CPS applications.
Download Paper (PDF; Only available from the DATE venue WiFi)

12:31 429 PHYLAX: SNAPSHOOT-BASED PROFILING OF REAL-TIME EMBEDDED DEVICES VIA JTAG INTERFACE
Speaker: Eduardo Chielle, New York University Abu Dhabi, BIR
Authors: Charalampos Konstantinou¹, Eduardo Chielle² and Michail Maniatakos²
¹New York University, US; ²New York University Abu Dhabi, AE
Abstract: Real-time embedded systems play a significant role in the functionality of critical infrastructure. Legacy microprocessor-based embedded systems, however, have not been developed with security in mind. Applying traditional security mechanisms in such systems is challenging due to computing constraints and/or real-time requirements. Their typical 20-30 year lifespan further exacerbates the problem. In this work, we propose PHYLAX, a plug-and-play solution to detect intrusions in already installed embedded devices. PHYLAX is an external monitoring tool which does not require code instrumentation. Also, our tool adapts and prioritizes intrusion detection based on the requirements of the underlying infrastructure (power grid, chemical factory, etc.) as well as the computing capabilities of the target embedded system (CPU model, memory size, etc.). PHYLAX can be employed on any legacy device which incorporates a JTAG interface. As a case study, we present the inclusion of PHYLAX on a power grid recloser controller.
Download Paper (PDF; Only available from the DATE venue WiFi)

12:32 463 CHARACTERIZING DISPLAY QOS BASED ON FRAME DROPPING FOR POWER MANAGEMENT OF INTERACTIVE APPLICATIONS ON SMARTPHONES
Speaker: Chung-Ta King, National Tsing Hua University, TW
Authors: Kuang-Ting Ho¹, Chung-Ta King¹, Bhaskar Das¹ and Yung-Ju Chang²
¹National Tsing Hua University, TW; ²National Chiao Tung University, TW
Abstract: User-centric power management in smartphones aims to conserve power without affecting user perceived quality of experience. Most existing works focus on periodically updated applications such as games and video players and use a fixed frame rate, measured in frame per second (FPS), as the metric to quantify the display quality of service (QoS). The idea is to adjust the CPU/GPU frequency just enough to maintain the frame rate at a user satisfactory level. However, when applied to aperiodically updated interactive applications, e.g. Facebook or Instagram, that draw the frame buffer at a varying rate in response to user inputs, such a power management strategy becomes too conservative. Based on real user experiments, we observe that users can tolerate a certain percentage of frame drops when running aperiodically updated applications without affecting their perceived display quality. Hence, we introduce a new metric to characterize display quality of service, called the frame drawn ratio (FDR), and propose a new CPU/GPU frequency governor based on the FDR metric. The experiments by real users show that the proposed governor can conserve 17.2% power in average when compared to the default governor, while maintaining the same or even better QoS rating.
Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018

- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

6.5 Special Session: Three Years of Low-Power Image Recognition Challenge

Date: Wednesday, March 21, 2018
Time: 11:00 - 12:30
Location / Room: Konf. 3

Chair: Yung-Hsiang Lu, Purdue University, US, Contact Yung-Hsiang Lu

Reducing power consumption has been one of the most important goals since the creation of electronic systems. Energy efficiency is increasingly important as battery-powered systems equipped with cameras (such as drones and body cameras) are widely used. It is desirable using the on-board computers to recognize objects in the images captured by these cameras. The Low-Power Image Recognition Challenge (LPIRC) is an annual competition started in 2015. LPIRC considers both energy consumption and the accuracy in detecting and locating objects in images. The special session includes presentations given by the winners of the first three years of LPIRC explaining their winning solutions.

6.5.1 THREE YEARS OF LOW-POWER IMAGE RECOGNITION CHALLENGE: INTRODUCTION TO SPECIAL SESSION

Speaker: Kent Gauen1, Ryan Daley1, Yung-Hsiang Lu1, Eunbyung Park2, Wei Lu3, Alexander Berg2 and Yiran Chen2

Authors: 1Purdue University, US; 2University of North Carolina at Chapel Hill, US; 3Duke University, US

Abstract: Reducing power consumption has been one of the most important goals since the creation of electronic systems. Energy efficiency is increasingly important as battery-powered systems equipped with cameras are widely used. It is desirable using the on-board computers to recognize objects in the images captured by these cameras. The Low-Power Image Recognition Challenge (LPIRC) is an annual competition started in 2015. The special session includes presentations given by the winners of the first three years of LPIRC. This paper explains the rules of the competition and the rationale, summarizes the teams’ scores, and describes the lessons learned in the first three years. The paper suggests possible improvements of future challenges.

Download Paper (PDF; Only available from the DATE venue WiFi)

6.5.2 REAL-TIME OBJECT DETECTION TOWARDS HIGH POWER EFFICIENCY

Speaker: Jincheng Yu, Tsinghua University, CN

Authors: Jincheng Yu1, Kaiyuan Guo1, Yiming Hu1, Xuefeng Ning1, Jiantao Qiu1, Huizhi Mao1, Song Yao6, Tianqi Tang1, Boxun Li1, Yu Wang1 and Huazhong Yang1

Abstract: In recent years, Convolutional Neural Network (CNN) has been widely applied in computer vision tasks and has achieved significant improvement in image object detection. The CNN methods consume more computation as well as storage, so GPU is introduced for real-time object detection. However, due to the high power consumption of GPU, it is difficult to adopt GPU in mobile applications like automatic driving. The previous work proposes some optimizing techniques to lower the power consumption of object detection on mobile GPU or FPGA. In the first Low-Power Image Recognition Challenge (LPIRC), our system achieved the best result with mAP/Energy on mobile GPU platforms. We further research the acceleration of detection algorithms and implement two more systems for real-time detection on FPGA with higher energy efficiency. In this paper, we will introduce the object detection algorithms and summarize the optimizing techniques in three of our previous energy efficient detection systems on different hardware platforms for object detection.

Download Paper (PDF; Only available from the DATE venue WiFi)
## A RETROSPECTIVE EVALUATION OF ENERGY-EFFICIENT OBJECT DETECTION SOLUTIONS ON EMBEDDED DEVICES

**Speaker:**  
Ying Wang, Chinese Academy of Sciences, CN

**Authors:**  
Ying Wang, Zhenyu Quan, Yinhe Han, Jiajun Li, Huawei Li and Xiaowei Li, Institute of Computing Technology Chinese Academy of Sciences, Beijing, CN

**Abstract**  
The field of image and video recognition has been propelled by the rapid development of deep learning in recent years. With its fascinating accuracy and generalization ability, deep CNNs have shown remarkable performance in large-scale and real-life image datasets. However, accommodating computation-intensive CNN-based image detection frameworks on power-constrained devices is considered more challenging than desktop or warehouse computing systems. Instead of emphasizing purely on detection accuracy, the Low Power Image Recognition Challenge (LPIRC) is initiated to highlight the energy-efficiency of different image recognition solutions, and it witnesses the advancement of cost-effective image recognition technology in aspects of both algorithmic and architecture innovation. This paper introduces the cost-effective CNN-based object detection solutions that reached an improved tradeoff between energy and accuracy for mobile CPU+GPU SoCs, which is the winner of LPIRC2016, and it also analyzes the implications of both recent hardware and algorithm advancement on such a technique. It is demonstrated in our evaluation that the performance growth of embedded SoCs and CNN models have clearly contributed to a sheer growth of mAP/WH in current CNN-based object detection solutions, and also shifted the balance between accuracy and energy-cost in the contest solution design when we seek to maximize the efficiency score defined by LPIRC through design parameter exploration.

Download Paper (PDF; Only available from the DATE venue WiFi)

## JOINT OPTIMIZATION OF SPEED, ACCURACY, AND ENERGY FOR EMBEDDED IMAGE RECOGNITION SYSTEMS

**Speaker:**  
Soonhoi Ha, Seoul National University, KR

**Authors:**  
Duseok Kang, Jintaek Kang, Donghyun Kang, Sungjoo Yoo and Soonhoi Ha, Seoul National University, KR

**Abstract**  
This paper presents the image recognition system that won the first prize in the LPIRC (Low Power Image Recognition Challenge) in 2017. The goal of the challenge is to maximize the ratio between the accuracy and energy consumption within a time limit of 10 minutes for the processing of 20,000 images. Among three conflicting goals of accuracy, speed, and energy consumption, we considered the trade-off between accuracy and speed first to select Nvidia Jetson TX2 as the hardware platform and TinyYOLO as the image recognition algorithm. Next, we applied a series of software optimization techniques to improve throughput, such as pipelining, multithreading, Tucker decomposition, and 16-bit quantization. Lastly, we explored the CPU and GPU frequencies to minimize the total energy consumption. As a result, we could achieve an accuracy of 0.24 mAP with energy consumption of 2.08Wh, which corresponds to the score of 0.11931, 2.7 times higher than the winner of LPIRC 2016.

Download Paper (PDF; Only available from the DATE venue WiFi)

## Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

### Lunch Break (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

**Tuesday, March 20, 2018**
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

**Wednesday, March 21, 2018**
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

**Thursday, March 22, 2018**
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

## 6.8 Innovative Products for Autonomous Driving (part 2)

**Date:** Wednesday, March 21, 2018  
**Time:** 11:00 - 12:30  
**Location / Room:** Exhibition Theatre

**Organiser:**  
Hans-Jürgen Brand, IDT/ZMDI, DE, Contact Hans-Jürgen Brand

The workshop on Innovative Products for Autonomous Driving includes 2 sessions (part 1: session 3.8). This session will highlight how to do ultra-low-voltage design, how to accelerate physical signoff and a 22 nm FDSOI System-on-Chip development for Advanced Driver Assistance System.
<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:00</td>
<td>6.8.1</td>
<td>22FDX ULTRA-LOW-VOLTAGE DESIGN BASED ON ADAPTIVE BODY BIAS</td>
<td>Holger Eisenreich, Racyics GmbH, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>22FDX body bias allows to compensate process, temperature and slow voltage variations. Applied in Adaptive Body Bias (ABB) scheme, this technology feature enables Ultra-Low-Voltage implementations down to 0.4V for IoT-like designs with unparalleled energy efficiency. Racyics will present its 22FDX ABB IP platform and the related ABB-aware implementation and sign-off methodology.</td>
</tr>
<tr>
<td>11:30</td>
<td>6.8.2</td>
<td>A NEW ADAS CHIP DESIGN IN 22 NM FDSOI TECHNOLOGY FOR AUTOMOTIVE COMPUTER VISION APPLICATIONS</td>
<td>Jens Benndorf, Dream Chip Technologies, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>The presentation explores the European collaboration on a 22 nm FDSOI System-on-Chip development for Advanced Driver Assistance System. It will highlight partner co-operation and key trade-offs necessary to deliver the target performance. The scope includes:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Short company introduction</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Project setup and target applications</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Chip architecture and performance requirements</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Hardware Demonstration System</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Outlook</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Summary</td>
<td></td>
</tr>
<tr>
<td>12:00</td>
<td>6.8.3</td>
<td>ACCELERATING PHYSICAL SIGNOFF FOR LEADING EDGE CHIP DESIGNS</td>
<td>David DeMarcos, Synopsys, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>Physical Verification with IC Validator in the Synopsys Design Platform provides technology-leading, production-proven signoff solutions for design rule checking (DRC), connectivity verification layout-vs.-schematic (LVS), metal fill insertion, and design-for-manufacturability (DFM) enhancements. IC Validator is supported by all major foundries as a signoff solution for established-node designs, as well as advanced emerging-node designs at 20nm and below. It includes productivity links to leading design tools such as IC Compiler™/IC Compiler II physical implementation, StarRC™ parasitic extraction, and Custom Compiler™ mixed-signal design. IC Validator’s In-Design physical verification speeds up design closure with timing-aware metal fill and DRC living within the IC Compiler and IC Compiler II environments.</td>
</tr>
<tr>
<td>12:30</td>
<td></td>
<td>End of session</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lunch Break in Großer Saal and Saal 1</td>
<td></td>
</tr>
</tbody>
</table>

**Coffee Breaks in the Exhibition Area**

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

**Lunch Breaks (Großer Saal + Saal 1)**

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

**Tuesday, March 20, 2018**
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

**Wednesday, March 21, 2018**
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

**Thursday, March 22, 2018**
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00
Abstract

Analog/Mixed-Signal design and verification strongly relies on more or less abstract models to make extensive simulations feasible. Maintaining consistent behavior between system model and implementation is crucial for a correct verification. This also involves the operating conditions: A faulty model might introduce false-positive verification results despite of e.g., an incorrect supply voltage or missing bias currents. We present an automated workflow for extracting these checks from a transistor-level implementation and transfer it into a given Verilog-AMS model. The correctness of our approach is proved by evaluating the model coverage between the implementation and the model. As a demonstration scenario, we use a demodulator component of a HF RFID communication system. We extract the acceptance region from the transistor-level schematic and automatically generate and integrate a model safe-guard unit for performing the operating condition check.

More information ...
The complexity of modern embedded system design is managed by advanced, high-level design methodologies such as IP-XACT. However, integrating IP-XACT as a part of an existing design flow and packaging legacy sources is often hindered due to interface differences between IP-XACT and the traditional hardware description languages. In this work, we present an existing Verilog implementation of a RISC-V microprocessor and package it with our open-source IP-XACT tool Kactus2. The resulting IP-XACT description will be publicly available and based on the modeling experience we report the observed pitfalls in the transition from HDL to IP-XACT.

More information...
KEYNOTE SPEAKER

Author:
Jelena Vuckovic, Stanford University, US

Abstract

It is estimated that nearly 10% of the world electricity is consumed in information processing and computing, including data centers [D.A.B. Miller, Journal of Lightwave Technology, 2017]. It is clear that the exponential growth in use of these technologies is not sustainable unless dramatic changes are made to computing hardware, in order to increase its speed and energy efficiency. Optical interconnects are considered a solution to these obstacles, with potential to reduce energy consumption in on-chip optical interconnects to atto-Joule per bit (aJ/bit), while increasing operating speed beyond 20GHz. However, the state of the art photonics is bulky, inefficient, sensitive to environment, lossy, and its performance is severely degraded in real-world environment as opposed to ideal laboratory conditions, which has prevented from using it in many practical applications, including interconnects. Therefore, it is clear that new approaches for implementing photonics are crucial. We have recently developed a computational approach to inverse-design photonics based on desired performance, with fabrication constraints and structure robustness incorporated in design process. Our approach performs physics guided search through the full parameter space until the optimal solution is reached. Resulting device designs are non-intuitive, but are fabricable using standard techniques, resistant to temperature variations of hundreds of degrees, typical fabrication errors, and they outperform state of the art counterparts by many orders of magnitude in footprint, efficiency and stability. This is completely different from conventional approach to design photonics, which is almost always performed by brute-force or intuition-guided tuning of a few parameters of known structures, until satisfactory performance is achieved, and which almost always leads to sub-optimal designs. Apart from integrated photonics, our approach is also applicable to any other optical and quantum optical devices and systems.

Coffee Break in Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018

- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

UB07 Session 7

Date: Wednesday, March 21, 2018
Time: 14:00 - 16:00
Location / Room: Booth 1, Exhibition Area

Label | Presentation Title | Authors
--- | --- | ---
UB07.1 | GENERATING FULL-CUSTOM SCHEMATICS IN A MIXED-SIGNAL TOP-DOWN DESIGN FLOW | Authors:
- Tobias Markus, Markus Mueller and Ulrich Bruening
  1University of Heidelberg, DE; 2Extol GmbH, DE

Abstract

Design time is one of the precious assets in the cycle of hardware design. The top down methodology has been used in digital designs very successfully and now we also apply it for analog and mixed signal designs. Generating most of the structures automatically saves time and avoids errors. A Top Down Design Flow for Mixed Signal Designs is used which generates the schematic structure from the system RNM representation. Since the structural verilog part of the system level design will automatically generate the schematic structure it is only the functional part which is missing and has to be implemented by the analog designer. Some often used blocks can be used as an entry point to partially generate parts of the design in the schematic and furthermore even parts of the layout. We will demonstrate this design method with an example project.
UB07.2 CLAVA-MARGOT: CLAVA + MARGOT = C/C++ TO C/C++ COMPILER AND RUNTIME AUTOTUNING FRAMEWORK

Authors: João Bispo¹, Davide Gadioli², Pedro Pinto¹, Emanuele Vitali², Hamid Arabnejad¹, Gianluca Palermo², Cristina Silvano², Jorge G. Barbosa¹ and João M. P. Cardoso¹
¹Porto University, PT; ²Politecnico di Milano (POLIMI), IT

Abstract
Current computing platforms consist of heterogeneous architectures. To efficiently target those platforms, compilers can be extended with code transformations and insertion of code to interface to runtime autotuning schemes, which tune application parameters according to: the actual execution, target architecture, and workload. We present an approach consisting of a C/C++ source-to-source compiler (Clava) and an autotuner (mARGOT). They are part of the toolflow of the FET-HPC ANTARES project and allow parallelization, multiversioning and code transformations in the context of runtime autotuning. mARGOT is an autotuner that allows application adaptation to changing conditions and goals. Clava is a source-to-source compiler to transform C/C++ programs, including code instrumentation and integration with components such as mARGOT. We will demonstrate how to use Clava to integrate the mARGOT autotuner in an example application, and several mARGOT functionalities exposed through a Clava API.

More information ...

UB07.3 OISC MULTICORE STENCIL PROCESSOR: ONE INSTRUCTION-SET COMPUTER-BASED MULTICORE PROCESSOR FOR STENCIL COMPUTING

Authors: Kaoru Saso, Jing Yuan Zhao and Yuko Hara-Azumi, School of Engineering, Tokyo Institute of Technology, JP

Abstract
Subtract and Branch on NEGative with 4 operands (SUBNEG4) is one of One Instruction-Set Computers that execute only one type of instruction. Thanks to its simplicity, SUBNEG4 has only 1/20th circuit area and 1/10th power consumption against MIPS processor. As SUBNEG4 is Turing-complete, it is suitable for parallel computing by multiple cores, while keeping its low-power feature. Our on-going project is seeking for effective use and deployment of SUBNEG4 cores on embedded systems. Our booth will demonstrate the significant speed-up by a SUBNEG4-based many-core processor against a conventional processor, for stencil computing. Our 64-core processor efficiently handles 2D von Neumann neighborhood stencils, e.g., wave simulation by Verlet integration and 2D Jacobi iteration, to compute 64 points simultaneously. We show that small many-core processors can be realized even with such large number of cores while achieving good speed-up for heaving computation.

More information ...

UB07.4 ROS X FPGA FOR ROBOT-CLOUD SYSTEM: ROBOT-CLOUD COOPERATIVE VISUAL SLAM PROCESSING USING ROS-COMPLIANT FPGA COMPONENT

Authors: Takeshi Chikawa, Yuhki Sugata, Aoi Soya, Kanemitsu Ootsu and Takashi Yokota, University of Tokyo, JP

Abstract
Distributed processing in robot-cloud cooperation system is discussed in terms of processing performance and communication performance. Cooperation of robots and cloud-servers is inevitable for realizing intelligent robots in the next generation society and industry. To improve processing performance of the cooperative system, we utilize ROS-compliant FPGA component as a robot-side embedded processing for low-power and high-performance image processing. We prepare two demonstrations. (1) Key-point Detection from camera image using Fully-harwdared ROS-Compliant FPGA component In the evaluation, the processing performance of the component is almost same as PC, while it operates at more than 10 times less power (SW) compared to PC (SW). (2) Distributed Visual SLAM using two wheeled robot (TurtleBot3) Distributed Visual SLAM (Simultaneous Localization and Mapping) are presented as a concrete example of the robot-cloud cooperative system.

More information ...

UB07.5 WIRELESS SENSOR SYSTEM WITH ELECTROMAGNETIC ENERGY HARVESTER FOR INDUSTRY 4.0 APPLICATIONS

Authors: Bianca Leistritz, Elena Chevakovka, Sven Engelhardt, Axel Schreiber and Wolfram Kattanek, Institut für Mikroelektronik- und Mechatronik-Systeme gemeinnützige GmbH, DE

Abstract
An energy-autonomous and adaptive wireless multi-sensor system for a wide range of Industry 4.0 applications is presented here. By taking a holistic view of the sensor system and of the specific interactions of its components, technological barriers of individual system elements can be overcome. The energy supply of the demonstrator is realized by an miniaturized electromagnetic energy harvester, which can be easily and quickly adapted to the application-specific boundary conditions with the help of a computer assisted design process. Variations in the available energy are monitored by advanced energy management functions. The modular hardware and software platform is demonstrated by an adaptive measurement and data transmission rate. Communication takes place by means of industry 4.0 compliant standard protocols. The demonstrator was developed in the research group Green-ISES funded by the Free State of Thuringia from the European Social Fund (ESF) under grant no. 2016 FGR 0055.

More information ...

UB07.6 EMBEDDED ACCELERATION OF IMAGE CLASSIFICATION APPLICATIONS FOR STEREO VISION SYSTEMS

Authors: Mohammad Loni¹, Carl Ahlberg², Masoud Daneshbozorg³, Mikael Ekström² and Mikael Sjödin²
¹MDH, SE; ²Mälardalen University, SE

Abstract
Autonomous systems are used in a broad range of applications from indoor utensils to medical application. Stereo vision cameras probably are the most flexible sensing way in these systems since they can extract depth, luminance, color, and shape information. However, stereo vision based applications suffer from huge image sizes, computational complexity and high energy consumption. To tackle these challenges, we first developed GIMME2 [1], a high-throughput, and cost efficient FPGA-based stereo vision system. In the next step, we present a novel approach accelerator which is also compatible with GIMME2. Our accelerator tries to map neural network (NN) based image classification algorithms to FPGA by using DeepMaker which is an evolutionary based module embed in our accelerator that reconfigures a near-optimal NN in term of accuracy. Then, the back-end side of DeepMaker maps the generated NN to FPGA. We will demo a GIMME2-based accelerator for image classification applications.

More information ...

UB07.7 T-CREST: THE OPEN-SOURCE REAL-TIME MULTICORE PROCESSOR

Authors: Martin Schoebel, Luca Pezzarossa and Jens Sparse, Technical University of Denmark, DK

Abstract
Future real-time systems, such as advanced control systems or real-time image recognition, need more powerful processors, but still a system where the worst-case execution time (WCET) can be statically predicted. Multicore processors are one answer to the need for more processing power. However, it is still an open research question how to best organize and implement time-predictable communication between processing cores. T-CREST is an open-source multicore processor for research on time-predictable computer architecture. It consists of several Patmos processors connected by various time-predictable communication structures: access to shared off-chip, access to shared on-chip memory, and the Aigo network-on-chip for fast inter-processor communication. T-CREST is supported by open-source development tools, such as compilation and WCET analysis. To best of our knowledge, T-CREST is the only fully open-source software for research on future real-time multicore architectures.

More information ...

UB07.8 IIP GENERATORS TO EASE ANALOG IC DESIGN

Authors: Benjamin Prautsch, Uwe Eichler and Torsten Reich, Fraunhofer Institute for Integrated Circuits IIS/EAS, DE

Abstract
Semiconductor technology has shown significant progress over the last decades. Digital EDA (electronic design automation) allowed that this progress could be converted to high-performance digital ICs. Analog components are part of Systems-on-Chip (SoC) too, but analog EDA lags far behind. Therefore, a lot of effort was spent to automate analog IC design. Major results are constraint-based layout-aware optimization tools using predefined layout templates or pure automation as well as analog generators containing expert knowledge. While optimization is a holistic top-down approach, generators allow parameterized and fast bottom-up generation of critical schematic and layout parts, pre-planned by experienced designers. With IIP Generators, we follow three use cases to ease analog design: 1) design on higher hierarchy levels, 2) development of hierarchical high-level IIPs, and 3) automated design porting due to highly technology-independent blocks down to 22nm.

More information ...
CIJTAG: CONCURRENT IJTAG DEMONSTRATOR

Author:
Krenz-Baath René, Hammi-Lippstadt University of Applied Sciences, DE

Abstract
The flexibility of on-chip instrument access enabled by IEEE 1687 (IJTAG) has shown tremendous improvements in modern industrial designs. Due to a constantly increasing spectrum of tasks performed through 1687 networks such as performing test operations during production test, on-line test operations as well as operating health monitors the test requirements in modern designs increase dramatically with respect to test performance, responsiveness and low power. These requirements have a major impact on the design of such test infrastructures. In complex designs with large test infrastructures it might be challenging to comply with the large spectrum of requirements. Concurrent IJTAG is a novel partitioning concept to a reconfigurable test infrastructure in order to enable an independent operation of different sections of the test infrastructure. The proposed demonstrator shows the first FPGA-based implementation of concurrent IJTAG test infrastructures.

EXPERIENCE-BASED AUTOMATION OF ANALOG IC DESIGN

Authors:
Florian Leber and Juergen Scheible, Reutlingen University, DE

Abstract
While digital design automation is highly developed, analog design automation still remains behind the demands. Previous circuit synthesis approaches, which are usually based on optimization algorithms, do not satisfy industrial requirements. A promising alternative is given by procedural approaches (also known as “generators”): They (a) emulate experts’ decisions, thus (b) make expert knowledge re-usable and (c) can consider all relevant aspects and constraints implicitly. Nowadays, generators are successfully applied in analog layout (Pcells, Pycells). We aim at an entire design flow completely based on procedural automation techniques. This flow will consist of procedures for the generation of schematics and layouts for every typical analog circuit class, such as amplifier, bandgap, filter a.s.o. In our presentation we give an overview on such a design flow and we show an approach for capturing an analog circuit designer’s strategy as an executable “expert design plan”.

Coffee Break in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

7.1 Special Day Session on Future and Emerging Technologies: Theoretical and practical aspects of verification of quantum computers

Date: Wednesday, March 21, 2018
Time: 14:30 - 16:00
Location / Room: Saal 2

Chair:
Naveh Yehuda, IBM Research, IL, Contact Yehuda Naveh

Quantum computing is emerging at a meteoric pace from a pure academic field to a fully industrial framework. Rapid advances are happening both in the physical realizations of quantum chips, and in their potential software applications. In contrast, we are not seeing that rapid growth in the design and verification methodologies for scaled-up quantum machines. In this session we describe the field of verification of quantum computers. We discuss the underlying concepts of this field, its theoretical and practical challenges, and state-of-the-art approaches to addressing those challenges. The goal of this session is to help facilitate early efforts to adapt and create verification methodologies for quantum computers and systems. Without such early efforts, a debilitating gap may form between the state-of-the-art of low level physical technologies for quantum computers, and our ability to build medium, large, and very large scale integrated quantum circuits (M/L/VLSIQ).
14:30 7.1.1 VERIFICATION OF QUANTUM COMPUTING
Speaker: Petros Wallden, School of Informatics, University of Edinburgh, GB
Author: Elham Kashefi, School of Informatics, University of Edinburgh, UK & CNRS LIP6, GB
Abstract: Quantum computers promise to efficiently solve not only problems believed to be intractable for classical computers, but also problems for which verifying the solution is also considered intractable. This raises the question of how one can check whether quantum computers are indeed producing correct results. This task, known as quantum verification, has been highlighted as a significant challenge on the road to scalable quantum computing technology. We review the most significant approaches to quantum verification and compare them in terms of structure, complexity and required resources. We also comment on the use of cryptographic techniques which, for many of the presented protocols, has proven extremely useful in performing verification. Finally, we discuss issues related to fault tolerance, experimental implementations and the outlook for future protocols.

14:50 7.1.2 GAINING INSIGHT INTO NEAR-TERM QUANTUM DEVICES WITH TAILOR-MADE APPLICATIONS
Author: James R. Wootton, University of Basel, CH
Abstract: Many interesting algorithms have been designed for large scale fault-tolerant quantum computers. However, most will not be suitable for the smaller and noisier devices of the next decade. To understand how these devices function, we must therefore use applications specifically designed for their capabilities. In this talk we briefly introduce two possibilities. One is quantum error correction, which allows us to directly analyze imperfections in a device, as well as determine how well we can control them. The other is games, which can provide general insights into the capabilities of a device in a widely relatable manner.

15:15 7.1.3 THE ENGINEERING CHALLENGES IN QUANTUM COMPUTING
Author: Koen Bertels, Delft University of Technology, NL
Abstract: In this presentation we will present the ongoing work that focuses on defining and building a micro-architecture for a quantum computer. We will present the essence of quantum computing, the challenges as well as our current long term (>5 years) and short term (<5 years) in this respect and we will discuss the system vision as well as the Transmon and Spinqubit processor prototypes that we have developed with the colleagues L. DiCarmo and L. Vandersypen at QuTech.

15:40 7.1.4 QUANTUM VERIFICATION: WHAT CAN WE ADOPT AND LEARN FROM CLASSICAL VERIFICATION
Author: Yehuda Naveh, IBM Research - Haifa, IL
Abstract: I will provide a view of the challenges of verifying quantum computers through the lenses of classical verification methodologies. I will argue that while the fields are inherently different (e.g., quantum verification is a challenge already at the 50-qubit chip levels, while classical verification challenges stem mostly from complex micro-architectural structures present only at multi-million transistor chips), many methods of classical verification may still be adapted to the quantum regime. These include abstracted simulation and modeling languages, directed constraint-based random benchmarking, coverage measures, and more. My hope is that learning from the long history of classical verification will make the process of reaching robust, efficient, and stable verification methodologies for quantum computers much faster and less painful than has been for the classical case.

16:00 End of session
Coffee Break in Exhibition Area

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
☐ Coffee Break 10:30 - 11:30
☐ Lunch Break 13:00 - 14:30
☐ Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
☐ Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
☐ Coffee Break 10:00 - 11:00
☐ Lunch Break 12:30 - 14:30
☐ Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
☐ Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
☐ Coffee Break 10:00 - 11:00
☐ Lunch Break 12:30 - 14:00
☐ Coffee Break 15:30 - 16:00

7.2 Run-time power estimation and optimization

Date: Wednesday, March 21, 2018
Time: 14:30 - 16:00
Location / Room: Konf. 6
Chair: Pascal Vivet, CEA-Leti, FR, Contact Pascal Vivet
Co-Chair:
Donghwa Shin, Yeungnam Univ, Daegu, KR, Contact Donghwa Shin

In this session, the first paper presents energy efficiency optimization for CPU-GPU heterogeneous architectures using machine-learning. The next two papers present run-time power modeling and estimation methods for embedded systems. Finally, the last paper presents an online reconfiguration method for photovoltaic power system.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30</td>
<td>7.2.1</td>
<td>AIRAVAT: IMPROVING ENERGY EFFICIENCY OF HETEROGENEOUS APPLICATIONS</td>
<td>Tmnyan Bazuah, Northeastern University, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Tmnyan Bazuah, Northeastern University, US</td>
<td>Authors: Tmnyan Bazuah, Yitan Sun, Shi Dong, David Kaeli, and Norm Rubin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: We are seeing an emerging class of applications that attempt to make use of both the CPU and GPU in a heterogeneous system. The peak performance for these applications is achieved when both the CPU and GPU are used collaboratively. However, with this increased gain in performance, power and energy management is a larger challenge. In this paper, we address the issue of executing applications that utilize both the CPU and GPU in an energy-efficient way. Towards this end, we propose a power management framework named Airavat that tunes the CPU, GPU, and memory frequencies, synergistically, in order to improve the energy efficiency of collaborative CPU-GPU applications. Airavat uses machine learning-based prediction models, combined with feedback based Dynamic Voltage and Frequency Scaling to improve the energy efficiency of such applications. We demonstrate our framework on the Jetson TX1 and observe an improvement in terms of Energy Delay Product (EDP) by 24% with only a minimal performance loss.</td>
<td></td>
</tr>
<tr>
<td>15:00</td>
<td>7.2.2</td>
<td>ALL-DIGITAL EMBEDDED METERS FOR ON-LINE POWER ESTIMATION</td>
<td>Daniele Pagliati, Politecnico di Torino, IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Daniele Pagliati, Politecnico di Torino, IT</td>
<td>Authors: Daniele Pagliati, Giuseppe Peluso, Yukai Chen, Andrea Calimera, Enrico Maci and Massimo Poncin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Modern low-power designs use multiple knobs for concurrent dynamic and leakage power optimization: supply voltage and threshold voltage are the most adopted. An efficient control of these knobs needs management policies aware of the power breakdown. This implies the availability of smart on-chip strategies for dynamic and leakage power estimation at runtime. In this paper, we address this issue proposing the implementation of embedded dynamic/static power meters that use an optimized regression model fed with data collected from in-situ activity monitors. The number of sensors, their bitwidth and optimal placement are obtained through an automated design flow. The methodology works for general logic and applies not just to processor cores, but also to application-specific designs. We apply our solution to a representative class of benchmarks, showing that it can achieve an average estimation error smaller than 3%, with limited area and power overheads.</td>
<td></td>
</tr>
<tr>
<td>15:30</td>
<td>7.2.3</td>
<td>POWERPROBE: RUN-TIME POWER MODELING THROUGH AUTOMATIC RTL INSTRUMENTATION</td>
<td>Davide Zoni, Politecnico di Milano, IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Davide Zoni, Politecnico di Milano, IT</td>
<td>Authors: Davide Zoni, Luca Cremona and Wiliam Fornaciari, Politecnico di Milano, IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Online power monitoring represents a de-facto solution to enable energy- and power-aware run-time optimizations for current and future computing architectures. Traditionally, the performance counters of the target architecture are used to feed a software-based, power model that is continuously updated to deliver the required run-time power estimates. The solution introduces a non-negligible performance and energy overhead. Moreover, it is limited to the availability of such performance counters that, however, are not primarily intended for online power monitoring. This paper introduces PowerProbe, a run-time power monitoring methodology that automatically extracts and implements a power model from the RTL description of the target architecture. The solution does not leverage any performance counter to ensure wide applicability. Moreover, the use of ad-hoc hardware that continuously updates the power estimate minimizes both the performance and the power overheads. We employ a fully compliant OpenRisc 1000 implementation to validate PowerProbe. The results highlight an average prediction error within 9% (standard deviation less than 2%), with a power and area overheads limited to 6.89% and 4.71%, respectively.</td>
<td></td>
</tr>
<tr>
<td>15:45</td>
<td>7.2.4</td>
<td>DESIGN OPTIMIZATION OF PHOTOVOLTAIC ARRAY ON A CURVED SURFACE</td>
<td>Sangyoung Park, Technical University of Munich, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Sangyoung Park, Technical University of Munich, DE</td>
<td>Authors: Sangyoung Park and Samsat Chakraborty, Technical University of Munich, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Flexible photovoltaic (PV) arrays often have to be mounted on surfaces that have a significant amount of curvature. These include solar-powered vehicles, planes, and also some wearable devices. However, this inevitably leads to non-uniform solar irradiance among connected PV cells. If one cell among series-connected PV cells receives significantly lower solar irradiance, the overall power generation of the string is reduced. While previous works dealt with this by employing sophisticated run-time techniques, we show that design-time approaches that determine the electrical series-parallel connection of a PV array could also significantly enhance the power output. In this paper, we propose a k-means clustering-based algorithm to group PV cells/modules with similar solar irradiance to form a PV string, even allowing irregular arrays, to maximize the power generation of the array for a given irradiance profile. Our experimental results show that the power generation of a PV array could be increased by 84% compared to usual PV array organizations that do not take the curvature of the mounted surface into account.</td>
<td></td>
</tr>
<tr>
<td>16:00</td>
<td>IP3-6</td>
<td>PREDICTION-BASED FAST THERMOELECTRIC GENERATOR RECONFIGURATION FOR ENERGY HARVESTING FROM VEHICLE RADIATORS</td>
<td>Xue Lin, Northeastern University, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Xue Lin, Northeastern University, US</td>
<td>Authors: Hanchen Yang, Feiyang Kang, Caiwen Ding, Ji Li, Jiajin Kim, Dongyu Baek, Shahin Nazarian, Xue Lin, Paul Bogdan and Naeyhuck Chang</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Thermoelectric generation has increasingly drawn attention for being environmentally friendly. However, only a few of the prior researches on thermoelectric generators (TEG) have focused on improving efficiency at system level. They attempt to capture the electrical property changes on TEG modules as the temperature fluctuates on vehicle radiators. The most recent reconfiguration algorithm shows large improvements on output performance but suffers from major drawback on computational time and energy overhead, and non-scalability in terms of array size and processing frequency. In this paper, we propose a novel TEG array reconfiguration algorithm that determines near-optimal configuration with an acceptable computational time. Our prediction-based fast TEG reconfiguration algorithm enables all modules to work at or near their maximum power points (MPP). Additionally, we incorporate prediction methods to further reduce the runtime and switching overhead during the reconfiguration process. Experimental results present 30% performance improvement, almost 100x reduction on switching overhead and 13x enhancement on computational speed compared to the baseline and prior work. The scalability of our algorithm makes it applicable to larger scale systems such as industrial boilers and heat exchangers.</td>
<td></td>
</tr>
</tbody>
</table>

Download Paper (PDF; Only available from the DATE venue WiFi)
A PARAMETERIZED TIMING-AWARE FLIP-FLOP MERGING ALGORITHM FOR CLOCK POWER REDUCTION

Chaochao Feng, National University of Defense Technology, CN

Authors:
Chaochao Feng1, Daheng Yue1, Zhenyu Zhao1 and Zhufan Liao2
1National University of Defense Technology, CN; 2Changsha University of Science and Technology, CN

Abstract
In modern integrated circuits, the clock power contributes a dominant part of the chip power. Clock power can be reduced effectively by utilizing multi-bit flip-flops. In this paper, a parameterized timing-aware flip-flop merging algorithm is proposed for clock power reduction. The single-bit flip-flops are merged into multi-bit flip-flops after placement & optimization and before clock network synthesis with consideration of function, scan chain information, distance and timing constraints. The algorithm can be configured with different parameters, such as the bit-number of MBFF, the setup timing margin and the distance margin. Experimental results under an industrial design show that compared with the basic design without MBFF, the design with 2-bit, 4-bit, 6-bit, and 8-bit MBFFs can save 7.5%, 12%, 11.8% and 11.1% total power consumption respectively. Using MBFF4 to replace 1-bit FFs is the best choice for the design optimization, which achieves minimum area and total power consumption. We also compare the designs with MBFF4 replacement under five different setup timing margins and distance margins. Without violating any timing constraint, it is better to set the setup timing margin as small as possible to achieve best power optimization. The distance margin (100µm, 30µm) is the best choice for this industry design to achieve minimum power consumption.

Download Paper (PDF; Only available from the DATE venue WiFi)

FAST CHIP-PACKAGE-PCB COANALYSIS METHODOLOGY FOR POWER INTEGRITY OF MULTI-DOMAIN HIGH-SPEED MEMORY: A CASE STUDY

Seungwon Kim, Ulsan National Institute of Science and Technology, KR

Authors:
Seungwon Kim1, Ki Jin Han2, Youngmin Kim3 and Seokhyeong Kang1
1Ulsan National Institute of Science and Technology (UNIST), KR; 2Dongguk University, KR; 3Kwangwoon University, KR

Abstract
The power integrity of high-speed interfaces is an increasingly important issue in mobile memory systems. However, because of complicated design variations such as adjacent VDD domain coupling, conventional case-specific modeling is limited in analyzing trends in results from parametric variations. Moreover, conventional industrial methods can be simulated only after the design layout is completed and it requires a lot of back-annotation processes, which result in delayed delays time to market. In this paper, we propose a chip-package-PCB coanalysis methodology applied to our multi-domain high-speed memory system model with a current generation method. Our proposed parametric simulation model can analyze the tendency of power integrity results from variable sweeps and Monte Carlo simulations, and it shows a significantly reduced runtime compared to the conventional EDA methodology under JEDEC LPPDR4 environment.

Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

7.3 Advances in Logic Synthesis and Technology Mapping

Date: Wednesday, March 21, 2018
Time: 14:30 - 16:00
Location / Room: Konf. 1
Chair: Luciano Lavagno, Politecnico di Torino, IT, Contact Luciano Lavagno
Co-Chair: Mathias Soeken, EPFL, CH, Contact Mathias Soeken

This session presents recent progress in logic synthesis and technology mapping. The first paper discusses improvements to Boolean resynthesis using a theory on Boolean filtering and a more general notion of permissible functions. The second paper applies methods based on Boolean relations for the optimization of combinational logic networks. The third paper proposes a technology mapping approach for silicon nanowire reconfigurable FETs. The fourth paper presents for the first time an approximate synthesis methods for threshold logic circuits through an iterative approach that guarantees an error bound.

14:30 7.3.1 IMPROVEMENTS TO BOOLEAN RESYNTHESIS
Speaker: Mathias Soeken, EPFL, CH
Authors: Luca Amaru1, Mathias Soeken2, Patrick Vuillod3, Jiong Luo1, Alan Mishchenko4, Janet Olson1, Robert Brayton4 and Giovanni De Micheli2
1Synopsys Inc., US; 2Integrated System Laboratory – EPFL, CH; 3Synopsys Inc., FR; 4UC Berkeley, US
Abstract
Boolean resynthesis techniques are increasingly used in electronic design automation, to improve quality of results where algebraic methods hit local minima. Boolean methods rely on complete functional properties of a logic circuit, eventually including don’t cares. In order to gather such properties, computationally expensive engines are required, e.g., truth tables, SAT and BDDs, which in turn determine the scalability of Boolean resynthesis. In this paper, we present theoretical and practical improvements to Boolean resynthesis, enabling more optimization opportunities to be found at the same, or smaller, runtime cost than state-of-the-art methods. Our contributions include: (i) a theory of Boolean filtering, to drastically reduce the number of gates processed and still retain all possible optimization opportunities, (ii) a weaker notion of maximum set of permissible functions, which can be computed efficiently via truth tables, (iii) a parallel package for truth table computation tailored to speedup Boolean methods, (iv) a generalized refactoring engine which supports multiple representation forms and (v) a practical Boolean resynthesis flow, which combines the techniques proposed so far. Using our Boolean resynthesis on the EPFL benchmarks, we improve 9 of the best known area results in the synthesis competition. Embedded in a commercial EDA flow for ASICs, our Boolean resynthesis flow reduces the area by 2.67%, and total negative slack by 5.48%, after physical implementation, at negligible runtime cost.
Download Paper (PDF; Only available from the DATE venue WiFi)
**Presentation Title:** RECONFIGURABLE IMPLEMENTATION OF $GF(2^m)$ BIT-PARALLEL MULTIPLIERS

**Speaker:** Chia-Cheng Wu, National Tsing Hua University, TW

**Authors:**
- Tung-Yuan Lee1, Chia-Cheng Wu1, Chia-Chun Lin1, Yung-Chih Chen2 and Chun-Yao Wang3
- 1National Tsing Hua University, TW; 2Yuan Ze University, TW; 3Dept. CS, National Tsing Hua University, TW

**Abstract:**

Hardware implementations of arithmetic operations over binary finite fields $GF(2^m)$ are widely used in several important applications, such as cryptography, digital signal processing and error-control codes. In this paper, efficient reconfigurable implementations of bit-parallel canonical basis multipliers over binary fields generated by type II irreducible pentanomials are presented. These pentanomials are important because all five binary fields recommended by NIST for ECDSA can be constructed using such polynomials. In this work, a new approach for $GF(2^m)$ multiplication based on type II pentanomials is given and several post-place and route implementation results in Xilinx Artix-7 FPGA are reported. Experimental results show that the proposed multiplier implementations improve the area by an average of 36.3% and reduce the maximum delay by an average of 38.2% compared to existing implementations. The proposed approach allows for automatic generation of approximate hardware wrt. accuracy for performance. The behavior of a circuit can be defined at an arithmetic level, by describing the input and output relation as a polynomial. Symbolic Computer Algebra (SCA) has been employed to verify that a given circuit netlist matches the behavior specified at the arithmetic level. In this paper, we present a method that relaxes the exactness requirement of the implementation. We propose a heuristic method to generate an approximation for a given netlist and use SCA to ensure that the result is within application-specific bounds for given error-metrics. In addition, our approach allows for automatic generation of approximate hardware with applicationspecific input probabilities. To the best of our knowledge, existing approximate adders and show that the results outperform state-of-the-art, handcrafted approximate hardware.

Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

7.4 DRAM and NVMs

Date: Wednesday, March 21, 2018
Time: 14:30 - 16:00
Location / Room: Konf. 2

Chair:
Francisco Cazorla, BSC, ES, Contact Francisco Cazorla

Co-Chair:
Olivier Sentieys, IRISA, FR, Contact Olivier Sentieys

Memory is one of the major bottlenecks for performance and power. The first paper addresses this inefficiency by proposing a unified LLC+DRAM memory controller to harvest row buffer hits and to increase memory bandwidth. The next paper adopts approximate computing to increase the performance of STT and to reduce the number of writes to PCM. Finally, the third paper proposes an adaptive write current scaling approach that adjusts the write current at runtime while considering the write rates of the running application.

7.4.1 ROW-BUFFER HIT HARVESTING IN ORCHESTRATED LAST-LEVEL CACHE AND DRAM SCHEDULING FOR HETEROGENEOUS MULTICORE SYSTEMS

Speaker:
Xun Jiao, University of California, San Diego, US

Authors:
Yang Song1, Olivier Alavoine2 and Bill Lin1

1University of California, San Diego, US; 2Qualcomm Inc., US

Abstract
In heterogeneous multicore systems, the memory subsystem, including the last-level cache and DRAM, is widely shared among the CPU, the GPU, and the real-time cores. Due to their distinct memory traffic patterns, heterogeneous cores result in more frequent cache misses at the last-level cache. As cache misses travel through the memory subsystem, two schedulers are involved for the last-level cache and DRAM respectively. Prior studies treated the scheduling of the last-level cache and DRAM as independent stages. However, with no orchestration and limited visibility of memory traffic, neither scheduling stage is able to ensure optimal scheduling decisions for memory efficiency. Unnecessary precharges and row activations happen in DRAM when the memory scheduler is ignorant of incoming cache misses and DRAM row-buffer states are invisible to the last-level cache. In this paper, we propose a unified memory controller for the the last-level cache and DRAM with orchestrated schedulers. The memory scheduler harvests row-buffer hit opportunities in cache request buffers during spare time without inducing significant implementation cost. Extensive evaluations show that the proposed controller improves the total memory bandwidth of DRAM by 16.8% on average and saves DRAM energy by up to 29.7% while achieving comparable CPU IPC. In addition, we explore the impact of last-level cache bypassing techniques on the proposed memory controller.

Download Paper (PDF; Only available from the DATE venue WiFi)
**Time** | **Label** | **Presentation Title** | **Authors** |
---|---|---|---|
15:00 | 7.4.2 | **ADAM: ADAPTIVE APPROXIMATION MANAGEMENT FOR THE NON-VOLATILE MEMORY HIERARCHIES**<br>**Speaker:** Muhammad Abdullah Hanif, TU Wien, AT<br>**Authors:** Mohammad Taghi Teimoori¹, Mohammad Abdullah Hanif², Alireza Ejlali¹ and Muhammad Sharique²<br>¹Sharif University of Technology, IR; TU Wien, AT<br>**Abstract**<br>Existing memory approximation techniques focus on employing approximations at an individual level of the memory hierarchy (e.g., cache, scratchpad, or main memory). However, to exploit the full potential of approximations, there is a need to manage different approximation knobs across the complete memory hierarchy. Towards this, we model a system including STT-RAM scratchpad and PCM main memory with different approximation knobs (e.g., read/write pulse magnitude/duration) in order to synthetically trade data accuracy for both STT-RAM access delay and PCM lifetime by means of an integer linear programming (ILP) problem at design-time. Furthermore, a run-time approach is proposed to adaptively tune the approximation knobs of both STT-RAM and PCM to obtain high energy savings while keeping the quality within acceptable ranges across the complete memory hierarchy. We evaluated our proposed technique in a baseline system consisting 1MB STT-RAM scratchpad and 1GB PCM main memory. The experimental results demonstrate that our proposed technique improves the execution time and the lifetime by up to 23% and 2.3X, respectively.<br>**Download Paper (PDF; Only available from the DATE venue WiFi)**<br>|
15:30 | 7.4.3 | **A CROSS-LAYER ADAPTIVE APPROACH FOR PERFORMANCE AND POWER OPTIMIZATION IN STT-MRAM**<br>**Speaker:** Nour Sayed, KIT - Karlsruhe Institute of Technology, DE<br>**Authors:** Nour Sayed, Rajendra Bishnoi, Fabian Oboril and Mehdi Tahoori, Karlsruhe Institute of Technology, DE<br>**Abstract**<br>Spin Transfer Torque Magnetic Random Access Memory (STT-MRAM) is a promising candidate as a universal on-chip memory technology due to non-volatility, high density and scalability. However, high write energy and latency are major challenges in this memory technology due to the asymmetry and stochastic nature of the write operation. Typically, the write current is set for the minimum energy point, which can further impact the write latency. To mitigate these issues, we propose an adaptive write current scaling technique that adjusts the write current, and hence the write latency and energy based on the performance needs at run-time. Using this technique, optimal energy and performance points for write current are obtained using detailed device and system level analysis. Furthermore, we use run-time adaptation of write current by predicting the write access rate for the next execution phase. We evaluate the efficiency of the proposed approach on SPEC2000 applications for STT-MRAM-based L1 and L2-cache levels. The results show that the effective write latency of L1 and L2 is reduced by 52.4% and 55.7% with 7.6% and 1.4% area overheads, respectively, corresponding to the overall system performance optimization of 15.5% while the total memory energy consumption is increasing by only 3.2%.<br>**Download Paper (PDF; Only available from the DATE venue WiFi)**<br>|
16:00 | IP3-11, 331 | **PROCESSING IN 3D MEMORIES TO SPEEDUP OPERATIONS ON COMPLEX DATA STRUCTURES**<br>**Speaker:** Luigi Carro, UFRGS, BR<br>**Authors:** Paulo Cesar Santos¹, Geraldo Francisco de Oliveira Junior¹, Joao Paulo Lima¹, Marco Antonio Zanata Akres², Luigi Carro¹ and Antonio Carlos Schneider Beck¹<br>¹UFRGS, BR; ²UFPR, BR<br>**Abstract**<br>Pointer chasing has been, for years, the kernel operation employed by diverse data structures, from graphs to hash tables and dictionaries. However, due to the bewildering growth in the volume of data that current applications have to deal with, performing pointer chasing operations have become a major source of performance and energy bottlenecks, due to its sparse memory access behavior. In this work, we aim to tackle this problem by taking advantage of the already available parallelism present in today’s 3D-stacked memories. We present a simple mechanism that can accelerate pointer chasing operations by making use of a state-of-the-art PIM design that executes in-memory vector operations. The key idea behind our design is to run speculative loads in parallel, based on a given memory address in a reconfigurable window of addresses. Our design can perform pointer-chasing operations on b-tree 4.9x faster when compared to modern baseline systems. Besides that, since our device avoids data movement and alleviates the memory hierarchy’s inefficiency due to poor spatial data locality, we can also reduce energy consumption by 85% when compared to the baseline.<br>**Download Paper (PDF; Only available from the DATE venue WiFi)**<br>|
16:00 |  | **Coffee Break in Exhibition Area**<br>**Coffee Breaks in the Exhibition Area**<br>On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).<br><br>**Lunch Breaks (Großer Saal + Saal 1)**<br>On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.<br><br>Tuesday, March 20, 2018<br>- Coffee Break 10:30 - 11:30<br>- Lunch Break 13:00 - 14:30<br>- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20<br>- Coffee Break 16:00 - 17:00<br><br>Wednesday, March 21, 2018<br>- Coffee Break 10:00 - 11:00<br>- Lunch Break 12:30 - 14:30<br>- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20<br>- Coffee Break 16:00 - 17:00<br><br>Thursday, March 22, 2018<br>- Coffee Break 10:00 - 11:00<br>- Lunch Break 12:30 - 14:30<br>- Keynote Lecture in “Saal 2” 13:20 - 13:50<br>- Coffee Break 15:30 - 16:00
This session covers various reliability modeling, characterization and mitigation approaches at different abstraction levels. The first paper uses deep learning for variability characterization. The second paper addresses program aging mitigation schemes for voltage regulators. The third paper addresses program vulnerability in GPU applications.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30</td>
<td>7.5.1</td>
<td>LOW-COST HIGH-ACCURACY VARIATION CHARACTERIZATION FOR NANOSCALE IC TECHNOLOGIES VIA NOVEL LEARNING-BASED TECHNIQUES</td>
<td>Zhijian Pan, Tsinghua University, CN; Miao Li, University of Minnesota, US; Paolo Rech, Universidade Federal do Rio Grande do Sul, BR</td>
</tr>
<tr>
<td>15:00</td>
<td>7.5.2</td>
<td>MITIGATION OF NBTI INDUCED PERFORMANCE DEGRADATION IN ON-CHIP DIGITAL LDOs</td>
<td>Longlei Wang, University of South Florida, US; Fritz Previlon, Northeastern University, US; Jian Yao, Yuan Ze University, TW</td>
</tr>
<tr>
<td>15:30</td>
<td>7.5.3</td>
<td>EVALUATING THE IMPACT OF EXECUTION PARAMETERS ON PROGRAM VULNERABILITY IN GPU APPLICATIONS</td>
<td>Fritz Previlon, Northeastern University, US; Charu Kaira, Paolo Rech, David Kaeli, Northeastern University, US; Yuan Ze University, TW</td>
</tr>
</tbody>
</table>

Download Paper (PDF; Only available from the DATE venue WiFi)

14:30 - 16:00 Location / Room: Kont. 3

AN EFFICIENT NBTI-AWARE WAKE-UP STRATEGY FOR POWER-GATED DESIGNS

Yu-Guang Chen, Yuan Ze University, TW

Kun-Wei Chu, Yu-Guang Chen and Ing-Chao Lin
National Cheng Kung University, TW; Yuan Ze University, TW

Abstract
The wake-up process of a power-gated design may induce an excessive surge current and threaten the signal integrity. A proper wake-up sequence should be carefully designed to avoid surge current violations. On the other hand, PMOS sleep transistors may suffer from the negative-bias temperature instability (NBTI) effect which results in decreased driving current. Conventional wake-up sequence decision approaches do not consider the NBTI effect, which may result in a longer or unacceptable wake-up time after circuit aging. Therefore, in this paper, we propose a novel NBTI-aware wake-up strategy to reduce the average wake-up time within a circuit lifetime. Our strategy first finds a set of proper wake-up sequences for different aging scenarios (i.e. after a certain period of aging), and then dynamically reconfigures the wake-up sequences at runtime. The experimental results show that compared to a traditional fixed wake-up sequence approach, our strategy can reduce average wake-up time by as much as 45.04% with only 3.7% extra area overhead for the reconfiguration structure.

Download Paper (PDF; Only available from the DATE venue WiFi)
DESIGNING RELIABLE PROCESSOR CORES IN ULTIMATE CMOS AND BEYOND: A DOUBLE SAMPLING SOLUTION

Nacer-Eddine Zergainoh, TIMA, FR

Thierry Bonnot, Fraidy Bouesse, Nacer-Eddine Zergainoh and Michael Nicolaidis, TIMA, FR

Abstract

The double sampling paradigm is an efficient method to protect the circuits against soft-errors. But the data that are going out of the area protected by double sampling are still vulnerable. To eliminate this weakness without having additional constraints on the datapaths, the most common solution adds a contaminable buffer stage between the two areas. Therefore, this stage avoids the propagation of the potentially corrupted data further in the circuit when an error is detected in the double sampling area. But the issue is that this stage must itself be protected against soft-errors, which drastically increases the cost of the solution. In this paper we characterize the additional implementation constraints due to this vulnerability. We proposed an architectural solution that uses three latches to remove those constraints and protect the area outside the double sampling domain without adding a buffer stage. We present an implementation of this solution on the LEON3 processor, and we compare the results in terms of additional cost and efficiency with other solutions.

Download Paper (PDF; Only available from the DATE venue WiFi)
Deep Neural Networks (DNNs) have emerged as a powerful and versatile set of techniques to address challenging artificial intelligence (AI) problems. Applications in domains such as image/video processing, natural language processing, speech synthesis and recognition, genomics and many others have embraced deep learning as the foundational technique. DNNs achieve superior accuracy for these applications using very large models which require 100s of MBs of data storage, ExaOps of computation and high bandwidth for data movement. Despite advances in computing systems, training state-of-the-art DNNs on large datasets takes several days/weeks, directly limiting the pace of innovation and adoption. In this paper, we discuss how these challenges can be addressed via approximate computing. Based on our earlier studies demonstrating that DNNs are resilient to numerical errors from approximate computing, we present techniques to reduce communication overhead of distributed deep learning training via adaptive residual gradient compression ([email protected][AdaComp]), and computation cost for deep learning inference via Parameterized clipping ACTivation ([email protected][PACT]) based network quantization. Experimental evaluation demonstrates order of magnitude savings in communication overhead for training and computational cost for inference while not compromising application accuracy.

In this paper, we discuss how these challenges can be addressed via approximate computing. Based on our earlier studies demonstrating that DNNs are resilient to numerical errors from approximate computing, we present techniques to reduce communication overhead of distributed deep learning training via adaptive residual gradient compression ([email protected][AdaComp]), and computation cost for deep learning inference via Parameterized clipping ACTivation ([email protected][PACT]) based network quantization. Experimental evaluation demonstrates order of magnitude savings in communication overhead for training and computational cost for inference while not compromising application accuracy.
Coffee Break in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

7.7 Rigorous design, analysis, and monitoring of dependable embedded systems

Date: Wednesday, March 21, 2018
Time: 14:30 - 16:00
Location / Room: Konf. 5

Chair:
Petru Eles, Linköping University, SE, Contact Petru Eles

Co-Chair:
Akash Kumar, Technische Universität Dresden, DE, Contact Akash Kumar

Dependability is a crucial aspect of embedded software systems. This session focuses on achieving dependability in different stages of the embedded software life cycle: requirements engineering, design, and maintenance. In particular, the papers presented in this session will address (1) contract based requirement engineering for cyber-physical systems, (2) formal analysis of code using SMT-based symbolic execution to deal with hardware faults, and (3) non-intrusive runtime trace analysis using FPGAs.

7.7.1 CHASE: CONTRACT-BASED REQUIREMENT ENGINEERING FOR CYBER-PHYSICAL SYSTEM DESIGN

Speaker:
Pier Luigi Nuzzo, University of Southern California, US

Authors:
Pier Luigi Nuzzo1, Michele Lora2, Yishai Feldman3 and Alberto Sangiovanni-Vincentelli4
1University of Southern California, US; 2University of Verona, IT; 3IBM Research, Haifa, IL; 4University of California at Berkeley, US

Abstract
This paper presents CHASE, a framework for requirement capture, formalization, and validation for cyber-physical systems. CHASE combines a practical front-end formal specification language based on patterns with a rigorous verification back-end based on assume-guarantee contracts. The front-end language can express temporal properties of networks using a declarative style, and supports automatic translation from natural-language constructs to low-level mathematical languages. The verification back-end leverages the mathematical formalism of contracts to reason about system requirements and determine inconsistencies and dependencies between them. CHASE features a modular and extensible software infrastructure that can support different domain-specific languages, modeling formalisms, and analysis tools. We illustrate its effectiveness on industrial design examples, including control of aircraft power distribution networks and arbitration of a mixed-criticality automotive bus.

Download Paper (PDF; Only available from the DATE venue WiFi)

7.7.2 RESILIENCE EVALUATION VIA SYMBOLIC FAULT INJECTION ON INTERMEDIATE CODE

Speaker:
Hoang M. Le, University of Bremen, DE

Authors:
Hoang M. Le1, Vladimir Herdt2, Daniel Grosse2 and Rolf Drechsler2
1University of Bremen, DE; 2University of Bremen/DFKI GmbH, DE

Abstract
There is a growing need for error-resilient software that can tolerate hardware faults as well as for new resilience evaluation techniques. For the latter, a promising direction is to apply formal techniques in fault injection-based evaluations to improve the coverage of evaluation results. Building on the recent development of Software-implemented Fault Injection (SWI) techniques on compiler's intermediate code, this paper proposes a novel resilience evaluation framework combining LLVM-based SWI and SMT-based symbolic execution. This novel combination offers significant advantages over state-of-the-art approaches with respect to accuracy and coverage.

Download Paper (PDF; Only available from the DATE venue WiFi)
ONLINE ANALYSIS OF DEBUG TRACE DATA FOR EMBEDDED SYSTEMS

Speaker: Philip Gottschling, TU Darmstadt, DE
Authors: Normann Decker1, Boris Deyer2, Philip Gottschling2, Christian Hochberger2, Alexander Lange2, Martin Leucker1, Torben Scheffel1, Simon Wegener4 and Alexander Weiss2
1Universität zu Lübeck, DE; 2TU Darmstadt, DE; 3Accemic Technologies GmbH, DE; 4AbsInt Angewandte Informatik GmbH, DE

Abstract
Modern multi-core Systems-on-Chip (SoC) provide very high computational power. On the downside, they are hard to debug and it is often very difficult to understand what is going on in these chips because of the limited observability inside the SoC. Chip manufacturers try to compensate this difficulty by providing highly compressed trace data from the individual cores. In the past, the common way to deal with this data was storing it for later offline analysis, which severely limits the time span that can be observed. In this contribution, we present an FPGA-based solution that is able to process the trace data in real-time, enabling continuous observation of the state of a core. Moreover, we discuss applications enabled by this technology.

Download Paper (PDF; Only available from the DATE venue WiFi)

Coffee Break in Exhibition Area

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

7.8 22FDX - the superior technology for IoT, RF, Automotive and Mobility: Advanced Design Methodologies for Ultra-low Power Solutions

Date: Wednesday, March 21, 2018
Time: 14:30 - 16:00
Location / Room: Exhibition Theatre

Organiser: Claudia Kretzschmar, GLOBALFOUNDRIES, DE, Contact Claudia Kretzschmar

22FDX is the choice for applications in mobility, IoT, RF and mmWave as well as Automotive applications. It provides low active and standby power at a very small area. It is equally suited for digital as well as analog/RF/mmWave applications. The back gate bias capability provides an additional degree of freedom to the designer allowing the usage of near-threshold operation. Back gate biasing opens the possibility for many innovative design features like boosting the operation speed when needed as well as compensating for aging and process, temperature and voltage variations. Compared to other advanced node technologies 22FDX has a very low mask count which makes the technology a perfect fit for low-cost applications.

This session will give an introduction into the technology and provide an overview over design methodology. Adaptive body biasing is one of the innovative design methods that will be presented in the third talk applied to extreme low-voltage MPSoC. This session will be concluded with the design of a SoC base on the open-source PULPissimo architecture, built around a 32-bit RISC-V core.

22FDX: A TECHNOLOGY ALTERNATIVE TO THE MAINSTREAM OPTIMIZED FOR IOT APPLICATIONS

Speaker: Jürgen Faul, GLOBALFOUNDRIES Fab1 LLC & Co. KG, DE

Abstract
Serving the new trend in semiconductor industries to connect everything with everything, computing power does not matter as much as low leakage and/or low dynamic power at low cost.

GLOBALFOUNDRIES offers a technology with less complexity than FinFET, same gate length scaling enabled by fully-depleted channels, but with additional features like back-gate biasing which perfectly suits the IoT market needs.

Back-biasing is unique to FDSOI technologies and provides an additional degree of freedom to circuit and chip designers. Prominent examples for back-biasing utilization are extremely low Vdd operation and chip-level global corner trimming, static by OTP or eFuse as well as dynamic for power and temperature compensation.

This talk will give an overview on technology capabilities and features.
15:10 7.8.3 ADAPTIVE BODY BIAS FOR A 0.4V OPERABLE MPSOC IN 22FDX AS AN EXAMPLE FOR BIG DATA HANDLING

Speaker: Christian Mayr, Technische Universität Dresden, DE

Abstract
One of the hottest buzzwords today is big data, i.e. the massive amounts of data that are produced by an ever expanding number of sensors across disciplines from archaeology to soccer. Some examples: in 2016, the amount of data transmitted globally for the first time exceeded a Zettabyte (10^21), using up about 10% of the world energy supply. In 2015, the number of image sensors has risen above the number of humans on earth.

Thus, there is a need for dedicated data processing/machine learning chips that handle this load automatically and reduce it to a data extract usable by humans. Deployment of these chips can be anywhere along the processing chain, e.g. integrated with the sensor interface to reduce data load at the source or as data aggregator in a server farm.

Prof. Mayr will give an overview of the multi-processor systems-on-chip (MPSoC) and sensor interfacing developed at his chair. As some of the first MPSoCs in 22nm FDSOI which use adaptive body biasing to compensate for process variability, they operate as low as 0.4V. Through a combination of dedicated accelerators (e.g. for machine learning) and conventional CPUs, these MPSoCs achieve an optimal compromise between energy efficiency and configurability. Applications pursued at the chair include: sensor nodes for the tactile internet, autonomous driving, neural implants and brain simulation.

15:30 7.8.4 QUENTIN: A NEAR-THRESHOLD SOC FOR ENERGY-EFFICIENT IOT END-NODES IN 22NM FDX TECHNOLOGY

Speaker: Davide Rossi, Università di Bologna, IT

Abstract
An increasing number of end-node IoT applications require high performance and extreme energy efficiency to deal with the high computational requirements of near-sensor data analytics algorithms, within a power envelope of few milliWatts for long battery lifetime. A significant improvement of energy efficiency for digital computing systems can be achieved exploiting near-threshold operation. We present Quentin: a near-threshold SoC based on the open-source PULPsimo architecture, implemented in 22nm FDX technology. The proposed SoC is built around a 32-bit RISC-V core “RISCY”, optimized for energy efficient digital signal processing, 512 KByte of L2 memory, and an autonomous I/O subsystem featuring an IO DMA coupled with a standard set of peripherals. RISCY features a 32-bit, 4 stages in-order pipeline implementing the RV32IMF C RISC-V ISA plus domain specific extensions for near-sensor data analytics such as packed-SIMD (additions, comparisons, logic, shuffle and dot product), bit manipulation, hardware loops, etc. This talk will present the implementation of the SoC in 22nm FDX, featuring a not less than 2 mm², a maximum operating frequency of 170 MHz (SSG, 0.59V, 40 C) and an estimated power consumption of 5 mW.

16:00 End of session
Coffee Break in Exhibition Area

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00
Interactive Presentations run simultaneously during a 30-minute slot. Additionally, each IP paper is briefly introduced in a one-minute presentation in a corresponding regular session.

**IP3-1**

**TESTBENCH QUALIFICATION FOR SYSTEMC-AMS TIMED DATA FLOW MODELS**

**Speaker:** Muhammad Hassan, DFKI GmbH, DE  
**Authors:** Muhammad Hassan¹, Daniel Grosse², Hoang M. Le³, Thilo Voelter⁴, Karsten Einwich⁵ and Roll Drechsle²  
¹Cyber Physical Systems, DFKI, DE; ²University of Bremen/DFKI GmbH, DE; ³University of Bremen, DE; ⁴COSEDA Technologies GmbH, DE

**Abstract**  
Analog Mixed Signal (AMS) circuits have become increasingly important for today’s SoCs. The Timed Data Flow (TDF) model of computation available in SystemC-AMS offers here a good tradeoff between accuracy and simulation speed at the system-level. One of the main challenges in system-level verification is the quality of the testbench. In this paper, we present a testbench qualification approach for SystemC-AMS TDF models. Our contribution is twofold: First, we propose specific mutation models for the class of filters implemented as TDF models. This requires to analyze the Laplace transfer function of the filter design. Second, we present the mutation based qualification approach based on the proposed specific mutations as well as standard behavioral mutations. This allows to find serious quality issues in the testbench. Our experimental results for a real-world AMS system demonstrate the applicability and efficacy of our approach.

**Download Paper (PDF; Only available from the DATE venue WiFi)**

**IP3-2**

**AN ALGEBRA FOR MODELING CONTINUOUS TIME SYSTEMS**

**Speaker:** José Medeiros, University of Brasilia, BR  
**Authors:** José E. G. de Medeiros¹, George Ungureanu² and Ingo Sander²  
¹University of Brasilia, BR; ²KTH Royal Institute of Technology, SE

**Abstract**  
Advancements on analog integrated design have led to new possibilities for complex systems combining both continuous and discrete time modules on a signal processing chain. However, this also increases the complexity any design flow needs to address in order to describe a synergy between the two domains, as the interactions between them should be better understood. We believe that a common language for describing continuous and discrete time computations is beneficial for such a goal and a step towards it is to gain insight and describe more fundamental building blocks. In this work we present an algebra based on the General Purpose Analog Computer, a theoretical model of computation recently updated as a continuous time equivalent of the Turing Machine.

**Download Paper (PDF; Only available from the DATE venue WiFi)**

**IP3-3**

**TTW: A TIME-TRIGGERED WIRELESS DESIGN FOR CPS**

**Speaker:** Roman Jacob, ETH Zurich, CH  
**Authors:** Roman Jacob¹, Licong Zhang¹², Marco Zimmerling³, Jan Beutel¹, Samarjit Chakraborty² and Lothar Thiele¹  
¹ETH Zurich, CH; ²Technical University of Munich, DE; ³Technical Universität Dresden, DE

**Abstract**  
Wireless fieldbuses have long been proven effective in supporting Cyber-Physical Systems (CPS). However, various domains are now striving for wireless solutions due to ease of deployment or novel functionality requiring the ability to support mobile devices. Low-power wireless protocols have been proposed in response to this need, but requirements of a large class of CPS applications can still not be satisfied. We thus propose Time-Triggered Wireless (TTW), a distributed low-power wireless system design that minimizes communication energy consumption and offers end-to-end timing predictability, runtime adaptability, reliability, and low latency. Evaluation shows a 2x reduction in communication latency and 33-40% lower radio-on time compared with DRP, the closest related work, validating the suitability of TTW for new exciting wireless CPS applications.

**Download Paper (PDF; Only available from the DATE venue WiFi)**

**IP3-4**

**PHYLAX: SNAPSHOT-BASED PROFILING OF REAL-TIME EMBEDDED DEVICES VIA JTAG INTERFACE**

**Speaker:** Eduardo Chielle, New York University Abu Dhabi, BR  
**Authors:** Chalambos Konstantinou¹, Eduardo Chielle² and Michail Maniatakos²  
¹New York University, US; ²New York University Abu Dhabi, AE

**Abstract**  
Real-time embedded systems play a significant role in the functionality of critical infrastructure. Legacy microprocessor-based embedded systems, however, have not been developed with security in mind. Applying traditional security mechanisms in such systems is challenging due to computing constraints and/or real-time requirements. Their typical 20-30 year lifespan further exacerbates the problem. In this work, we propose PHYLAX, a plug-and-play solution to detect intrusions in already installed embedded devices. PHYLAX is an external monitoring tool which does not require code instrumentation. Also, our tool adapts and prioritizes intrusion detection based on the requirements of the underlying infrastructure (power grid, chemical factory, etc.) as well as the computing capabilities of the target embedded system (CPU model, memory size, etc.). PHYLAX can be employed on any legacy device which incorporates a JTAG interface. As a case study, we present the inclusion of PHYLAX on a power grid recloser controller.

**Download Paper (PDF; Only available from the DATE venue WiFi)**

**IP3-5**

**CHARACTERIZING DISPLAY QOS BASED ON FRAME DROPPING FOR POWER MANAGEMENT OF INTERACTIVE APPLICATIONS ON SMARTPHONES**

**Speaker:** Chung-Ta King, National Tsing Hua University, TW  
**Authors:** Kuan-Ting Ho¹, Chung-Ta King¹, Bhaskar Das¹ and Yung-Ju Chang²  
¹National Tsing Hua University, TW; ²National Chiao Tung University, TW

**Abstract**  
User-centric power management in smartphones aims to conserve power without affecting user’s perceived quality of experience. Most existing works focus on periodically updated applications such as games and video players and use a fixed frame rate, measured in frame per second (FPS), as the metric to quantify the display quality of service (QoS). The idea is to adjust the CPU/GPU frequency just enough to maintain the frame rate at a user satisfactory level. However, when applied to aperiodically-updated interactive applications, e.g. Facebook or Instagram, that draw the frame buffer at a varying rate in response to user inputs, such a power management strategy becomes too conservative. Based on real user experiments, we observe that users can tolerate a certain percentage of frame drops when running aperiodically updated applications without affecting their perceived display quality. Hence, we introduce a new metric to characterize display quality of service, called the frame drawn ratio (FDR), and propose a new CPU/GPU frequency governor based on the FDR metric. The experiments by real users show that the proposed governor can conserve 17.2% power in average when compared to the default governor, while maintaining the same or even better QoE rating.

**Download Paper (PDF; Only available from the DATE venue WiFi)**
Abstract

Hardware implementations of arithmetic operations over binary finite fields $\mathbb{GF}(2^m)$ are widely used in several important applications, such as cryptography, digital signal processing, and error-control codes. In this paper, efficient reconfigurable implementations of bit-parallel canonical basis multipliers over binary fields generated by type II irreducible pentanomials $f(y) = y^m + y^{n+2} + y^{n+1} + y^n + 1$ are presented. These pentanomials are important because all five binary fields recommended by NIST for ECDSA can be generated from them. Experimental results show that the proposed multiplier implementations improve the area$ \times$ time parameter when compared with similar multipliers found in the literature.

Download Paper (PDF; Only available from the DATE venue WiFi)
IP3-11  PROCESSING IN 3D MEMORIES TO SPEEDUP OPERATIONS ON COMPLEX DATA STRUCTURES
Speaker:
Luigi Camo, UFRGS, BR
Authors:
Paulo Cesar Santos1, Geraldo Francisco de Oliveira Junior1, Joao Paulo Lima1, Marco Antonio Zanata Akex2, Luigi Camo1 and Antonio Carlos Schneider Beck1
1UFRGS, BR; 2UFFR, BR
Abstract
Pointing chasing has been, for years, the kernel operation employed by diverse data structures, from graphs to hash tables and dictionaries. However, due to the bewildering growth in the volume of data that current applications have to deal with, performing pointer chasing operations have become a major source of performance and energy bottleneck, due to its sparse memory access behavior. In this work, we aim to tackle this problem by taking advantage of the already available parallelism present in today’s 3D-stacked memories. We present a simple mechanism that can accelerate pointer chasing operations by making use of a state-of-the-art PIM design that executes in-memory vector operations. The key idea behind our design is to run speculative loads, in parallel, based on a given memory address in a reconfigurable window of addresses. Our design can perform pointer-chasing operations on b-tree 4.9x faster when compared to modern baseline systems. Besides that, since our device avoids data movement and alleviates the memory hierarchy’s inefficiency due to poor spatial data locality, we can also reduce energy consumption by 85% when compared to the baseline.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP3-12  AN EFFICIENT NBTI-AWARE WAKE-UP STRATEGY FOR POWER-GATED DESIGNS
Speaker:
Yu-Guang Chen, Yuan Ze University, TW
Authors:
Korn-Wei Chiu1, Yu-Guang Chen2 and Ing-Chao Lin1
1National Cheng Kung University, TW; 2Yuan Ze University, TW
Abstract
The wake-up process of a power-gated design may induce an excessive surge current and threaten the signal integrity. A proper wake-up sequence should be carefully designed to avoid surge current violations. On the other hand, PMOS sleep transistors may suffer from the negative bias temperature instability (NBTI) effect which results in decreased driving current. Conventional wake-up sequence decision approaches do not consider the NBTI effect, which may result in a longer or unacceptable wake-up time after circuit aging. Therefore, in this paper, we propose a novel NBTI-aware wake-up strategy to reduce the average wake-up time within a circuit Werime. Our strategy first finds a set of proper wake-up sequences for different aging scenarios (i.e. after a certain period of aging), and then dynamically reconfigures the wake-up sequences at runtime. The experimental results show that compared to a traditional fixed wake-up sequence approach, our strategy can reduce average wake-up time by as much as 45.04% with only 3.7% extra area overhead for the reconfiguration structure.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP3-13  DESIGNING RELIABLE PROCESSOR CORES IN ULTIMATE CMOS AND BEYOND: A DOUBLE SAMPLING SOLUTION
Speaker:
Nacer-Eddine Zergainoh, TIMA, FR
Authors:
Thierry Bonnot, Frady Bouesse, Nacer-Eddine Zergainoh and Michael Nicolaidis, TIMA, FR
Abstract
The double sampling paradigm is an efficient method to protect the circuits against soft-errors. But the data that are going out of the area protected by double sampling are still vulnerable. To eliminate this weakness without having additional constraints on the datapaths, the most common solution adds a controllable buffer stage between the two areas. Therefore, this stage avoids the propagation of the potentially corrupted data further in the circuit when an error is detected in the double sampling area. But the issue is that this stage must itself be protected against soft-errors, which drastically increases the cost of the solution. In this paper we characterize the additional implementation constraints due to this vulnerability. We proposed an architectural solution that uses three latches to remove those constraints and protect the area outside the double sampling domain without adding a buffer stage. We present an implementation of this solution on the LEON3 processor, and we compare the results in terms of additional cost and efficiency with other solutions.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP3-14  DESIGN OF A TIME-PREDICTABLE MULTICORE PROCESSOR: THE T-CREST PROJECT
Speaker and Author:
Martin Schroeber, Technical University of Denmark, DK
Abstract
Real-time systems need to deliver results in time and often this timely production of a result needs to be guaranteed. Static timing analysis can be used to bound the worst-case execution time of tasks. However, this timing analysis is only possible if the processor architecture is analysis friendly. This paper presents the T-CREST processor, a real-time multicore processor developed to be time-predictable and an easy target for static worst-case execution time analysis. We present how to achieve time-predictability at all levels of the architecture, from the processor pipeline, via a network-on-chip, up to the memory controller. The main architectural feature to provide time predictability is to use static arbitration of shared resources in a time division multiplexing way.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP3-15  ERROR RESILIENCE ANALYSIS FOR SYSTEMATICALLY EMPLOYING APPROXIMATE COMPUTING IN CONVOLUTIONAL NEURAL NETWORKS
Speaker:
Muhammad Abdullah Hanif, Vienna University of Technology, Vienna, AT
Authors:
Muhammad Abdullah Hanif1, Rehan Hafiz2 and Muhammad Shafique1
1TU Wien, AT; 2ITU, PK
Abstract
Approximate computing is an emerging paradigm for error resilient applications as it leverages accuracy loss for improving power, energy, area, and/or performance of an application. The spectrum of error resilient applications includes the domains of image and video processing, Artificial intelligence (AI) and Machine Learning (ML), data analytics, and other Recognition, Mining, and Synthesis (RMS) applications. In this work, we address one of the most challenging question, i.e., how to systematically employ approximate computing in Convolutional Neural Networks (CNNs), which are one of the most compute-intensive and the pivotal part of AI. Towards this, we propose a methodology to systematically analyze error resilience of deep CNNs and identify parameters that can be exploited for improving performance-efficiency of these networks for inference purposes. We also present a case study for significance-driven classification of filters for different convolutional layers, and propose to prune those having the least significance, and thereby enabling accuracy vs. efficiency tradeoffs by exploiting their resilience characteristics in a systematic way.
Download Paper (PDF; Only available from the DATE venue WiFi)
DEMAS: AN EFFICIENT DESIGN METHODOLOGY FOR BUILDING APPROXIMATE ADDERS FOR FPGA-BASED SYSTEMS

Speaker:
Semeen Rehman, Vienna University of Technology (TU Wien), AT

Authors:
Bharath Srinivas Prabakaran1, Semeen Rehman1, Muhammad Abdullah Hanif1, Salim Ullah2, Ghazal Mazaheri2, Akash Kumar2 and Muhammad Shafique1
1TU Wien, AT; 2Technische Universität Dresden, DE; 3UC Riverside, US

Abstract:
The current state-of-the-art approximate adders are mostly ASIC-based, i.e., they focus solely on gate and/or transistor level approximations (e.g., through circuit simplification or truncation) to achieve latency, power, and/or energy savings at the cost of accuracy loss. However, when these designs are synthesized for FPGA-based systems, they do not offer similar reductions in area, latency, and power/energy due to the underlying architectural differences between ASICs and FPGAs. In this paper, we present a novel design methodology to synthesize and implement approximate adders for any FPGA-based system by considering the underlying resources and architectural differences. Using our methodology, we have designed, analyzed and presented eight different multi-bit adder architectures. Compared to the 16-bit accurate adder, our designs are successful in achieving area, latency and power-delay product gains of 55%, 28%, and 53%, respectively. We also compare our approximate adders to state-of-art adders specialized for ASIC and FPGA fabrics and demonstrate the benefits of our approach. We will make the RTL and behavioral models of our and state-of-the-art designs open-source at https://sourceforge.net/projects/approxfpgas/ to further fuel the research and development in the FPGA community and to ensure reproducible research.

Download Paper (PDF; Only available from the DATE venue WiFi)

GAIN SCHEDULED CONTROL FOR NONLINEAR POWER MANAGEMENT IN CMPS

Speaker:
Nikil Dutt, University of California, Irvine, US

Authors:
Bryan Donyanavard, Amir M. Rahmani, Tiago Muck, Kasra Moazzeni and Nikil Dutt, University of California, Irvine, US

Abstract:
Dynamic voltage and frequency scaling (DVFS) is a well-established technique for power management of thermal- or energy-sensitive chip multiprocessors (CMPs). In this context, linear control theoretic solutions have been successfully implemented to control the voltage-frequency knobs. However, modern CMPs with a large range of operating frequencies and multiple voltage levels display nonlinear behavior in the relationship between frequency and power. State-of-the-art linear controllers therefore leave room for opportunity in optimizing DVFS operation. We propose a Gain Scheduled Controller (GSC) for nonlinear runtime power management of CMPs that simplifies the controller implementation of systems with varying dynamic properties by utilizing an adaptive control theoretic approach in conjunction with static linear controllers. Our design improves the stability, accuracy, settling time, and overshoot of the controller over a linear controller with minimal overhead. We implement our approach on an Exynos platform containing ARM's big.LITTLE-based heterogeneous multi-processor (HMP) and demonstrate that the system's response to changes in target power is improved by 2x while operating up to 12% more efficiently.

Download Paper (PDF; Only available from the DATE venue WiFi)
Abstract
Augmented Reality (AR) currently require large form factors, weight, cost and frequent recharging cycles that reduce usability. Connectivity, image processing, localization, and direction evaluation lead to high processing and power requirements. A multi-antenna system, patented by the industrial partner, enables a new generation of smart eye-wear that elegantly requires less hardware, connectivity, and power to provide AR functionalities. They will allow users to directionally locate nearby radio emitting sources that highlight objects of interest (e.g., people or retail items) by using existing standards like Bluetooth Low Energy, Apple’s iBeacon and Google’s Eddystone. This booth will report the current level of research addressed by the Computer Science Department of University of Verona, Wagoo LLC, and Wagoo Italia srls. In the presented demo, different objects emit an “I am here” signal and a prototype of the smart glasses shows the information related to the observed object.

More information ...

T-CREST: THE OPEN-SOURCE REAL-TIME MULTICORE PROCESSOR

Authors:
Martin Schoebert, Luca Pezzarossa and Jens Sparsø, Technical University of Denmark, DK

Abstract
Future real-time systems, such as advanced control systems or real-time image recognition, need more powerful processors, but still a system where the worst-case execution time (WCET) can be statically predicted. Multicore processors are one answer to the need for more processing power. However, it is still an open research question how to best organize and implement time-predictable communication between processing cores. T-CREST is an open-source multicore processor for research on time-predictable computer architecture. It consists of several Patmos processors connected by various time-predictable communication structures: access to shared off-chip, access to shared on-chip memory, and the Ago network-on-chip for fast inter-processor communication. T-CREST is supported by open-source development tools, such as compilation and WCET analysis. To best of our knowledge, T-CREST is the only fully open-source architecture for research on future real-time multicore architectures.

More information ...

EXPERIENCE-BASED AUTOMATION OF ANALOG IC DESIGN

Authors:
Florent Leber and Juergen Scheible, Reutlingen University, DE

Abstract
While digital design automation is highly developed, analog design automation still remains behind the demands. Previous circuit synthesis approaches, which are usually based on optimization algorithms, do not satisfy industrial requirements. A promising alternative is given by procedural approaches (also known as “generators”): They (a) emulate experts’ decisions, thus (b) make expert knowledge re-usable and (c) can consider all relevant aspects and constraints implicitly. Nowadays, generators are successfully applied in analog layout (Pcells, Pycells). We aim at an entire design flow completely based on procedural automation techniques. This flow will consist of procedures for the generation of schematics and layouts for every typical analog circuit class, such as amplifier, bandgap, filter a.s.o. In our presentation we give an overview on such a design flow and we show an approach for capturing an analog circuit designer's strategy as an executable “expert design plan”.

More information ...

3D NANOSYSTEMS: THE PATH TO 1,000X ENERGY EFFICIENCY

Author:
Max Shulaker, MIT, US

Abstract
While trillions of sensors connected to the “Internet of Everything” (IoE) promise to transform our lives, they simultaneously pose major obstacles which we are already encountering today. The massive amount of generated raw data (i.e., the “data deluge”) is quickly exceeding computing capabilities of existing systems, and cannot be overcome by isolated improvements in sensors, transistors, memories or architectures alone. Rather, an end-to-end approach is needed, whereby the unique benefits of new emerging nanotechnologies - for sensors, memories and transistors - are exploited to realize new nanosystem architectures that are not possible using today’s technologies. However, emerging nanomaterials and nanodevices suffer from significant imperfections and variations. Thus, realizing working circuits, let alone transistors, has been infeasible. In this talk, I present a path towards realizing future nanosystems, and show how recent progress in several emerging nanotechnologies (carbon nanotubes for logic, non-volatile memories for data storage, and new materials for sensing) enables us to realize such nanosystems today. As a case-study, I will discuss how by leveraging emerging nanotechnologies, we have realized the first monolithically-integrated three-dimensional (3D) nanosystem architectures with vertically-integrated layers of logic, memory, and sensing circuits. With dense and fine-grained connectivity between millions of on-chip sensors, data storage, and embedded computation, such nanosystems can capture terabytes of data from the outside world every second, and produce “processed information” by performing in-situ classification of the sensor data using on-chip accelerators. As a demonstration, we tailor a demo system for gas classification, for real-time health monitoring from breath.

RESISTIVE RAM FOR NEW COMPUTING SYSTEMS: FROM DEEP LEARNING TO BIOMIMICRY

Author:
Elisa Vianello, CEA LETI, FR

Abstract
Resistive random-access memory (RRAM) is a memory technology that promises high-capacity, non-volatile data storage, low voltages, fast programming and reading time (few 10's of ns, even <1ns), single bit alterability, execution in place, good cycling performance (higher than Flash), density. Moreover RRAM can be easily integrated in the Back-End-Of-Line of advanced CMOS logic. This will revolutionize traditional memory hierarchy and facilitate the implementation of in-memory computing architectures and Deep Learning accelerators. To further improve the connectivity between memory arrays and computing, a combination of logic 3D Sequential Integration (3DSI) and memory arrays is a promising solution. Thanks to low processing thermal budget (>400 °C), thermal stability (>500 °C) and low cost (few additional masks), RRAM technologies are good candidates to be inserted in between sequentially stacked MOSFET tiers. RRAMs are also promising candidates for implementing energy-efficient bioinspired synapses, creating a path towards online real time unsupervised learning and life-long learning abilities. We will also explore the use of RRAM for future circuits and systems inspired by the emerging paradigm of biominervy.
The tremendous value computation has shown across applications is driving its expansion from cyber systems to systems that pervade every aspect of our lives. This is being fueled especially by algorithms from artificial intelligence, leading to systems qualified for such integration in our lives, with cognitive capabilities approaching those of humans. A fascinating consequence for system designers is that a tight coupling now results between the data sensed from the physical world and the computations performed on that data. This enforces a unification of design spaces, where new sensing technologies open up new algorithmic opportunities, which in turn open up new architectural options, bringing the potential to overcome traditional bottlenecks in computing. But, a conceptual unification is not enough, a technological unification is also needed. This talk explores such a unification, via hybrid systems based on Large-Area Electronics (LAE) and silicon-CMOS technologies. LAE enables diverse, expansive, and form-fitting sensors, which can be associated with physical objects. This yields semantic structure in the sensor data, which can be exploited towards simpler machine-learning models that can fuel especially by algorithms from artificial intelligence, leading to systems qualified for such integration in our lives, with cognitive capabilities approaching those of humans. A fascinating consequence for system designers is that a tight coupling now results between the data sensed from the physical world and the computations performed on that data. This enforces a unification of design spaces, where new sensing technologies open up new algorithmic opportunities, which in turn open up new architectural options, bringing the potential to overcome traditional bottlenecks in computing. But, a conceptual unification is not enough, a technological unification is also needed. This talk explores such a unification, via hybrid systems based on Large-Area Electronics (LAE) and silicon-CMOS technologies. LAE enables diverse, expansive, and form-fitting sensors, which can be associated with physical objects. This yields semantic structure in the sensor data, which can be exploited towards simpler machine-learning models that bring significant improvements of WCET estimates (up to 2.7x) provided that the WCET analysis process is guided with automatically generated flow annotations obtained using polyhedral counter techniques.

This article is a review of the current progress and results obtained in the European H2020 CONNECT project. Amongst all the research on carbon nanotube interconnects, those discussed here cover 1) process & growth of carbon nanotube interconnects compatible with back-end-of-line integration, 2) modeling and simulation from atomic to circuit-level benchmarking and performance prediction, and 3) characterization and electrical measurements. We provide an overview of the current advancements on carbon nanotube interconnects and also regarding the prospects for designing energy efficient integrated circuits. Each selected category is presented in an accessible manner aiming to serve as a review and informative cornerstone on carbon nanotube interconnects.

Download Paper (PDF; Only available from the DATE venue WiFi)
8.3 Real time intelligent methods for energy-efficient approaches in CNN and biomedical applications

**Date:** Wednesday, March 21, 2018  
**Time:** 17:00 - 18:30  
**Location / Room:** Konf. 1

**Chair:** Theocharis Theocharides, University of Cyprus, CY; Contact Theocharis Theocharides

**Co-Chair:** Jose L. Ayala, Depto Arquitectura de Computadores - UCM, ES; Contact Jose L. Ayala

Mobile devices and wearables require increased integration of technology for real-time applications, in particular in health and transport technology. This enables the possibility to implement machine-learning techniques directly on board. This session will firstly outline applications to detect and predict pathological health conditions, before examining real-time applications in UAVs.

### 8.3.1 ONLINE EFFICIENT BIO-MEDICAL VIDEO TRANSCODING ON MPSOCs THROUGH CONTENT-AWARE WORKLOAD ALLOCATION

**Speaker:** Arman Iranfar, Embedded Systems Lab (ESL), EPFL, CH

**Authors:** Arman Iranfar1, Ali Pahlevan1, Marina Zapater1, Martin Zaga2, Mario Kovač2 and David Atienza1

1Embedded Systems Lab (ESL), EPFL, CH; 2University of Zagreb, HR

**Abstract**

Bio-medical image processing in the field of telemedicine, and in particular the definition of systems that allow medical diagnostics in a collaborative and distributed way is experiencing an undeniable growth. Due to the high quality of bio-medical videos and the subsequent large volumes of data generated, to enable medical diagnosis on-the-go it is imperative to efficiently transcode and stream the stored videos on real time, without quality loss. However, online video transcoding is a high-demanding computationally-intensive task and its efficient management in Multiprocessor Systems-on-Chip (MPSoCs) poses an important challenge. In this work we propose an efficient motion- and texture-aware frame-level parallelization approach to enable online medical imaging transcoding on MPSoCs for next generation video encoders. By exploiting the unique characteristics of bio-medical videos and the medical procedure that enable diagnosis, we split frames into tiles based on their motion and texture, deciding the most adequate level of parallelization. Then, we employ the available encoding parameters to satisfy the required video quality and compression. Moreover, we propose a new fast motion search algorithm for bio-medical videos that allows to drastically reduce the computational complexity of the encoder, thus achieving the frame rates required for online transcoding. Finally, we heuristically allocate the threads to the most appropriate available resources and set the operating frequency of each one. We evaluate our work on an enterprise multicore server achieving online medical imaging with 1.6x higher throughput and 44% less power consumption when compared to the state-of-the-art techniques.

**Download Paper** (PDF: Only available from the DATE venue WiFi)

### 8.3.2 HIGHLY EFFICIENT AND ACCURATE SEIZURE PREDICTION ON CONSTRAINED IOT DEVICES

**Speaker:** Farzad Samee, Karlsruhe Institute of Technology (KIT), DE

**Authors:** Farzad Samee, Sebastian Paul, Lars Bauer and Joerg Henkel, Karlsruhe Institute of Technology, DE

**Abstract**

In this paper we present an efficient and accurate algorithm for epileptic seizure prediction on low-power and portable IoT devices. State-of-the-art algorithms suffer from two issues: computation intensive features and large internal memory requirement, which makes them inapplicable for constrained devices. We reduce the memory requirement of our algorithm by reducing the size of data segments (i.e. the window of input stream data on which the processing is performed), and the number of required EEG channels. To respect the limitation of the computing power, we reduce the complexity of our exploited features by only considering the simple features, which also contributes to reducing the memory requirements. Then, we provide new relevant features to compensate the information loss due to the simplifications (i.e. less number of channels, simpler features, shorter segment, etc.). We measured the energy consumption (12.41 mJ) and execution time (565 ms) for processing each segment (i.e. 5.12 seconds of EEG data) on a low-power MSP432 device. Even though the state-of-art does not fit to IoT devices, we evaluate the classification performance and show that our algorithm achieves the highest AUC score (0.79) for the held-out data and outperforms the state-of-the-art.

**Download Paper** (PDF: Only available from the DATE venue WiFi)
### 8.4 Efficient and reliable memory and computing architectures

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>18:30</td>
<td>IP3-15</td>
<td>ERROR RESILIENCE ANALYSIS FOR SYSTEMATICALLY EMPLOYING APPROXIMATE COMPUTING IN CONVOLUTIONAL NEURAL NETWORKS</td>
<td>Muhammad Abdullah Hanif, Vienna University of Technology, Vienna, AT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Muhammad Abdullah Hanif, Vienna University of Technology, Vienna, AT</td>
<td>Authors: Muhammad Abdullah Hanif, Rehan Hafiz and Muhammad Shafique</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Approximate computing is an emerging paradigm for error resilient applications as it leverages accuracy loss for improving power, energy, area, and/or performance of an application. The spectrum of error resilient applications includes the domains of image and video processing, Artificial intelligence (AI) and Machine Learning (ML), data analytics, and other Recognition, Mining, and Synthesis (RMS) applications. In this work, we address one of the most challenging question, i.e., how to systematically employ approximate computing in Convolutional Neural Networks (CNNs), which are one of the most compute-intensive and the pivotal part of AI. Towards this, we propose a methodology to systematically analyze error resilience of deep CNNs and identify parameters that can be exploited for improving performance/efficiency of these networks for inference purposes. We also present a case study for significance-driven classification of filters for different convolutional layers, and propose to prune those having the least significance, and thereby enabling accuracy vs. efficiency tradeoffs by exploiting their resilience characteristics in a systematic way.</td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
</tbody>
</table>

18:30 End of session
**Abstract**

HiVE is a Hybrid Vertex-Edge memory hierarchy. We propose HiVE to address the shortcomings of existing memory hierarchies by avoiding random access and data written to ReRAM modules. HiVE is process variation-aware and aging-aware. It comprises of two levels: (1) it identifies a region of cores suitable for mapping, and (2) it maps threads in the region and intersperses deep cores for thermal mitigation while considering the current heat of the cores. Both the levels strive to reduce aging variance across the chip. HiVE outperforms the current state-of-the-art memory hierarchies by up to 2 years at the end of 3.25 years of use, as compared to the latest FPGA and CPU implementations.

**Download Paper (PDF; Only available from the DATE venue WiFi)**
8.5 From NBTI to IoT security: industrial experiences

Date: Wednesday, March 21, 2018
Time: 17:00 - 18:30
Location / Room: Kont. 3

Chair: Doris Keitel-Schulz, Infineon Technologies, DE; Contact Doris Keitel-Schulz
Co-Chair: Norbert Wehn, University of Kaiserslautern, DE; Contact Norbert Wehn

This session covers industrial experiences such as NBTI mitigation and adaptive voltage scaling to system level aspects including safety-critical applications and IoT security.

17:00 8.5.1 NBTI AGED CELL REJUVENATION WITH BACK BIASING AND RESULTING CRITICAL PATH REORDERING FOR DIGITAL CIRCUITS IN 28NM FDSOI
Speaker: Lorena Anghel, TIMA Labs, FR
Authors: Ajith Sivadasan1, Riddhi Jitendrakumar Shah2, Vincent Huard3, Florian Cados3 and Lorena Anghel4
1TIMA Labs & ST Microelectronics, FR; 2TIMA Labs, FR; 3STMicroelectronics, FR; 4Grenoble-Alpes University, FR

Abstract — Increasing demands from Autonomous Driving and IoT markets are pushing the need for products with advanced CMOS nodes that guarantee a high level of performance and at the same time having to comply with industrial/regulatory standards like ISO26262, AEC-Q100 etc. Implementation of NBTI & DMR Reliability models for 28nm FDSOI, developed in-house, are fundamental means to evaluate the reliability of digital IPs during the design phase. Process, Temperature, Voltage, Workload based Aging in the form of process profile and aging profile parameters traditionally taken into account for design margin evaluations and critical path pruning. A precise critical path selection methodology is highly important considering the In-situ monitor Insertion and Critical Path Replica generation strategies to be applied to Runtime Reliability assessment with the vision to move towards dynamic wear out management solutions. This paper recommends the consideration of the silicon technology feature of back biasing as an important parameter while selecting Critical Paths for circuits fabricated with FDSOI process. Back biasing is an add-on feature of this technology with ABB (adaptive back biasing) techniques having been shown to be effective in reducing PVT variations or aimed at a gain in the overall digital circuit performance. This technique is now being increasingly applied to aging mitigation. The back-biasing gain for an aged digital IP is quantified while performing design stage Gate Level Analysis yielding interesting insights on its impact on the operational frequency determining critical path rankings. Keywords— NBTI, Back/Body Biasing, Aging, Critical Path, Reliability

Download Paper (PDF; Only available from the DATE venue WiFi)

17:15 8.5.2 AN INDUSTRIAL CASE STUDY OF LOW COST ADAPTIVE VOLTAGE SCALING USING DELAY TEST PATTERNS
Speaker: Mahroo Zandrahimi, TU Delft, NL
Authors: Mahroo Zandrahimi1, Philippe Debaud2, Amrand Castillo3 and Zaid Al-Ars1
1Delft University of Technology, NL; 2STMicroelectronics, FR

Abstract — In deep sub-micron technologies, the increasing effect of process and environmental variations has lead chip manufacturers to use adaptive voltage scaling techniques in order to adapt operation parameters exclusively to each chip. The increasing effect of process variation is limiting the effectiveness of current chip monitoring approaches, such as on-chip performance monitor boxes (PMBs), which results in yield loss and high design margins, thus high power consumption. This paper proposes an alternative solution for adaptive voltage scaling using delay test patterns, which is able to eliminate the need for PMBs, and thus the long expensive characterization phase of tuning PMBs to each design, while improving the yield as well as power optimization. Results show, using an industrial grade 28nm FD- SOI library developed for low power devices, that delay testing for performance prediction reduces the inaccuracy down to 1.85%.

Download Paper (PDF; Only available from the DATE venue WiFi)
17:30 8.5.3 A CASE STUDY FOR USING DYNAMIC PARTITIONING BASED SOLUTION IN VOLUME DIAGNOSIS

Speaker: Wu Yang, Mentor, A Siemens Business, US
Authors: Tao Wang¹, Zhanghun Shi¹, Junlin Huang¹, Huaxing Tang¹, Wu Yang² and Junna Zhong²
¹Hisilicon Technologies Co., Ltd, CN; ²Mentor, A Siemens Business, US; ³Mentor, A Siemens Business, CN

Abstract
Diagnosis driven yield analysis (DDYA) has been widely adopted for advanced technology node product yield ramp [1]. However, gigantic design size and high pattern count demand intense computation resources to diagnose volume failure data, and the diagnosis throughput becomes the bottleneck of the DDYA flow. This paper presents a case study which uses the fully automated dynamic partitioning based diagnosis solution to dramatically improve the throughput. Experimental results based on real silicon data manufactured by a 16nm FinFET technology shows more than 3X reduction for memory footprint and more than 4X improvement for runtime, which eliminates the throughput bottleneck.

Download Paper (PDF; Only available from the DATE venue WiFi)

17:45 8.5.4 ON-LINE RF BUILT-IN SELF-TEST USING NOISE INJECTION AND TRANSMITTER SIGNAL MODULATION BY PHASE SHIFTER

Speaker and Author: Jan Schat, NXP Semiconductors, DE

Abstract
For on-chip self-test of radar ICs, loopback test using a signal feedback path from transmitter to receiver is state-of-the-art. Usually, such a loopback test is performed periodically after a number of application-mode chips. The traditional loopback test has two drawbacks, however: It is performed intermittent to the application mode, not within the application mode. Moreover, it cannot detect the case that the attenuation from transmitter to receiver becomes too low due to defects on the IC, or due to targets very near to the antennas. This paper proposes an advanced loopback test not intermittent to the application, but during application mode. That way, spurious defects like transient faults (also known as Single Event Upsets) can be detected; moreover, an error-prone plausibility check of the received signal is avoided. To detect receiver saturation due to near targets, modulating the transmitter output signal using a phase shifter is proposed.

Download Paper (PDF; Only available from the DATE venue WiFi)

18:00 8.5.5 NEURAL NETWORKS FOR SAFETY-CRITICAL APPLICATIONS - CHALLENGES, EXPERIMENTS AND PERSPECTIVES

Speaker: Chih-Hong Cheng, fortiss, DE
Authors: Chih-Hong Cheng¹, Frederik Diehl², Yassine Hamza², Gereon Hinz², Georg Nührenberg², Markus Ricken², Harald Ruess² and Michael Troung-Le²
¹fortiss - Landesforschungsinstitut des Freistaats Bayern, DE; ²Rottis GmbH, DE

Abstract
We propose a methodology for designing dependable Artificial Neural Networks (ANNs) by extending the concepts of understandability, correctness, and validity that are crucial ingredients in existing certification standards. We apply the concept in a concrete case study for designing a highway ANN-based motion predictor to guarantee safety properties such as impossibility for the ego vehicle to suggest moving to the right lane if there exists another vehicle on its right.

Download Paper (PDF; Only available from the DATE venue WiFi)

18:15 8.5.6 IOT SECURITY ASSESSMENT THROUGH THE INTERFACES P-SCAN TEST BENCH PLATFORM

Speaker: Thomas Maurin, CEA, Leti, Univ. Grenoble Alpes, FR
Authors: Thomas Maurin¹, Laurent-Frédéric Ducreux¹, George Caramia² and Philippe Sissoko²
¹CEA Leti, Univ. Grenoble Alpes, FR; ²LCIE Bureau Veritas, FR

Abstract
The recent, massive and always-growing usage of communicating objects exchanging data over interconnected networks makes these objects vulnerable to cyber-attacks. Ranging from mainstream industrial devices to IoT products, the P-SCAN test platform is designed as a convenient solution to democratize connected objects security assessment. Associated to guidelines easing the definition of a device security target, the platform provides a library of test suites which enables automating the process of testing security features on the device’s communication interfaces. As technologies evolve, the platform is designed to be scalable and customizable (new interfaces, new standard test suites, specific test cases with respect to new Common Vulnerabilities and Exposures) to detect potential vulnerabilities. This paper explains the identified business needs and market segment, the related value proposition and gives an overview of the provided technical solution.

Download Paper (PDF; Only available from the DATE venue WiFi)

18:30 End of session

8.6 Designing reliable embedded architectures under uncertainty

Date: Wednesday, March 21, 2018
Time: 17:00 - 18:30
Location / Room: Kont. 4

Chair: Oliver Bringmann, Universität Tübingen, DE; Contact Oliver Bringmann
Co-Chair: Amit Singh, University of Essex, GB; Contact Amit Kumar Singh

Reliability is a target which can be reached in some different ways, by using, for instance, fault-tolerant architectures or by exploiting adaptable architecture. The session presents original contributions in both directions. On the first paper, reconfigurable VLIW processors are targeted by means of dynamic binary translation to explore a performance-energy trade-off. The following two papers propose solutions for fault prevention, detection and isolation, without compromising performance. In the last paper, the potential use of approximate, low-power functional units is targeted while remaining within the overall error budget of an application.
### SUPPORTING RUNTIME RECONFIGURABLE VLIW CORES THROUGH DYNAMIC BINARY TRANSLATION

**Speaker:** Simon Rokicki, Univ Rennes, INRIA, CNRS, IRISA, FR  
**Authors:** Simon Rokicki\(^1\), Erven Rohou\(^2\) and Steven Derrien\(^2\)  
\(^1\)Inria, FR; \(^2\)Inria, FR; \(^3\)University of Rennes 1/IRISA, FR  
**Abstract**  
Single ISA Heterogeneous multi-cores such as the ARM big.LITTLE have proven to be an attractive solution to explore different energy/performance trade-offs. Such architectures combine Order of cores with smaller in-order ones to offer different power/energy profiles. They however do not really exploit the characteristics of workloads (compute intensive vs control dominated). In this work, we propose to enrich these architectures with runtime configurable VLIW cores, which are very efficient at compute intensive kernels. To preserve the single ISA programming model, we resort to Dynamic Binary Translation, and use this technique to enable dynamic code specialization for runtime reconfigurable VLIW cores. Our proposed DBT framework targets the RISC-V ISA, for which both OoO and in-order implementations exist. Our experimental results show that our approach can lead to best-case performance and energy efficiency when compared against static VLIW configurations.  
**Download Paper (PDF; Only available from the DATE venue WiFi)**

### USF-ULTRA-LIGHTWEIGHT SOFTWARE FAULT ISOLATION FOR IOT-CLASS DEVICES

**Speaker:** Zelalem Aweke, University of Michigan, US  
**Authors:** Zelalem Birhanu Aweke and Todd Austin, University of Michigan, US  
**Abstract**  
Embedded device security is a particularly difficult challenge, as the quantity of devices makes them attractive to attackers, while their cost-sensitive design leads to less-than-desirable security implementations. Most current low-end embedded devices do not include any form of security or only include simple memory protection support. One line of research in crafting low-cost security for low-end embedded devices has focused on sand-boxing trusted code from untrusted code using both hardware and software techniques. These previous attempts suffer from large trusted code bases (e.g., including the entire kernel), high runtime overheads (e.g., due to code instrumentation), partial protection (e.g., only provide write protection), or heavyweight hardware modifications. In this work, we leverage the runtime memory protection support found in modern IoT-class microcontrollers to build a light-weight, low-overhead, flexible sandboxing mechanism that can provide isolation between tightly-coupled software modules. With our approach, named uSF1, only the trust management code must be trusted. Through the use of a static verifier and monitored inter-module transitions, module code at all privileges (including the kernel) is able to run uninstrumented and untrusted code. We implemented uSF1 on an ARMv7 M-based processor, both bare metal running the freeRTOS kernel, and analyzed the performance using the MiBench embedded benchmark suite and two additional highly detailed applications. We found that performance overheads were minimal, with at most 1.1% slowdown, and code size overheads were also low, at a maximum of 10%. In addition, our trusted code base is trivially small at only 150 lines of code.  
**Download Paper (PDF; Only available from the DATE venue WiFi)**

### CONVERGING SAFETY AND HIGH-PERFORMANCE DOMAINS: INTEGRATING OPENMP INTO ADA

**Speaker:** Sara Royuela, Barcelona Supercomputing Center, ES  
**Authors:** Sara Royuela\(^1\), Eduardo Queiroz\(^2\) and Luis Miguel Pinho\(^3\)  
\(^1\)Barcelona supercomputing center, ES; \(^2\)politecnico institute of porto, PT  
**Abstract**  
The use of parallel heterogeneous embedded architectures is needed to implement the level of performance required in advanced safety-critical systems. Hence, there is a demand for using high level parallel programming models capable of efficiently exploiting the performance opportunities. In this paper, we evaluate the incorporation of OpenMP, a parallel programming model used in HPC, into Ada, a language spread in safety-critical domains. We demonstrate that the execution model of OpenMP is compatible with the recently proposed Ada tasklet model, meant to exploit fine-grain structured parallelism. Moreover, we show the compatibility of the OpenMP and tasklet models, enabling the use of OpenMP directives in Ada to further exploit unstructured parallelism and heterogeneous computation. Finally, we state the safety properties of OpenMP and analyze the interoperability between the OpenMP and Ada runtimes. Overall, we conclude that OpenMP can be effectively incorporated into Ada without jeopardizing its safety properties.  
**Download Paper (PDF; Only available from the DATE venue WiFi)**

### COMPILER-DRIVEN ERROR ANALYSIS FOR DESIGNING APPROXIMATE ACCELERATORS

**Speaker:** Jorge Castro-Godínez, Chair for Embedded Systems (CES), Karlsruhe Institute of Technology (KIT), DE  
**Authors:** Jorge Castro-Godínez\(^1\), Sven Esser\(^2\), Muhammad Shafique\(^2\), Santiago Pagani\(^3\) and Joerg Henkel\(^4\)  
\(^1\)Karlsruhe Institute of Technology, DE; \(^2\)TU Wien, AT; \(^3\)Irisa, FR; \(^4\)University of Rennes 1/IRISA, FR  
**Abstract**  
Approximate Computing has emerged as a design paradigm suitable to applications with inherent error resilience. This paradigm aims to reduce the associated computing costs (such as execution time, area, or energy) of exact calculations by reducing the quality of their results. Several approximate arithmetic circuits have been proposed, which can be used to implement hardware blocks such as approximate accelerators. However, to satisfy quality constraints in these accelerators, it is imperative to assess how the errors introduced by approximate circuits propagate through other exact and approximate computations, and finally accumulate at the output. This is, in particular, crucial to enable high-level synthesis of approximate accelerators. This work proposes a compiler-driven error analysis methodology to evaluate the behavior of errors generated from approximate adders in the design of approximate accelerators. We present GEDA, a tool to perform a static analysis of the error propagation. This tool uses #pragma-based annotated C/C++ source code as input. With these annotations, exact additions are replaced by approximate ones during the code analysis to estimate the error at the output. The error estimations produced by our tool are comparable to those obtained through simulations.  
**Download Paper (PDF; Only available from the DATE venue WiFi)**

### PARALLEL CODE GENERATION OF SYNCHRONOUS PROGRAMS FOR A MANY-CORE ARCHITECTURE

**Speaker:** Simon Rokicki, Univ Rennes, INRIA, CNRS, IRISA, FR  
**Authors:** Simon Rokicki\(^1\), Matthieu Moyal\(^2\), Pascal Raymond\(^3\) and Beno\^{i}t Dupont de Dinechin\(^4\)  
\(^1\)Verimag - Univ. Grenoble Alpes, FR; \(^2\)Univ. Grenoble Alpes, FR; \(^3\)IRISA, FR; \(^4\)Verimag/CNRS, FR; \(^5\)Kalray, FR  
**Abstract**  
AmEmbedded systems tend to require more and more computational power. Many-core architectures are good candidates since they offer power and are considered more time predictable than classical multi-cores. Data-flow synchronous languages such as Lustre or Scade are widely used for avionic critical software. Programs are described by networks of computational nodes. Implementation of such programs on a many-core architecture must ensure a bounded static response time and preserve the functional behavior by taking interference into account. We consider the top-level node of a Lustre application as a software architecture description where each sub-node corresponds to a potential parallel task. Given a mapping (tasks to cores), we automatically generate code suitable for the targeted many-core architecture. This minimizes memory interferences and allows usage of a framework to compute the Worst-Case Response Time.  
**Download Paper (PDF; Only available from the DATE venue WiFi)**
Increasingly higher digital clock frequencies are feasible. Another outstanding application is N-Path Filters for "Software Defined Radios" which are being addressed in the third talk. N-path filters benefit from CMOS scaling as switch parasitics improve, minimum power consumption and smallest silicon area.

The second presentation highlights the strengths of 22FDX regarding noise and linearity demonstrating smart mixed-signal calibration techniques in order to meet the performance targets at frequency band, energy efficiency and area.

The first talk of this session will present RF integrated circuits for multi-Gbps communication and provide two examples which are best in class regarding frequency of operation, broadest 22FDX combines the best of RF performance of SiGe and PDSOI into one process technology, giving designers the opportunity to design best in class switches, LNA's and PA's onto a single chip.

Claudia Kretzschmar, GLOBALFOUNDRIES, DE,

Organiser: Claudia Kretzschmar, GLOBALFOUNDRIES, DE, Contact Claudia Kretzschmar

This session focuses on the RF and mmWave capabilities of 22FDX where the technology has a great advantage over bulk or FinFET technologies: 22FDX is the best choice for any application where lowest analog and RF/mmWave circuit power consumption is desired. It offers a high peak frequency performance (fT, fMAX), enables great integration of PA due to a high stacking efficiency and has one of the best CMOS switch behaviors due to low Ron and better Ron*COFF than FinFET or bulk technologies and a low-loss BEOL.

22FDX combines the best of RF performance of SiGe and PDSOI into one process technology, giving designers the opportunity to design best in class switches, LNA's and PA's onto a single die, integrated with transceiver and digital baseband.

The first talk of this session will present RF integrated circuits for multi-Gbps communication and provide two examples which are best in class regarding frequency of operation, broadest frequency band, energy efficiency and area.

The second presentation highlights the strengths of 22FDX regarding noise and linearly demonstrating smart mixed-signal calibration techniques in order to meet the performance targets at minimum power consumption and smallest silicon area.

Another outstanding application are N-Path Filters for "Software Defined Radios" which are being addressed in the third talk. N-path filters benefit from CMOS scaling as switch parasitics improve, and increasingly higher digital clock frequencies are feasible.

The fourth presentation concludes this session describing a mmWave circuit design example on how to utilize the 22FDX features.

Time | Label | Presentation Title |
--- | --- | --- |
18:32 | IP4-3.272 | SOCirates - A Seamless Online Compiler and System Runtime Autotuning Framework for Energy-Aware Applications |

**Abstract**

Configuring program parallelism and selecting optimal compiler options according to the underlying platform architecture is a difficult task if completely demanded to the programmer or done by using a default one-fits-all policy generated by the compiler or runtime system. Given the dynamics of the problem, a runtime selection of the best configuration is obviously the desirable solution. However, implementing this solution into the application requires the insertion of a lot of glue code for profiling and runtime selection. This represents a programming wall to actually make it feasible. This paper presents a structured approach called SOCirates, based on a Domain Specific Language (LARA) and a runtime autotuner (mARGOt), to alleviate this effort. LARA has been used to hide the glue code insertion, thus separating the pure functional application description from extra-functional requirements. mARGOt has been used for the automatic selection of the best configuration according to the runtime evolution of the application. To demonstrate the effectiveness of the proposed approach, we evaluated SOCirates by varying the application workloads, hardware resources and energy efficiency requirements for 12 OpenMP Polybench/C with respect to a standard one-fits-all solution.

Download Paper (PDF; Only available from the DATE venue WiFi)

18:33 | IP4-4.377 | Non-Intrusive Program Tracing of Non-Preemptive Multitasking Systems Using Power Consumption |

**Abstract**

System tracing, runtime monitoring, execution reconstruction are useful techniques for protecting the safety and integrity of systems. Furthermore, with time-aware or overhead-aware techniques being available, these techniques can also be used to monitor and secure production systems. As operating systems gain in popularity, even in deeply embedded systems, these techniques face the challenge to support multitasking. In this paper, we propose a novel non-intrusive technique, which efficiently reconstructs the execution trace of non-preemptive multitasking system by observing power consumption characteristics. Our technique uses the control-flow graph (CFG) of the application program to identify the most likely block of code that the system is executing at any given point in time. For the purpose of the experimental evaluation, we instrument the source code to obtain power consumption information for each basic block, which is used as the training data for our Dynamic Time Warping and k-Nearest Neighbours (k-NN) classifier. Once the system is trained, this technique is used to identify live code-block execution (LCEB). We show that the technique can reconstruct the execution flow of programs in a multi-tasking environment with high accuracy.

Download Paper (PDF; Only available from the DATE venue WiFi)

8.8 22FDX - the superior technology for IoT, RF, Automotive and Mobility: Best-in-Class RF, 5G and mmWave designs

**Time:** Wednesday, March 21, 2018  
**Date:** 17:00 - 18:30  
**Location / Room:** Exhibition Theatre

**Organiser:** Claudia Kretzschmar, GLOBALFOUNDRIES, DE; Contact Claudia Kretzschmar

This session focuses on the RF and mmWave capabilities of 22FDX where the technology has a great advantage over bulk or FinFET technologies: 22FDX is the best choice for any application where lowest analog and RF/mmWave circuit power consumption is desired. It offers a high peak frequency performance (fT, fMAX), enables great integration of PA due to a high stacking efficiency and has one of the best CMOS switch behaviors due to low Ron and better Ron*COFF than FinFET or bulk technologies and a low-loss BEOL.

22FDX combines the best of RF performance of SiGe and PDSOI into one process technology, giving designers the opportunity to design best in class switches, LNA's and PA's onto a single die, integrated with transceiver and digital baseband.

The first talk of this session will present RF integrated circuits for multi-Gbps communication and provide two examples which are best in class regarding frequency of operation, broadest frequency band, energy efficiency and area.

The second presentation highlights the strengths of 22FDX regarding noise and linearly demonstrating smart mixed-signal calibration techniques in order to meet the performance targets at minimum power consumption and smallest silicon area.

Another outstanding application are N-Path Filters for "Software Defined Radios" which are being addressed in the third talk. N-path filters benefit from CMOS scaling as switch parasitics improve, and increasingly higher digital clock frequencies are feasible.

The fourth presentation concludes this session describing a mmWave circuit design example on how to utilize the 22FDX features.

Time | Label | Presentation Title |
--- | --- | --- |
17:00 | 8.8.1 | Best-in-Class RF Integrated Circuits for Multi-Gbps Communication in 22FDX |

**Abstract**

This talk presents an overview of the ongoing circuit design activities in cooperation with GLOBALFOUNDRIES. Several RF integrated circuits for multi-Gbps communication have been demonstrated in 22FDX. Two examples will be presented in details: a Travelling Wave Amplifier (TWA) and a Mach-Zehnder Modulator (MZM) driver.

The gain cell employed for the TWAs is based on a conventional cascode topology and its layout is optimized to avoid performance degradation and instability. The TWA delivers a maximum gain of 10 dB over a 3-dB band of 110 GHz. The measured output power at 1 dB compression of the gain (P1dBcp) is 12.5 dBm at 20 GHz. Compared against the state of the art for TWAs, the presented design achieves the highest frequency of operation, as well as the broadest frequency band, and the highest P1dBcp. This circuit was awarded the best student paper award at the 2017 IEEE Asia Pacific Microwave Conference (APMC).

The high-voltage MZM driver is realized with stacked, low-impedance and low-loss switches which allow a voltage swing significantly larger than the breakdown voltage of the high-voltage MZM driver. The output power stage is configured with a high-voltage MZM, providing high gain and bandwidth. The power amplifier stage is designed to deliver a maximum output power of 40 dBm at 50 GHz. Compared against the state of the art for MZM drivers, the presented design achieves the highest frequency of operation, as well as the broadest frequency band, and the highest output power. This circuit was awarded the best student paper award at the 2017 IEEE Asia Pacific Microwave Conference (APMC).

The gain cell employed for the TWAs is based on a conventional cascode topology and its layout is optimized to avoid performance degradation and instability. The TWA delivers a maximum gain of 10 dB over a 3-dB band of 110 GHz. The measured output power at 1 dB compression of the gain (P1dBcp) is 12.5 dBm at 20 GHz. Compared against the state of the art for TWAs, the presented design achieves the highest frequency of operation, as well as the broadest frequency band, and the highest P1dBcp. This circuit was awarded the best student paper award at the 2017 IEEE Asia Pacific Microwave Conference (APMC).

The high-voltage MZM driver is realized with stacked, low-impedance and low-loss switches which allow a voltage swing significantly larger than the breakdown voltage of the high-voltage MZM driver. The output power stage is configured to deliver a maximum output power of 40 dBm at 50 GHz. Compared against the state of the art for MZM drivers, the presented design achieves the highest frequency of operation, as well as the broadest frequency band, and the highest output power. This circuit was awarded the best student paper award at the 2017 IEEE Asia Pacific Microwave Conference (APMC).

The gain cell employed for the TWAs is based on a conventional cascode topology and its layout is optimized to avoid performance degradation and instability. The TWA delivers a maximum gain of 10 dB over a 3-dB band of 110 GHz. The measured output power at 1 dB compression of the gain (P1dBcp) is 12.5 dBm at 20 GHz. Compared against the state of the art for TWAs, the presented design achieves the highest frequency of operation, as well as the broadest frequency band, and the highest P1dBcp. This circuit was awarded the best student paper award at the 2017 IEEE Asia Pacific Microwave Conference (APMC).

The high-voltage MZM driver is realized with stacked, low-impedance and low-loss switches which allow a voltage swing significantly larger than the breakdown voltage of the high-voltage MZM driver. The output power stage is configured to deliver a maximum output power of 40 dBm at 50 GHz. Compared against the state of the art for MZM drivers, the presented design achieves the highest frequency of operation, as well as the broadest frequency band, and the highest output power. This circuit was awarded the best student paper award at the 2017 IEEE Asia Pacific Microwave Conference (APMC).

The high-voltage MZM driver is realized with stacked, low-impedance and low-loss switches which allow a voltage swing significantly larger than the breakdown voltage of the high-voltage MZM driver. The output power stage is configured to deliver a maximum output power of 40 dBm at 50 GHz. Compared against the state of the art for MZM drivers, the presented design achieves the highest frequency of operation, as well as the broadest frequency band, and the highest output power. This circuit was awarded the best student paper award at the 2017 IEEE Asia Pacific Microwave Conference (APMC).
complex tasks however it requires novel hardware and software solutions to achieve the required performances. This session introduces the knowledge chain for embedded machine learning. It

Ahmed Jerraya, CEA, FR,
Chair:

Location / Room:

Time:

Date:

Price for extra ticket: EUR 70.00 per person.
for the DATE Party (which needs to be booked during the online registration process though). Additional tickets can be purchased on-site at the registration desk (subject to availability of tickets).

All delegates, exhibitors and their guests are invited to attend the party. Please be aware that entrance is only possible with a valid party ticket. Each full conference registration includes a ticket

During the evening, all delegates will have the chance to visit the different expositions for free.

The party is scheduled on March 21, 2018, from 19:00 to 23:00, and will take place in the Deutsches Hygiene-Museum Dresden. During the evening, all delegates will have the chance to visit the different expositions for free.

At delegates, exhibitors and their guests are invited to attend the party. Please be aware that entrance is only possible with a valid party ticket. Each full conference registration includes a ticket
for the DATE Party (which needs to be booked during the online registration process though). Additional tickets can be purchased on-site at the registration desk (subject to availability of tickets). Price for extra ticket: EUR 70.00 per person.

DATE-Party DATE Party | Networking Event; A bus shuttle from the congress centre to the Hygiene-Museum Dresden will be organized, starting at 19:00 from the main entrance of the ICC Dresden.

Date: Wednesday, March 21, 2018
Time: 19:30 - 23:00
Location / Room: Deutsches Hygiene-Museum Dresden

The highlight of the DATE week will again be the DATE Party, which offers the perfect occasion to meet friends and colleagues in a relaxed atmosphere while enjoying local amenities. Thus, it states one of the main networking opportunities during the DATE week.

The party is scheduled on March 21, 2018, from 1900 to 2300, and will take place in the Deutsches Hygiene-Museum Dresden.

In this presentation, we will discuss the latest ADC and DAC architectures and their respective design challenges required for 5G RF systems, optical communication and automotive systems using 22FDX. The stringent performance requirements in terms of noise and linearity demand smart mixed-signal calibration techniques in order to meet the performance targets at minimum power consumption and smallest silicon area. In addition we highlight technological benefits of 22FDX, which enable power savings on both system and circuit level.

17:20 8.8.2 SMART DATA CONVERTERS FOR WIRELINE AND WIRELESS SYSTEMS USING 22FDX
Speaker: Friedel Gerfers, Technical University Berlin, DE
Abstract
High-performance, high precision, energy-efficient data converters are indispensable in mixed-signal and RF ICs that enable next generation wireless, mobile computing, automotive, medical, and IoT applications. These applications demand the development of data conversion including mixed-signal calibration techniques that emphasize performance, accuracy, robustness, and energy efficiency with reduced silicon area / cost.

In this presentation, we will discuss the latest ADC and DAC architectures and their respective design challenges required for 5G RF systems, optical communication and automotive systems using 22FDX. The stringent performance requirements in terms of noise and linearity demand smart mixed-signal calibration techniques in order to meet the performance targets at minimum power consumption and smallest silicon area. In addition we highlight technological benefits of 22FDX, which enable power savings on both system and circuit level.

17:40 8.8.3 N-PATH FILTERS AND MIXERS CONTROLLABLE BY A DIGITAL MULTI-PHASE CLOCK
Speaker: Eric Klumperink, University of Twente, NL
Abstract
FDSOI technology offers both power-efficient and high-performance digital, amongst others because SOI-MOSFET switches have lower parasitic capacitances compared to bulk-technologies. These benefits are not only relevant for digital signal processing, but can also benefit analog radio frequency circuits. For “Software Defined Radios”, very selective Radio Frequency bandpass-filters are wanted with a flexibly programmable center frequency to choose the channel. Also, highly linear mixers for frequency down-conversion to baseband before A/D conversion are needed. It turns out that these functions can both be implemented exploiting switches, combined with linear capacitors and resistors, realising so called “N-path filters” or “Frequency Translated filters”. Moreover, the reception frequency is defined by the frequency of a digital clock, which can be implemented using digital dividers and logic. The resulting N-path filters benefit from CMOS scaling as switch parasitics improve, and increasingly higher digital clock frequencies are feasible. This contribution will review the developments in CMOS N-path filters over the last decade, highlighting promising achieved results, while also discussing some implementation aspects and simulation results in 22nm FDSOI.

18:00 8.8.4 MM-WAVE CIRCUIT DESIGN USING GLOBALFOUNDRIES 22FDX
Speaker: Janne Aikio, University of Oulu, FI
Abstract
Co-authors: Janne P. Aikio1, Mikko Hetanen2, Henri Hurskainen2, Timo Rahnknonen1, Aarno Pasanen2
1. Circuits and systems Research unit, University of Oulu; 2. Center for Wireless Communication - Radio Technologies, University of Oulu

Design complexity is increasing in all aspects when moving towards 5G wireless systems. Enhanced mobile broadband (eMBB) communications require increased bandwidth and thus also higher carrier frequencies starting from lower mmW regime. Solutions call for increased speed of transistors and better passives that are not easily available in bulk CMOS technologies. 22FDX has many potential features to support complete transceiver implementation from mmW antenna interface down to baseband processing.

22FDX provides several enhancements such as wide variety of devices: fast devices for PA and LNA circuits and slower devices for static switches, for example. Another interesting feature is the back-gating option, which we used to decrease the knee voltage of the transistors of a divider circuit. The circuit library is extensive and devices for mm-wave application contain modeling and layout up to fifth metal layer of the stack. This approach simplifies routing and reduces the extraction and simulation time significantly.

We will provide design experience from the first time access to this technology and how we used different modeling approaches precharacterized from library cells to EM simulations with Momentum, and designed test structures including active and passive devices for verification as well as complete amplifiers (LNA & PA) at 28 GHz based mostly on standard cells provided by foundry.

By the time of writing the chip fabrication is still on-going and thus we cannot guarantee if initial measurement results would be available by the time of the conference.

18:30 End of session

9.1 Special Day Session on Designing Autonomous Systems: Embedded Machine Learning
Date: Thursday, March 22, 2018
Time: 08:30 - 10:00
Location / Room: Saal 2

Chair: Ahmed Jerraya, CEA, FR, Contact Ahmed Jerraya

Autonomous systems will use machine learning techniques to deal with uncertainty and to accomplish intelligent tasks. In the case of cyberphysical systems, machine learning allow to solve complex tasks however it requires novel hardware and software solutions to achieve the required performances. This session introduces the knowledge chain for embedded machine learning. It
9.2 Emerging architectures and technologies for ultra low power and efficient embedded systems

Date: Thursday, March 22, 2018
Time: 08:30 - 10:00
Location / Room: Konf. 6

Chair: Johanna Sepulveda, Technical University of Munich, DE, Contact Johanna Sepulveda

includes talks on application needs, existing approaches and design tools.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30</td>
<td>9.1.1</td>
<td>AUTOMOTIVE APPLICATION REQUIREMENTS FOR EMBEDDED MACHINE LEARNING</td>
<td>Dirk Ziegenbein, Robert Bosch GmbH, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker and Author:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Liliana Andrade, Adrien Post-Boude and Frédéric Pétrot, Univ. Grenoble Alpes, CNRS, Grenoble INP, FR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.1.2</td>
<td>OVERVIEW OF THE STATE OF THE ART IN EMBEDDED MACHINE LEARNING</td>
<td>Frédéric Pétrot, TIMA Lab, Univ. Grenoble Alpes, FR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td></td>
</tr>
<tr>
<td>09:00</td>
<td>9.1.3</td>
<td>PNEURO: A SCALABLE ENERGY-EFFICIENT PROGRAMMABLE HARDWARE ACCELERATOR FOR NEURAL NETWORKS</td>
<td>Nicolas Ventroux, CEA LIST, FR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td></td>
</tr>
<tr>
<td>09:30</td>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td>Alexandre Carbon1, Jean-Marc PHILIPPE1, Olivier Bichler1, Renaud Schmit1, David Brand1, Benoit Tain1, Nicolas Ventroux1, Michel Paindavoine2 and Olivier Brousse2</td>
</tr>
</tbody>
</table>

Abstract

Artificial intelligence and especially Machine Learning recently gained a lot of interest from the industry. Indeed, new generation of neural networks built with a large number of successive computing layers enables a large amount of new applications and services implemented from smart sensors to data centers. These Deep Neural Networks (DNN) can interpret signals to recognize objects or situations to drive decision processes. However, their integration into embedded systems remains challenging due to their high computing needs. This paper presents PNeuro, a scalable energy-efficient hardware accelerator for the inference phase of DNN processing chains. Simple programmable processing elements architectured in SIMD clusters perform all the operations needed by DNN (convolutions, pooling, non-linear functions, etc.). An FDSOI 28nm prototype shows an energy efficiency of 700GMACS/s/W at 800 MHz. These results open important perspectives regarding the development of smart energy-efficient solutions based on Deep Neural Networks.

Download Paper (PDF; Only available from the DATE venue WiFi)
New waves of architectures and technologies are emerging with the potential of bringing high efficiency and ultra low power in future embedded systems. This session on one hand focuses on the datapath advancements in neural networks, deep learning, and mix-precision accelerators. On the other hand it presents new technologies for non-volatile memories, and bus coding techniques for volatile memories.

**Presentation Title:** OPTIMAL DC/AC DATA BUS INVERSION CODING  
**Speaker:** Jan Lucas, TU Berlin, DE  
**Authors:** Jan Lucas, Sohan Lal and Ben Juurlink, TU Berlin, DE

**Abstract**

GDDR5 and DDR4 memories use data bus inversion (DBI) coding to reduce termination power and decrease the number of output transitions. Two main strategies exist for encoding data using DBI: DBI DC minimizes the number of outputs transmitting a zero, while DBI AC minimizes the number of signal transitions. We show that neither of these strategies is optimal and reduction of interface power of up to 6% can be achieved by taking both the number of zeros and the number of signal transitions into account when encoding the data. We then demonstrate that a hardware implementation of optimal DBI coding is feasible, results in a reduction of system power and requires only an insignificant additional die area.

Download Paper (PDF; Only available from the DATE venue WiFi)
### ENERGY-PERFORMANCE DESIGN EXPLORATION OF A LOW-POWER MICROPROGRAMMED DEEP-LEARNING ACCELERATOR

**Speaker:** Andrea Calimera, Politecnico di Torino, IT  
**Authors:** Andrea Calimera¹, Mario R. Casu², Giulia Santoro¹, Valentino Peluso¹ and Massimo Aiello³  
¹Politecnico di Torino, IT; ²Politecnico di Torino, Department of Electronics and Telecommunications, IT; ³National University of Singapore, SG  
**Abstract**  
This paper presents the design space exploration of a novel microprogrammable accelerator in which PEs are connected with a Network-on-Chip and benefit from low-power features enabled through a practical implementation of a Dual-Vdd assignment scheme. An analytical model, fitted with postlayout data obtained with a 28nm FDSOI design kit, returns implementations with optimal energy-performance tradeoff by taking into consideration all the key design-space variables. The obtained Pareto analysis helps us infer optimization rules aimed at improving quality of design.

Download Paper (PDF; Only available from the DATE venue WiFi)

---

### GENPIM: GENERALIZED PROCESSING IN-MEMORY TO ACCELERATE DATA INTENSIVE APPLICATIONS

**Speaker:** Tajana Rosing, UC San Diego, US  
**Authors:** Mohsen Imani, Saransh Gupta and Tajana Rosing, University of California, San Diego, US  
**Abstract**  
Big data has become a serious problem as data volumes have been skyrocketing for the past few years. Storage and CPU technologies are overwhelmed by the amount of data they have to handle. Traditional computer architectures show poor performance which processing such huge data. Processing in memory is a promising technique to address data movement issue by locally processing data inside memory. However, there are two main issues with stand-alone PIM designs: (i) PIM is not always computationally faster than CMOS logic, (ii) PIM cannot process all operations in many applications. Thus, not many applications can benefit from PIM. To generalize the use of PIM, we designed GenPIM, a general processing in-memory architecture consisting of the conventional processor as well as the PIM accelerators. GenPIM supports basic PIM functionalities in specialized non-volatile memory including: bitwise operations, search operation, addition and multiplication. For each application, GenPIM identifies the part which uses PIM operations, and processes the rest of non-PIM operations or not data intensive part of applications in general purpose cores. GenPIM also enables configurable PIM approximation by relaxing in-memory computation. We test the efficiency of proposed design over different emerging machine learning, compression and security applications. Our experimental evaluation shows that our design can achieve 10.9x improvement in energy efficiency and 6.4x speedup as compared to processing data in conventional cores. The results can be improved by 21.0% in energy consumption and 30.6% in performance by enabling PIM approximation while ensuring less than 2% quality loss.

Download Paper (PDF; Only available from the DATE venue WiFi)

---

**Coffee Break in Exhibition Area**

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

**Lunch Breaks (Großer Saal + Saal 1)**  
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

- **Tuesday, March 20, 2018**  
  - 10:30 - 11:30 Coffee Break  
  - 13:00 - 14:30 Lunch Break  
  - 13:50 - 14:20 Awards Presentation and Keynote Lecture in “Saal 2”  
  - 16:00 - 17:00 Coffee Break

- **Wednesday, March 21, 2018**  
  - 10:00 - 11:00 Coffee Break  
  - 12:30 - 14:30 Lunch Break  
  - 13:30 - 14:20 Awards Presentation and Keynote Lecture in “Saal 2”  
  - 16:00 - 17:00 Coffee Break

- **Thursday, March 22, 2018**  
  - 10:00 - 11:00 Coffee Break  
  - 12:30 - 14:00 Lunch Break  
  - 15:30 - 16:00 Coffee Break

---

### Advances in Reconfigurable Computing

**Date:** Thursday, March 22, 2018  
**Time:** 08:30 - 10:00  
**Location / Room:** Konf. 1

**Chair:**  
Jürgen Teich, Friedrich-Alexander Universität, DE, Contact Jürgen Teich

**Co-Chair:**  
Florent de Dinechin, INSA-Lyon, FR, Contact Florent de Dinechin

This session presents four papers advancing the current state of the art in Coarse Grain Reconfigurable Architectures and two interactive presentations dealing with posit arithmetic and convolutional neural networks.
Coarse-Grained Reconfigurable Arrays (CGRAs) are popular accelerators predominantly used in streaming, filtering, and decoding applications. Due to their high performance and high power-efficiency, CGRAs can be a promising solution to accelerate the loops of general purpose applications also. However, the loops in general purpose applications are often complicated, like loops with perfect and imperfect nests and loops with nested if-then-else’s (conditionals). We argue that the existing hardware-software solutions to execute branches and conditions are inefficient. In order to efficiently execute complicated loops on CGRAs, we present a hardware-software hybrid solution: LASER – a comprehensive technique to accelerate compute-intensive loops of applications. In LASER, compiler transforms complex loops, maps them to the CGRA, and lays them out in the memory in a specific manner, such that the hardware can fetch and execute the instructions from the right path at runtime. LASER achieves a geometric performance improvement of 40.91% and utilization of 43.43% with 46% lower energy consumption.

Download Paper (PDF; Only available from the DATE venue WiFi)
10:01 | IP4_8-347 | BLOCK CONVOLUTION: TOWARDS MEMORY-EFFICIENT INFERENCE OF LARGE-SCALE CNNS ON FPGA

Speaker:
Gang Li, Institute of Automation, Chinese Academy of Sciences, CN

Authors:
Gang Li, Fanrong Li, Tianli Zhao and Jian Cheng, Institute of Automation, Chinese Academy of Sciences, CN

Abstract
FPGA-based CNN accelerators are gaining popularity due to high energy efficiency and great flexibility in recent years. However, as the networks grow in depth and width, the great volume of intermediate data is too large to store on chip, data transfers between on-chip memory and off-chip memory should be frequently executed, which leads to unexpected off-chip memory access latency and energy consumption. In this paper, we propose a block convolution approach, which is a memory-efficient, simple yet effective block-based convolution to completely avoid intermediate data from streaming out to off-chip memory during network inference. Experiments on the very large VGG-16 network show that the improved top-1/5 accuracy of 72.60%/91.10% can be achieved on the ImageNet classification task with the proposed approach. As a case study, we implement the VGG-16 network with block convolution on Xilinx Zynq ZC706 board, achieving a frame rate of 12.18fps under 150MHz working frequency, with all intermediate data staying on chip.

Download Paper (PDF; Only available from the DATE venue WiFi)

10:00 | End of session

Coffee Break in Exhibition Area

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

9.4 EU projects: Novel Platforms - from Self-Aware MPSoCs to Server Ecosystems

Date: Thursday, March 22, 2018
Time: 08:30 - 10:00
Location / Room: Konf. 2

Chair:
Martin Schoeberl, Technical University of Denmark, DK, Contact Martin Schoeberl

Co-Chair:
Flavius Gruian, Lund University, SE, Contact Flavius Gruian

This session presents three EU projects. The EU projects are: dReDBox—developing the next generation, low-power, across form-factor datacenters, enabling the creation of A338:AN521-as-a-unit, UniServer—developing a universal system architecture and software ecosystem for servers targeting cloud data centers as well as upcoming edge-computing markets and OPRECOMP—developing concepts, methods, hardware and software building blocks for practical transprecision computing systems.
DREDDBOX: MATERIALIZING A FULL-STACK RACK-SCALE SYSTEM PROTOTYPE OF A NEXT-GENERATION DISAGREGATED DATACENTER

Authors: Dimitris Syvdekis, IBM Research, Ireland, GR

Abstract: Current datacenters are based on server machines, whose mainboard and hardware components form the baseline, monolithic building block that is the rest of the system middleware and application stack are built upon. This leads to the following limitations: (i) resource proportionality of a multi-tier system is bounded by the basic building block (mainboard), (ii) resource allocation to processes or virtual machines (VM) is bounded by the available resources within the boundary of the mainboard, leading to spare resource fragmentation and inefficiencies, and (iii) updates must be applied to each and every server even when only a specific component needs to be upgraded. The dRedBox project (Disaggregated Recursive Datacentre-in-a-Box) addresses the above limitations, and proposes the next generation, low-power, across form-factor datacenters, departing from the paradigm of the mainboard-as-a-unit and enabling the creation of function-block-as-a-unit. Hardware-level disaggregation and software-defined wiring of resources is supported by a full-fledged Type-1 hypervisor that can execute commodity virtual machines, which communicate over a low-latency and high-throughput software-defined optical network. To evaluate its novel approach, dRedBox will demonstrate application execution in the domains of network functions virtualization, infrastructure analytics, and real-time video surveillance.

AN ENERGY-EFFICIENT AND ERROR-RESILIENT SERVER ECOSYSTEM EXCEEDING CONSERVATIVE SCALING LIMITS

Authors: Georgios Karakonstantis, Queen, GB

Abstract: The explosive growth of Internet-connected devices will soon result in a flood of generated data, which will increase the demand for network bandwidth as well as compute power to process the generated data. Consequently, there is a need for more energy efficient servers to empower traditional centralized Cloud datacenters as well as emerging decentralized data-centers at the Edges of the Cloud. In this paper, we present our approach, which aims at developing a new class of micro-servers - the UniServer - that exceed the conservative energy and performance scaling limits by introducing novel mechanisms at all layers of the design stack. The main idea lies on the realization of the intrinsic hardware heterogeneity and the development of mechanisms that will automatically expose the unique varying capabilities of each hardware component within commercial micro-servers and allow their operation at new extended operating points. Low-overhead schemes are employed to monitor and predict the hardware behavior and report it to the system software. The system software including a virtualization and resource management layer is responsible for optimizing the system operation in terms of energy or performance, while guaranteeing non-disruptive operation under the extended operating points. Our characterization results on a 64-bit ARM® micro-server in 28nm process reveal large voltage margins in terms of Vmin variation among the 8 cores of the CPU chip, among 3 different sigma chips, and among different benchmarks with the potential to obtain up to 38.8% energy savings. Similarly, DRAM characterizations show that refresh rate and voltage can be relaxed by 43x and 5%, respectively, leading to 23.2% power savings on average.

THE TRANSPRECISION COMPUTING PARADIGM: CONCEPT, DESIGN, AND APPLICATIONS

Authors: Dionyssios Diamantopoulos, IBM Research, Zurich, CH

Abstract: Guaranteed numerical precision of each elementary step in a complex computation has been the mainstay of traditional computing systems for many years. This era, fueled by Moore's law and the constant exponential improvement in computing efficiency, is at its twilight: from tiny nodes of the Internet-of-Things, to large HPC computing centers, sub-picoJoule operation energy efficiency is essential for practical realizations. To overcome the power wall, a shift from traditional computing paradigms is now mandatory. In this paper we present the driving motivations, roadmap, and expected impact of the European project OPRECOMP. OPRECOMP aims to (i) develop the first complete transprecision computing framework, (ii) apply it to a wide range of hardware platforms, from the sub-mWatt up to the MegaWatt range, and (iii) demonstrate impact in a wide range of computational domains, spanning IoT, Big Data Analytics, Deep Learning, and HPC simulations. By combining together into a seamless design transprecision advances in devices, circuits, software tools, and algorithms, we expect to achieve major energy efficiency improvements, even when there is no freedom to relax end-to-end application quality of results. Indeed, OPRECOMP aims at demolishing the ultra-conservative “precise” computing abstraction, replacing it with a more flexible and efficient one, namely transprecision computing.
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

9.5 Physical Attacks

Date: Thursday, March 22, 2018
Time: 08:30 - 10:00
Location / Room: Konf. 3

Chair:
Bilge Kavun Elf, Infineon Technologies, DE, Contact Bilge Kavun Elf

Co-Chair:
Batina Lejla, Radboud University, NL, Contact Lejla Batina

Electronic circuits are increasingly processing sensitive confidential data, such as personal information. In this session, new types of attacks to extract such data out of circuits are discussed in-depth. They encompass passive side-channel attacks and active manipulations of circuits.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30</td>
<td>9.5.1</td>
<td>AN INSIDE JOB: REMOTE POWER ANALYSIS ATTACKS ON FPGAS</td>
<td>Falk Schellenberg, Ruhr-Universität Bochum, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Falk Schellenberg⁠¹, Dennis Gnadt⁠², Amr Moradi² and Mehdi Tahoori²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>¹Ruhr University Bochum, DE; ²Karlsruhe Institute of Technology, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>Hardware Trojans have gained increasing interest during the past few years. Undeniably, the detection of such malicious designs needs a deep understanding of how they can practically be built and developed. In this work we present a design methodology dedicated to FPGAs which allows measuring a fraction of the dynamic power consumption. More precisely, we develop internal sensors which are based on FPGA primitives, and transfer the internally-measured side-channel leakages outside. These are distributed and calibrated delay sensors which can indirectly measure voltage fluctuations due to power consumption. By means of a cryptographic core as a case study, we present different settings and parameters for our employed sensors. Using their side-channel measurements, we further exhibit practical key-recovery attacks confirming the applicability of the underlying measurement methodology. This opens a new door to integrate hardware Trojans in a) applications where the FPGA is remotely accessible and b) FPGA-based multi-user platforms where the reconfigurable resources are shared among different users. This type of Trojan is highly difficult to detect since there is no signal connection between targeted (cryptographic) core and the internally-deployed sensors.</td>
</tr>
<tr>
<td>09:00</td>
<td>9.5.2</td>
<td>CONFIDENT LEAKAGE DETECTION - A SIDE-CHANNEL EVALUATION FRAMEWORK BASED ON CONFIDENCE INTERVALS</td>
<td>Florian Bache¹, Christina Plump¹ and Tim Güneysu²</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>¹University of Bremen, DE; ²University of Bremen &amp; DFKI, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>Cryptographic devices that potentially operate in hostile physical environments need to be secured against side-channel attacks. In order to ensure the effectiveness of the required countermeasures, scientists, developers, and evaluators need efficient methods to test the level of security of a device. In this paper we propose a new framework based on confidence intervals that extends established t-test based approaches for test-vector leakage assessment (TVLA). In comparison to previous TVLA approaches the new methodology does not only enable the detection of leakage but can also assert its absence. The framework is robust against noise in the evaluation system and thereby avoids false negatives. These improvements can be achieved without overhead in measurement complexity and with a minimum of additional computational costs compared to previous approaches. We evaluate our method under realistic conditions by applying it to a protected implementation of AES.</td>
</tr>
</tbody>
</table>

Download Paper (PDF; Only available from the DATE venue WiFi)
Abstract

Time variation during program execution can leak sensitive information. Time variations due to program control flow and hardware resource contention have been used to steal encryption keys in cipher implementations such as AES and RSA. A number of approaches to mitigate timing-based side-channel attacks have been proposed including cache partitioning, control-flow obfuscation and injecting timing noise into the outputs of code. While these techniques make timing-based side-channel attacks more difficult, they do not eliminate the risks. Prior techniques are either too specific or too expensive, and all leave remnants of the original timing side channel for later attackers to attempt to exploit. In this work, we show that the state-of-the-art techniques in timing side-channel protection, which limit timing leakage but do not eliminate it, still have significant vulnerabilities to timing-based side-channel attacks. To provide a means for total protection from timing-based side-channel attacks, we develop Ozone, the first zero timing leakage execution resource for a modern microarchitecture. Code in Ozone executes under a special hardware thread that gains exclusive access to a single core’s resources for a fixed (and limited) number of cycles during which it cannot be interrupted. Memory access under Ozone thread execution is limited to pre-allocated cache lines that can not be evicted, and all Ozone threads begin execution with a known fixed microarchitectural state. We evaluate Ozone using a number of security sensitive kernels that have previously been targets of timing side-channel attacks, and show that Ozone eliminates timing leakage with minimal performance overhead.

Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

IP4 Interactive Presentations

Date: Thursday, March 22, 2018
Time: 10:00 - 10:30
Location / Room: Conference Level, Foyer

Interactive Presentations run simultaneously during a 30-minute slot. Additionally, each IP paper is briefly introduced in a one minute presentation in a corresponding regular session.

<table>
<thead>
<tr>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP4-1</td>
<td>EFFICIENT MAPPING OF QUANTUM CIRCUITS TO THE IBM QX ARCHITECTURES</td>
<td>Alwin Zulehner, Johannes Kepler University Linz, AT; Alexandru Paler and Robert Wille, Johannes Kepler University Linz, AT</td>
</tr>
<tr>
<td>IP4-2</td>
<td>PARALLEL CODE GENERATION OF SYNCHRONOUS PROGRAMS FOR A MANY-CORE ARCHITECTURE</td>
<td>Amaury Graillat, Verimag - Univ. Grenoble Alpes, FR; Matthieu Moy, Pascal Raymond, Benoit Dupont de Dinechin</td>
</tr>
</tbody>
</table>

In March 2017, IBM launched the project IBM Q with the goal to provide access to quantum computers for a broad audience. This allowed users to conduct quantum experiments on a 5-qubit and, since June 2017, also on a 16-qubit quantum computer (called IBM QX2 and IBM QX3, respectively). In order to use these, the desired quantum functionality (e.g. provided in terms of a quantum circuit) has to properly be mapped so that the underlying physical constraints are satisfied - a complex task. This demands for solutions to automatically and efficiently conduct this mapping process. In this paper, we propose such an approach which satisfies all constraints given by the architecture and, at the same time, aims to keep the overhead in terms of additionally required quantum gates minimal. The proposed approach is generic and can easily be configured for future architectures.

Experimental evaluations show that the proposed approach clearly outperforms IBM's own mapping solution with respect to runtime as well as resulting costs.

Download Paper (PDF; Only available from the DATE venue WiFi)

AmEmbedded systems tend to require more and more computational power. Many-core architectures are good candidates since they offer power and are considered more time predictable than classical multi-cores. Data-flow synchronous languages such as Lustre or Scade are widely used for avionic critical software. Programs are described by networks of computational nodes. Implementation of such programs on a many-core architecture must ensure a bounded response time and preserve the functional behavior by taking interference into account. We consider the top-level node of a Lustre application as a software architecture description where each sub-node corresponds to a potential parallel task. Given a mapping (tasks to cores), we automatically generate code suitable for the targeted many-core architecture. This minimizes memory interferences and allows usage of a framework to compute the Worst-Case Response Time.

Download Paper (PDF; Only available from the DATE venue WiFi)
IP4-3
Socrates - A Seamless Online Compiler and System Runtime Autotuning Framework for Energy-Aware Applications
Speaker:
Gianluca Palermo, Politecnico di Milano, IT
Authors:
Davide Gaddoli1, Ricardo Nobre2, Emanuele Vitali3, Amir H. Ashouri4, Gianluca Palermo4, Cristina Silvano5 and João M. P. Cardoso5
1Politecnico di Milano, IT; 2University of Porto / INESC TEC, PT; 3Faculty of Engineering, University of Porto, PT; 4University of Toronto, Canada, CA; 5University of Porto, PT
Abstract
Configuring program parallelism and selecting optimal compiler options according to the underlying platform architecture is a difficult task if completely demanded to the programmer or done by using a default one-size-all policy generated by the compiler or runtime system. Given the dynamics of the problem, a runtime selection of the best configuration is obviously the desirable solution. However, implementing this solution into the application requires the insertion of a lot of glue code for profiling and runtime selection. This represents a programming wall to actually make it feasible. This paper presents a structured approach called Socrates, based on a Domain Specific Language (LARA) and a runtime autotuner (mARGOt), to alleviate this effort. LARA has been used to hide the glue code insertion, thus separating the pure functional application description from extra-functional requirements. mARGOt has been used for the automatic selection of the best configuration according to the runtime evolution of the application. To demonstrated the effectiveness of the proposed approach, we evaluated Socrates by varying the application workloads, hardware resources and energy efficiency requirements for 12 OpenMP Polybench/C with respect to a standard one-size-all solution.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP4-4
Non-Intrusive Program Tracing of Non-Preemptive Multitasking Systems Using Power Consumption
Speaker:
Kamal Lamichhane, University of Waterloo, CA
Authors:
Kamal Lamichhane, Carlos Moreno and Sebastian Fischmeister, University of Waterloo, CA
Abstract
System tracing, runtime monitoring, execution reconstruction are useful techniques for protecting the safety and integrity of systems. Furthermore, with time-aware or overhead-aware techniques being available, these techniques can also be used to monitor and secure production systems. As operating systems gain in popularity, even in deeply embedded systems, these techniques face the challenge to support multitasking. In this paper, we propose a novel non-intrusive technique, which effectively reconstructs the execution trace of non-preemptive multitasking system by observing power consumption characteristics. Our technique uses the control-flow graph (CFG) of the application program to identify the most likely block of code that the system is executing at any given point in time. For the purpose of the experimental evaluation, we first instrument the source code to obtain power consumption information for each basic block, which is used as the training data for our Dynamic Time Warping and k-Nearest Neighbours (k-NN) classifier. Once the system is trained, this technique is used to identify live code-block execution (LCBE). We show that the technique can reconstruct the execution flow of programs in a multi-tasking environment with high accuracy.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP4-5
Energy-Performance Design Exploration of a Low-Power Microprogrammed Deep-Learning Accelerator
Speaker:
Andrea Calimer1, Politecnico di Torino, IT
Authors:
Andrea Calimer1, Mario R. Casu2, Giulia Santoro3, Valentina Peluso1 and Massimo Alotto3
1Politecnico di Torino, IT; 2Politecnico di Torino, Department of Electronics and Telecommunications, IT; 3National University of Singapore, SG
Abstract
This paper presents the design space exploration of a novel microprogrammable accelerator in which PEs are connected with a Network-on-Chip and benefit from low-power features enabled through a practical implementation of a Dual-Vdd assignment scheme. An analytical model, fitted with postlayout data obtained with a 28nm FDSOI design kit, returns implementations with optimal energy-performance tradeoff by taking into consideration all the key design-space variables. The obtained Pareto analysis helps us infer optimization rules aimed at improving quality of design.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP4-6
GenPIM: Generalized Processing In-Memory to Accelerate Data Intensive Applications
Speaker:
Tajana Rosing, UC San Diego, US
Authors:
Mohsen Imani, Saransh Gupta and Tajana Rosing, University of California, San Diego, US
Abstract
Big data has become a serious problem as data volumes have been skyrocketing for the past few years. Storage and CPU technologies are overwhelmed by the amount of data they have to handle. Traditional computer architectures show poor performance when processing such huge data. Processing in-memory is a promising technique to address data movement issue by locally processing data inside memory. However, there are two main issues with stand-alone PIM designs: (i) PIM is not always computationally faster than CMOS logic; (ii) PIM cannot process all operations in many applications. Thus, not many applications can benefit from PIM. To generalize the use of PIM, we designed GenPIM, a general processing in-memory architecture consisting of the conventional processor as well as the PIM accelerators. GenPIM supports basic PIM functionalities in specialized non-volatile memory including bitwise operations, search operation, addition and multiplication. For each application, GenPIM identifies the part which uses PIM operations, and processes the rest of non-PIM operations or not data intensive part of applications in general purpose cores. GenPIM also enables configurable PIM approximation by relaxing in-memory computation. We test the efficiency of proposed design over different emerging machine learning, compression and security applications. Our experimental evaluation shows that our design can achieve 10.9x improvement in energy efficiency and 6.4x speedup as compared to processing data in conventional cores. The results can be improved by 21.0% in energy consumption and 30.6% in performance by enabling PIM approximation while ensuring less than 2% quality loss.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP4-7
Universal Number Posit Arithmetic Generator on FPGA
Speaker:
Hayden K. H. So, The University of Hong Kong, HK
Authors:
Manish Kumar Jaiswal and Hayden So, The University of Hong Kong, HK
Abstract
Posit number system format includes a run-time varying exponent component, defined by a combination of regime-bit (with run-time varying length) and exponent-bit (with size of up to 8 bits, the exponent size). This also leads to a run-time variation in its mantissa field size and position. This run-time variation in posit format poses a hardware design challenge. Being a recent development, posit lacks for its adequate hardware arithmetic architectures. Thus, this paper is aimed towards the posit arithmetic algorithmic development and their generic hardware generator. It is focused on basic posit arithmetic (floating-point to posit conversion, posit to floating point conversion, addition/subtraction and multiplication). These are also demonstrated on a FPGA platform. Target is to develop an open-source solution for generating basic posit arithmetic architectures with parameterized choices. This contribution would enable further exploration and evaluation of post-system.
Download Paper (PDF; Only available from the DATE venue WiFi)
### IP4-8

**BLOCK CONVOLUTION: TOWARDS MEMORY-EFFICIENT INFERENCE OF LARGE-SCALE CNNS ON FPGA**

**Speaker:**
Gang Li, Institute of Automation, Chinese Academy of Sciences, CN

**Authors:**
Gang Li, Fanrong Li, Tanli Zhao and Jian Cheng, Institute of Automation, Chinese Academy of Sciences, CN

**Abstract:**
FPGA-based CNN accelerators are gaining popularity due to high energy efficiency and great flexibility in recent years. However, as the networks grow in depth and width, the great volume of intermediate data is too large to store on chip, data transfers between on-chip memory and off-chip memory should be frequently executed, which leads to unexpected off-chip memory access latency and energy consumption. In this paper, we propose a block convolution approach, which is a memory-efficient, simple yet effective block-based convolution to completely avoid intermediate data from streaming out to off-chip memory during network inference. Experiments on the very large VGG-16 network show that the improved top-1/top-5 accuracy of 72.60%/91.10% can be achieved on the ImageNet classification task with the proposed approach. As a case study, we implement the VGG-16 network with block convolution on Xilinx Zynq ZC706 board, achieving a frame rate of 12.15fps under 150MHz working frequency, with all intermediate data staying on chip.

Download Paper (PDF; Only available from the DATE venue WiFi)

### IP4-9

**EXAMINING THE CONSEQUENCES OF HIGH-LEVEL SYNTHESIS OPTIMIZATIONS ON THE POWER SIDE CHANNEL**

**Speaker:**
Lu Zhang, Northwestern Polytechnical University, CN

**Authors:**
Lu Zhang1, Wei Hu2, Armali Ardeshrichiam2, Yu Tai1, Jeremy Blackstone2, Dejun Mu1 and Ryan Kastner2
1Northwestern Polytechnical University, Ch, 2University of California, San Diego, US

**Abstract:**
High-level synthesis (HLS) allows hardware designers to think algorithmically and not have to worry about low-level, cycle-by-cycle details. This provides the ability to quickly explore the architectural design space and tradeoff between resource utilization and performance. Unfortunately, evaluating the security is not a standard part of the HLS design flow. In this work, we aim to understand the effects of HLS optimizations with respect to power side-channel leakage. We use Vivado HLS to develop different cryptographic cores, implement them on a Xilinx Spartan 6 FPGA, and collect power traces. We evaluate the designs with respect to resource utilization, performance, and side-channel leakage through power consumption. Furthermore, we analyze the first-order leakage of the HLS-based designs alongside well-known register transfer level (RTL) cryptographic cores. We describe an evaluation procedure for hardware designers and use it to make insightful recommendations on how to design the best architecture in cryptographic domain.

Download Paper (PDF; Only available from the DATE venue WiFi)

### IP4-10

**DFARP: DIFFERENTIAL FAULT ATTACK RESISTANT PHYSICAL DESIGN AUTOMATION**

**Speaker:**
Debdeep Mukhopadhyay, Indian Institute of Technology Kharagpur, IN

**Authors:**
Mustafa Khairallah1, Rajat Sadhukhan2, Radhamanjari Samanta3, Jakub Breier4, Shivam Bhasin2, Rajat Subhra Chakrabarty2, Anupam Chattopadhyay2 and Debdeep Mukhopadhyay
1Indian Institute of Technology Kharagpur, IN; 2Temasek Laboratories, Nanyang Technological University, SG; 3Binghamton University, US; 4Indian Institute of Technology Kharagpur, IN

**Abstract:**
Differential Fault Analysis (DFA), aided by sophisticated mathematical analysis techniques for ciphers and precise fault injection methodologies, has become a potent threat to cryptographic implementations. In this paper, we propose, to the best of the our knowledge, the first “DFA-aware” physical design automation methodology, that effectively mitigates the threat posed by DFA. We first develop a novel heuristics technique, which resists the simultaneous corruption of ciphers states necessary for successful fault attack, by exploiting the fact that most fault injections are localized in practice. Our technique results in the computational complexity of the fault attack to shoot up to exhaustive search levels, making them practically infeasible. In the second part of the work, we develop a routing mechanism, which tackles more precise and costly fault injection techniques, like laser and electromagnetic guns. We propose a routing technique by integrating a specially designed ring oscillator based sensor circuit around the potential fault attack targets without incurring any performance overhead. We demonstrate the effectiveness of our technique by applying it on state of the art ciphers.

Download Paper (PDF; Only available from the DATE venue WiFi)

### IP4-11

**AN ENERGY-EFFICIENT STOCHASTIC COMPUTATIONAL DEEP BELIEF NETWORK**

**Speaker:**
Yidong Liu, University of Alberta, CA

**Authors:**
Yidong Liu1, Yanzhi Wang2, Fabrizio Lombardi2 and Jie Han1
1University of Alberta, CA; 2Syracuse university, US; 3Northeastern University, US

**Abstract:**
Deep neural networks (DNNs) are effective machine learning models to solve a large class of recognition problems, including the classification of nonlinearly separable patterns. The applications of DNNs are, however, limited by the large size and high energy consumption of the networks. Recently, stochastic computation (SC) has been considered to implement DNNs to reduce the hardware cost. However, it requires a large number of random number generators (RNGs) that lower the energy efficiency of the network. To overcome these limitations, we propose the design of an energy-efficient deep belief network (DBN) based on stochastic computation. An approximate SC activation unit (A-SCAU) is designed to implement different types of activation functions in the neurons. The A-SCAU is immune to signal correlations, so the RNGs can be shared among all neurons in the same layer with no accuracy loss. The area and energy of the proposed design are 5.27% and 3.31% (or 26.55% and 29.89%) of a 32-bit floating-point (or an 8-bit fixed-point) implementation. It is shown that the proposed SC-DBN design achieves a higher classification accuracy compared to the fixed-point implementation. The accuracy is only lower by 0.12% than the floating-point design at a similar computation speed, but with a significantly lower energy consumption.

Download Paper (PDF; Only available from the DATE venue WiFi)

### IP4-12

**PUSHING THE NUMBER OF QUBITS BELOW THE "MINIMUM": REALIZING COMPACT BOOLEAN COMPONENTS FOR QUANTUM LOGIC**

**Speaker:**
Alwin Zulehner, Johannes Kepler University Linz, AT

**Authors:**
Alwin Zulehner and Robert Wille, Johannes Kepler University Linz, AT

**Abstract:**
Research on quantum computers has gained attention since they are able to solve certain tasks significantly faster than classical machines (in some cases, exponential speed-ups are possible). Since quantum computations typically contain large Boolean components, design automation techniques are required to realize the respective Boolean functions in quantum logic. They usually introduce a significant amount of additional qubits - a highly limited resource. In this work, we propose an alternative method for the realization of Boolean components for quantum logic. In contrast to the current state-of-the-art, we dedicatedly address the main reasons causing the additionally required qubits (namely the number of the most frequently occurring output pattern as well as the number of primary outputs of the function to be realized) and propose to manipulate the function so that both issues are addressed. The resulting methods allow to push the number of required qubits below what is currently considered the minimum.

Download Paper (PDF; Only available from the DATE venue WiFi)
Abstract
Convolutional neural networks (CNNs) have been proposed to be widely adopted to make predictions on a large amount of data in modern embedded systems. Prior studies have shown that convolutional computations which consist of numbers of multiply and accumulate (MAC) operations, serve as the most computationally expensive portion in CNNs. Compared to the manner of executing MAC operations in GPU and FPGA, CNN implementation in the RRAM crossbar-based computing system (RCS) demonstrates the outstanding advantages of high performance and low power. However, the current design is energy-unbalanced among the three parts of RRAM crossbar computation, peripheral circuits and memory accesses, the latter two factors can significantly limit the potential gains of RCS. Addressing the problem of high power overhead of peripheral circuits in RCS, this paper adopts the Peripheral Circuit Unit (PeriCU) Reuse scheme to meet a certain power budget. The underlying idea is to put the expensive AD/DA outputs onto spotlight and arrange multiple convolution layers to be sequentially served by the same PeriCU. Furthermore, it is observed that memory accesses can be bypassed if two adjacent layers are assigned in the different PeriCUs. Then a loop tiling technique is proposed to further improve the energy and throughput of RCS. The experiments of two convolutional applications validate that the PeriCU-Reuse scheme integrated with the loop tiling techniques can efficiently meet power requirement, and further reduce energy consumption by 61.7%.

Download Paper (PDF; Only available from the DATE venue WiFi)
UB09.1 CCF: A CGRA COMPILATION FRAMEWORK
Authors: Shail Dave and Avarul Shrivastava, Arizona State University, US
Abstract
Coarse-grained reconfigurable array (CGRA) can efficiently accelerate even non-parallel loops. Although scores of techniques have been developed in the past decade to map loops on CGRA PEs, several challenges in enabling acceleration of general-purpose applications on CGRAs remained unresolved. In particular, the automatic code generation for the CGRA accelerator coupled with modern processor cores. In this demonstration, we showcase CCF - CGRA compiler framework. CCF is implemented in LLVM 4.0 and includes a set of transformation and analysis passes. We show that given performance-critical loops annotated in embedded applications, how CCF extracts the loop, constructs the data dependency graph (DDG), maps it onto CGRA architecture, off-loads necessary configuration instructions for CGRA PEs, and automatically communicates data between the CPU and CGRA.

More information ...

UB09.2 TOPOLINANO & MAGCAD: A DESIGN AND SIMULATION FRAMEWORK FOR THE EXPLORATION OF EMERGING TECHNOLOGIES
Authors: Umberto Garlando and Fabrizio Riente, Politecnico di Torino, IT
Abstract
We developed a design framework that enables the exploration and analysis of emerging beyond-CMOS technologies. It is composed of two powerful tools: ToPoLinano and MagCAD. Different technologies are supported, and new ones could be added thanks to their modular structure. ToPoLinano starts from a VHDL description of a circuit and performs the place-and-route following the technological constraints. The resulting circuit can be simulated both at logical or physical level. MagCAD is a layout editor where the user can design custom circuits, by placing basic elements of the selected technology. The tool can extract a VHDL netlist based on compact models of placed elements derived from experiments or physical measurements. The circuits can be verified with standard VHSIC simulators. The design workflow will be demonstrated at the U-booth to show how those tools could be a valuable help in the studying and development of emerging technologies and to obtain feedbacks from the scientific community.

More information ...

UB09.3 CONSTRAINED RANDOM APPLICATION GENERATION FOR FIRMWARE-BASED POWER MANAGEMENT VALIDATION
Authors: Vladimir Herdt1, Hoang M. Le1, Daniel Große2 and Rolf Drechsler2
1University of Bremen, DE; 2University of Bremen, DFKI GmbH, DE
Abstract
Efficient power management (PM) is very important for modern SoCs. To handle the ever rising complexity of embedded system design, power-aware virtual prototypes (VPs) are employed to enable an early power analysis. Most modern SoC implement the PM strategy in firmware (FW) due to ease of development. Validation of these strategies at VP level is crucial as undetected flaws will propagate. However, existing validation approaches are based on engineered software (SW), which might miss rare corner cases. We propose a demonstrator based on a novel approach to assess the power-versus-performance trade-off of FW-based PM. Instead of executing real SW applications, our approach makes use of workload scenarios described by a set of constraints to automatically generate SW with a specific power consumption profile. The main novelty is the modeling of scenarios based on constrained random techniques that are very successful in the area of SOC/HW functional validation.

More information ...

UB09.4 POWER-AWARE SOFTWARE MAPPING OF PARALLEL APPLICATIONS ONTO HETEROGENEOUS MIPSOCs
Authors: Gereon Onnebrink and Rainer Leupers, RWTH Aachen University, DE
Abstract
Heterogeneous multi- and many processor systems-on-chip provide the best trade-off between performance, cost, and power. One of the biggest hurdles to exploit multicore architectures from the SW side, Considering an application that has been properly partitioned into multiple concurrent tasks, and programmed in a parallel language, the process of mapping those tasks onto the processors with optimal DVFS is a huge challenge for a certain design goal. An automatic approach is needed that determines the optimal decision. A great amount of research has been conducted aiming to optimise the performance of a parallelised application. Another research track is the ESL power estimation methodology. Combining both, a novel power-aware software mapping heuristic has been implemented to develop performance and power co-optimized parallel software. This algorithm can be used to identify the gain of sophisticated power management techniques by providing the power-performance trade-off.

More information ...

UB09.5 VIRTUAL PROTOTYPE MAKANI: ANALYZING THE USAGE OF POWER MANAGEMENT TECHNIQUES AND EXTRA-FUNCTIONAL PROPERTIES BY USING VIRTUAL PROTOTYPING
Author: Silen Scheiner, OFFIS – Institute for Information Technology, DE
Abstract
My PhD work consists of analyzing the correct usage of power management techniques, as well as the analysis of extra-functional properties, including power and timing properties, in MPSoC's. Especially in safety-critical environments the power management gets safely critical too, since it is able to influence the overall system behavior. To demonstrate my methodologies a mixed-critical multi-processor and its corresponding virtual prototype is used. The multi-processor system avionics is served by a Xilinx Zynq 7000 MPSoC. The hardware architecture includes ARM and MicroBlaze cores, a NoC for communication and peripherals. The MPSoC processes the flight algorithms with triple modular redundancy and a mission-critical video processing task. The virtual prototype consists of a virtual platform and an environmental model. The virtual platform is equipped with my measuring tool libraries to generate traces of the observed power management techniques and extra-functional properties.

More information ...

UB09.6 WARE: WEARABLE ELECTRONICS DIRECTIONAL AUGMENTED REALITY
Authors: Gabriele Morandi1, Walter Vendraminetti2, Federico Fraccaroli2, Davide Quaglia1 and Gianluca Benedettith
1University of Verona, IT; 2REDalab Srl, IT; 3Wagoo LLC, IT; 4Wagoo Italia srls, IT
Abstract
Augmented Reality (AR) currently requires large form factors, weight, cost and frequent recharging cycles that reduce usability. Connectivity, image processing, localization, and direction evaluation lead to high processing and power requirements. A multi-antenna system, patented by the industrial partner, enables a new generation of smart eye-wear that elegantly requires less hardware, connectivity, and power to provide AR functionalities. They will allow users to directionally locate nearby radio emitting sources that highlight objects of interest (e.g., people or retail items) by using existing standards like Bluetooth Low Energy, Apple’s iBeacon and Google’s Eddystone. This booth will report the current level of research addressed by the Computer Science Department of University of Verona, Wagoo LLC, and Wagoo Italia srls. In the presented demo, different objects emit an “I am here” signal and a prototype of the smart glasses shows the information related to the observed object.

More information ...

UB09.7 OTPG: SPECIFICATION-BASED CONSTRUCTION OF ONLINE TPGS FOR MICROPROCESSORS
Authors: Mikhail Chupilko, Alexander Kamkin and Andrei Tatarnikov, ISP RAS, RU
Abstract
This work presents an approach to construction of online test program generators (TPGs). The approach is intended to use specifications of ISA presented in nML/mmuSL specification languages. They are processed by a meta-generator to obtain their binary representations supplied with meta information and a test generation core compatible with the target microprocessor. The test generation core is loaded as a binary image into the target microprocessor’s memory (for experiments we’re using QEMU for MIPS) and produces test cases to be processed (incl. results checking) by an executor. It should be noticed that the meta-generator and the executor are not obligatory run at the same microprocessor (especially, if it is highly incomplete). The final goal of the project is to propose a method of obtaining online TPGs for a wide range of ISAs, and to develop a mature tool implementing this method.

More information ...
IIP GENERATORS TO EASE ANALOG IC DESIGN

Authors:
Benjamin Prautsch, Uwe Eichler and Torsten Reich, Fraunhofer Institute for Integrated Circuits IIS/EAS, DE

Abstract
Semiconductor technology has shown significant progress over the last decades. Digital EDA (electronic design automation) allowed that this progress could be converted to high-performance digital ICs. Analog components are part of Systems-on-Chip (SoC) too, but analog EDA lags far behind. Therefore, a lot of effort was spent to automate analog IC design. Mayor results are constraint-based layout-aware optimization tools using predefined layout templates or pure automation as well as analog generators containing expert knowledge. While optimization is a holistic top-down approach, generators allow parameterized and fast bottom-up generation of critical schematic and layout parts, pre-planned by experienced designers. With IIP Generators, we follow three use cases to ease analog design: 1) design on higher hierarchy levels, 2) development of hierarchical high-level IIPs, and 3) automated design porting due to highly technology-independent blocks down to 22nm.

More information ...

EXPERIENCE-BASED AUTOMATION OF ANALOG IC DESIGN

Authors:
Florian Leber and Juergen Scheible, Reutlingen University, DE

Abstract
ABSYNTH was first presented in CEBIT 2014 where complete, practical circuit sizing approaches have been shown using meta-heuristics on trusted simulators. This tool was then proven by its use in design of several cells in a research project. Here, we present the extension to our nested optimization approach that creates a symmetric and well matched layout in every step for every instance in the population of the swarm, that is extracted in our flow to provide feedback to the cost function impacting on the population update for more viable and robust circuits. The layout optimization presented in this DEMO works with Cadence Layout design tools. Our initial focus is, motivated by Industry 4.0, IoT, on cells for signal conditioning electronics with reconfigurability and Self-X features.[1] Abhaya C. Kammara, L. Palanichamy, and A. König, "Multi-Objective optimization and visualization for analog automation", Complex. Intel. Syst, Springer, DOI 10.1007/s40747-016-0027-3, 2016

More information ...

IIP GENERATORS TO EASE ANALOG IC DESIGN

Authors:
Abhaya Chandra Kammara S.1, Sidney Pontes-Filho2 and Andreas König2
1ISE, TU Kaiserslautern, DE; 2University of Kaiserslautern, DE

Abstract
ABS SYNTH is a comprehensive approach to front to back analog block design automation. In order to make expert knowledge reusable, we have developed IIP Generators that can be used for a variety of applications. The presented IIP Generators, which are freely available, are able to generate high-level IIPs which can be customized to meet the specific requirements of a specific application. The IIP Generators are based on a modular design approach and can be used for a wide range of applications. The presented IIP Generators are able to handle a wide range of analog circuits, from simple amplifiers to complex analog filters.

More information ...

10.1 Special Day Session on Designing Autonomous Systems: Digitalization in automotive and industrial systems

Date: Thursday, March 22, 2018
Time: 11:00 - 12:30
Location / Room: Saal 2

Chair:
Matthias Traub, BMW, DE, Contact Matthias Traub

Autonomous systems are an important part of todays and future solutions for the automotive and industrial sector. The research and development activities to enable highly/fully automated driving and industry 4.0 have to deal with a lot of new requirements (e.g. fail operational, cyber security), technologies (connectivity over 5G, neuronal networks, future computing platforms) and topics (data analytics, artificial intelligence). Furthermore processes, methods und tools lag behind and need to speed up to cope with all the consequences in validation and verification. The short paper will give an overview over these challenges and the actual state of research and the development in the field of digital autonomous systems.
10.2 Neural Networks and Neurotechnology

Date: Thursday, March 22, 2018
Time: 11:00 - 12:30
Location / Room: Konf. 6
Chair:

---

Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00
New approaches to energy efficiency in neural networks using approximate computing and non-volatile FeFET memory are presented. A novel inductive coupling interconnect approach for
neurotechnology applications and an optimization strategy for efficiently mapping spiking neural networks to neuromorphic hardware are also presented.

DESIGN AND OPTIMIZATION OF FEFET-BASED CROSSBARS FOR BINARY CONVOLUTION NEURAL NETWORKS
Speaker: Xiaoming Chen, Institute of Computing Technology, Chinese Academy of Sciences, CN
Authors: Xiaoming Chen, Xunzhao Yin, Michael Niemier and Xiaobao Sharon Hu, University of Notre Dame, US
Abstract
Binary convolution neural networks (CNNs) have attracted much attention for embedded applications due to low hardware cost and acceptable accuracy. Nonvolatile, resistive random-access memories (RRAMs) have been adopted to build crossbar accelerators for binary CNNs. However, RRAMs still face fundamental challenges such as sneak paths, high write energy, etc. We exploit another emerging nonvolatile device—ferroelectric field-effect transistor (FeFET), to build crossbars to improve the energy efficiency for binary CNNs. Due to the three-terminal transistor structure, an FeFET can function as both a nonvolatile storage element and a controllable switch, such that both write and read power can be reduced. Simulation results demonstrate that compared with two RRAM-based crossbar structures, our FeFET-based design improves write power by 560X and 395X, and read power by 4.1X and 3.1X. We also tackle an important challenge in crossbar-based CNN accelerators: when a crossbar array is not large enough to hold the weights of one convolution layer, how do we partition the workload and map computations to the crossbar array? We introduce a hardware-software co-optimization solution for this problem that is universal for any crossbar accelerators.

Download Paper (PDF; Only available from the DATE venue WiFi)

LOW-POWER 3D INTEGRATION USING INDUCTIVE COUPLING LINKS FOR NEUROTECHNOLOGY APPLICATIONS
Speaker: Benjamin Fletcher, University of Southampton, GB
Authors: Benjamin Fletcher1, Shidhartha Das2, Chi-Sang Poon3 and Terrence Mak1
1University of Southampton, GB; 2ARM Ltd., GB; 3Massachusetts Institute of Technology, US
Abstract
Three dimensional system integration offers the ability to stack multiple dies, fabricated in disparate technologies, within a single IC. For this reason, it is gaining popularity for use in sensor devices which perform concurrent analogue and digital processing, as both analogue and digital dies can be coupled together. One such class of devices are closed-loop neuromodulators; neurostimulators which perform real-time digital signal processing (DSP) to deliver bespoke treatment. Due to their implantable nature, these devices are inherently governed by very strict volume constraints, power budgets, and must operate with high reliability. To address these challenges, this paper presents a low-power inductive coupling link (ICL) transceiver for 3D integration of digital CMOS and analogue BiCMOS dies for use in closed-loop neuromodulators. The use of an ICL, as opposed to through silicon vias (TSVs), ensures high reliability and fabrication yield in addition to circumventing the use of voltage level conversion between disparate dies, improving power efficiency. The proposed transceiver is experimentally evaluated using SPICE as well as nine traditional TSV baseline solutions. Results demonstrate that, whilst the achievable bandwidth of the TSV-based approaches is much higher, for the typical data rates demanded by neuromodulator applications (0.5 - 1 Gbps) the ICL design consumes on average 36.7% less power through avoiding the use of voltage level shifters.

Download Paper (PDF; Only available from the DATE venue WiFi)

MAPPING OF LOCAL AND GLOBAL SYNAPSES ON SPIKING NEUROMORPHIC HARDWARE
Speaker: Fransky Catthoor, IMEC Fellow, BE
Authors: Anup Das1, Yuefeng Wu1, Khanh Huynh1, Francesco Dell Anna2, Fransky Catthoor2 and Siebren Schaafsma1
1IMEC, NL; 2IMEC, BE
Abstract
Spiking neural networks (SNNs) are widely deployed to solve complex pattern recognition, function approximation and image classification tasks. With the growing size and complexity of these networks, hardware implementation becomes challenging because scaling up the size of a single array (crossbar) of fully connected neurons is no longer feasible due to strict energy budget. Modern neuromorphic hardware integrates small-sized crossbars with time-multiplexed interconnects. Partitioning SNNs becomes essential in order to map them on neuromorphic hardware with the major aim to reduce the global communication latency and energy overhead. To achieve this goal, we propose our instantiation of particle swarm optimization, which paritions SNNs into local synapses (mapped on crossbars) and global synapses (mapped on time-multiplexed interconnects), with the objective of reducing spike communication on the interconnect. This improves latency, power consumption as well as application performance by reducing inter-spike interval distortion and spike disorders. Our framework is implemented in Python, interfacing CARLsim, a GPU-accelerated application-level spiking neural network simulator with an extended version of Noxim, for simulating time-multiplexed interconnects. Experiments are conducted with realistic and synthetic SNN-based applications with different computation models, topologies and spike coding schemes. Using power numbers from in-house neuromorphic chips, we demonstrate significant reductions in energy consumption and spike latency over PACMAN, the widely-used partitioning technique for SNNs on SpiNNaker.

Download Paper (PDF; Only available from the DATE venue WiFi)

ENERGY-EFFICIENT NEURAL NETWORKS USING APPROXIMATE COMPUTATION REUSE
Speaker: Xun Jiao, University of California San Diego, US
Authors: Xun Jiao1, Vahidken Akhlaghi1, Yu Jiang2 and Rajesh Gupta1
1University of California, San Diego, US; 2Tsinghua University, CN
Abstract
As a problem-solving method, neural networks have shown broad success for medical applications, speech recognition, and natural language processing. Current hardware implementations of neural networks exhibit high energy consumption due to the intensive computing workloads. This paper proposes a methodology to design an energy-efficient neural network that effectively exploits computation reuse opportunities. To do so, we use Bloom filters (BFs) by tightly integrating them with computation units. BFs store and recall frequently occurring input patterns to reuse computations. We expand the opportunities for computation reuse by storing frequent input patterns specific to a given layer and using approximate pattern matching with hashing for limited data precision. This reconfigurable matching is key to achieving a “controllable approximation” for neural networks. To lower the energy consumption of BFs, we also use low-power memristor arrays to implement BFs. Our experimental results show that for convolutional neural networks, the BFs enable 47.5% energy saving of multiplication operations while incurring only 1% accuracy drop. While the actual savings will vary depending upon the extent of approximation and reuse, this paper presents a method for reducing computing workloads and improving energy efficiency.

Download Paper (PDF; Only available from the DATE venue WiFi)
AN ENERGY-EFFICIENT STOCHASTIC COMPUTATIONAL DEEP BELIEF NETWORK

Speaker:
Yidong Liu, University of Alberta, CA

Authors:
Yidong Liu¹, Yanzhi Wang², Fabrizio Lombardi³ and Jie Han¹
¹University of Alberta, CA; ²Syracuse university, US; ³Northeastern University, US

Abstract
Deep neural networks (DNNs) are effective machine learning models to solve a large class of recognition problems, including the classification of nonlinearly separable patterns. The applications of DNNs are, however, limited by the large size and high energy consumption of the networks. Recently, stochastic computation (SC) has been considered to implement DNNs to reduce the hardware cost. However, it requires a large number of random number generators (RNGs) that lower the energy efficiency of the network. To overcome these limitations, we propose the design of an energy-efficient deep belief network (DBN) based on stochastic computation. An approximate SC activation unit (A-SCAU) is designed to implement different types of activation functions in the neurons. The A-SCAU is immune to signal correlations, so the RNGs can be shared among all neurons in the same layer with no accuracy loss. The area and energy of the proposed design are 5.27% and 3.31% (or 26.55% and 29.89%) of a 32-bit floating-point (or an 8-bit fixed-point) implementation. It is shown that the proposed SC-DBN design achieves a higher classification accuracy compared to the fixed-point implementation. The accuracy is only lower by 0.12% than the floating-point design at a similar computation speed, but with a significantly lower energy consumption.

Download Paper (PDF; Only available from the DATE venue WiFi)

PUSHING THE NUMBER OF QUBITS BELOW THE "MINIMUM": REALIZING COMPACT BOOLEAN COMPONENTS FOR QUANTUM LOGIC

Speaker:
Alwin Zulehner, Johannes Kepler University Linz, AT

Authors:
Alwin Zulehner and Robert Wille, Johannes Kepler University Linz, AT

Abstract
Research on quantum computers has gained attention since they are able to solve certain tasks significantly faster than classical machines (in some cases, exponential speed-ups are possible). Since quantum computations typically contain large Boolean components, design automation techniques are required to realize the respective Boolean functions in quantum logic. They usually introduce a significant amount of additional qubits - a highly limited resource. In this work, we propose an alternative method for the realization of Boolean components for quantum logic. In contrast to the current state-of-the-art, we dedicatedly address the main reasons causing the additionally required qubits (namely the number of the most frequently occurring output pattern as well as the number of primary outputs of the function to be realized) and propose to manipulate the function so that both issues are addressed. The resulting methods allow to push the number of required qubits below what is currently considered the minimum.

Download Paper (PDF; Only available from the DATE venue WiFi)

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

10.3 From Non-Volatile Flip-Flops to Storage Systems

Date: Thursday, March 22, 2018
Time: 11:00 - 12:30
Location / Room: Konf. 1

Chair:
Alexandre Levisse, EPFL, CH, Contact Alexandre Levisse

Co-Chair:
Weisheng Zhao, Beihang University, CN, Contact Weisheng Zhao

This session combines research from circuit to system level on non-volatile memories. The first paper proposes an STT-MRAM-based multi-bit non-volatile flip-flop. The other papers address system-level challenges, such as write disturbance mitigation, wear levelling scheme and latency reduction, for various technologies (PCM, SSD-flash).
**MULTI-BIT NON-VOLATILE SPINTRONIC FLIP-FLOP**

Speaker: Fei Wu, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, CN

Authors: Xi'an Jiaotong University, CN; University of Pittsburgh, US

Abstract: We present a multi-bit non-volatile flip-flop architecture using STT devices to reduce the area and energy consumption associated with non-volatile components. Our architecture is developed based on the resource sharing principle using a custom designing technique that enables the optimization for the area and energy consumption. We have developed a framework in which we have replaced the conventional neighbor flip-flop layouts in the layout with our proposed multi-bit non-volatile designs. Results show that, at system-level, using our proposed multi-bit flip-flop architecture, we significantly improve the area and energy compared to the standard single bit non-volatile flip-flop designs.

Download Paper (PDF; Only available from the DATE venue WiFi)

**ADAM: ARCHITECTURE FOR WRITE DISTURBANCE MITIGATION IN SCALED PHASE CHANGE MEMORY**

Speaker: Shihvim Swami, University of Pittsburgh, US

Authors: Shihvim Swami and Kartik Mohanram, University of Pittsburgh, US

Abstract: With technology scaling, phase change memory (PCM) has become highly vulnerable to write disturbance (WD) errors. A PCM WD error occurs when a cell write dissipates heat to idle cells in the same/adjacent word lines (WLs), disturbing the states of those cells. Whereas state-of-the-art solutions, e.g., data insolation (DIN) and super dense PCM (SD-PCM), have successfully addressed WL PCM WD errors, reducing (i) bit line (BL) WD error recovery and (ii) performance penalties of aggregate (WL+BL) WD error recovery remain areas of active research and development. Architecture for Write Disturbance Mitigation, ADAM, is a low cost, high performance pattern-based data compression and alignment solution to reduce the aggregate (WL+BL) WD error rate in PCM. At no impact to inter-cell spacing, ADAM increases the lateral separation between the cells storing useful data in adjacent WLs, ensuring that the heat dissipated to adjacent WLs minimally impacts the cells storing useful data. For one compression tag bit per 512-bit cache line, ADAM provides an effective solution to reduce the number of WL and BL cells vulnerable to WD errors. ADAM also integrates a novel Deferred WD Correction scheme, DEFT, that opportunistically defers latency-intensive WD error recovery of cached data in the adjacent WLs without impacting memory reliability. ADAM is evaluated on single-/multi-level cell (SLC/MLC) PCM using the SPEC CPU2006 benchmarks. Results for SLC (MLC) PCM show that in comparison to state-of-the-art SD-PCM, ADAM reduces the aggregate WD error rate by 32% (60%); this translates to a 50% (81%) reduction in error correction energy and a 7% (15%) improvement in system performance.

Download Paper (PDF; Only available from the DATE venue WiFi)

**PROGRAM ERROR RATE-BASED WEAR LEVELING FOR NAND FLASH MEMORY**

Speaker: Fei Wu, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, CN

Authors: Xi'an Jiaotong University, CN; University of Pittsburgh, US

Abstract: Wear leveling schemes have become a fundamental issue in the design of Solid State Disk (SSD) based on NAND Flash memory. Existing schemes aim to equalize the number of programming/erase (P/E) cycles and memory raw bit error rates (BER) among all the flash blocks. However, due to fabrication process variation, different blocks of the same flash chip usually have largely different endurance in terms of BER and program error rate (PER). Such conventional design cannot obtain the wear status of flash blocks precisely. This paper proposes PER-WL, an efficient PER-based wear leveling scheme that uses PER statistics as the measurement of flash block wear-out pace, and performs block data swapping to improve the wear-leveling efficiency. In our evaluation with four realistic workloads, PER-based wear leveling scheme can achieve 17% and 9% variance of program error rate reduction, 8% and 3% program error rate reduction with 5% and 2% system performance degradation when compared to two state-of-the-art wear leveling schemes on average.

Download Paper (PDF; Only available from the DATE venue WiFi)

**SHADOWGC: COOPERATIVE GARBAGE COLLECTION WITH MULTI-LEVEL BUFFER FOR PERFORMANCE IMPROVEMENT IN NAND FLASH-BASED SSDS**

Speaker: Jinhua Cui, Xi'an Jiaotong University, CN

Authors: Jinhua Cui1, Youtao Zhang2, Jianhang Huang3, Weiguo Wu4 and Jun Yang5

1Xi'an Jiaotong University, CN; 2University of Pittsburgh, US

Abstract: Garbage collection, an essential background activity in NAND flash based Solid-State Drives, often introduces large runtime overhead. Recent studies showed that it is beneficial to separate the flash pages that have dirty copies in the write buffers from those that do not. However, the existing schemes exploring this observation have limitations, which prevent them from maximizing the performance improvement. In this paper, we address the above challenge through ShadowGC, a novel GC design that exploits the pages in both host-side and device-side write buffers and adopts different read and write strategies to minimize the GC overhead. When garbage collecting flash pages that have dirty copies in the device-side write buffer, ShadowGC reads data from the write buffer such that the page relocation operations in GC merge with the write-back operations of the buffer. When garbage collecting flash pages that are written to the device-side write buffer, ShadowGC moves them to dedicated blocks and speeds up the movement with fast-write operations. Our experimental results show that, on average, ShadowGC reduces the write amplification by 16.2% and the GC latency by 20.5% over the state-of-the-art.

Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Coffee Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

Download Paper (PDF; Only available from the DATE venue WiFi)
This session presents novel ideas realized in hardware for cryptographic systems. The contributions range from implementations of leakage resilient cryptography in ASICs, to FPGA realizations of novel public-key primitives as well as optimization of FPGA resources used by random number generation schemes.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:00</td>
<td>10.4.1</td>
<td>BINARY RING-LWE HARDWARE WITH POWER SIDE-CHANNEL COUNTERMEASURES</td>
<td>Ye Wang, The University of Texas at Austin, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Ye Wang, The University of Texas at Austin, US</td>
<td>Aydin Ayatu, Mohit Tiwari and Michael Orshansky, University of Texas at Austin, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: We describe the first hardware implementation of a quantum-secure encryption scheme along with its low-cost power side-channel countermeasures. The encryption uses an implementation-friendly Binary-Ring-Learning-with-Errors (B-RLWE) problem with binary errors that can be efficiently generated in hardware. We demonstrate that a direct implementation of B-RLWE exhibits vulnerability to power side-channel attacks, even to Simple Power Analysis, due to the nature of binary coefficients. We mitigate this vulnerability with a redundant addition and memory update. To further protect against Differential Power Analysis (DPA), we use a B-RLWE specific opportunity to construct a lightweight yet effective countermeasure based on randomization of intermediate states and masked threshold decoding. On a SAKURA-G FPGA board, we show that our method increases the required number of measurements for DPA attacks by 40X compared to unprotected design. Our results also quantify the trade-off between side-channel security and hardware area-cost of B-RLWE.</td>
<td></td>
</tr>
<tr>
<td>11:30</td>
<td>10.4.2</td>
<td>HIGH SPEED ASIC IMPLEMENTATIONS OF LEAKAGE-RESILIENT CRYPTOGRAPHY</td>
<td>Thomas Unterluggauer, Graz University of Technology, AT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Thomas Unterluggauer, Graz University of Technology, AT</td>
<td>Robert Schilling1, Thomas Unterluggauer2, Stefan Mangard1, Frank Gürkan1, Michael Muehlberghuber1 and Luca Benini5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors: Robert Schilling1, Thomas Unterluggauer2, Stefan Mangard1, Frank Gürkan1, Michael Muehlberghuber1 and Luca Benini5</td>
<td>1Graz University of Technology / Know Center GmbH, AT; 2Graz University of Technology, AT; 3ETH Zurich, CH; 4Integrated Systems Laboratory (ETH Zurich), CH; 5Università di Bologna, IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Embedded devices in the Internet-of-Things require encryption functionalities to secure their communication. However, side-channel attacks and in particular differential power analysis (DPA) attacks pose a serious threat to cryptographic implementations. While state-of-the-art countermeasures like masking slow down the performance and can only prevent DPA up to a certain order, leakage-resilient schemes are designed to stay secure even in the presence of side-channel leakage. Although several leakage-resilient schemes have been proposed, there are no hardware implementations to demonstrate their practicality and performance on measurable silicon. In this work, we present an ASIC implementation of a multi-core System-on-Chip extended with a software-programmable accelerator for leakage-resilient cryptography. The accelerator is deeply embedded in the shared memory architecture of the many-core system, supports different configurations, contains a high-throughput implementation of the 2PRG primitive based on AES-128, offers two side-channel protected re-keying functions, and is the first fabricated design of the side-channel secure authenticated encryption scheme ISAP. The accelerator reaches a maximum throughput of 7.49 Gbps and a best-case energy efficiency of 137 Gbps/W making this accelerator suitable for high-speed secure IoT applications.</td>
<td></td>
</tr>
<tr>
<td>12:00</td>
<td>10.4.3</td>
<td>OPTIMIZATION OF THE PLL CONFIGURATION IN A PLL-BASED TRNG DESIGN</td>
<td>Ele Neumon Allini, Laboratoire Hubert Curien, University of Saint-Etienne, FR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Ele Neumon Allini, Laboratoire Hubert Curien, University of Saint-Etienne, FR</td>
<td>Ele Neumon Allini, Oto Petura, Viktor Fischer and Florent Bernard, Hubert Curien Laboratory, Jean Monnet University, FR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors: Ele Neumon Allini, Oto Petura, Viktor Fischer and Florent Bernard, Hubert Curien Laboratory, Jean Monnet University, FR</td>
<td>Several recent designs show that the phase-locked loops (PLLs) are well suited for building true random number generators (TRNGs) in logic devices and especially in FPGAs, in which PLLs are physically isolated from the rest of the device. However, the setup of the PLL configuration for the PLL-based TRNG is a challenging task. Indeed, the designer has to take into account physical constraints of the hardened block, when trying to achieve required performance (bit rate) and security (entropy rate per bit). In this paper, we introduce a method aimed at choosing PLL parameters (e.g. input frequency, multiplication and division factors of the PLL) that satisfy hardware constraints, while achieving the highest possible bit rate or entropy rate according to application requirements. The proposed method is fast enough to produce all possible configurations in a short time. Comparing to the previous method based on a generic algorithm, which was able to find only a locally optimized solution and only for one PLL in tens of seconds, the new method finds exhaustive set of possible configurations of one- or two-PLL TRNG in few seconds, while the found configurations can be ordered depending on their performance or sensitivity to jitter.</td>
</tr>
<tr>
<td>12:30</td>
<td>IP4-15, 187</td>
<td>ERASMUS: EFFICIENT REMOTE ATTESTATION VIA SELF-MEASUREMENT FOR UNATTENDED SETTINGS</td>
<td>Norrathep Rattanavipanon, University of California, Irvine, TH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Norrathep Rattanavipanon, University of California, Irvine, TH</td>
<td>Xavier Cartet1, Norrathep Rattanavipanon2 and Gene Tsudik2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors: Xavier Cartet1, Norrathep Rattanavipanon2 and Gene Tsudik2</td>
<td>1UC Irvine, US; 2UCI, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Remote attestation (RA) is a popular means of detecting malware in embedded and IoT devices. RA is usually realized as a protocol via which a trusted verifier measures software integrity of an untrusted remote device called prover. All prior RA techniques require on-demand operation. We identify two drawbacks of this approach in the context of unattended devices: First, it fails to detect mobile malware that enters and leaves the prover between successive RA instances. Second, it requires the prover to keep track of RA queries, which can only prevent DPA up to a certain order, leakage-resilient schemes are designed to stay secure even in the presence of side-channel leakage. Although several leakage-resilient schemes have been proposed, there are no hardware implementations to demonstrate their practicality and performance on measurable silicon. In this work, we present an ASIC implementation of a multi-core System-on-Chip extended with a software-programmable accelerator for leakage-resilient cryptography. The accelerator is deeply embedded in the shared memory architecture of the many-core system, supports different configurations, contains a high-throughput implementation of the 2PRG primitive based on AES-128, offers two side-channel protected re-keying functions, and is the first fabricated design of the side-channel secure authenticated encryption scheme ISAP. The accelerator reaches a maximum throughput of 7.49 Gbps and a best-case energy efficiency of 137 Gbps/W making this accelerator suitable for high-speed secure IoT applications.</td>
<td></td>
</tr>
</tbody>
</table>
NON-INTRUSIVE TESTING TECHNIQUE FOR DETECTION OF TROJANS IN ASYNCHRONOUS CIRCUITS

Speaker:
Rodrigo Possamai Bastos, TIMA Laboratory, CNRS/Grenoble INP/LIF, FR

Authors:
Leonel Acunha Guimarães, Thiago Ferreira Paiva Leite, Rodrigo Possamai Bastos and Laurent Fesquet, TIMA - Grenoble Institute of Technology, FR

Abstract
Asynchronous circuits, as any IC, are vulnerable to hardware Trojans (HTs), which might be maliciously implanted in IC designs during outsourced fabrication phases. In this paper, a new testing technique to detect HTs by exploiting the regular side-channel properties of quasi-delay insensitive (QDI) asynchronous circuits is proposed. The technique does not need neither additional circuitry nor significant adjustments in the post-fabrication testing phase. Simulation results show that the proposed technique is able to detect HTs with dimensions smaller than 1% of the original circuit.

Download Paper (PDF; Only available from the DATE venue WiFi)

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

10.5 Mixed-Criticality and Fault-Tolerant Real-Time Embedded Systems

Date: Thursday, March 22, 2018
Time: 11:00 - 12:30
Location / Room: Konf. 3

Chair:
Leandro Indrusiak, Univ. of York, GB, Contact Leandro Soares Indrusiak

Co-Chair:
Andy Pimentel, University of Amsterdam, DE, Contact Andy Pimental

The session presents advances in mixed criticality systems related to Availability, Memory Bandwidth and Fault-Tolerance. The first paper details on service degradation in mixed criticality systems. The second paper handles mixed-critical workloads in the presence of memory contention. The third paper considers fault-tolerance to be incorporated into control algorithms.

Time Label Presentation Title Authors

11:00 10.5.1 AVAILABILITY ENHANCEMENT AND ANALYSIS FOR MIXED-CRITICALITY SYSTEMS ON MULTI-CORE
Speaker:
Roberto Medina, Télécom ParisTech, FR

Authors:
Roberto Medina, Etienne Borde and Laurent Pautet, Télécom ParisTech, FR

Abstract
In the critical systems domain, Mixed Criticality Systems (MCS) improve considerably the usage of computation resources by running tasks with different levels of criticality on multi-core processors. To ensure the safety of MCS, services provided by low criticality tasks are degraded or stopped whenever high criticality tasks need more computation time than initially credited. The evaluation of this degradation is hardly considered in the literature although low criticality services are of prime importance for the quality of service (QoS) of critical systems. In this paper, we propose a method to evaluate the availability of low criticality services, i.e. how often these services are delivered in MCS. We also propose a task model that improves this availability, demonstrated thanks to our evaluation method on an illustrative example of MCS.

Download Paper (PDF; Only available from the DATE venue WiFi)
<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:30</td>
<td>10.5.2</td>
<td>MIXED-CRITICALITY SCHEDULING WITH MEMORY BANDWIDTH REGULATION</td>
<td>Muhammad Ali Awan, CISTER/INESC-TEC and ISEP/IPP, Porto, Portugal, PT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Muhammad Ali Awan, CISTER/INESC-TEC and ISEP/IPP, Porto, Portugal, PT</td>
<td>Authors: Muhammad Ali Awan(^1), Pedro Soulø(^5), Konstantinos Bletsas(^3), Benny Akesson(^1) and Eduardo Tovar(^1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1CISTER/INESC-TEC, ISEP, PT; 2Faculty of Engineering of the University of Porto, PT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Mixed criticality (MC) multicore system design must reconcile safety guarantees and high performance. The interference among cores on shared resources in such systems leads to unpredictable temporal behaviour. Memory bandwidth regulation among different cores can be a useful tool to mitigate the interference when accessing main memory. However, for mixed criticality systems conforming to the (well-established) Vestal model, the existing schedulability analyses are oblivious to memory stalling effects, including stalls from memory bandwidth regulation. This makes it unsafe. In this paper, we address this issue by formulating a schedulability analysis for mixed criticality fixed priority-scheduled multicore systems using per-core memory access regulation. We also propose multiple heuristics for memory bandwidth allocation and task-to-core assignment. We implement our analysis and heuristics in a tool and evaluate them, performance-wise, through extensive experiments. Our experiments show that stall-oblivious schedulability analysis may be optimistic due to contention on shared memory resources. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
<tr>
<td>12:00</td>
<td>10.5.3</td>
<td>DESIGN AND VALIDATION OF FAULT-TOLERANT EMBEDDED CONTROLLERS</td>
<td>Soumyajit Dey, IIT Kharagpur, IN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Soumyajit Dey, IIT Kharagpur, IN</td>
<td>Authors: Saurav Kumar Ghosh(^1), Soumyajit Dey(^2), Dip Goswami(^3), Daniel Mueller-Gritschneider(^4) and Samarjit Chakraborty(^4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1Dept. of CSE, IIT Kharagpur, IN; 2Indian Institute of Technology Kharagpur, IN; 3Eindhoven University of Technology, NL; 4Technical University of Munich, DE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Embedded control systems are an important and often safety-critical class of applications that need to operate reliably even in the presence of faults. We show that intermittent fault scenarios caused by wear-out effects due to a higher density and a smaller geometry of the embedded electronic components may become a reliability concern for real-time embedded control applications. To mitigate the effects of such intermittent faults, we propose a novel fault-tolerant controller design method such that the resulting controllers ensure closed loop stability (i.e. guarantee safety) with only possibly degraded performance under such fault scenarios. In order to measure the amortized performance offered by the software implementations of such fault-tolerant controllers, we provide a program analysis methodology that statically estimates the quality of control guaranteed by the C code implementation of the fault-tolerant control law. This combination of fault-tolerant controller design followed by performance feedback computed using a formal analysis is illustrated with a case study from the automotive domain. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
<tr>
<td>12:30</td>
<td>10.5.3</td>
<td>END-TO-END LATENCY ANALYSIS OF CAUSE-EFFECT CHAINS IN AN ENGINE MANAGEMENT SYSTEM</td>
<td>Junchul Choi, Donghyun Kang and Soonhoi Ha, Seoul National University, KR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Junchul Choi, Seoul National University, KR</td>
<td>Authors: Junchul Choi, Donghyun Kang and Soonhoi Ha, Seoul National University, KR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: An engine management system consists of periodic or sporadic real-time tasks. A task is a set of runnables that may be fully preemptive or partially at runnable boundaries. A cause-effect chain is defined as a chain of runnables that are connected by the read/write dependency. We propose a novel analytical technique to estimate the end-to-end latency of a cause-effect chain by considering conservatively estimated schedule time bounds of associated runnables. The proposed approach is verified with an industrial-strength automotive benchmark. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
<tr>
<td>12:31</td>
<td>10.5.3</td>
<td>TOWARDS FULLY AUTOMATED TLM-TO-RTL PROPERTY REFINEMENT</td>
<td>Vladimir Herdt, University of Bremen, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Vladimir Herdt, University of Bremen, DE</td>
<td>Authors: Vladimir Herdt(^1), Hoang M. Le(^1), Daniel Grosse(^2) and Rolf Drechsler(^2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1University of Bremen, DE; 2University of Bremen/DFKI GmbH, DE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: An ESL design flow starts with a TLM description, which is thoroughly verified and then refined to a RTL description in subsequent steps. The properties used for TLM verification are refined alongside the TLM description to serve as starting point for RTL property checking. However, a manual transformation of properties from TLM to RTL is error prone and time consuming. Therefore, in this paper we propose a fully automated TLM-to-RTL property refinement based on a symbolic analysis of transactors. We demonstrate the applicability of our property refinement approach using a case study. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
</tbody>
</table>
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 13:00 - 14:00
- Coffee Break 15:30 - 16:00

10.6 Special Session: Computing with Ferroelectric FETs - Devices, Models, Systems, and Applications

Date: Thursday, March 22, 2018
Time: 11:00 - 12:30
Location / Room: Konf. 4

Chair:
Michael Niemier, University of Notre Dame, US, Contact Michael Niemier

Co-Chair:
Ian O'Connor, Ecole Centrale de Lyon, FR, Contact Ian O'Connor

In this session, we consider devices, circuits, and systems comprised of transistors with integrated ferroelectrics. Said structures are actively being considered by various semiconductor manufacturers as they can address a large and unique design space. Transistors with integrated ferroelectrics could (i) enable a better switch (i.e., offer steeper subthreshold swings), (ii) be CMOS compatible, (iii) have multiple operating modes (i.e., I-V characteristics can also enable compact, 1-transistor, non-volatile storage elements, as well as analog synaptic behavior), and (iv) have been experimentally demonstrated (i.e., with respect to all of the aforementioned operating modes). These device-level characteristics offer unique opportunities at the circuit, architectural, and system-level, and are considered from device, circuit/architecture, and foundry-level perspectives.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:00</td>
<td>10.6.1</td>
<td>OUTLOOK FOR LOW-POWER BEYOND-CMOS DEVICES</td>
<td>Author: An Chen, Semiconductor Research Corp, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>tbd</td>
</tr>
<tr>
<td>11:30</td>
<td>10.6.2</td>
<td>EXPLOITING FERROELECTRIC FETS: FROM LOGIC-IN-MEMORY TO NEURAL NETWORKS AND BEYOND</td>
<td>Speaker and Author: Xiaobo Sharon Hu, University of Notre Dame, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>tbd</td>
</tr>
<tr>
<td>12:00</td>
<td>10.6.3</td>
<td>FEFETS: FROM NON-VOLATILE MEMORY TO NON-VOLATILE COMPUTING</td>
<td>Author: Stefan Slesazeck, NaMLab gGmbH, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>tbd</td>
</tr>
</tbody>
</table>
12:30 End of session

Lunch Break in Großer Saal and Saal 1

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

10.8 An Industry Approach to FPGA and SOC System Development and Verification

**Date:** Thursday, March 22, 2018  
**Time:** 11:00 - 12:30  
**Location / Room:** Exhibition Theatre

**Organiser:**  
Alexander Scheiber, The MathWorks, DE, Contact Alexander Scheiber

**Speaker:**  
John Zhao, MathWorks, US, Contact John Zhao

MATLAB and Simulink provide a rich environment for embedded-system development, with libraries of proven, specialized algorithms ready to use for specific applications. The environment enables a model-based design workflow for fast prototyping and implementation of the algorithms on heterogeneous embedded targets, such as MPSoC. A system-level design approach enables architectural exploration and partitioning, as well as coordination between SW and HW development workflows. Functional verification throughout the design process improves coverage and test-case generation while reducing the time and resources required.

In this exhibition theater session, you will learn:

- Automatically generate synthesizable RTL code from your MATLAB and Simulink algorithms targeting FPGA, ASIC or Programmable SoC
- A HW/SW co-design workflow that combines system level design and simulation with automatic code generation
- Functional verification using MATLAB and Simulink in a SystemVerilog workflow illustrated by a detailed example
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 12:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:30
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:30
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

UB10 Session 10

Date: Thursday, March 22, 2018
Time: 12:00 - 14:30
Location / Room: Booth 1, Exhibition Area

<table>
<thead>
<tr>
<th>Label</th>
<th>Presentation Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>UB10.1</td>
<td>ARCHON: AN ARCHITECTURE-OPEN RESOURCE-DRIVEN CROSS-LAYER MODELLING FRAMEWORK</td>
</tr>
<tr>
<td>Authors</td>
<td>Fei Xia1, Ashur Rafiev1, Mohammed Al-Hayanni2, Alexei Illasov3, Rishad Shafik1, Alexander Romanovsky1 and Alex Yakovlev1</td>
</tr>
<tr>
<td>1 Newcastle University, GB; 2 Newcastle University, UK and University of Technology and HCED, IQ</td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>This demonstration showcases a modeling method for large complex computing systems focusing on many-core types and concentrating on the crosslayer aspects. The resource-driven models aim to help system designers reason about, analyze, and ultimately design such systems across all conventional computing and communication layers, from application, operating system, down to the finest hardware details. The framework and tool support the notion of selective abstraction and are suitable for studying such non-functional properties such as performance, reliability and energy consumption.</td>
</tr>
<tr>
<td>More information ...</td>
<td></td>
</tr>
</tbody>
</table>

| UB10.2 | FPGA-BASED HARDWARE ACCELERATOR FOR DRUG DISCOVERY |
| Authors | Ghaith Tarawneh, Alessandro de Gennaro, Georgy Lukyanov and Andrey Mokhov, Newcastle University, GB |
| Abstract | We present an FPGA-based hardware accelerator for drug discovery, developed during the EPSRC programme grant POETS (EP/N031768/1) in partnership with e-Therapeutics, an Oxford-based drug discovery company. e-Therapeutics is pioneering a novel form of drug discovery based on analyzing protein interactome networks (https://www.youtube.com/watch?v=wOQPuUrzgA). This approach can discover suitable drug candidates much more efficiently compared to wet lab testing but requires considerable computing power, particularly because commodity computers are generally inefficient at analyzing large-scale networks. The presented accelerator, consisting of an FPGA board with a silicon-mapped protein interactome plus accompanying software formalisms and tools, can deliver a 1000x speed up in this application compared to software running on commodity computers. We will showcase demos in which we run in-silico analysis of protein interactomes to test drug effects and visualize the results in real-time. |
| More information ... |

| UB10.3 | ADVANCED SIMULATION OF QUANTUM COMPUTATIONS |
| Authors | Zulehner Aiken and Robert Wille, Johannes Kepler University Linz, AT |
| Abstract | Quantum computation is a promising emerging technology which allows for substantial speed-ups compared to classical computation. Since physical realizations of quantum computers are in their infancy, most research in this domain still relies on simulations on classical machines. This causes an exponential overhead which current simulators try to tackle with straightforward array-based representations and massive hardware power. There also exist solutions based on decision diagrams (graph-based approaches) that try to tackle the complexity by exploiting redundancies in quantum states and operations. However, they did not get established since they yield speedups only for certain benchmarks. Here, we demonstrate a new graph-based simulation approach which clearly outperforms state-of-the-art simulators. By this, users can efficiently execute quantum algorithms even if the respective quantum computers are not broadly available yet. |
| More information ... |
UB10.4 RISC-V PROCESSOR MODELING IN IP-XACT USING KACTUS2
Authors:
Esko Peikkarinen and Timo Hämäläinen, Tampere University of Technology, FI
Abstract
The complexity of modern embedded system design is managed by advanced, high-level design methodologies such as IP-XACT. However, integrating IP-XACT as a part of an existing design flow and packaging legacy sources is too often inhibited by the inherent differences between IP-XACT and the traditional hardware description languages. In this work, we present an existing Verilog implementation of a RISC-V microprocessor and package it with our open-source IP-XACT tool Kactus2. The resulting IP-XACT description will be publicly available and based on the modeling experience we report the observed pitfalls in the transition from HDL to IP-XACT.
More information ...

UB10.5 RECONFIGURABLE SELF-TIMED DATAFLOW ACCELERATOR
Authors:
Danil Sokolov, Alessandro de Gennaro and Andrey Mokhov, Newcastle University, GB
Abstract
Many applications require reconfigurable pipelines to handle incoming data items differently depending on their values or the operating mode. Currently, reconfigurable synchronous pipelines are the mainstream of dataflow accelerators. However, there are certain advantages to be gained from self-timed dataflow processing, e.g., robustness to unstable power supply, data-dependent performance, etc. To become attractive for industry, reconfigurable asynchronous pipelines need a formal behavioural model and design automation. This demo will present a design flow for the specification, verification and synthesis of reconfigurable self-timed pipelines using Dataflow Blitoune formalism in Workcraft(https://workcraft.org/). As a case study we will use an asynchronous accelerator for Ordinal Pattern Encoding(OPE) with reconfigurable pipeline depth. We will exhibit the resultant OPE chip fabricated in TSMC90nm to show the benefits of reconfigurability and asynchrony for dataflow processing.
More information ...

UB10.6 TOOL/OMC: OPTIMIZED COMPILATION OF EXECUTABLE UML/SYSML DIAGRAMS FOR THE DESIGN OF DATA-FLOW APPLICATIONS
Authors:
Andrea Entrò1, Julien Lallet1, Renaud Pacalet2 and Ludovic Aptrelle2
1Nokia Bell Labs, FR; 2Télécom ParisTech, FR
Abstract
Future 5G networks are expected to increase data rates by a factor of 10x. To meet this requirement, baseband stations will be equipped with both programmable (e.g., CPUs, DSPs) and reconfigurable components (e.g., FPGA). Efficiently programming these architectures is not trivial due to the inner complexity and interactions of these two types of components. This raises the need for unified design flows capable of rapidly partitioning and programming these mixed architectures. Our demonstration will show the complete system-level design and Design Space Exploration, based on UML/SysML diagrams, of a 5G data-link layer receiver, that is partitioned onto both programmable and reconfigurable hardware. We realize an implementation of such a UML/SysML design by compiling it into an executable C application whose memory footprint is optimized with respect to a given scheduling. We will validate the effectiveness of our solution by comparing automated vs manual designs.
More information ...

UB10.7 USING FORMAL METHODS FOR AUTOMATIC PLATFORM-INDEPENDENT CODE GENERATION OF RUN-TIME MANAGEMENT
Authors:
MohammadSadegh Dalvandi, Michael Butler and Asieh Salehi Fathabadi, University of Southampton, GB
Abstract
Run-Time Management (RTM) systems are used in embedded systems to dynamically adapt hardware performance to minimise energy consumption. In this demonstration, we present a framework for automatic generation of RTM implementations from platform-independent formal models. The methodology in designing the RTM systems uses a high-level mathematical language, Event-B, which can describe systems at different abstraction levels. A code generation tool is used to translate platform-independent Event-B RTM models to platform-specific implementations in C. Formal verification is used to ensure correctness of the Event-B models. The portability offered by our methodology is demonstrated by modelling a Reinforcement Learning (RL) based RTM and generating implementations for two different platforms that all achieve energy savings on the respective platforms. The generated RTM code has been integrated with the PRIME framework, a cross-layer framework for embedded power management.
More information ...

UB10.8 IIP GENERATORS TO EASE ANALOG IC DESIGN
Authors:
Benjamin Prautsch, Uwe Eichler and Torsten Reich, Fraunhofer Institute for Integrated Circuits IIS/EAS, DE
Abstract
Semiconductor technology has shown significant progress over the last decades. Digital EDA (electronic design automation) allowed that this progress could be converted to high-performance digital ICs. Analog components are part of Systems-on-Chip (SoC) too, but analog EDA lags far behind. Therefore, a lot of effort was spent to automate analog IC design. Major results are constraint-based layout-aware optimization tools using predefined layout templates or pure automation as well as analog generators containing expert knowledge. While optimization is a holistic top-down approach, generators allow parameterized and fast bottom-up generation of critical schematic and layout parts, pre-planned by experienced designers. With IIP Generators, we follow three use cases to ease analog design: 1) design on higher hierarchy levels, 2) development of hierarchical high-level IIPs, and 3) automated design porting due to highly technology-independent blocks down to 22nm.
More information ...

UB10.9 ABSYNTH: A COMPREHENSIVE APPROACH TO FRONT TO BACK ANALOG BLOCK DESIGN AUTOMATION
Authors:
Abhaya Chandra Kammara S.1, Sidney Pontes-Filho2 and Andreas König2
1IIS, TU Kaiserslautern, DE; 2University of Kaiserslautern, DE
Abstract
ABSYNTH was first presented in CEBIT 2014 where complete, practical circuit sizing approaches have been shown using meta-heuristics on trusted simulators. This tool was then proven by its use in design of several cells in a research project. Here, we present the extension to our nested optimization approach that creates a symmetric and well matched layout in every step for every instance in the population of the swarm, that is extracted in our flow to provide feedback to the cost function impacting on the population update for more viable and robust circuits. The layout optimization presented in this DEMO works with Cadence Layout design tools. Our initial focus is, motivated by Industry 4.0, IoT, on cells for signal conditioning electronics with reconfigurability and SatX features.1 Abhaya C. Kammara, L. Palamichamy, and A. König, “Multi-Objective optimization and visualization for analog automation”. Complex. Intel. Syst, Springer, DOI 10.1007/s40747-016-0027-3, 2016
More information ...

14:30 End of session
Coffee Break in Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

11.0 LUNCH TIME KEYNOTE SESSION: Autonomous Driving: Ready to Market? Which are the Remaining Top Challenges?

Date: Thursday, March 22, 2018
Location / Room: Saal 2

Chair:
Ayse Coskun, Boston University, US, Contact Ayse Coskun

During the last years a lot of prototypes for automated/autonomous driving vehicles have been presented to the public. Depending on the use case car manufacturers or tech companies have used an evolutionary or a revolutionary approach. While the evolutionary way should be more reasonable applied for owned cars due to cost restraints and the need for the functionality to work more or less by “something everywhere”, the revolutionary approach following the strategy “everything somewhere” seems to be the better solution for fleets of autonomous cabs or shuttles. Although we have seen a lot of functional concepts for both approaches to automation, there are still some big challenges to be solved. On one hand the whole automation function has to be designed redundantly to ensure a sufficient functional safety level. In this context the use of Artificial Intelligence based networks could be a solution in particular neuronal networks based on deep learning. On the other hand there is still the question “how good is good enough” having in mind that perfectly working systems cannot be realized and how can the necessary verification/validation process be implemented. The public funded project PEGASUS is working to provide first answers. However: do we have considered all impacts of automated mobility?

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:20</td>
<td>11.0.1</td>
<td>KEYNOTE SPEAKER</td>
<td>Thomas Form,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Head of Electronics and Vehicle Research, Volkswagen AG, and co-ordinator of the Pegasus research project on safety of automated driving, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>During the last years a lot of prototypes for automated/autonomous driving vehicles have been presented to the public. Depending on the use case car manufacturers or tech companies have used an evolutionary or a revolutionary approach. While the evolutionary way should be more reasonable applied for owned cars due to cost restraints and the need for the functionality to work more or less by “something everywhere”, the revolutionary approach following the strategy “everything somewhere” seems to be the better solution for fleets of autonomous cabs or shuttles. Although we have seen a lot of functional concepts for both approaches to automation, there are still some big challenges to be solved. On one hand the whole automation function has to be designed redundantly to ensure a sufficient functional safety level. In this context the use of Artificial Intelligence based networks could be a solution in particular neuronal networks based on deep learning. On the other hand there is still the question “how good is good enough” having in mind that perfectly working systems cannot be realized and how can the necessary verification/validation process be implemented. The public funded project PEGASUS is working to provide first answers. However: do we have considered all impacts of automated mobility?</td>
<td></td>
</tr>
<tr>
<td>13:50</td>
<td></td>
<td>End of session</td>
<td></td>
</tr>
</tbody>
</table>
Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

11.1 Special Day Session on Designing Autonomous Systems: Smart Vision Systems

Date: Thursday, March 22, 2018
Time: 14:00 - 15:30
Location / Room: Saal 2
Chair: Bernhard Rinner, Alpen-Adria-Universität Klagenfurt, AT, Contact Rinner Bernhard

Smart vision systems that capture data in both private and public environments are now ubiquitous and have applications in security, disaster response, robotics, and smart environments, among others. Processing this data manually is an immensely tedious - and for some applications - an infeasible task, and an enhanced level of automation and self-awareness in the overall system is a key to overcome the design challenges. This special session addresses design aspects of smart vision systems realized at different levels: the image sensor, the camera node and the system level.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00</td>
<td>11.1.1</td>
<td>THE CAMEL APPROACH TO STACKED SENSOR SMART CAMERAS</td>
<td>Marilyn Wolf, Georgia Institute of Technology, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Marilyn Wolf, Georgia Institute of Technology, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>Saiyal Mukhopadhyay, Marilyn Wolf and Evan Gebahrdt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1Georgia Institute of Technology, US; 2School of ECE, Georgia Institute of Technology, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>Stacked image sensor systems combine an image sensor, memory, and processors using 3D technology. Stacking camera components that have traditionally been packaged separately provides several benefits: very high bandwidth out of the image sensor, allowing for higher frame rates; very low latency, providing opportunities for image processing and computer vision algorithms which can adapt at very high rates; and lower power consumption. This paper will describe the characteristics of stacked image sensor systems and novel algorithmic and systems concepts that are made possible by these stacked sensors.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
<tr>
<td>14:18</td>
<td>11.1.2</td>
<td>A DESIGN TOOL FOR HIGH PERFORMANCE IMAGE PROCESSING ON MULTICORE PLATFORMS</td>
<td>Shuvra Bhattacharya, University of Maryland, College Park, MD, USA and Tampere University of Technology, Finland, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td>Shuvra Bhattacharya, University of Maryland, College Park, MD, USA and Tampere University of Technology, Finland, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>Jahhao Wu, Timothy Blattner, Wald Keyrouz and Shuvra S. Bhattacharya</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1University of Maryland, US; 2National Institute of Standards and Technology, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>Design and implementation of smart vision systems often involve the mapping of complex image processing algorithms into efficient, real-time implementations on multicore platforms. In this paper, we describe a novel design tool that is developed to address this important challenge. A key component of the tool is a new approach to hierarchical dataflow scheduling that integrates a global scheduler and multiple local schedulers. The local schedulers are lightweight modules that work independently. The global scheduler interacts with the local schedulers to optimize overall memory usage and execution time. The proposed design tool is demonstrated through a case study involving an image stitching application for large scale microscopy images.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
</tbody>
</table>
### 11.2 Timing and Power Driven Physical Design

**Date:** Thursday, March 22, 2018  
**Time:** 14:00 - 15:30  
**Location / Room:** Kont 6  
**Chair:** Miguel Silveira, INESC-ID/IST, PT, Contact Luis Miguel Silveira  
**Co-Chair:**

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
<th>Authors</th>
<th>Abstract</th>
<th>Download Paper (PDF; Only available from the DATE venue WiFi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:36</td>
<td>11.1.3</td>
<td>QUASAR, A HIGH-LEVEL PROGRAMMING LANGUAGE AND DEVELOPMENT ENVIRONMENT FOR DESIGNING SMART VISION SYSTEMS ON EMBEDDED PLATFORMS</td>
<td>Bart Goossens, Ghent University - imec, BE</td>
<td>Bart Goossens, Hélène Luong, Jan Aelterman and Wilfried Philips, Ghent University, Dept. of Telecommunications and Information Processing, BE</td>
<td>We present Quasar, a new programming framework that handles many complex aspects in the design of smart vision systems on embedded platforms. The principle of parallelization, data flow management, scheduling, and load balancing is a highlevel programming language, which has a low barrier of entry and is therefore well suited for algorithm design and rapid prototyping. Through several benchmarks and application use cases we demonstrate that programs written in Quasar have a performance that is on a par with (or better than) hand-tuned CUDA and OpenACC code while the development requires much less time and is future-proof.</td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
<tr>
<td>14:54</td>
<td>11.1.4</td>
<td>CONCURRENT FOCAL-PLANE GENERATION OF COMPRESSED SAMPLES FROM TIME-ENCODED PIXEL VALUES</td>
<td>Ricardo Camona-Galan, Instituto de Microelectronica de Sevilla (CSIC-Univ. de Sevilla), ES</td>
<td>Mano Trevisi, Héctor C Bandala, Jorge Fernández-Benítez, Ricardo Camona-Galán and Ángel Rodríguez-Vázquez</td>
<td>Compressive sampling allows wrapping the relevant content of an image in a reduced set of data. This exploits the sparsity of natural images. This principle can be employed to deliver images over a network under a restricted data rate and still receive enough meaningful information. An efficient implementation of this principle lies in the generation of the compressed samples right at the imager. Otherwise, i.e. digitizing the complete image and then compressing the compressed samples in the digital plane, the required memory and processing resources can seriously compromise the budget of an autonomous camera node. In this paper we present the design of a pixel architecture that encodes light intensity into time, followed by a global strategy to pseudo-randomly combine pixel values and generate, on-chip and on-line, the compressed samples.</td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
<tr>
<td>15:12</td>
<td>11.1.5</td>
<td>CONTACTLESS FINGER AND FACE CAPTURING ON A SECURE HANDHELD EMBEDDED DEVICE</td>
<td>Axel Weissenfeld, AIT Austrian Institute of Technology GmbH, AT</td>
<td>Axel Weissenfeld and Bernhard Strobil, Austrian Institute of Technology, AT</td>
<td>Traveler flows and crossings at the external borders of the EU are increasing and are expected to increase even more in the future; trends which encompass great challenges for travelers, border guards and the border infrastructure. In this paper we present a new handheld device, which enables border control authorities to check European, visa-holding and frequent third country travelers in a comfortable, fast and secure way. The mobile solution incorporates new multimodal biometric capturing and matching units for face and 4-finger authentication. Thereby, the focus is on the capturing unit and fingerprint verification, which is evaluated in detail. On the other hand, the use in border control requires high security measurements and trustworthy use of credentials, which are also presented. Tests of the handheld device at a land border indicate great acceptance by travelers and border guards.</td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
</tbody>
</table>

**Coffee Breaks in the Exhibition Area**

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

**Lunch Breaks (Großer Saal + Saal 1)**

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

**Tuesday, March 20, 2018**

- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

**Wednesday, March 21, 2018**

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

**Thursday, March 22, 2018**

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00
The first two papers in this session present timing analysis algorithms to handle non-ideal physical conditions. In particular, the first paper extends the involution model adding non-deterministic delay variations. The second paper presents a method for estimating the worst case delay based on the extreme value theory. The remaining two papers in this session deal with floorplanning and placement. One paper addresses 3D ICs for multiple supply voltages. The other shows a way to accelerate the analytical placement employing GPU cores.

**A FAITHFUL BINARY CIRCUIT MODEL WITH ADVERSARIAL NOISE**

**Speaker:** Jürgen Maier, TU Wien, AT

**Authors:** Matthias Függer¹, Jürgen Maier², Robert Nayer³, Thomas Nowak² and Ulrich Schmid²

1LSV, CNRS & ENS Paris-Saclay, FR; 2TU Wien, AT; 3Université Paris Sud, FR

**Abstract**

Accurate delay models are important for static and dynamic timing analysis of digital circuits, and mandatory for formal verification. However, Függer et al. [IEEE TC 2016] proved that pure and inertial delays, which are employed for dynamic timing analysis in state-of-the-art tools like ModelSim, NC-Sim and VCS, do not yield faithful digital circuit models. Involution delays, which are based on delay functions that are mathematical involutions depending on the previous-output-to-input time offset, were introduced by Függer et al. [DATE’15] as a faithful alternative (that can easily be used with existing tools). Although involution delays were shown to predict real signal traces reasonably accurately, any model with a deterministic delay function is naturally limited in its modeling power. In this paper, we thus extend the involution model, by adding non-deterministic delay variations (random or even adversarial), and prove analytically that faithfulness is not impaired by this generalization. Also the amount of non-determinism must be considerably restricted to ensure this property, the result is surprising: the involution model differs from non-faithful models mainly in handling fast glitch trains, where small delay shifts have large effects. This originally suggested that adding even small variations should break the faithfulness of the model, which turned out not to be the case. Moreover, the results of our simulations also confirm that this generalized involution model has larger modeling power and, hence, applicability.

**Download Paper (PDF; Only available from the DATE venue WiFi)**

**EVT-BASED WORST CASE DELAY ESTIMATION UNDER PROCESS VARIATION**

**Speaker:** Charalampos Antoniadis, University of Thessaly, GR

**Authors:** Charalampos Antoniadis, Dimitrios Gargalionis, Nestor Ermolopoulou and Georgios Stamoulis, University of Thessaly, GR

**Abstract**

Manufacturing process variation in sub-20nm processes has introduced ever increasing overhead in Static Timing Analysis (STA) in order to guarantee the reliable operation of the circuit. Chip designers apply corner-based analysis and add guard-bands to design parameters in order to take into account the impact of process variation on timing. However, the aforementioned techniques are either too slow as the number of design parameters proliferates with the integration of more components into a chip or inaccurate due to the assumption that the worst case delay resides at the corners of design parameters. In this paper, we present a novel statistical methodology, which relies on Extreme Value Theory (EVT), to estimate the worst case delay of VLSI circuits under variations in gate/interconnect parameters. Despite the previous statistical approaches toward maximum delay estimation, our methodology can be applied regardless of the underlying gate/interconnect delay model or any assumption about the distribution of the Arrival Time (AT) at every circuit node, making it very appealing for integration to any level of timing analysis abstraction (from spice-to-gate level) and provide fast yet accurate results. Experimental results on ISCAS85/ISCAS89 circuits show that the estimated maximum AT at the Primary Outputs (POs) can be within 5% of the true maximum AT, at the cost of a few thousand Monte Carlo simulations.

**Download Paper (PDF; Only available from the DATE venue WiFi)**

**CO-SYNTHESIS OF FLOORPLANNING AND POWERPLANNING IN 3D ICS FOR MULTIPLE SUPPLY VOLTAGE DESIGNS**

**Speaker:** Ji-Jhi-Ying Yang, Department of Electrical Engineering, National Cheng Kung University, TW

**Authors:** Jai-Ming Lin, Chien-Yu Huang and Jihih-Ying Yang, Department of Electrical Engineering, National Cheng Kung University, TW

**Abstract**

This paper addresses a 3D floorplanning methodology, which considers floorplanning and powerplanning at the same time for Multiple Supply Voltage (MSV) circuits. Physical design becomes more complex for MSV designs since modules with the same power domain have to be placed at close locations in 3D space to facilitate powerplanning and reduce IR-drops, which would deteriorate wirelength. By properly partitioning modules of the same power domain into several voltage islands and increasing overlap area of the voltage islands in contiguous dies, we can reduce routing resource usage without increasing wirelength significantly. Furthermore, unlike previous works, our approach not only can handle a netlist with soft modules and hard modules but also can meet the fixed-outline constraint. The experimental results show that our methodology gets better results than other approaches in designs with single voltage domain and is also promising for MSV designs.

**Download Paper (PDF; Only available from the DATE venue WiFi)**

**ACCELERATE ANALYTICAL PLACEMENT WITH GPU: A GENERIC APPROACH**

**Speaker:** Martin D. F. Wong, University of Illinois Urbana-Champaign, US

**Authors:** Chun-Xun Lin and Martin Wong, University of Illinois at Urbana-Champaign, US

**Abstract**

This paper presents a generic approach of exploiting GPU parallelism to speed up the essential computations in VLSI nonlinear analytical placement. We consider the computational overhead and which are widely used as cost and constraint in nonlinear analytical placement. For wirelength gradient computing, we utilize the sparse characteristic of circuit graph to transform the compute-intensive portions into sparse matrix multiplications, which effectively optimizes the memory access pattern and mitigates the imbalance workload. For density, we introduce a computation flattening technique to achieve load balancing among threads and a High-Precision representation is integrated into our approach to guarantee the reproducibility. We have evaluated our method on a set of contest benchmarks from industry. The experimental results demonstrate our GPU method achieves a better performance over both the CPU methods and the straightforward GPU implementation.

**Download Paper (PDF; Only available from the DATE venue WiFi)**

**GENERAL FLOORPLANNING METHODOLOGY FOR 3D ICS WITH AN ARBITRARY BONDING STYLE**

**Speaker:** Chien-Yu Huang, Department of Electrical Engineering, National Chung Kung University, TW

**Authors:** Jai-Ming Lin and Chien-Yu Huang, Department of Electrical Engineering, National Chung Kung University, TW

**Abstract**

This paper proposes a general floorplanning methodology which can be applied to 3D ICs with an arbitrary bonding style. Some researches have shown that a 3D IC with the hybrid bonding style, which includes face-to-back and face-to-face, may obtain better results than that simply using the face-to-back bonding style. We respectively present an approach to assign modules to tiers for each kind of bonding style. Further, a new utilization function, called cosine-shaped function, is proposed to estimate the utilization of bins required by the analytical-based approach. Our experimental results show the cosine shaped function can obtain a little better result than the hel-shaped function on IBM benchmarks for 2D floorplanning. We also show that the proposed 3D floorplanning methodology consumes less TSVs and induces shorter wirelength compared to previous work in the hybrid bonding style.

**Download Paper (PDF; Only available from the DATE venue WiFi)**
11.3 More than Moore Interconnects

Presentation Title: A PLACEMENT ALGORITHM FOR SUPERCONDUCTING LOGIC CIRCUITS BASED ON CELL GROUPING AND SUPER-CELL PLACEMENT

Authors: Massoud Pedram, University of Southern California, US

Abstract: This paper presents a novel clustering-based placement algorithm for single flux quantum (SFO) family of superconductive electronic circuits. In these circuits, nearly all cells receive a clock signal and a placement algorithm that ignores the clock routing cost will not produce high-quality solutions. To address this issue, proposed approach simultaneously minimizes the total wirelength of the signal nets and area overhead of the clock routing. Furthermore, construction of a perfect H-tree in SFO logic circuits is not viable solution due to the resulting very high routing overhead and the infeasibility of building exact zero-skew clock routing trees. Instead, a hybrid clock tree must be used whereby higher levels of the clock tree (i.e., closer to the clock source) are based on H-tree construction whereas lower levels of the clock tree follow a linear (i.e., chain-like) structure. The proposed approach is able to reduce the overall half-perimeter wirelength by 15% and area by 8% compared with state-of-the-art techniques.

Download Paper (PDF; Only available from the DATE venue WiFi)

Presentation Title: LESAR: A DYNAMIC LINE-END SPACING AWARE DETAILED ROUTER

Authors: Yih-Lang Li, Computer Science Department, NCTU, TH

Abstract: LESAR is a dynamic line-end spacing aware detailed router. It supports single-tier 2D or 3D for multiple, logic-on-logic 3D IC tiers, efficient look-ahead legalization of intermediate Global Placement (GP) iterations, Hard Macros, Blockages, row density constraints and multiple local cell displacement functions and cell orderings. For 3D-IC, Abax can produce multi-tier 3D-IC placements by performing Legalization-based Partitioning. For efficient look-ahead legalization, Abax supports two new local displacement cost functions, multi-cell mean and multi-cell total. We show that the classical single-cell displacement and multi-cell total can result in artifacts when legalizing early intermediate GPs, and that multi-cell mean is the best candidate for look-ahead legalization. Obstructions, i.e. hard macros and blockages are handled by using two strategies. We present legalization results for the ISPD2014 and ISPD2015 benchmarks, by using GP generated from Eh?Placer, and HPWL measurement by using RippleDP. For 3D, two-tier legalization we illustrate a ~30% reduction in HPWL for a set of ISPD2014 benchmarks. For 2D legalization on the ISPD2015 benchmarks, our average HPWL increase over GP is 3.03%, compared to 7.21% of the Eh?Placer legaliser, and 43.16% of the RippleDP legaliser.

Download Paper (PDF; Only available from the DATE venue WiFi)

Coffee Breaks in the Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Coffee Break 15:00 - 16:00

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (buffet style) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00
The third paper shows how to leverage the NoC to manage resource allocations in chip multiprocessors. The final paper applies silicon photonics to the design of chip-scale interconnection networks for high-performance computing systems.

In this session, the application of emerging technologies such as 3D Integration and Silicon Photonics broadens the capabilities of chip-scale interconnects and on-chip resource allocation mechanisms. The first two papers apply 3D integration to the design of networks-on-chip by providing enhanced collective communication mechanisms and resiliency to soft errors, respectively. However, conventional 3D NoCs are not efficient in handling collective communication. Existing works mainly explore Path and Tree multicast distribution schemes for 3D NoCs. However, both these mechanisms involve high network latency and lack scalability. In this work, we propose a SMART (Single-cycle Multi-hop Asynchronous Repeated Traversal) 3D NoC architecture that is capable of achieving high-performance collective communication. The proposed High-Performance SMART (HP-SMART) 3D NoC achieves 65% and 31% latency improvements compared to the existing Path and Tree multicast-based 3D NoCs respectively. HP-SMART 3D NoC also achieves significant improvement in message latency compared to its 2D counterpart.

**Authors**

Peng Yang, Hong Kong University of Science and Technology, CN

**Abstract**

The increasing demand for more computational power from scientific computing, big data processing, and machine learning is pushing the development of HPC (high-performance computing) systems. As the basic HPC building blocks, modularized server racks with a large number of multicore nodes are facing performance and energy efficiency challenges. This paper proposes RSON, an optical network for rack-scale computing systems. RSON connects processor cores, caches, local memories, and remote memories through a novel inter/intra-chip silicon photonic network architecture. We develop a low-latency scalable channel partition and low-power dynamic path priority control scheme for RSON. Experimental results show that RSON can help rack-scale computing systems achieve up to 6.8X higher performance under the same energy consumption than state-of-the-art systems under the latest APEX (application performance at extreme scale) benchmarks.

**Download Paper**

(PDF; Only available from the DATE venue WiFi)

---

**TIME**

**LABEL**

**PRESENTATION TITLE**

**AUTHORS**

---

**14:00**

**11.3.1**

**HIGH PERFORMANCE COLLECTIVE COMMUNICATION-AWARE 3D NETWORK-ON-CHIP ARCHITECTURES**

**Speaker:**

Bishesh Joardar, Washington State University, US

**Authors:**

Bishesh Joardar, Karthi Duraisamy and Pantha Pande, Washington State University, US

**Abstract**

3D Network-on-Chip (NoC) architectures are capable of better performance and lower energy consumption compared to their planar counterparts. However, conventional 3D NoCs are not efficient in handling collective communication. Existing works mainly explore Path and Tree multicast distribution schemes for 3D NoCs. However, both these mechanisms involve high network latency and lack scalability. In this work, we propose a SMART (Single-cycle Multi-hop Asynchronous Repeated Traversal) 3D NoC architecture that is capable of achieving high-performance collective communication. The proposed High-Performance SMART (HP-SMART) 3D NoC achieves 65% and 31% latency improvements compared to the existing Path and Tree multicast-based 3D NoCs respectively. HP-SMART 3D NoC also achieves significant improvement in message latency compared to its 2D counterpart.

**Download Paper**

(PDF; Only available from the DATE venue WiFi)

---

**14:30**

**11.3.2**

**A SOFT-ERROR RESILIENT ROUTE COMPUTATION UNIT FOR 3D NETWORKS-ON-CHIPS**

**Speaker:**

Alexandre Coelho, TIMA Laboratory, FR

**Authors:**

Alexandre Coelho, Amr Charif, Nacer-Eddine Zergainoh, Jean Fraire and Raoul Velasco, Université Grenoble Alpes, CNRS, Grenoble INP, FR

**Abstract**

Three-dimensional Networks-on-Chips (3D-NoCs) have emerged as an alternative to further enhance the performance, functionality, and packaging density of 2D-NoCs. However, the increasing complexity of NoC routers, the continuous miniaturization of silicon technology, the lower-operating voltages, and the higher-operating frequencies have made the NoC increasingly vulnerable to soft errors. In particular, transient faults occurring in the route computation unit (RCU) can provoke misrouting which may lead to severe effects such as deadlocks or packet loss, consuming the operation of the entire chip. By combining a reliable fault detection circuit leveraging circuit-level double-sampling, with a cost-effective rerouting mechanism, we develop a full fault-tolerance solution that can efficiently detect and correct such fatal errors before the affected packets leave the router. To validate the proposed solution, we also introduce a novel method for simulation-based fault-injection based on the NoC's gate-level netlist.

**Download Paper**

(PDF; Only available from the DATE venue WiFi)

---

**15:00**

**11.3.3**

**SPA: SIMPLE POOL ARCHITECTURE FOR APPLICATION RESOURCE ALLOCATION IN MANY-CORE SYSTEMS**

**Speaker:**

Iraitis Anagnostopoulos, Southern Illinois University Carbondale, US

**Authors:**

Jayantika Sai Koduri and Iraitis Anagnostopoulos, Southern Illinois University Carbondale, US

**Abstract**

The technology push by Moore’s law brings a paradigm shift in the adoption of many core systems which replace high frequency superscalar processors with more simpler ones. On the software side, in order to utilize the available computational power, applications are following the high performance parallel multi-threading model. Thus, many-core systems raise the challenges of resource allocation and fragmentation making necessary efficient run-time resource management techniques. In this paper, we propose SPA, a Simple Pool Architecture for managing resource allocation in many-core systems. The proposed framework follows a distributed approach in which cores are organized into clusters and multiple clusters form a pool. Clusters are created based on system’s characteristics and the allocation of cores is performed in a distributed manner so as to take advantage of spatial features, shared resources and reduce scattering of cores. Experimental results show that SPA produces on average 15% better application response time while waiting time is reduced by 45% on average compared to other state-of-art methodologies.

**Download Paper**

(PDF; Only available from the DATE venue WiFi)

---

**15:15**

**11.3.4**

**RSON: AN INTER/INTRA-CHIP SILICON PHOTONIC NETWORK FOR RACK-SCALE COMPUTING SYSTEMS**

**Speaker:**

Peng Yang, Hong Kong University of Science and Technology, CN

**Authors:**

Peng Yang1, Zhening Pang2, Zhehui Wang2, Zhenhai Wang1, Min Xie2, Xuangui Chen1, Luan H.K. Duong1 and Jiang Xu1

1Hong Kong University of Science and Technology, HK; 2National University of Defense Technology, CN

**Abstract**

The increasing demand for more computational power from scientific computing, big data processing, and machine learning is pushing the development of HPC (high-performance computing) systems. As the basic HPC building blocks, modularized server racks with a large number of multicore nodes are facing performance and energy efficiency challenges. This paper proposes RSON, an optical network for rack-scale computing systems. RSON connects processor cores, caches, local memories, and remote memories through a novel inter/intra-chip silicon photonic network architecture. We develop a low-latency scalable channel partition and low-power dynamic path priority control scheme for RSON. Experimental results show that RSON can help rack-scale computing systems achieve up to 6.8X higher performance under the same energy consumption than state-of-the-art systems under the latest APEX (application performance at extreme scale) benchmarks.

**Download Paper**

(PDF; Only available from the DATE venue WiFi)

---

**15:30**

**IP5-4**

**UNDERSTANDING TURN MODELS FOR ADAPTIVE ROUTING: THE MODULAR APPROACH**

**Speaker:**

Eduardo Fusella, Department of Electrical Engineering and Information Technologies, University of Naples Federico II, IT

**Authors:**

Eduardo Fusella and Alessandro Cilardo, University of Naples Federico II, IT

**Abstract**

Routing algorithms were extensively studied first in multi-computer systems, then in multi- and many-core architectures. Among the commonly used routing techniques, the turn model seems the most promising solution when targeting adaptiveness. Based on the turn model, several alternative approaches with different turn prohibition schemes were proposed. This paper gives a new theoretical background for designing deadlock-free partially adaptive logic-based distributed routing algorithms that are based on the turn model. Two properties are presented, including a necessary and sufficient condition to prove that a routing algorithm is deadlock-free as long as turn restrictions follow a modular distribution. Existing approaches can be considered a subset of the solution space identified by this work. Finally, we propose a novel routing algorithm exhibiting encouraging performance improvements over state-of-the-art approaches.

**Download Paper**

(PDF; Only available from the DATE venue WiFi)
QUATER-IMAGINARY BASE FOR COMPLEX NUMBER ARITHMETIC CIRCUITS

Speaker: Souradip Sarkar, Nokia Bell Labs, BE
Authors: Souradip Sarkar and Manil Dev Gomony, Nokia Bell Labs, BE

Abstract
Arithmetic operations involving complex numbers are widely used in the signal processing functions in the physical layer of modern wireless and wireline communication systems, electronic instrumentation and control systems. With the ever increasing throughput requirements of such systems, the power consumption of the hardware realization is increasing beyond the allowed budget. Arithmetic circuits based on binary numeral system that have been optimized rigorously over the past few decades are currently being used for the computation involving complex numbers. In this paper, we present the potential of arithmetic circuits for complex number computations based on the Quater-imaginary (QI) base numeral system to reduce power consumption. We show that for a simple multiplier implementation in the QI base, the savings in power and area consumption could be up to 40% when synthesized in 28nm TSMC standard cell technology node.

Download Paper (PDF; Only available from the DATE venue WiFi)
14:30  11.4.2 VERC3: A LIBRARY FOR EXPLICIT STATE SYNTHESIS OF CONCURRENT SYSTEMS
Speaker: Marco Elver, University of Edinburgh, GB
Authors: Marco Elver, Christopher J. Banks, Paul Jackson and Vijay Nagarajan, University of Edinburgh, GB
Abstract
We propose an alternative, explicit state only, approach to concurrent system synthesis. In particular, the focus of this work is on the synthesis of distributed protocols. Given a correctness specification and a protocol skeleton (i.e. incomplete with holes), the goal is to synthesize the holes. At the heart of our technique is a dynamic programming based algorithm that prunes inferred failure candidates. The algorithm exploits the fact that typically only a few transitions are needed to reach an erroneous state in a faulty distributed protocol. Therefore, it is unlikely that every hole to be synthesized is contributing towards the error; thus, faulty protocol candidates where only a subset of holes were used can be used to infer failures of later candidates with a superset of holes. We evaluate the tool using a cache coherence protocol synthesis case study. Specifically, we study a directory based MSI protocol, assuming an unordered interconnect which gives rise to numerous race conditions which must be resolved via introducing transient states—a common cause of complexity and bugs in such protocols. In the case study, we therefore focus on synthesizing the transient state actions (we consider up to 12 holes out of possible 35). With the proposed candidate pruning optimization, we report up to 43x improvement over a naive candidate enumeration scheme. We make available the tool and C++ library, VerC3.
Download Paper (PDF; Only available from the DATE venue WiFi)

15:00  11.4.3 PROMETHEUS: PROCESSING-IN-MEMORY HETEROGENEOUS ARCHITECTURE DESIGN FROM A MULTI-LAYER NETWORK THEORETIC STRATEGY
Speaker: Yao Xiao, Shahin Nazarian and Paul Bogdan, University of Southern California, US
Authors: Yao Xiao, Shahin Nazarian and Paul Bogdan, University of Southern California, US
Abstract
With increasing demand for distributed intelligent physical systems performing big data analytics on the field and in real-time, processing-in-memory (PIM) architectures integrating 3D-stacked memory and logic layers could provide higher performance and energy efficiency. Towards this end, the PIM design requires principled and rigorous optimization strategies to identify interactions and manage data movement across different vaults. In this paper, we introduce Prometheus, a novel PIM-based framework that constructs a comprehensive model of computation and communication (MoCC) based on a static and dynamic compilation of an application. Firstly, by adopting a low level virtual machine (LLVM) intermediate representation (IR), an input application is modeled as a two-layered graph consisting of (i) a computation layer in which the nodes denote computation IR instructions and edges denote data dependencies among instructions, and (ii) a communication layer in which the nodes denote memory operations (e.g., load/store) and edges represent memory dependencies detected by alias analysis. Secondly, we develop an optimization framework that partitions the multi-layer network into processing communities within which the computational workload is maximized while balancing the load among computational clusters. Thirdly, we propose a community-to-vault mapping algorithm for designing a scalable hybrid memory cube (HMC)-based system where vaults are interconnected through a network-on-chip (NoC) approach rather than a crossbar architecture. This ensures scalability to hundreds of vaults in each cube. Experimental results demonstrate that Prometheus consisting of 64 HMC-based vaults improves system performance by 9.8x and achieves 2.3x energy reduction, compared to conventional systems.
Download Paper (PDF; Only available from the DATE venue WiFi)

15:15  11.4.4 ADVANCING SOURCE-LEVEL TIMING SIMULATION USING LOOP ACCELERATION
Speaker: Joscha Benz, University of Tuebingen, DE
Authors: Joscha Benz1, Christoph Gerum3 and Oliver Bringmann2
1University of Tuebingen, DE; 2University of Tuebingen / FZI, DE
Abstract
Source-level timing simulation (STLS) is an important technique for early examination of timing behavior, as it is very fast and accurate. A factor occasionally more important than precision is simulation speed, especially in design space exploration or very early phases of development. Additionally, practices like rapid prototyping also benefit from high-performance timing simulation. Therefore, we propose to further reduce simulation run-time by utilizing a method called loop acceleration. Accelerating a loop in the context of SLTS means deriving the timing of a loop prior to simulation to increase simulation speed of that loop. We integrated this technique in our SLTS framework and conducted a comprehensive evaluation using the Mälardalen benchmark suite. We were able to reduce simulation time by up to 43% of the original time, while the introduced accuracy loss did not exceed 8 percentage points.
Download Paper (PDF; Only available from the DATE venue WiFi)

15:30  11.4.6 IN-MEMORY COMPUTING USING PATHS-BASED LOGIC AND HETEROGENEOUS COMPONENTS
Speaker: Alvaro Velasquez, University of Central Florida, US
Authors: Alvaro Velasquez and Sumit Kumar Jha, University of Central Florida, US
Abstract
In-memory crossbar computing. In this paper, we propose a framework for synthesizing logic-in-memory circuits based on the behavior of paths of electric current throughout the memory. Limitations of using only bidirectional components with this approach are also established. We demonstrate the effectiveness of our approach by generating n-bit addition circuits that can compute using a constant number of read and write cycles.
Download Paper (PDF; Only available from the DATE venue WiFi)
Coffee Break in Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

11.5 Microfluidic Devices and Inexact Computing

Date: Thursday, March 22, 2018
Time: 14:00 - 15:30
Location / Room: Konf. 3

Chair: Martin Trefzer, University of York, GB, Contact Martin Albrecht Trefzer
Co-Chair: Lukas Sekanina, University of Brno, CZ, Contact Lukas Sekanina

The first two presentations cover applications for microfluidic devices. The first one considers sample preparation, i.e. how to efficiently prepare certain dilutions and mixtures of fluids with a given amount of storages. The second one considers programmable versions of these devices that allow for the realization of general purpose applications. The last two presentations introduce new circuit structures for computing technologies that rely on approximation and probabilities. More precisely, an adaptive approximated divider design and manipulating circuits for stochastic computing are presented.

14:00 11.5.1 STORAGE-AWARE SAMPLE PREPARATION USING FLOW-BASED MICROFLUIDIC LAB-ON-CHIP

Speaker: Robert Wille, Institute for Integrated Circuits, Johannes Kepler University Linz, 4040 Linz, Austria, AT
Authors: Sukanta Bhattacharjee1, Robert Wille2, Juinn-Dar Huang3 and Bhargab Bhattacharya1
1Indian Statistical Institute, Kolkata, IN; 2Johannes Kepler University Linz, AT; 3National Chiao Tung University, Hsinchu, TW

Abstract
Recent advances in microfluidics have been the major driving force behind the ubiquity of Labs-on-Chip (LoC) in biochemical protocol automation. The preparation of dilutions and mixtures of fluids is a basic step in sample preparation for which several algorithms and chip-architectures are well known. Dilution and mixing are implemented on biocips through a sequence of basic fluid-mixing and splitting operations performed in certain ratios. These steps are abstracted using a mixing graph. During this process, on-chip storage-units are needed to store intermediate fluids to be used later in the sequence. This allows to optimize the reactant-costs, to reduce the sample-preparation time, and/or to achieve the desired ratio. However, the number of storage-units is usually limited in given LoC architectures. Since this restriction is not considered by existing methods for sample preparation, the results that are obtained are often found to be useless (in the case when more storage-units are required than available) or more expensive than necessary (in the case when storage-units are available but not used, e.g., to further reduce the number of mixing operations or reactant-cost). In this paper, we present a storage-aware algorithm for sample preparation with flow-based LoCs which addresses these issues. We present a SAT-based approach to construct a mixing graph that enables the best usage of available storage-units while optimizing sample-preparation cost and/or time. Experimental results on several test cases reveal the scope, effectiveness, and the flexibility of the proposed method.

Download Paper (PDF; Only available from the DATE venue WiFi)
15:30  15.5.2 PUMP-AWARE FLOW ROUTING ALGORITHM FOR PROGRAMMABLE MICROFLUIDIC DEVICES
Speaker: Tsung-Yi Ho, National Tsing Hua University, TW
Authors: Guan-Ru Lai1, Chun-Yu Lin2, and Tsung-Yi Ho1
1ISEM, TW; 2National Tsing Hua University, TW
Abstract
We address the above limitation by investigating the fault tolerance of the valve-based routing fabric. We develop a theory of failure assessment and introduce a design platform has been proposed to efficiently automate the analysis of a heterogeneous sequence of cells. In this design, a valve-based routing fabric based on transposers is

15:00  15.5.3 ADAPTIVE APPROXIMATION IN ARITHMETIC CIRCUITS: A LOW-POWER UNSIGNED DIVIDER DESIGN
Speaker: Honglan Jiang, University of Alberta, CA
Authors: Honglan Jiang1, Leibo Liu2, Fabrizio Lombardi1, and Je Han1
1University of Alberta, CA; 2Tsinghua University, CN; 3Rutgers University, US
Abstract
Many approximate arithmetic circuits have been proposed for high-performance and low-power applications. However, most designs are either hardware-efficient with a low accuracy or very accurate with a limited hardware saving, mostly due to the use of a static approximation. In this paper, an adaptive approximation approach is proposed for the design of a divider. In this design, division is computed by using a reduced-width divider and a shifter by adaptively pruning the input bits. Specifically, for a 2n/2k division 2k/k bits are selected starting from the most significant '1' in the dividend/divisor. At the same time, redundant least significant bits (LSBs) are truncated or if the number of remaining LSBs is smaller than 2k for the dividend or k for the divisor, '0's are appended to the LSBs of the input. To avoid overflow, a 2(k+1)/k(1+k) divider is used to compute the division of the 2k-bit dividend and the k-bit divisor, both with the most significant bit being '0'. Thus, k+n is a key variable that determines the size of the divider and the accuracy of the approximate design. Finally, an error correction circuit is proposed to recover the error caused by the shifter by using OR gates. The synthesis results in an industrial 28nm CMOS process show that the proposed 16/8 approximate divider using an 8/4 accurate divider is 2.5x as fast and consumes 34.42% of the power of the accurate 16/8 design. Compared with the other approximate dividers, the proposed design is significantly more accurate at a similar power-delay product. Moreover, simulation results show that the proposed approximate divider outperforms the other designs in two image processing applications.

15:15  15.5.4 CORRELATION MANIPULATING CIRCUITS FOR STOCHASTIC COMPUTING
Speaker: Vincent Lee, University of Washington, US
Authors: Vincent T. Lee, Amin Alaghi and Luis Ceze, University of Washington, US
Abstract
Stochastic computing (SC) is an emerging computing technique that promises high density, low power, and error tolerant solutions. In SC, values are encoded as unary bitstreams and SC arithmetic circuits operate on one or more bitstreams. In many cases, the input bitstreams must be correlated or uncorrelated for SC arithmetic to produce accurate results. As a result, a key challenge for designing SC accelerators is manipulating the impact of correlation across SC operations. This paper presents and evaluates a set of novel correlation manipulating circuits to manage correlation in SC computation: a synchronizer, desynchronizer, and decorrelator. We then use these circuits to propose improved SC maximum, minimum, and saturating adder designs. Compared to existing correlation manipulation techniques, our circuits are more accurate and up to 3x more energy efficient. In the context of an image processing pipeline, these circuits can reduce the total energy consumption by up to 24%.

15:30  15.5.5 IPS-6, 637 FAULT-TOLERANT VALVE-BASED MICROFLUIDIC ROUTING FABRIC FOR DROPLET BARCODING IN SINGLE-CELL ANALYSIS
Speaker: Yasamin Moradi, Technical University of Munich (TUM), DE
Authors: Yasamin Moradi1, Mohamed Ibrahim2, Krishnendu Chakrabarty3, and Ul Schlichtmann1
1Technical University of Munich, DE; 2Duke University, US
Abstract
High throughput single-cell genomics is used to gain insights into diseases such as cancer. Motivated by this important application, microfluidics has emerged as a key technology for developing comprehensive biochemical procedures for studying DNA, RNA, and proteins, and many other cellular components. Recently, a hybrid microfluidic platform has been proposed to efficiently automate the analysis of a heterogeneous sequence of cells. In this design, a valve-based routing fabric based on transposers is used to label/barcode the target cells. However, the design proposed in prior work overlooked defects that are likely to occur during chip fabrication and system integration. We address the above limitation by investigating the fault tolerance of the valve-based routing fabric. We develop a theory of failure assessment and introduce a design technique for achieving fault tolerance. Simulation results show that the proposed method leads to a slight increase in the fabrication cost and decrease in cell-analysis throughput, but this is only a small price to pay for the added assurance of fault tolerance in the new design.

15:31  15.5.6 IPS-7, 408 OPTIMIZING POWER-ACCURACY TRADE-OFF IN APPROXIMATE ADDERS
Speaker: Celia Dharmaraj, Indian Institute of Technology Madras, IN
Authors: Celia Dharmaraj, Vinita Vasudevan and Nitin Chandrachoodan, Indian Institute of Technology Madras, IN
Abstract
Approximate circuit design has gained significance in recent years targeting applications like media processing where full accuracy is not required. In this paper, we propose an approximate adder in which the approximate part of the sum is obtained by finding a single optimal level that minimizes the mean error distance. Therefore hardware needed for the approximate part computation can be removed, which effectively results in very low power consumption. We compare the proposed adder with various approximate adders in the literature in terms of power and accuracy metrics. The power savings of our adder is shown to be 17% to 55% more than power savings of the existing approximate adders over a significant range of accuracy values. Further, in an image addition application, this adder is shown to provide the best trade-off between PSNR and power.
Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

11.6 Memory: new technologies and reliability-related issues

Date: Thursday, March 22, 2018
Time: 14:00 - 15:30
Location / Room: Konf. 4

Chair:
Carles Hernandez, Barcelona Supercomputing Center (BSC), ES, Contact Carles Hernández

Co-Chair:
Shahar Kvatinisky, Technion, IL, Contact Shahar Kvatinisky

The session covers computation using emerging memory technologies, investigating techniques to protect against process variation and soft errors.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:00</td>
<td>11.6.1</td>
<td>XNOR-RRAM: A SCALABLE AND PARALLEL RESISTIVE SYNAPTIC ARCHITECTURE FOR BINARY NEURAL NETWORKS</td>
<td>Shimeng Yu, Arizona State University, CN, Xiaoyu Sun, Shihui Yin, Xiaochen Peng, Rui Liu, Jae-sun Seo and Shimeng Yu, Arizona State University, US</td>
</tr>
</tbody>
</table>

Abstract
Recent advances in deep learning have shown that Binary Neural Networks (BNNs) are capable of providing a satisfying accuracy on various image datasets with significant reduction in computation and memory cost. With both weights and activations binarized to +1 or -1 in BNNs, the high-precision multiply-and-accumulate (MAC) operations can be replaced by XNOR and bit-counting operations. In this work, we propose a RRAM synaptic architecture (XNOR-RRAM) with a bit-cell design of complementary word lines that implements equivalent XNOR and bit-counting operation in a parallel fashion. For large-scale matrices in fully connected layers or when the convolution kernels are unrolled in multiple channels, the array partition is necessary. Multi-level sense amplifiers (MLSAs) are employed as the intermediate interface for accumulating partial weighted sum. However, a low bit-level MLSA and intrinsic offset of MLSA may degrade the classification accuracy. We investigate the impact of sensing offsets on classification accuracy and analyze various design options with different sub-array sizes and sensing bit-levels. Experimental results with RRAM models and 65nm CMOS PDK show that the system with 128×128 sub-array size and 3-bit MLSA can achieve accuracies of 98.43% for MLP on MNIST and 86.08% for CNN on CIFAR-10, showing 0.34% and 2.39% degradation respectively compared to the accuracies of ideal BNN algorithms. The projected energy-efficiency of XNOR-RRAM is 141.18 TOPS/W, showing ~33X improvement compared to the conventional RRAM synaptic architecture with sequential row-by-row read-out.

Download Paper (PDF; Only available from the DATE venue WiFi)
15:15 11.6.2 A NOVEL FAULT TOLERANT CACHE ARCHITECTURE BASED ON ORTHOGONAL LATIN SQUARES THEORY

Speaker:
Kira Kraft, University of Kaiserslautern, DE; Chirag Sudarshan, Technische Universität Darmstadt, DE

Authors:
Kira Kraft1 and Chirag Sudarshan1
1University of Kaiserslautern, DE; 2Technische Universität Darmstadt, DE

Abstract
In this paper, we present a new communication theoretic channel model for Dynamic Random Access Memory (DRAM) error correction, that relies on the fully asymmetric bit-stuffing error behavior of DRAM cells. This model allows us to create a decoder with minimum overhead. The proposed decoder is able to detect and correct all single-bit errors and any number of errors in the same or different bits. The decoder is designed to be implemented in hardware using standard logic gates.

Download Paper (PDF; Only available from the DATE venue WiFi)

15:00 11.6.3 TECHNOLOGY-AWARE LOGIC SYNTHESIS FOR RERAM BASED IN-MEMORY COMPUTING

Speaker:
Debjyoti Bhattacharjee, Nanyang Technological University, SG

Authors:
Debjyoti Bhattacharjee1, Luca Aman2 and Anupam Chattopadhyay1
1Nanyang Technological University, SG; 2Synopsis, US

Abstract
Resistive RAMs (ReRAMs) have gained prominence for design of logic-in-memory circuits and architectures due to fast read/write speeds, high endurance, density and logic operation capabilities. ReRAM crossbar arrays allow constrained bit-level parallel operations. In this paper, we propose optimization techniques during logic synthesis, which are specifically targeted for leveraging the parallelism offered by ReRAM crossbar arrays. Our method uses Majority-Inverter Graph (MIG) for the internal representation of the Boolean functions. The novel optimization techniques, when applied to the MIG, exposes the bit-level parallelism, and is further coupled with an efficient technology mapping flow. The entire synthesis process is benchmarked exhaustively over large arithmetic functions using a representative ReRAM crossbar architecture, while varying the crossbar dimensions. For the hard benchmarks, we obtained 10% reduction in the number of nodes with 16% reduction in delay on average.

Download Paper (PDF; Only available from the DATE venue WiFi)

15:15 11.6.4 SMARTAG: ERROR CORRECTION IN CACHE TAG ARRAY BY EXPLOITING ADDRESS LOCALITY

Speaker:
Hamed Farbeh, School of Computer Science, Institute for Research in Fundamental Sciences (IPM), IR

Authors:
Seyedeh Golsana Ghaemi1, Iman Ahmadpour2, Mehdi Ardebili2 and Hamed Farbeh1
1Sharif University of Technology, IR; 2Sharif University of technology, IR; 3Tehran University, IR; 4School of Computer Science, Institute for Research in Fundamental Sciences (IPM), IR

Abstract
Soft errors on on-chip caches are the major cause of processors failure. Partitioning the cache into data and tag arrays, recent reports show that the vulnerability of the latter is as high as or even higher than that of the former. Although Error-Correcting Codes (ECCs) are widely used to protect the data array, their overheads are not affordable in the tag array and its protection is conventionally limited to parity code. In this paper, we propose Simultary-Managed Robust Tag (SMARTag) technique to provide the error correction capability in parity-protected tags. SMARTag exploits the inherent similarity between the upper parts of the tags in a cache set to share these parts between addresses and ECCs. Using SMARTag, the cache access time is intact since the ECC part is bypassed in normal cache operation and no extra memory is required since ECCs are stored in available tag space. The simulation results show that SMARTag is capable of correcting more than 98% of errors in the tag array, on average, and its energy consumption, area, and performance overhead is less than 0.2%.

Download Paper (PDF; Only available from the DATE venue WiFi)

15:30 11.6.5 IMPROVING THE ERROR BEHAVIOR OF DRAM BY EXPLOITING ITS Z-CHANNEL PROPERTY

Speaker:
Kira Kraft, University of Kaiserslautern, DE

Authors:
Kira Kraft1, Matthias Jung2, Chirag Sudarshan1, Deepak M. Mathew1, Christian Weiss1 and Norbert Wehn1
1University of Kaiserslautern, DE; 2Fraunhofer IESE, DE

Abstract
In this work, we propose a new communication theoretic channel model for Dynamic Random Access Memory (DRAM) error correction, that relies on the fully asymmetric retention error behavior of DRAM cells. This model allows us to create a decoder with minimum overhead. The proposed decoder is able to detect and correct all single-bit errors and any number of errors in the same or different bits. The decoder is designed to be implemented in hardware using standard logic gates.

Download Paper (PDF; Only available from the DATE venue WiFi)

15:31 11.6.6 ARCHITECTURE AND OPTIMIZATION OF ASSOCIATIVE MEMORIES USED FOR THE IMPLEMENTATION OF LOGIC FUNCTIONS BASED ON NANOELECTRONIC 1S1R CELLS

Speaker:
Ame Heittman, RWTH-Aachen University, DE

Authors:
Ame Heittman and Tobias G. Noll, RWTH Aachen University, DE

Abstract
A neuromorphic architecture based on Binary Associative memories and nanoelectronic resistive switches is proposed for the realization of arbitrary logic/arithmetic functions. Subsets of non-trivial code sets based on error detecting 2-out-of-n-codes are thoroughly used to encode operands, results, and intermediate states in order to enhance the circuit reliability by mitigating the impact of device variability. 2 ary functions can be implemented by cascading a mixer memory, a correlator memory, and a response memory. By introduction of a new cost function based on class-specific word line-coverage, stochastic optimization is applied with the aim to minimize the overall number of active amplifiers. For various exemplary functions optimized architectures are compared against solutions obtained using a standard-cost function. It is shown that the consideration of word-line-coverage results in a significant circuit compaction.

Download Paper (PDF; Only available from the DATE venue WiFi)
EXPLORING NON-VOLATILE MAIN MEMORY ARCHITECTURES FOR HANDHELD DEVICES

Speaker:
Virendra Singh, Indian Institute of Technology Bombay, IN

Authors:
Sneha Ved and Manu Awasthi, Indian Institute of Technology Gandhinagar, IN

Abstract
As additional functionality is being added to contemporary handheld devices, the SoCs inside these devices are becoming increasingly complex. Similarly, the applications executing on these handhelds are beginning to exhibit an ever increasing memory footprint. To support these trends, main memory capacity of these SoCs has been increasing over time. Due to these developments, memory system’s contribution to the overall system power has increased dramatically. Non-volatile memories have been used in server architectures to increase capacity as well as keep memory system’s power consumption in check. However, in the handheld domain, where user experience and battery life are of paramount importance, the applicability of such technologies has not been widely studied. In this paper, we propose and evaluate a number of hybrid memory architectures using mobile DRAM and PCM. We show that intelligent memory architectures, cognizant of workload’s memory access patterns can provide significant energy savings without compromising on user-experience. Using proposed approach, we can devise architectures that exhibit significant energy savings with only a 2.8% performance loss.

Download Paper (PDF; Only available from the DATE venue WiFi)

Coffee Break in Exhibition Area

Coffee Breaks in the Exhibition Area
On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Lunch Breaks (Großer Saal + Saal 1)
On all conference days (Tuesday to Thursday), a seated lunch (buffet) will be offered in the rooms “Großer Saal” and “Saal 1” (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018
- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in “Saal 2” 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018
- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- Coffee Break 15:30 - 16:00

11.7 Building Resistant Systems: From Temperature Awareness to Attack Resistance

Date: Thursday, March 22, 2018
Time: 14:00 - 15:30
Location / Room: Konf. 5

Chair:
Marina Zapater, EPFL, CH, Contact Marina Zapater Sancho

Co-Chair:
Georg Becker, ESMT Berlin, DE, Contact Georg Georg Becker

This session explores new methods in building reliable and secure systems, especially in larger SoCs. Temperature-induced stress can impact the reliability of digital systems. Temperature fluctuations can also be exploited as side channels. This session first explores the two side of this coin by discussing a novel temperature-aware chiplet placing algorithm in 2.5D systems and by showing how transmission bandwidth encoded as a temperature signal can be maximized. Then, the rest of the session highlights advances in PUFs that are resistant against lastest-generation attacks, and particularly integration of PUFs in larger systems.
14:00  11.7.1 LEVERAGING THERMALLY-AWARE CHIPLET ORGANIZATION IN 2.5D SYSTEMS TO RECLAIM DARK SILICON

**Speaker:** Yenai Ma, Boston University, US

**Authors:** Yenai Ma1, Ajay Joshi1, Andrew B. Kahng2, Yenai Ma1, Saiful Moonjamr1 and Tiansheng Zhang1

1Boston University, US; 2UCSD, US

**Abstract**

As on-chip power densities of manycore systems continue to increase, one cannot simultaneously run all the cores due to thermal constraints. This phenomenon, known as the 'dark silcon' problem, leads to inactive regions on the chip and limits the performance of manycore systems. This paper proposes to reclaim dark silicon through a thermally-aware chiplet organization technique in 2.5D manycore systems. The proposed technique adjusts the interposer size and the spacing between adjacent chiplets to reduce the peak temperature of the overall system. In this way, a system can operate with a larger number of active cores at a higher frequency without violating thermal constraints, thereby achieving higher performance. To determine the chiplet organization that jointly maximizes performance and minimizes manufacturing cost, we formulate and solve an optimization problem that considers temperature and interposer size constraints of 2.5D systems. We design a multi-start greedy approach to find near-optimal solutions efficiently. Our analysis demonstrates that by using our proposed technique, an optimized 2.5D manycore system improves performance by 41% and 16% on average and by up to 87% and 39% for temperature thresholds of 85°C and 105°C, respectively, compared to a traditional single-chip system at the same manufacturing cost. When maintaining the same performance as an equivalent single-chip system, our approach is able to reduce the system manufacturing cost by 36%.

Download Paper (PDF; Only available from the DATE venue WiFi)

14:30  11.7.2 ISING-PUF: A MACHINE LEARNING ATTACK RESISTANT PUF FEATURING LATTICE LIKE ARRANGEMENT OF ARBITER-PUFs

**Speaker:** Hiromitsu Awano, The University of Tokyo, JP

**Authors:** Hiromitsu Awano1 and Takashi Sato2

1The University of Tokyo, JP; 2Kyoto University, JP

**Abstract**

A concept of Ising-PUF, a novel PUF structure that utilizes chaotic behavior of mutually interacting small PUFs, is proposed. Ising-PUF consists of a lattice like arrangement of small PUFs, each of which contains a spin register that stores the response of the small PUF, which also serves as a challenge of its neighbors. The spin patterns that develop along time determine the 1-bit response of the Ising-PUF. Unlike state-memorizing nature of the spin registers, Ising-PUF attains a challenge hysteresis, i.e., allowing sequence of challenge inputs that continuously stimulate its chaotic behavior, which provides the drastically large challenge-to-response space. Experimental results demonstrate nearly ideal metrics; inter-chip Hamming distance (HD) of 50.1% and inter-environment HD of 2.26%. Further, Ising-PUF is remarkably tolerant to machine learning attacks, demonstrating that, even with a deep neural network using a 50k training CRPs, the prediction accuracy remains 50%, which is comparable to a random guess.

Download Paper (PDF; Only available from the DATE venue WiFi)

15:00  11.7.3 EFFICIENT HELPER DATA REDUCTION IN SRAM PUFs VIA LOSSY COMPRESSSION

**Speaker:** Ye Wang, University of Texas at Austin, US

**Authors:** Ye Wang and Michael Orshansky, University of Texas at Austin, US

**Abstract**

Fuzzy extractors used in PUF-based key generation require storage of helper data in non-volatile memory (NVM). The challenge of using SRAM PUF-based key generation on FPGAs is that high-capacity NVM, such as Flash, is not available on chip. Only expensive one-time-programmable (OTP) memory with limited capacity, such as e-fuses, can be utilized to store helper data. Our work allows a significant reduction of helper data size (HDS) through two innovative techniques. The first uses bit-error-rate (BER)-aware lossy compression: by treating a fraction of reliable bits as unreliable, it effectively reduces the size of the reliability mask. Considering practical costs of error characterization, the second technique permits across-temperature HDS minimization strategies based on bit-selection (with or without subsequent compression) using room-temperature only characterization. The method is based on stochastic concentration theory and allows efficiently forming confidence intervals for true worst-case BER. We use it to enable lossy compression and key reconstruction with success arbitrarily close to certainty. Results show that compared to maskless alternative, the proposed algorithm achieves an up to 4.5X HDS reduction with only 60% raw bits. Compared to lossless compression, we achieve a further 25% (total) HDS reduction, at the cost of doubling the number of raw PUF bits, for a 128-bit key. When bit-specific across-temperature characterization is not possible, our method achieves a significant 2.4X helper data reduction compared to the maskless alternative for extracting a 128-bit key and a 3X reduction for a 256-bit key.

Download Paper (PDF; Only available from the DATE venue WiFi)

15:15  11.7.4 IMPROVING THE EFFICIENCY OF THERMAL COVERT CHANNELS IN MULTI-/MANY-CORE SYSTEMS

**Speaker:** Zijun Long, South China University of Technology, CN

**Authors:** Zijun Long1, Xiaohang Wang1, Yingtao Jiang2, Guofeng Cui1, Yiming Zhao1, Li Zhang1 and Terrence Mark2

1South China University of Technology, CN; 2University of Nevada, Las Vegas, USA; 3University of Southampton, UK, GB

**Abstract**

In many-core chips seen in mobile computing, data center, AI, and elsewhere, thermal covert channels could be established to transmit data (e.g., passwords), supposedly to be kept secret and private. Effectiveness of a thermal covert channel, measured by its transmission rate and bit error rate (BER), is so much dependent on the thermal noise/interference imposed on the channel. In this paper, we present a few techniques to improve the capacity of thermal covert channel by overcoming the thermal interference. In particular, data in a thermal covert channel are encoded and represented following a new thermal signaling scheme where logic value, 0 or 1, modules the noise/interference imposed on the channel. Experimental results demonstrate nearly ideal metrics; inter-chip Hamming distance (HD) of 50.1% and inter-environment HD of 2.26%. Further, Ising-PUF is remarkably tolerant to machine learning attacks, demonstrating that, even with a deep neural network using a 50k training CRPs, the prediction accuracy remains 50%, which is comparable to a random guess.

Download Paper (PDF; Only available from the DATE venue WiFi)

15:30  11.7.5 ACCURATE PREDICTION OF SMARTPHONES’ SKIN TEMPERATURE BY CONSIDERING EXOTHERMIC COMPONENTS

**Speaker:** Furkan Eris, Boston University, US

**Authors:** Furkan Eris1, Xiaohang Wang1, Yingtao Jiang2, Guofeng Cui1, Yiming Zhao1, Li Zhang1 and Terrence Mark2

1University of Southampton, UK, GB

**Abstract**

Smartphones’ surface temperature, also called skin temperature, can rapidly heat up in certain cases, and this causes a variety of safety problems. Therefore, the thermal management of smartphones should consider the skin temperature, and its accurate prediction is important. However, due to the complicated relationship among the many exothermic components in the device, predicting skin temperature is extremely difficult. In this paper, we develop a thermal prediction model that accurately predicts the skin temperature of a mobile device. In an experiment with smartphones, we show that the proposed model achieves an accuracy of 98%, with a ±0.4 °C margin of error. To the best of our knowledge, our work is the first to reveal the complex relationship between the various components inside of a smartphone and its skin temperature.

Download Paper (PDF; Only available from the DATE venue WiFi)
In the program of the technical conference, on Thursday is the Special Day for “Autonomous Driving”. In the Exhibition Theatre this will be complemented by a workshop on how to design chips and electronic systems fulfilling the functional safety requirements of devices used in the car. This a challenge which is absolutely key for the automotive industry - no autonomous driving without functional safety, no matter how cool the driving algorithms might be.

The Internet of Things (IoT) is envisaged to consist of billions of connected devices coupled with sensors which generate huge volumes of data enabling control-and-command in this paradigm. However, integrity of this data is of utmost concern, and is prominently addressed leveraging the inherent unreliability of Physically Unclonable Functions (PUFs) w.r.t. ambient parameter variations, using the concept of Virtual Proofs (VPs). Advantage of these protocols is that they do not use explicit keys and aim at proving the authenticity of the sensor. Since the existing PUF-based protocols do not use the sensor data as a part of challenge (i.e. input) to PUFs, there is no guarantee of uniqueness of PUFs challenge-response behavior over multiple levels of ambient parameters. Few of these protocols needs to sequential search in the challenge-response database. To alleviate these issues, we develop a new class of authenticated sensing protocols where the sensor data is combined with the external challenge by utilizing the Strict Avalanche Criterion of the PUF. We validate the proposed protocol through actual experiments on FPGA using Double Arbiter PUFs (DAPUFs), which are implemented with superior uniformity, uniqueness, and reliability on Xilinx Artix-7 FPGAs. According to the FPGA-based validation, the proposed protocol with DAPUF can be effectively used to authenticate wide variations of temperature from −20◦C to 80◦C.

True random number generators (TRNGs) are fundamental constituents of secure embedded cryptographic systems. In this paper, we introduce a general methodology for porting TRNG across different FPGA vendor families. In order to demonstrate our methodology, we applied it to the delay-chain based TRNG (DC-TRNG) on Intel Cyclone IV and Cyclone V FPGAs. We examine vendor-agnostic generality of the underlying DC-TRNG principle and propose modifications to address differences in structure of FPGAs. Implementation of the DC-TRNG on Cyclone IV uses 149 LEs (<0.1% of available resources) and has a throughput of 5Mbps, while on Cyclone V it occupies 230 ALMs (<1.5% of resources) with an output rate of 12.5 Mbps. The quality of the random bits produced by the DC-TRNG on Intel Cyclone IV and V is further confirmed by using NIST statistical test suite.

The Internet of Things (IoT) is envisaged to consist of billions of connected devices coupled with sensors which generate huge volumes of data enabling control-and-command in this paradigm. However, integrity of this data is of utmost concern, and is prominently addressed leveraging the inherent unreliability of Physically Unclonable Functions (PUFs) w.r.t. ambient parameter variations, using the concept of Virtual Proofs (VPs). Advantage of these protocols is that they do not use explicit keys and aim at proving the authenticity of the sensor. Since the existing PUF-based protocols do not use the sensor data as a part of challenge (i.e. input) to PUFs, there is no guarantee of uniqueness of PUFs challenge-response behavior over multiple levels of ambient parameters. Few of these protocols needs to sequential search in the challenge-response database. To alleviate these issues, we develop a new class of authenticated sensing protocols where the sensor data is combined with the external challenge by utilizing the Strict Avalanche Criterion of the PUF. We validate the proposed protocol through actual experiments on FPGA using Double Arbiter PUFs (DAPUFs), which are implemented with superior uniformity, uniqueness, and reliability on Xilinx Artix-7 FPGAs. According to the FPGA-based validation, the proposed protocol with DAPUF can be effectively used to authenticate wide variations of temperature from −20◦C to 80◦C.

The Internet of Things (IoT) is envisaged to consist of billions of connected devices coupled with sensors which generate huge volumes of data enabling control-and-command in this paradigm. However, integrity of this data is of utmost concern, and is prominently addressed leveraging the inherent unreliability of Physically Unclonable Functions (PUFs) w.r.t. ambient parameter variations, using the concept of Virtual Proofs (VPs). Advantage of these protocols is that they do not use explicit keys and aim at proving the authenticity of the sensor. Since the existing PUF-based protocols do not use the sensor data as a part of challenge (i.e. input) to PUFs, there is no guarantee of uniqueness of PUFs challenge-response behavior over multiple levels of ambient parameters. Few of these protocols needs to sequential search in the challenge-response database. To alleviate these issues, we develop a new class of authenticated sensing protocols where the sensor data is combined with the external challenge by utilizing the Strict Avalanche Criterion of the PUF. We validate the proposed protocol through actual experiments on FPGA using Double Arbiter PUFs (DAPUFs), which are implemented with superior uniformity, uniqueness, and reliability on Xilinx Artix-7 FPGAs. According to the FPGA-based validation, the proposed protocol with DAPUF can be effectively used to authenticate wide variations of temperature from −20◦C to 80◦C.

The Internet of Things (IoT) is envisaged to consist of billions of connected devices coupled with sensors which generate huge volumes of data enabling control-and-command in this paradigm. However, integrity of this data is of utmost concern, and is prominently addressed leveraging the inherent unreliability of Physically Unclonable Functions (PUFs) w.r.t. ambient parameter variations, using the concept of Virtual Proofs (VPs). Advantage of these protocols is that they do not use explicit keys and aim at proving the authenticity of the sensor. Since the existing PUF-based protocols do not use the sensor data as a part of challenge (i.e. input) to PUFs, there is no guarantee of uniqueness of PUFs challenge-response behavior over multiple levels of ambient parameters. Few of these protocols needs to sequential search in the challenge-response database. To alleviate these issues, we develop a new class of authenticated sensing protocols where the sensor data is combined with the external challenge by utilizing the Strict Avalanche Criterion of the PUF. We validate the proposed protocol through actual experiments on FPGA using Double Arbiter PUFs (DAPUFs), which are implemented with superior uniformity, uniqueness, and reliability on Xilinx Artix-7 FPGAs. According to the FPGA-based validation, the proposed protocol with DAPUF can be effectively used to authenticate wide variations of temperature from −20◦C to 80◦C.
INTRODUCTION

Speaker: Dirk Hansen, Mentor, DE

Abstract: As semiconductor value in a modern car expands, reliability and safety of electronics must improve dramatically. If simple electronics in Bluetooth and power seats cause the most problems in cars today, as indicated by various reliability and dependability surveys, how are we going to make the shift to much more complex electronics systems that are needed for self-driving cars? It will become imperative to improve the quality of semiconductors going forward and we must get much better in verifying and validating these complex automotive systems knowing that lives will be at risk with autonomous driving. This will increase the test cycles, visibility and coverage to improve the safety and reliability.

In this session, we will present technologies and methodologies allowing us to handle an explosion of test scenarios to verify electronics and algorithms of driverless cars. We will explain a mature Development Process and show how Requirement driven development provides proof that design was built and tested as intended.

IC VERIFICATION: SHIFT-LEFT THE PATH TO ISO 26262 COMPLIANCE FOR DIGITAL IC DEVELOPMENT

Speaker: Dirk Hansen, Mentor, DE

Abstract: If you are developing IP or semiconductors targeting ADAS or autonomous driving, you must develop in accordance with ISO 26262 to ensure safety of your products. The challenge is that this imposes additional development practices, flows and verification needs beyond your normal IC development. In this presentation we will provide an overview of Mentor's solution, both today and tomorrow, to address the functional safety needs for IC development (focused primarily on the digital side) and how we are helping customers "shift-left" their path to compliance. We will also discuss why Mentor + Siemens is the perfect match to address these automotive challenges.

DFT PART: TEST SOLUTIONS FOR THE AUTOMOTIVE MARKET

Speaker: Ralph Sommer, Mentor, DE

Abstract: The amount of electronic content in passenger cars continues to grow rapidly, driven largely by the integration of various ADAS and autonomous driving capabilities. It is of course critical that these devices adhere to the highest possible quality and reliability requirements. Meeting the functional safety requirements mandated by the ISO 26262 standard requires the integration of advanced self-test and monitoring capabilities throughout the vehicle's electronics. The capabilities must not only have the ability to fully test all electronics during power-up, but more importantly must provide the ability to perform periodic tests throughout the functional operation of the vehicle. The Mentor Tessent product family offers a new generation of test solutions to address these evolving challenges.

Coffee Break in Exhibition Area

On all conference days (Tuesday to Thursday), coffee and tea will be served during the coffee breaks at the below-mentioned times in the exhibition area (Terrace Level of the ICCD).

Coffee Breaks in the Exhibition Area

Lunch Breaks (Großer Saal + Saal 1)

On all conference days (Tuesday to Thursday), a seated lunch (lunch buffet) will be offered in the rooms "Großer Saal" and "Saal 1" (Saal Level of the ICCD) to fully registered conference delegates only. There will be badge control at the entrance to the lunch break area.

Tuesday, March 20, 2018

- Coffee Break 10:30 - 11:30
- Lunch Break 13:00 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:50 - 14:20
- Coffee Break 16:00 - 17:00

Wednesday, March 21, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:30
- Awards Presentation and Keynote Lecture in "Saal 2" 13:30 - 14:20
- Coffee Break 16:00 - 17:00

Thursday, March 22, 2018

- Coffee Break 10:00 - 11:00
- Lunch Break 12:30 - 14:00
- Coffee Break 15:30 - 16:00

UB11 Session 11

Date: Thursday, March 22, 2018
Time: 14:30 - 16:30
Location / Room: Booth 1, Exhibition Area
**UB11.1 ARCHON: AN ARCHITECTURE-OPEN RESOURCE-DRIVEN CROSS-LAYER MODELLING FRAMEWORK**

**Authors:**
Fei Xia¹, Ashur Rafiei², Mohammad Al-Hayam², Alexei Ilarion¹, Rishad Shatik¹, Alexander Romanovsky¹ and Alex Yakovlev¹

¹Newcastle University, GB; ²Newcastle University, UK and University of Technology and HCED, IQ

**Abstract**
This demonstration showcases a modeling method for large complex computing systems focusing on many-core types and concentrating on the crosslayer aspects. The resource-driven models aim to help system designers reason about, analyse, and ultimately design such systems across all conventional computing and communication layers, from application, operating system, down to the finest hardware details. The framework and tool support the notion of selective abstraction and are suitable for studying such non-functional properties such as performance, reliability and energy consumption.

[More information ...]

**UB11.2 OSC MULTICORE STENCIL PROCESSOR: ONE INSTRUCTION-SET COMPUTER-BASED MULTICORE PROCESSOR FOR STENCIL COMPUTING**

**Authors:**
Kaoru Saso, Jing Yuan Zhao and Yuki Hara-Azumi, School of Engineering, Tokyo Institute of Technology, JP

**Abstract**
Subtract and branch on NE/Gatew with 4 operands (SUBNEG4) is one of One Instruction-Set Computers that execute only one type of instruction. Thanks to its simplicity, SUBNEG4 has only 1/20x circuit area and 1/10x power consumption against MIPS processor. As SUBNEG4 is Turing-complete, it is suitable for parallel computing by multiple cores, while keeping its low-power feature. Our on-going project is seeking for effective use and deployment of SUBNEG4 cores on embedded systems. Our booth will demonstrate the significant speed-up by a SUBNEG4-based many-core processor against a conventional processor, for stencil computing. Our 64-core processor efficiently handles 2D von-Neumann neighborhood stencils, e.g., wave simulation by Verlet integration and 2D Jacobi iteration, to compute 64 points simultaneously. We show that small many-core processors can be realized even with such large number of cores while achieving good speed-up for heavy computation.

[More information ...]

**UB11.3 OISP SPANNER: SELF-REPAIRING SPIKING NEURAL NETWORK CONTROLLER FOR AN AUTONOMOUS ROBOT**

**Authors:**
Alan Millard¹, Anju Johnson¹, James Hider¹, David Halliday¹, Andy Tyrell¹, Jon Timmis¹, Junxu Liu², Shwan Karr², Jim Hankin² and Liam McDaid²

¹University of York, GB; ²Ryder University, GB

**Abstract**
The human brain is remarkably resilient, and is able to self-repair following injury or a stroke. In contrast, electronic systems typically exhibit limited self-repair capabilities, and cannot recover from faults. We demonstrate a bio-inspired approach to self-repair that allows an autonomous robot to recover from faults in its artificial brain. Astrocytes are support cells in the human brain that interact with neurons to regulate synaptic activity. We have modelled this interaction to create a spiking neural network that can self-repair when synapses between neurons are damaged, by strengthening redundant pathways. We demonstrate a robot platform controlled by a self-repairing spiking neural network that is implemented on an FPGA. We demonstrate that injecting faults into the synapses of the network initially causes the robot to behave erratically, but that the neural controller is able to automatically repair itself, thus allowing the robot to resume normal function.

[More information ...]

**UB11.4 USING FORMAL METHODS FOR AUTOMATIC PLATFORM-INDEPENDENT CODE GENERATION OF RUN-TIME MANAGEMENT**

**Authors:**
Mohammadseadeh Dalvandi, Michael Butler and Ashel Salehi Fathabadi, University of Southampton, GB

**Abstract**
Run-Time Management (RTM) systems are used in embedded systems to dynamically adapt hardware performance to minimise energy consumption. In this demonstration, we present a framework for automatic generation of RTM implementations from platform-independent formal models. The methodology in designing the RTM systems uses a high-level mathematical language, Event-B, which can describe systems at different abstraction levels. A code generation tool is used to translate platform-independent Event-B RTM models to platform-specific implementations in C. Formal verification is used to ensure correctness of the Event-B models. The portability offered by our methodology is demonstrated by modelling a Reinforcement Learning (RL) based RTM and generating implementations for two different platforms that allow different energy savings on the respective platforms. The generated RTM code has been integrated with the PRME framework, a cross-layer framework for embedded power management.

[More information ...]

**UB11.5 ABSYNTH: A COMPREHENSIVE APPROACH TO FRONT TO BACK ANALOG BLOCK DESIGN AUTOMATION**

**Authors:**
Abhaya Chandra Kammara S., ¹, Sidney Pontes-Filho² and Andreas König²

¹ISE, TU Kaiserslautern, DE; ²University of Kaiserslautern, DE

**Abstract**
ABSYNTH was first presented in CEBIT 2014 where complete, practical circuit sizing approaches have been shown using meta-heuristics on trusted simulators. This tool was then proven by its use in design of several circuits. Here, we present the extension to our nested optimization approach that creates a symmetric and well matched layout in every step for every instance in the population of the swarm, that is extracted in our flow to provide feedback to the cost function impacting on the population update for more viable and robust circuits. The layout optimization presented in this DEMO works with Cadence Layout design tools. Our initial focus is, motivated by Industry 4.0, IoT, on cells layout in every step for every instance in the population of the swarm, that is extracted in our flow to provide feedback to the cost function impacting on the population update for.

[More information ...]

**UB11.6 OISP SPANNER: SELF-REPAIRING SPIKING NEURAL NETWORK CONTROLLER FOR AN AUTONOMOUS ROBOT**

**Authors:**
Alan Millard¹, Anju Johnson¹, James Hider¹, David Halliday¹, Andy Tyrell¹, Jon Timmis¹, Junxu Liu², Shwan Karr², Jim Hankin² and Liam McDaid²

¹University of York, GB; ²Ryder University, GB

**Abstract**
The human brain is remarkably resilient, and is able to self-repair following injury or a stroke. In contrast, electronic systems typically exhibit limited self-repair capabilities, and cannot recover from faults. We demonstrate a bio-inspired approach to self-repair that allows an autonomous robot to recover from faults in its artificial brain. Astrocytes are support cells in the human brain that interact with neurons to regulate synaptic activity. We have modelled this interaction to create a spiking neural network that can self-repair when synapses between neurons are damaged, by strengthening redundant pathways. We demonstrate a robot platform controlled by a self-repairing spiking neural network that is implemented on an FPGA. We demonstrate that injecting faults into the synapses of the network initially causes the robot to behave erratically, but that the neural controller is able to automatically repair itself, thus allowing the robot to resume normal function.

[More information ...]

**UB11.7 USING FORMAL METHODS FOR AUTOMATIC PLATFORM-INDEPENDENT CODE GENERATION OF RUN-TIME MANAGEMENT**

**Authors:**
Mohammadseadeh Dalvandi, Michael Butler and Ashel Salehi Fathabadi, University of Southampton, GB

**Abstract**
Run-Time Management (RTM) systems are used in embedded systems to dynamically adapt hardware performance to minimise energy consumption. In this demonstration, we present a framework for automatic generation of RTM implementations from platform-independent formal models. The methodology in designing the RTM systems uses a high-level mathematical language, Event-B, which can describe systems at different abstraction levels. A code generation tool is used to translate platform-independent Event-B RTM models to platform-specific implementations in C. Formal verification is used to ensure correctness of the Event-B models. The portability offered by our methodology is demonstrated by modelling a Reinforcement Learning (RL) based RTM and generating implementations for two different platforms that allow different energy savings on the respective platforms. The generated RTM code has been integrated with the PRME framework, a cross-layer framework for embedded power management.

[More information ...]

**UB11.8 OISP SPANNER: SELF-REPAIRING SPIKING NEURAL NETWORK CONTROLLER FOR AN AUTONOMOUS ROBOT**

**Authors:**
Alan Millard¹, Anju Johnson¹, James Hider¹, David Halliday¹, Andy Tyrell¹, Jon Timmis¹, Junxu Liu², Shwan Karr², Jim Hankin² and Liam McDaid²

¹University of York, GB; ²Ryder University, GB

**Abstract**
The human brain is remarkably resilient, and is able to self-repair following injury or a stroke. In contrast, electronic systems typically exhibit limited self-repair capabilities, and cannot recover from faults. We demonstrate a bio-inspired approach to self-repair that allows an autonomous robot to recover from faults in its artificial brain. Astrocytes are support cells in the human brain that interact with neurons to regulate synaptic activity. We have modelled this interaction to create a spiking neural network that can self-repair when synapses between neurons are damaged, by strengthening redundant pathways. We demonstrate a robot platform controlled by a self-repairing spiking neural network that is implemented on an FPGA. We demonstrate that injecting faults into the synapses of the network initially causes the robot to behave erratically, but that the neural controller is able to automatically repair itself, thus allowing the robot to resume normal function.

[More information ...]

**UB11.9 ABSYNTH: A COMPREHENSIVE APPROACH TO FRONT TO BACK ANALOG BLOCK DESIGN AUTOMATION**

**Authors:**
Abhaya Chandra Kammara S., ¹, Sidney Pontes-Filho² and Andreas König²

¹ISE, TU Kaiserslautern, DE; ²University of Kaiserslautern, DE

**Abstract**
ABSYNTH was first presented in CEBIT 2014 where complete, practical circuit sizing approaches have been shown using meta-heuristics on trusted simulators. This tool was then proven by its use in design of several circuits. Here, we present the extension to our nested optimization approach that creates a symmetric and well matched layout in every step for every instance in the population of the swarm, that is extracted in our flow to provide feedback to the cost function impacting on the population update for more viable and robust circuits. The layout optimization presented in this DEMO works with Cadence Layout design tools. Our initial focus is, motivated by Industry 4.0, IoT, on cells layout in every step for every instance in the population of the swarm, that is extracted in our flow to provide feedback to the cost function impacting on the population update for.

[More information ...]

**16:30 End of session**

**IP5 Interactive Presentations**

**Date:** Thursday, March 22, 2018
**Time:** 15:30 - 16:00
**Location / Room:** Conference Level, Foyer

Interactive Presentations run simultaneously during a 30-minute slot. Additionally, each IP paper is briefly introduced in a one minute presentation in a corresponding regular session

**Label** | **Presentation Title** | **Authors**
--- | --- | ---
IP5-1 | A PLACEMENT ALGORITHM FOR SUPERCONDUCTING LOGIC CIRCUITS BASED ON CELL GROUPING AND SUPER-CELL PLACEMENT | Massoud Pedram, University of Southern California, US

**Authors:**
Schei Nazar Shahsavani, Arefza Shahsaei Bejestan and Massoud Pedram, University of Southern California, US

**Abstract**
This paper presents a novel clustering based placement algorithm for single flux quantum (SFQ) family of superconductive electronic circuits. In these circuits nearly all cells receive a clock signal and a placement algorithm that ignores the clock routing cost will not produce high quality solutions. To address this issue, proposed approach simultaneously minimizes the total wirelength of the signal nets and area overhead of the clock routing. Furthermore, construction of a perfect H-tree in SFQ logic circuits is not viable solution due to the resulting very high routing overhead and the in-feasibility of building exact zero-skew clock routing trees. Instead a hybrid clock tree must be used whereby higher levels of the clock tree (i.e., those closer to the clock source) are based on H-tree construction whereas lower levels of the clock tree follow a linear (i.e., chain-like) structure. The proposed approach is able to reduce the overall half perimeter wirelength by 15% and area by 6% compared with state-of-the-art techniques.

[Download Paper (PDF; Only available from the DATE venue WiFi)]
IP5-2
ABAX: 2D/3D LEGALISER SUPPORTING LOOK-AHEAD LEGALISATION AND BLOCKAGE STRATEGIES

Speaker:
Nikolaos Sketopoulou, University of Thessaly, GR

Authors:
Nikolaos Sketopoulou, Christos Sotiropoulos, and Stavros Simoglou, Department of Electrical and Computer Engineering, University of Thessaly, GR

Abstract
Abax is a 2D/3D incremental version of the classical Abacus, minimum displacement, greedy legaliser. Abax supports single-tier 2D or 3D legalisation for multiple, logic-on-logic 3D IC layers, efficient look-ahead legalisation of intermediate Global Placement (GP) iterations, Hard Macros, Blockages, row density constraints and multiple local cell displacement functions and cell orderings. For 3D IC, Abax can produce multi-tier 3D IC placements by performing Legalisation-based Partitioning. For efficient Look-ahead Legalisation, Abax supports two new local placement cost functions, multi-cell mean and multi-cell total. We show that the classical single-cell placement and multi-cell total can result in artifacts when legalising early intermediate GPs, and that multi-cell mean is the best candidate for Look-ahead Legalisation. Obstructions, i.e. Hard Macros and Blockages are handled by using two strategies. We present legalisation results for the I3P2014 and I3P2015 benchmarks, by using GP generated from Eh?Placer, and HPWL measurement by using RippleDP. For 3D, two-tier legalisation we illustrate a ~30% reduction in HPWL for a set of I3P2014 benchmarks. For 2D legalisation on the ISP2015 benchmarks, our average HPWL increase over GP is 3.03%, compared to 7.21% of the Eh?Placer legaliser, and 43.16% of the RippleDP legaliser.

Download Paper (PDF; Only available from the DATE venue WiFi)

IP5-3
LESAR: A DYNAMIC LINE-END SPACING AWARE DETAIL ROUTER

Speaker:
Yin-Lang Li, Computer Science Department, NCTU, TH

Authors:
Ying-Chi Wei, Radhamanjari Samanta and Yin-Lang Li, National Chiao-Tung University, TW

Abstract
As the VLSI technology scales down, 193nm optical lithography reaches the limit and one-dimensional (1D) unidirectional style lithography technique emerges as one of the most promising solutions for coming advanced technology nodes. The 1D process flow generates unidirectional dense metal lines and then use line-end cutting to form the target patterns with cut masks. If cuts are too close, they will lead to conflicts. Line-end spacing rules become dynamic rather than static because of cut mask and also now need to be followed strictly. Line-end spacing check between two line-end pairs in the same mask has also been regarded as compulsory line-end spacing constraints that have not discussed in previous works yet. Complying with these rules during APR has become a new bottleneck. In this work, we propose to make the router aware of the dynamic line-end spacing rules, including end-end spacing and parity spacing constraints. Experimental results of our proposed router demonstrates that it can effectively expel all end-end spacing violations as well as 75% of parity spacing violations in a reasonable runtime increase of 14%.

Download Paper (PDF; Only available from the DATE venue WiFi)

IP5-4
UNDERSTANDING TURN MODELS FOR ADAPTIVE ROUTING: THE MODULAR APPROACH

Speaker:
Eduardo Fusella, Department of Electrical Engineering and Information Technologies, University of Naples Federico II, IT

Authors:
Eduardo Fusella and Alessandro Glerdo, University of Naples Federico II, IT

Abstract
Routing algorithms were extensively studied first in multi-computer systems, then in multi- and many-core architectures. Among the commonly used routing techniques, the turn model seems the most promising solution when targeting adaptiveness. Based on the turn model, several alternative approaches with different turn prohibition schemes were proposed. This paper gives a new theoretical background for designing deadlock-free partially adaptive logic-based distributed routing algorithms that are based on the turn model. Two properties are presented, including a necessary and sufficient condition to prove that a routing algorithm is deadlock-free as long as turn restrictions follow a modular distribution. Existing approaches can be considered a subset of the solution space identified by this work. Finally, we propose a novel routing algorithm exhibiting encouraging performance improvements over state-of-the-art approaches.

Download Paper (PDF; Only available from the DATE venue WiFi)

IP5-5
QUATER-IMAGINARY BASE FOR COMPLEX NUMBER ARITHMETIC CIRCUITS

Speaker:
Souradip Sarkar, Nokia Bell Labs, BE

Authors:
Souradip Sarkar and Manil Dev Gomony, Nokia Bell Labs, BE

Abstract
Arithmetic operations involving complex numbers are widely used in the signal processing functions in the physical layer of modern wireless and wireline communication systems, electronic instrumentation and control systems. With the ever increasing throughput requirements of such systems, the power consumption of the hardware realization is increasing beyond the allowed budget. Arithmetic circuits based on binary numeral system that have been optimized rigorously over the past few decades are currently being used for the computation involving complex numbers. In this paper, we present the potential of arithmetic circuits for complex number computations based on the Quater-imaginary (QI) base numeral system to reduce power consumption. We show that for a simple multiplier implementation in the QI base, the savings in power and area consumption could be up to 40% when synthesized in 28nm TSMC standard cell technology node.

Download Paper (PDF; Only available from the DATE venue WiFi)

IP5-6
FAULT-TOLERANT VALVE-BASED MICROFLUIDIC ROUTING FABRIC FOR DROPLET BARCODING IN SINGLE-CELL ANALYSIS

Speaker:
Yasamin Moradi, Technical University of Munich (TUM), DE

Authors:
Yasamin Moradi1, Mohamed Ibrahim2, Krishnendu Chakrabarty2 and Ulf Schlichtmann1
1Technical University of Munich, DE; 2Duke University, US

Abstract
High throughput single-cell genomics is used to gain insights into diseases such as cancer. Motivated by this important application, microfluidics has emerged as a key technology for developing comprehensive biochemical procedures for studying DNA, RNA, proteins, and many other cellular components. Recently, a hybrid microfluidic platform has been proposed to efficiently automate the analysis of a heterogeneous sequence of cells. In this design, a valve-based routing fabric based on transposers is used to label/barcode the target cells. However, the design proposed in prior work overlooked defects that are likely to occur during chip fabrication and system integration. We address the above limitation by investigating the fault tolerance of the valve-based routing fabric. We develop a theory of failure assessment and introduce a design technique for achieving fault tolerance. Simulation results show that the proposed method leads to a slight increase in the fabrication size and decrease in cell-analysis throughput, but this is only a small price to pay for the added assurance of fault tolerance in the new design.

Download Paper (PDF; Only available from the DATE venue WiFi)

IP5-7
OPTIMIZING POWER-ACCURACY TRADE-OFF IN APPROXIMATE ADDERS

Speaker:
Celia Dhamara, Indian Institute of Technology Madras, IN

Authors:
Celia Dhamara, Vinuta Vasudevan and Nithin Chandranchoodan, Indian Institute of Technology Madras, IN

Abstract
Approximate circuit design has gained significance in recent years targeting applications like media processing where full accuracy is not required. In this paper, we propose an approximate adder in which the approximate part of the sum is obtained by finding a single optimal angle that minimizes the mean error distance. Therefore hardware needed for the approximate part computation can be removed, which effectively results in very low power consumption. We compare the proposed adder with various approximate adders in the literature in terms of power and accuracy metrics. The power savings of our adder is shown to be 17% to 55% more than power savings of the existing approximate adders over a significant range of accuracy values. Further, in an image addition application, this adder is shown to provide the best trade-off between PSNR and power.

Download Paper (PDF; Only available from the DATE venue WiFi)
TOWARDS FULLY AUTOMATED TLM-TO-RTL PROPERTY REFINEMENT

Speaker: Vladimir Herdt1, Hoang M. Le1, Daniel Grosse2 and Rolf Drechsler2
Authors: Vladimir Herdt1, University of Bremen, DE; Hoang M. Le1, Daniel Grosse2 and Rolf Drechsler2, University of Bremen/DFKI GmbH, DE

Abstract
An ESL design flow starts with a TLM description, which is thoroughly verified and then refined to a RTL description in subsequent steps. The properties used for TLM verification are refined alongside the TLM description to serve as starting point for RTL property checking. However, a manual transformation of properties from TLM to RTL is error prone and time consuming. Therefore, in this paper we propose a fully automated TLM-to-RTL property refinement based on a symbolic analysis of transactors. We demonstrate the applicability of our property refinement approach using a case study.

Download Paper (PDF; Only available from the DATE venue WiFi)
Non-intrusive Testing Technique for Detection of Trojans in Asynchronous Circuits

Speaker: Rodrigo Possamai Bastos, TIMA Laboratory, CNRS/Grenoble INP/ULF, FR

Authors: Leonel Acunha Guimardes, Thiago Ferreira de Paiva Leite, Rodrigo Possamai Bastos and Laurent Fesquet, TIMA - Grenoble Institute of Technology, FR

Abstract: Asynchronous circuits, as any IC, are vulnerable to hardware Trojans (HTs), which might be maliciously implanted in IC designs during outsourced fabrication phases. In this paper, a new testing technique to detect HTs by exploiting the regular side-channel properties of quasi-delay insensitive (QDI) asynchronous circuits is proposed. The technique does not need neither additional circuitry nor significant adjustments in the post-fabrication testing phase. Simulation results show that the proposed technique is able to detect HTs with dimensions smaller than 1% of the original circuit.

Download Paper (PDF; Only available from the DATE venue WiFi)

Towards Inter-Vendor Compatibility of True Random Number Generators for FPGAs

Speaker: Milos Giuji, imec-COSIC, KU Leuven, BE

Authors: Milos Giuji, Bohan Yang, Vladimir Rozic and Ingrid Verbauwhede, imec-COSIC, KU Leuven, BE

Abstract: True random number generators (TRNGs) are fundamental constituents of secure embedded cryptographic systems. In this paper, we introduce a general methodology for porting TRNG across different FPGA vendor families. In order to demonstrate our methodology, we applied it to the delay-chain based TRNG (DC-TRNG) on Intel Cyclone IV and Cyclone V FPGAs. We examine vendor-agnostic generality of the underlying DC-TRNG principle and propose modifications to address differences in structure of FPGAs. Implementation of the DC-TRNG on Cyclone IV uses 149 LEs (<0.1% of available resources) and has a throughput of 5Mbps, while on Cyclone V it occupies 230 ALMs (<1.5% of resources) with an output rate of 12.5 Mbps. The quality of the random bits produced by the DC-TRNG on Intel Cyclone IV and V is further confirmed by using NIST statistical test suite.

Download Paper (PDF; Only available from the DATE venue WiFi)

Efficient Wear Leveling for Inodes of File Systems on Persistent Memories

Speaker: Xianzhang Chen, Chongqing University, CN

Authors: Xianzhang Chen¹, Edwin Sha², Yuansong Zang², Chaoshu Yang³, Weiken Jiang³ and Qingfeng Zhuge³

¹Chongqing University, CN; ²Chongqing University, US; ³East China Normal University, CN

Abstract: Existing persistent memory file systems achieve high-performance file accesses by exploiting advanced characteristics of persistent memories (PMs), such as PCM. However, they ignore the limited endurance of PMs. Particularly, the frequently updated inodes are stored on fixed locations throughout their lifetime, which can easily damage PM with common file operations. To address such issues, we propose a new mechanism, Virtualized Inode (VInode), for the wear leveling of inodes of persistent memory file systems. In VInode, we develop an algorithm called Pages as Communicating Vessels (PCV) to efficiently find and migrate the heavily written inodes. We implement VInode in SIMFS, a typical persistent memory file system. Experiments are conducted with well-known benchmarks. Compared with original SIMFS, experimental results show that VInode can reduce the maximum value and standard deviation of the write counts of pages to 1800x and 6200x lower, respectively.

Download Paper (PDF; Only available from the DATE venue WiFi)

Exploring Non-Volatile Main Memory Architectures for Handheld Devices

Speaker: Virenda Singh, Indian Institute of Technology Bombay, IN

Authors: Sneha Ved and Manu Awasthi, Indian Institute of Technology Gandhinagar, IN

Abstract: As additional functionality is being added to contemporary handheld devices, the SoCs inside these devices are becoming increasingly complex. Similarly, the applications executing on these handhelds are beginning to exhibit an ever increasing memory footprint. To support these trends, main memory capacity of these SoCs has been increasing over time. Due to these developments, memory system’s contribution to the overall system power has increased dramatically. Non-volatile memories have been used in server architectures to increase capacity as well as keep memory system’s power consumption in check. However, in the handheld domain, where user experience and battery life are of paramount importance, the applicability of such technologies has not been widely studied. In this paper, we propose and evaluate a number of hybrid memory architectures using mobile DRAM and PCM. We show that intelligent memory architectures, cognizant of workloads’ memory access patterns can provide significant energy savings without compromising on user experience. Using proposed approach, we can devise architectures that exhibit significant energy savings with only a 2.8% performance loss.

Download Paper (PDF; Only available from the DATE venue WiFi)

12.1 Special Day Session on Designing Autonomous Systems: Self-awareness for Autonomous Systems

Date: Thursday, March 22, 2018
Time: 16:00 - 17:30
Location / Room: Saal 2

Chair: Nikil Dutt, University of California at Irvine, US, Contact Nikil Dutt

With increasing interest in the deployment of autonomous vehicles and robots, a critical open challenge is to empower these systems with self-awareness for achieving truly autonomous operation. Self-awareness principles hold the promise to manage effectively continuous change and evolution, application interference, environment dynamics and system uncertainty, thereby adhering to safety, availability, and security guarantees as needed. The goal of this special session is to make the audience appreciate the benefits of self-awareness for systems autonomy, highlight challenges for self-awareness using two application contexts (unmanned aerial systems and autonomous vehicles), and outline EDA and HW/SW challenges to support self-awareness for systems autonomy.

Download Paper (PDF; Only available from the DATE venue WiFi)
This session presents innovative solutions to identify challenging situations in RTL designs. The first work presents a fully automated and scalable approach for concolic generation of direct test generation method that can efficiently generate a test to activate a given target. Our experimental results demonstrate that our approach is both efficient and scalable.

Existing approaches are tuned for improving overall coverage, rather than covering a specific target. We developed a Control Flow Graph (CFG) assisted directed test approach for generating directed tests using concolic testing of RTL models. While application of concolic testing on hardware designs has shown some promising results, scenarios. While formal methods are promising in such cases, it is infeasible to apply them on large designs. In this paper, we propose a fully automated and scalable approach to overcome from the perspective of the German Research Priority Program "Cooperative Interactive Automobiles".

This paper deals with challenges and possible solutions for incorporating self-awareness principles in EDA design flows for autonomous systems. We present a holistic approach that enables self-awareness across the software/hardware stack, from systems-on-chip to systems-of-systems (autonomous car) contexts. We use the Information Processing Factory (IPF) metaphor as an exemplar to show how self-awareness can be achieved across multiple abstraction levels, and discuss new research challenges. The IPF approach represents a paradigm shift in platform design by envisioning the move towards a consequent platform-centric design in which the combination of self-organizing learning and formal reactive methods guarantee the applicability of such cyber-physical systems in safety-critical and high-availability applications.

17:30 12.1.3 DESIGN METHODOLOGIES FOR ENABLING SELF-AWARENESS IN AUTONOMOUS SYSTEMS

Speaker:
Andreas Herkersdorf, Technical University Munich (TUM), DE

Authors:
Amin Sadigh1, Bryan Donyanavard, Thawra Kadeed, Kasra Moazzem, Tiago Muck, Ahmed Nassar, Amir Rahmani, Thomas Wiel1, Nik Dutt1, Rolf Ernst, Andreas Herkersdorf and Fadi Kurdahi

Abstract
This paper deals with challenges and possible solutions for incorporating self-awareness principles in EDA design flows for autonomous systems. We present a holistic approach that enables self-awareness across the software/hardware stack, from systems-on-chip to systems-of-systems (autonomous car) contexts. We use the Information Processing Factory (IPF) metaphor as an exemplar to show how self-awareness can be achieved across multiple abstraction levels, and discuss new research challenges. The IPF approach represents a paradigm shift in platform design by envisioning the move towards a consequent platform-centric design in which the combination of self-organizing learning and formal reactive methods guarantee the applicability of such cyber-physical systems in safety-critical and high-availability applications.

Download Paper (PDF; Only available from the DATE venue WiFi)
16:30 12.2.2 IMPROVING AND EXTENDING THE ALGEBRAIC APPROACH FOR VERIFYING GATE-LEVEL MULTIPLIERS
Speaker: Armin Biere, Johannes Kepler Universität Linz, AT
Authors: Daniela Ritirc, Armin Biere, and Manuel Kauers, Johannes Kepler University Linz, AT
Abstract: The currently most effective approach for verifying gate-level multipliers uses Computer Algebra. It reduces a word-level multiplier specification by a Gröbner basis derived from a gate-level implementation. This reduction produces zero if and only if the circuit is a multiplier. We improve this approach by extracting full- and half-adder constraints to reduce the Gröbner basis, which speeds up computation substantially. Refactoring the specification in terms of partial products instead of inputs yields further improvements. As a third contribution we extend these algebraic techniques to verify the equivalence of bit-level multipliers without using a word-level specification.
Download Paper (PDF; Only available from the DATE venue WiFi)

16:30 12.2.3 RECONFIGURABLE ASYNCHRONOUS PIPELINES: FROM FORMAL MODELS TO SILICON
Speaker: Daniil Sokolov, Newcastle University, GB
Authors: Daniil Sokolov, Alessandro de Gennaro and Andrey Mokhov, Newcastle University, GB
Abstract: Datatflow pipelines are widely used in the design of high-throughput computation systems. Real-life applications often require dynamically reconfigurable pipelines to differently process data items or adjust to the current operating mode. Reconfigurable synchronous pipelines are known since 1980s and are well supported by formal models and tools. Reconfigurable asynchronous pipelines on the other hand, have neither a formal behavioural model, nor mature EDA support, making them unattractive to industry. This paper presents a model and an open-source tool for the design and verification of reconfigurable asynchronous pipelines, and validates this approach in silicon.
Download Paper (PDF; Only available from the DATE venue WiFi)

17:00 12.3 Verification and Formal Synthesis
Date: Thursday, March 22, 2018
Time: 16:00 - 17:30
Location / Room: Kont. 1
Chair: Christoph Scholl, University of Freiburg, DE; Contact Christoph Scholl
Co-Chair: Gianpiero Cabodi, Politecnico di Torino, IT; Contact Gianpiero Cabodi

This session explores formal design methodology for asynchronous pipelines, reasons about decomposing assume/guarantee contracts, links high-level models and RTL designs, and also improves arithmetic circuit verification.

17:00 12.3.3 SYMBOLIC ASSERTION MINING FOR SECURITY VALIDATION
Speaker: Alessandro Danese, University of Verona, IT
Authors: Alessandro Danese, Valeria Bertacco, and Graziano Pravadelli
Abstract: This paper presents DOVE, a validation framework to identify points of vulnerability inside IP firmwares. The framework relies on the symbolic simulation of the firmware to search for corner cases in its computational paths that may hide vulnerabilities. Then, DOVE automatically mine a compact set of formal assertions representing these unlikely paths to guide the analysis of the verification engineers. Experimental results on two case studies show the effectiveness of the generated assertions in pinpointing actual vulnerabilities and its efficiency in terms of execution time.
Download Paper (PDF; Only available from the DATE venue WiFi)

17:30 End of session
17:15 12.3.4 SPECIFICATION DECOMPOSITION FOR SYNTHESIS FROM LIBRARIES OF LTL ASSUME/GUARANTEE CONTRACTS
Speaker: Antonio Iannopollo, UC Berkeley, IT
Authors: Antonio Iannopollo, Stavros Tripakis and Alberto Sangiovanni-Vincentelli, University of California, Berkeley, US
Abstract: Contract-Based Design is a methodology that allows for compositional design of complex systems. Given a contract representing a specification, it is possible to formally satisfy it by composing a number of simpler contracts. When these simpler contracts are chosen from a library of existing solutions, we talk about synthesis from contract libraries. There are techniques to automate the synthesis process, but they are computationally intensive, especially for complex specifications. In this paper, we describe an efficient technique to partition a specification, i.e., an LTL-based Assume/Guarantee contract, in a number of simpler sub-specifications which can be satisfied independently. Once all these smaller problems are solved, it is possible to safely merge their solutions to satisfy the original specification. We show the effectiveness of our technique in an industrial case study.
Download Paper (PDF; Only available from the DATE venue WiFi)

17:30 End of session

12.4 Hardware-assisted Security
Date: Thursday, March 22, 2018
Time: 16:00 - 17:30
Location / Room: Konf. 2
Chair: Ilia Polian, University of Stuttgart, DE; Contact Ilia Polian
Co-Chair: Nele Mentens, KU Leuven, BE; Contact Nele Mentens
Security of today’s system cannot be achieved by software techniques alone. This session presents hardware anchors which provide security at circuit and system level and allow its efficient verification.

16:00 12.4.1 HARDWARE-ASSISTED ROOTKIT DETECTION VIA ON-LINE STATISTICAL FINGERPRINTING OF PROCESS EXECUTION
Speaker: Yiorgos Makris, University of Texas at Dallas, US
Authors: Liwei Zhou and Yiorgos Makris, The University of Texas at Dallas, US
Abstract: Kernel rootkits generally attempt to maliciously tamper kernel objects and surreptitiously distort program execution flow. Herein, we introduce a hardware-assisted hierarchical on-line system which detects such kernel rootkits by identifying deviation of dynamic intra-process execution profiles based on architecture-level semantics captured directly in hardware. The underlying key insight is that, in order to take effect, malicious manipulation of kernel objects must distort the execution flow of benign processes, thereby leaving abnormal traces in architecture-level semantics. While traditional detection methods rely on software modules to collect such traces, their implementations are susceptible to being compromised through software attacks. In contrast, our detection system maintains immunity to software attacks by resorting to hardware for trace collection. The proposed method is demonstrated on a Linux-based operating system running on a 32-bit x86 architecture, implemented in Simics. Experimental results, using real-world kernel rootkits, corroborate the effectiveness of this method, while a predictive 45nm PDK is used to evaluate hardware overhead.
Download Paper (PDF; Only available from the DATE venue WiFi)

16:30 12.4.2 SECURING CONDITIONAL BRANCHES IN THE PRESENCE OF FAULT ATTACKS
Speaker: Robert Schilling, Graz University of Technology, AT
Authors: Robert Schilling1, Mario Werner2 and Stefan Mangard2
1Graz University of Technology / Know Center GmbH, AT; 2Graz University of Technology, AT
Abstract: In typical software, many comparisons and subsequent branch operations are highly critical in terms of security. Examples include password checks, signature checks, secure boot, and user privilege checks. For embedded devices, these security-critical branches are a preferred target of fault attacks as a single bit flip or skipping a single instruction can lead to complete access to a system. In the past, numerous redundancy schemes have been proposed in order to provide control-flow-integrity (CFI) and to enable error detection on processed data. However, current countermeasures for general purpose software do not provide protection mechanisms for conditional branches. Hence, critical branches are in practice often simply duplicated. We present a generic approach to protect conditional branches, which links an encoding-based comparison result with the redundancy of CFI protection mechanisms. The presented approach can be used for all types of data encodings and CFI mechanisms and maintains their error-detection capabilities throughout all steps of a conditional branch. We demonstrate our approach by realizing an encoded comparison based on AV-codes, which is a frequently used encoding scheme to detect errors on data during arithmetic operations. We extended the LLVM compiler so that standard code and conditional branches can be protected automatically and analyze its security. Our design shows that the overhead in terms of size and runtime is lower than state-of-the-art duplication schemes.
Download Paper (PDF; Only available from the DATE venue WiFi)

17:00 12.4.3 TOWARDS PROVABLY-SECURE PERFORMANCE LOCKING
Speaker: Abhrajit Sengupta, Texas A&M University, US
Authors: Monir Zaman1, Abhrajit Sengupta2, Danqing Liu3, Ozgur Sinanoglu4, Yiorgos Makris1 and Jeyavijayan Rajendran3
1The University of Texas at Dallas, US; 2New York University, US; 3Texas A&M University, US; 4New York University Abu Dhabi, AE
Abstract: Locking the functionality of an integrated circuit (IC) thrwarts attacks such as intellectual property (IP) piracy, hardware Trojans, overbuilding, and counterfeiting. Although functional locking has been extensively investigated, locking the performance of an IC has been little explored. In this paper, we develop provably-secure performance locking, where only on applying the correct key the IC shows superior performance; for an incorrect key, the performance of the IC degrades significantly. This leads to a new business model, where the companies can design a single IC capable of different performances for different users. We develop mathematical definitions of security and theoretically, and experimentally prove the security against the state-of-the-art attacks. We implemented performance locking on a FabScalar microprocessor, achieving a degradation in instructions per clock cycle (IPC) of up to 77% on applying an incorrect key, with an overhead of 0.6%, 0.2%, and 0% for area, power, and delay, respectively.
Download Paper (PDF; Only available from the DATE venue WiFi)
address mapping.

This session concentrates on methods for prolonging the lifetime of persistent main memory. Based on reference frequencies, papers in this session promote novel approaches to mitigate the adverse impact of writes from prospective cache replacement policy, frequent pattern compression, SLC/MLC hybrid organization, inode virtualization, as well as Android application specific enhancements. The proposed architecture achieves up to 10^5 times lower soft error rate with considerably less ECC overhead. With simple ECC scheme, about 22% performance reduction is achieved. The drift-induced error rate is considerably reduced. By alternating each cell operation between SLC and 4LC over time, the overall lifetime can also be significantly enhanced.

In this paper, we propose a new PCM memory architecture with heterogeneous PCM arrays to increase reliability, performance and lifetime. The basic storage unit in the proposed architecture consists of two single-level cells (SLCs) and one four-level cell (4LC). Using the reduced number of 4LCs compared to conventional homogeneous 4LC PCM arrays, the drift-induced error rate is considerably reduced. By alternating each cell operation between SLC and 4LC over time, the overall lifetime can also be significantly enhanced. The proposed architecture achieves up to 10^5 times lower soft error rate with considerably less ECC overhead. With simple ECC scheme, about 22% performance improvement is achieved and additionally, the overall lifetime is also enhanced by about 57%.

Download Paper (PDF; Only available from the DATE venue WiFi)
AN EFFICIENT PCM-BASED MAIN MEMORY SYSTEM VIA EXPLOITING FINE-GRAINED DIRTINESS OF CACHELINES

Speaker: 
Jie Xu, Huazhong University of Science and Technology, CN

Authors: 
Jie Xu1, Dan Feng2, Yu Hua3, Wei Tong4, Jingping Liu5, Chunyan LiP and Zheng LP
1Wuhan National Lab for Optoelectronics, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China, CN; 2Wuhan National Lab for Optoelectronics, CN; 3NANO, CN

Abstract
Phase Change Memory (PCM) has the potential to replace traditional DRAM memory due to its better scalability and non-volatility. However, PCM also suffers from high write latency and energy consumption. To mitigate the write overhead of PCM-based main memory, we propose a Fine-grained Dirtiness Aware (FDA) last-level cache (LLC) victimization scheme. The key idea of FDA is to preferentially evict cachelines with fewer dirty words when victimizing dirty cachelines. The modified word is defined to be dirty. FDA exploits two key observations. First, the write service time of a cacheline is proportional to the number of dirty words. Second, a cacheline with fewer dirty words has the same or lower reference frequency compared with other dirty cachelines. Therefore, existing cachelines with fewer dirty words can reduce the write service time of cachelines and will not increase the miss rate. To reduce the write service time of cachelines, FDA evicts the cacheline with the fewest dirty words when victimizing dirty cachelines. We also present FDARP to decrease the miss rate by further synergizing the number of dirty words with Re-reference Prediction Value. Experimental results show that FDA (FDARP) can improve the IPC performance by 8.3% (14.8%), decrease the write service time of cachelines by 37.0% (36.3%) and reduce write energy consumption of PCM by 27.0% (32.5%) under the mixed benchmarks.

Download Paper (PDF; Only available from the DATE venue WiFi)

DFPC: A DYNAMIC FREQUENT PATTERN COMPRESSION SCHEME IN NVMe-BASED MAIN MEMORY

Speaker: 
Yuncheng Guo, Huazhong University of Science and Technology, CN

Authors: 
Yuncheng Guo, Yu Hua and Pengfei Zuo, Huazhong University of Science and Technology, CN

Abstract
Non-volatile memory technologies (NVMs) are promising candidates as the next-generation main memory due to high scalability and low energy consumption. However, the performance bottlenecks, such as high write latency and low cell endurance, still exist in NVMs. To address these problems, frequent pattern compression schemes have been widely used, which however suffer from the lack of flexibility and adaptability. In order to overcome these shortcomings, we propose a well-adaptive NVM write scheme, called Dynamic Frequent Pattern Compression (DFPC), to significantly reduce the amount of write units and extend the lifetime. Instead of only using static frequent patterns in existing FPGA schemes, which are pre-defined and not always efficient for all applications, the idea behind DFPC is to exploit the characteristics of data distribution in execution to obtain dynamic patterns, which often appear in the real-world applications. To further improve the compression ratio, we exploit the value locality in a cache line to extend the granularity of dynamic patterns. Hence DFPC can encode the contents of cache lines with more kinds of frequent data patterns. We implement DFPC in GEMS with NVMain and execute 8 applications from SPEC CPU2006 to evaluate our scheme. Experimental results demonstrate the efficacy and efficiency of DFPC.

Download Paper (PDF; Only available from the DATE venue WiFi)

12.6 Special Session: Computing with Emerging Memories: How Good can it be?

Date: Thursday, March 22, 2018

Time: 16:00 - 17:30

Location / Room: Konr. 4

Chair: 
Pierre-Emmanuel Gaillardon, University of Utah, US, Contact Pierre-Emmanuel Gaillardon

Co-Chair: 
Ian O'Connor, Ecole Centrale de Lyon, FR, Contact Ian O'Connor

With the recent evolutions of nanometer transistor technologies, power consumption emerged as the most critical limitation. Within advanced processors and computing architectures, the processor-memory communication accounts for a significant part of the energy requirement. While alternative design approaches, such as the use of optimized accelerators or advanced power management techniques are successfully employed in contemporary designs, the trend keeps worsening due to the ever-increasing gap between on-chip and off-chip memory data rates. This trend, known as Von Neumann bottleneck, not only limits the system performance, but also acts nowadays as a limiter of the energy scaling. The quest towards more energy-efficiency requires solutions that solve the Von Neumann bottleneck by tightly intertwining computing with memories. In this hot topic session, we intend to elaborate on in-memory computing by identifying and comparing the latest computing models in light of conventional, e.g., SRAMs, and emerging memory technologies, e.g., RRAMs, STT-MRAMs. In-memory computing is considered here in the general sense of computing information locally within large data storage.

PRACTICAL CHALLENGES IN DELIVERING THE PROMISES OF REAL PROCESSING-IN-MEMORY MACHINES

Speaker: 
Nishit Talati, Technion - Israel Institute of Technology, IL

Authors: 
Nishit Talati1, Ameer Haj Ali2, Rotem Ben Hur3, Nimrod Wald4, Ronny Ronen5, Pierre-Emmanuel Gaillardon2 and Shahar Kvatsinsky1
1Technion, IL; 2Technion - Israel Institute of Technology, IL; 3University of Utah, US

Abstract
Processing-in-Memory (PIM) machines promise to overcome the von Neumann bottleneck in order to further scale performance and energy efficiency of computing systems by reducing the extent of data transfer and offering ample parallelism. In this paper, we take the memristive Memory Processing Unit (mMPU) as a case study of a PIM machine and scrutinize it in practical scenarios. Specifically, we explore the limitations of parallelism and data transfer-elimination. We argue that lack of operand locality and management might make data transfer inevitable in the mMPU. We then devise techniques to move data within the mMPU, without transferring it off-chip, and quantify their costs. Additionally, we present electrical parameters that might limit the parallelism offered by the mMPU and evaluate their impact. Using benchmarks from the LGsynth91 suite, their vector extensions, and a few synthetic data parallel workloads, we show that the internal data transfer results in an increase of up to 1.5x in the execution time, while the limited parallelism increases it by 1.1x to 2x.

Download Paper (PDF; Only available from the DATE venue WiFi)
<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 16:30 | 12.6.2 | SMART INSTRUCTION CODES FOR IN-MEMORY COMPUTING ARCHITECTURES COMPATIBLE WITH STANDARD SRAM INTERFACES | Maha Kooli, CEA-Leti, FR  
Authors:  
Maha Kooli¹, Henri-Pierre CHARLES², Bastien Giraud³ and Jean-Philippe Noel²  
¹CEA/LETI, FR; ²CEA, FR; ³CEA LETI, FR |
|       |       | **Abstract**                                                                     | This paper presents the computing model for the In-Memory Computing architecture based on SRAM memory that embeds computing abilities. This memory concept offers significant performance gains in terms of energy consumption and execution time. To handle the interaction between the memory and the CPU, new memory instruction codes were designed. Those instructions are communicated by the CPU to the memory, using standard SRAM buses. This implementation allows (1) to embed In-Memory Computing capabilities on a system without Instruction Set Architecture (ISA) modification, and (2) to finely interlace CPU instructions and in-memory computing instructions. |
| 17:00 | 12.6.4 | MEMRISTIVE DEVICES FOR COMPUTATION-IN-MEMORY                                    | Said Hamdioui, Delft University of Technology, NL  
Authors:  
Jintao Yu, HoangAnh DuNguyen, Mottaqiallah Tasiull and Said Hamdioui, TU Delft, NL |
|       |       | **Abstract**                                                                     | CMOS technology and its continuous scaling have made electronics and computers accessible and affordable for almost everyone on the globe; in addition, they have enabled the solutions of a wide range of societal problems and applications. Today, however, both the technology and the computer architectures are facing severe challenges/walls making them incapable of providing the demanded computing power with tight constraints. This motivates the need for the exploration of novel architectures based on new device technologies; not only to sustain the financial benefit of technology scaling, but also to develop solutions for extremely demanding emerging applications. This paper presents two computation-in-memory based accelerators making use of emerging memristive devices; they are Memristive Vector Processor and RRAM Automata Processor. The preliminary results of these two accelerators show significant improvement in terms of latency, energy and area as compared to today's architectures and design. |
| 17:15 | 12.6.3 | COMPUTING-IN-MEMORY WITH SPINTRONICS                                         | Shubham Jain, Purdue University, US  
Authors:  
Shubham Jain¹, Sachin Sapatnekar², Jian-Ping Wang², Kaushik Roy¹ and Anand Raghunathan¹  
¹Purdue University, US; ²Department of Electrical and Computer Engineering, University of Minnesota, US |
|       |       | **Abstract**                                                                     | In-memory computing is a promising approach to alleviating the processor-memory data transfer bottleneck in computing systems. While spintronics has attracted great interest as a non-volatile memory technology, recent work has shown that its unique properties can also enable in-memory computing. We summarize efforts in this direction, and describe three different designs that enhance STT-MRAM to perform logic, arithmetic, and vector operations and evaluate transcendental functions within memory arrays. |
| 17:30 |       | End of session                                                                   |                                                                       |

Source URL: https://past.date-conference.com/date18/booklet/proof_reading