Goto Session:

- **1.1 Opening Session: Plenary, Awards Ceremony & Keynote Addresses**
  - UB01 Session 1
- **2.1 Executive Panel: How Electronics May Change Our Lives, and the World**
- **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 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 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

Time Label Presentation Title Authors

08:30 1.1.1 WELCOME ADDRESSES Speakers: Jan Madsen¹ and Ayse Coskun²
¹DTU, DK; ²Boston University, US

08:45 1.1.2 PRESENTATION OF AWARDS Abstract
- 2018 EDAA Achievement Award
- EDAA Outstanding Dissertations Award 2017
- DATE Fellow Award
- IEEE Fellow Award
- IEEE CEDA and CS TTTC Outstanding Service Contribution Award 2017

09:15 1.1.3 KEYNOTE ADDRESS: THE RESPONSIBILITY SENSITIVE SAFETY (RSS) FORMAL MODEL TOWARD SAFETY GUARANTEES FOR AUTONOMOUS VEHICLES Speaker: Amnon Shashua, Intel Corporation, US

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 fenced 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 “ethical dilemmas” by providing a formal framework for the discussion.
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.

10:30 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 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

Label | Presentation Title | Authors
--- | --- | ---
UB01.1 | ARCHON: AN ARCHITECTURE-OPEN RESOURCE-DRIVEN CROSS-LAYER MODELLING FRAMEWORK | Fei Xia¹, Ashur Rafiev¹, Mohammed Al-Hayanni², Alexei Iliasov³, Rishad Shafik¹, Alexander Romanovsky¹ and Alex Yakovlev¹
¹Newcastle University, GB; ²Newcastle University, UK and University of Technology and HCED, IQ

Abstract
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.

More information ...

UB01.2 | TOPOLINO & MAGCAD: A DESIGN AND SIMULATION FRAMEWORK FOR THE EXPLORATION OF EMERGING TECHNOLOGIES | 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: ToPoNANO and MagCAD. Different technologies are supported, and new ones could be added thanks to their modular structure. ToPoNANO 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 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 ...
ADVANCED SIMULATION OF QUANTUM COMPUTATIONS

Authors: Zulehner Anlen 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 straight forward 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 ...

OTPGEN: 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 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 microprocessors memory (for experiments we’re using UEMK 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 ...

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

Authors: Abhaya Chandra Kammara S., Sidney Pontes-Filho and Andreas König

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 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. Int. Syst. Springer, DOI 10.1007/s40477-016-0027-3, 2016

More information ...

WARE: WEARABLE ELECTRONICS DIRECTIONAL AUGMENTED REALITY

Authors: Gabriele Morand, Walter Vendraminetti, Federico Fiacaro, Davide Quaglia and Gianluca Benedetti

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 ...

IDEAA: DESIGN SPACE EXPLORATION FOR FUNCTIONAL-LEVEL APPROXIMATION

Authors: Marcello Traiola, Mario Barbareschi, Marcello Traiola and Alberto Bosio

Abstract: Approximate Computing (AxC) aims at enabling the production of computing systems which can satisfy 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++ code. We designed and implemented IDEAA, 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++ code.

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. 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 ...

CIJTAG: CONCURRENT IJTAG DEMONSTRATOR

Author: Kienz-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 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 FFPGA-based implementation of concurrent IJTAG test infrastructures.
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 1

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

Innovation runs strong in our industry. A 17 qubits Quantum Computer (QC) has been shipped to Delft 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, and the challenges to be overcome

Panelists:
- Loic Lietar, Greenwaves Technologies, FR
- Martin Roetteler, 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: Konf. 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. Download Paper (PDF; Only available from the DATE venue WiFi)</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 16–38% 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 9–22% energy savings. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
</tbody>
</table>
MODNN: MEMORY OPTIMAL DNN TRAINING ON GPUs

Speaker:
Xiaoming Chen, Institute of Computing Technology, Chinese Academy of Sciences, CN

Authors:
Xiaoming Chen, Danny Z. Chen and Xiaobo Sharon Hu, University of Notre Dame, US

Abstract
Graphics processing units (GPUs) are widely adopted to accelerate the training of deep neural networks (DNNs). However, the limited GPU memory size restricts the maximum scale of DNNs that can be trained on GPUs, which presents serious challenges. This paper proposes an moDNN framework to optimize the memory usage in DNN training. moDNN supports automatic tuning of DNN training code to match any given memory budget (not smaller than the theoretical lower bound). By taking full advantage of overlapping computations and data transfers, we have developed heuristics to judiciously schedule data unloading and prefetching, together with training algorithm selection, to optimize the memory usage. We further introduce a new sub-batch size selection method which also greatly reduces the memory usage. moDNN can save the memory usage up to 50X, compared with the ideal case which assumes that the GPU memory is sufficient to hold all data. When executing moDNN on a GPU with 12GB memory, the performance loss is only 8%, which is much lower than that caused by the best known existing approach. vDNN, moDNN is also applicable to multiple GPUs and attains 1.84X average speedup on two GPUs.

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

HYPERPOWER: POWER- AND MEMORY-CONSTRAINED HYPER-PARAMETER OPTIMIZATION FOR NEURAL NETWORKS

Speaker:
Dimitrios Stamoulis, Carnegie Mellon University, US

Authors:
Dimitrios Stamoulis1, Ermao Cai1, Da-Cheng Juan2 and Diana Marculescu1
1Carnegie Mellon University, US; 2Google Research, US

Abstract
While selecting the hyper-parameters of Neural Networks (NNs) has been so far treated as an art, the emergence of more complex, deeper architectures poses increasingly more challenges to designers and Machine Learning (ML) practitioners, especially when power and memory constraints need to be considered. In this work, we propose HyperPower, a framework 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. HyperPower is the first work (i) to show that power consumption can be used as a low-cost, a priori known constraint, and (ii) to propose predictive models for the power and memory of NNs executing on GPUs. Thanks to HyperPower, the number of function evaluations and the best test error achieved by a constraint-unaware method are reached up to 112.99X and 30.12X faster, respectively, while never considering invalid configurations. HyperPower significantly speeds up the hyper-parameter optimization, achieving up to 57.20X more function evaluations compared to constraint-unaware methods for a given time interval, effectively yielding significant accuracy improvements by up to 67.6%.

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

RECOM: AN EFFICIENT RESISTIVE ACCELERATOR FOR COMPRESSED DEEP NEURAL NETWORKS

Speaker:
Houxiang Ji, Shanghai Jiao Tong University, CN

Authors:
Houxiang Ji1, Linghao Song2, Li Jiang1, Hai (Helen) Li2 and Yiran Chen2
1Shanghai Jiao Tong University, CN; 2Duke University, US. Download Paper (PDF; Only available from the DATE venue WiFi)

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 inessential 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)

SPARSENN: 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, Jingbo Jiang, Xizi Chen and Chi-Ying Tsui, Hong Kong University of Science and Technology, HK

Abstract
Contemporary Deep Neural Network (DNN) contains millions of synaptic connections with tens to hundreds of layers. 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, SparseNN, is proposed to exploit both the input and output sparsity. SparseNN is a 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, SparseNN can achieve a 10%-70% improvement in throughput and a power reduction of around 50%.

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

2.3 High-Level Synthesis

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

### 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)**

### 2.3.2 FallS: 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, sush 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)**
The session consists of four papers on model checking. It ranges from model checking of multiple properties, improving bit-level model checking, checking specific properties in processor design to software model checking of the Linux kernel.
VERIFICATION OF TREE-BASED HIERARCHICAL READ-COPY UPDATE IN THE LINUX KERNEL

Speaker: Lihao Liang
Authors: Paul McKenney, IBM Linux Technology Center, US

Abstract
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.

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

2.5 GPU and GPU-based heterogeneous system management

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

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

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).

Time Label Presentation Title Authors

11:30 2.5.1 HVSM: HARDWARE-VARIABILITY AWARE STREAMING PROCESSORS’ MANAGEMENT POLICY IN GPUS
Speaker: Jingweijia Tan, Jilin University, CN
Authors: Jingweijia Tan\(^1\) and Kaige Yan\(^2\)
\(^1\)Jilin University, CN; \(^2\)College 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 desirers 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 SP’s performance. We further observe unbalanced SP utilization, e.g., some SPs are idle when 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)
12:00  2.5.2 THROUGHPUT OPTIMIZATION AND RESOURCE ALLOCATION ON GPUs UNDER MULTI-APPLICATION EXECUTION
Speaker: Isaral Ahmad, Sanjoy Patil and Smruti R. Sarangi, IIT Delhi, IN
Authors: Isaral Ahmad, Sanjoy 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 and
parse tags 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,
we describe how to extend this to parse full XML documents. We compare our results with the existing works and find that we outperform them 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)

12:30  2.5.3 SET VARIATION-AWARE SHARED LAST-LEVEL CACHE MANAGEMENT FOR CPU-GPU HETEROGENEOUS ARCHITECTURE
Speaker: Xin Li, Shandong University, CN
Authors: Zhaoying Li, Lei Ju, Hongjun Dai, Xin Li, Mengying Zhao and Zhiping Jia, Shandong University, CN
Abstract
Heterogeneous CPU-GPU multiprocessor systems-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 lightweight and fine-grained cache management policy to cope with the CPU-GPU 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 LSP by
12.6% and 10.01%, respectively.
Download Paper (PDF; Only available from the DATE venue WiFi)

13:00  IP1-4, 464
HPXA: A HIGHLY PARALLEL XML PARSER
Speaker: Smruti Sarangi, IIT Delhi, IN
Authors: Isaral Ahmad, Sanjoy 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 and
tags 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)

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
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></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>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.</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 1TU München (TUM), DE; 2TU München, DE; 3The Chinese University of Hong Kong, HK; 4University of Texas at Austin, US; 5Technical University of Munich, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>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.</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 1NEW YORK UNIVERSITY, US; 2New York University, US; 3New York University Abu Dhabi (NYUAD), AE; 4New York University Abu Dhabi, AE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract</strong></td>
<td>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 camouflaging all 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.</td>
</tr>
</tbody>
</table>

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

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. 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. In 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.

**2.7 Special Session: Spintronics based New Computing Paradigms and Applications**

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

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

**Location / Room:** Konf. 5

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

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

n 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**
--- | --- | ---
11:30 | 2.7.1 | MAIN MEMORY ORGANIZATION TRADE-OFFS WITH DRAM AND STT-MRAM OPTIONS BASED ON EXTENDED GEMS/NVMAIN SIMULATION FRAMEWORK

**Speaker:** Manu Komalan, IMEC, BE

**Authors:** Manu Komalan1, Oh Hyung Rock1, Matthias Hartmann1, Sushil Sakhane1, Christian Tenllado2, Jose Ignacio Gomez2, Gouri Sankar Kar3, Arnaud Furnemont1, Francky Cattaneo1, Sophie Senn2, David Novo1, Abdoulaye Gamatie5 and Lionel Torres6

1IMEC, BE; 2Rheinische Friedrich wilhelms-Universitat Bonn (RUB), DE; 3French National Centre for Scientific Research (CNRS), FR; 4LIRMM / 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)**
EXPLORING THE OPPORTUNITY OF IMPLEMENTING NEUROMORPHIC COMPUTING SYSTEMS WITH SPINTRONIC DEVICES

Speaker: Hai Li, Duke University, US
Authors: Bionan Yan1, Fan Chen1, Yaojun Zhang2, Chang Song1, Hai (Helen) Li1 and Yiran Chen1
1Duke University, US; 2University of Pittsburgh, US; 3Duke University/TUM-IAS, US

Abstract

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 also been invented to support this effort by leveraging emerging nano-devices. 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)

SPINTRONIC NORMALLY-OFF HETEROGENEOUS SYSTEM-ON-CHIP DESIGN

Speaker: Rajendra Bishnoi, Karlsruhe Institute of Technology, DE
Authors: Anteneh Gebregiorgis, Rajendra Bishnoi and Mehdi Tahoori, Karlsruhe Institute of Technology, DE

Abstract

One of the major challenges in device down-scaling is the increase in the leakage power, which becomes a major component in the overall system power consumption. One way to deal with this problem is to introduce the concept of normally-off instant-on computing architectures, in which the system components are powered off when they are not active. An associated challenge is the back-up and restoration of system states, which in turn can introduce additional costs that erode some of the gains. A promising alternative is the use of non-volatile storage elements in the System-on-Chip (SoC) design which can instantly power-down and retain their values. In this work, we show how we can design a normally-off SoC by exploiting non-volatile latches, flip-flops and registers. The idea is to design a hybrid architecture containing conventional CMOS bistables as well as different flavors of spintronic-based non-volatile storage elements, to balance performance, area, and energy efficiency.

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

MAGNETIC SKYRMIONS FOR FUTURE POTENTIAL MEMORY AND LOGIC APPLICATIONS: ALTERNATIVE INFORMATION CARRIERS

Speaker and Author: Wang Kang, Beihang University, CN

Abstract

Magnetic skyrmions are swirling topological configurations, which are mostly induced by chiral interactions between atomic spins in non-centrosymmetric magnetic bulks or in thin films with broken inversion symmetry. They hold promise as information carriers in future ultra-dense, low-power memory and logic devices owing to the nanoscale size and extremely low spin-polarized currents needed to move them. To date, an intense research effort has led to the identification, creation/annihilation, motion and manipulation of skyrmions at room temperature. Meanwhile, a rich variety of skyrmion-based device concepts and prototypes have been proposed, indicating the considerable potential of magnetic skyrmions in future electronic applications. However, current studies mainly focus on physical or principle investigations, whereas the electrical design methodology, implementation and evaluations are still lacking. In this paper, we will bring the readers in the “design, automation and test (DAT) society” the current status and outlook of magnetic skyrmions in future electronic applications. However, current studies mainly focus on physical or principle investigations, whereas the electrical design methodology, implementation and evaluations are still lacking. In this paper, we will bring the readers in the “design, automation and test (DAT) society” to address the challenges, e.g., a-forementioned.

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

NOVEL APPLICATION OF SPINTRONICS IN COMPUTING, SENSING, STORAGE AND CYBERSECURITY

Speaker: Anirudh Iyengar, Pennsylvania State University, US
Author: Swapna Ghosh, Pennsylvania State University, US

Abstract

With conventional Von Neumann computing struggling to match the energy-efficiency of biological systems, there is pressing need to explore alternative computing models. CMOS switches, although universal, fails to offer additional features to meet this end goal. Recent experimental studies have revealed that spintronics possess many promising features that can not only enable Von Neumann compute models but also high-density storage, sensing of environmental parameters and protection from cybersecurity threats. This talk will provide an in-depth study of spintronics and its relation to these novel aspects from device, circuit and system standpoint.

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

LARGE SCALE, HIGH DENSITY INTEGRATION OF ALL SPIN LOGIC

Speaker and Author: Jacques-Oliver Klein, C2N, Univ. Paris-Sud, FR

Abstract

Spintronics brings new features that make it a viable candidate technology to implement non-conventional processing for new computing paradigms in an efficient way. The first milestone of the spintronics roadmap was the fabrication of hybrid systems where the data processing relies mostly on charge-based electronics devices (CMOS), while the memory hierarchy is partially or totally replaced by MRAM. In the next step, spintronics can also be used for data processing, still in conjunction with CMOS. Nevertheless, replacing all the processing by pure spintronic circuits, without any charge current, remains the ultimate objective of spintronics. All spin logic (ASL) paves the way towards that goal, even if some CMOS control circuits are still necessary. However, as ASL does not rely on the same computing principle as CMOS, it is necessary to address some specific issues. Pure spin current propagates in every direction, including backwards in the presence of multiple inputs; and is divided when crossings are encountered. It combines mainly linearly, while logic operations require non-linear binary decisions. Interconnect between logic gates requires directionality from inputs to outputs, and fanout with negligible signal attenuation. In this context, we develop new strategies for ASL modeling and logic design. We propose an architecture and a design strategy based on a high-density array to address the specific issues of directionality, attenuation and linearity. Moreover, the feasibility is supported through the modeling and the simulation of its basic block. This implies modularity to simulate complex circuits, even when they are ahead of today’s experimental demonstrations.

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 16:30 - 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 falls 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.
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, impactful 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.

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.

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.

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.

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
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- 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 Rafievi1, Mohammed Al-Hayanni², Alexei Iliasov¹, Rishad Shatik¹, Alexander Romanovsky¹ and Alex Yakovlev¹</td>
</tr>
<tr>
<td></td>
<td><strong>Abstract</strong></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, 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><strong>More information ...</strong></td>
<td></td>
</tr>
<tr>
<td>UB02.2</td>
<td>GENERATING FULL-CUSTOM SCHEMATICS IN A MIXED-SIGNAL TOP-DOWN DESIGN FLOW</td>
<td>Tobias Markus¹, Markus Mueller² and Ulrich Bruening¹</td>
</tr>
<tr>
<td></td>
<td><strong>Authors</strong></td>
<td>¹University of Heidelberg, DE; ²Extoll GmbH, DE</td>
</tr>
<tr>
<td></td>
<td><strong>Abstract</strong></td>
<td>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 venug 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.</td>
</tr>
<tr>
<td></td>
<td><strong>More information ...</strong></td>
<td></td>
</tr>
<tr>
<td>UB02.3</td>
<td>OISC MULTICORE STENCIL PROCESSOR: ONE INSTRUCTION-SET COMPUTER-BASED MULTICORE PROCESSOR FOR STENCIL COMPUTING</td>
<td>Kaoru Saso, Jing Yuan Zhao and Yoko Hara-Azumi, School of Engineering, Tokyo Institute of Technology, JP</td>
</tr>
<tr>
<td></td>
<td><strong>Authors</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Abstract</strong></td>
<td>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 1/20x 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 heaving computation.</td>
</tr>
<tr>
<td></td>
<td><strong>More information ...</strong></td>
<td></td>
</tr>
</tbody>
</table>
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 reusable and (c) can consider all relevant aspects and constraints implicitly. Nowadays, generators are successfully applied in analog layout (Pcells, 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 ...

Comprehensive Approach to Front to Back Analog Block Design Automation

Authors: Abhaya Chandra Kammara S.1, Siddney Pontes-Filho2 and Andreas König2

1ISE, TU Kaiserslautern, DE; 2University of Kaiserslautern, DE

Abstract

ABSYNTH was first presented in CEBIT 2014 where complete, parallel 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. Palanchamy, 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 ...

Tool/OMC: Optimized Compilation of Executable UML/SysML Diagrams for the Design of Data-Flow Applications

Authors: Andrea Enrico1, Julien Lallet1, Renaud Paclele2 and Ludovic Apvrille2

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., FPGAs). 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 ...

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 establish 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 ...

Risc-V Processor Modeling in IP-XACT Using Kactus2

Authors: Esko Pekkarinen 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 take 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 ...

Embedded Acceleration of Image Classification Applications for Stereo Vision Systems

Authors: Mohammad Loni1, Carl Ahlberg2, Masoud Daneshlou2, Mikael Ekström2 and Mikael Sjödin2

1MDH, SE; 2Mälardalen University, SE

Abstract

Autonomous systems are used in a broad range of applications from indoor utenels 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 images 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 approximation 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 re-generates 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 ...

Clava-MARGOT: Clava + MARGOT + C++ to C++ Compiler and Runtime Autotuning Framework

Authors: João Bispo1, Davide Gadioli2, Pedro Pinto1, Emanuele Vitali1, 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++ source-to-source compiler (Clava) and an autotuner (mARGOT). They are part of the toolflow of the FET-HPC ANTAREX project and allow parallelization, multithreading 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++ 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 ...

End of session: 15:00
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

3.1 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" 13:30 - 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

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.
Abstract

Computation kernels in emerging recognition, mining, and synthesis (RMS) applications are inherently error-resilient, where approximate computing can be applied to improve their energy efficiency by trading off computational effort and output quality. One promising approximate computing technique is to perform approximate computing with memory, which stores a subset of function responses in a lookup table (LUT), and avoids redundant computation when encountering similar input patterns. Limited by the memory space, most existing solutions simply store values for those frequently-appeared input patterns, without considering output quality and/or intrinsic characteristic of the target kernel. In this paper, we propose a novel LUT allocation technique for approximate computing with memory, which is able to dramatically improve the hit rate of LUT and hence achieves significant energy savings under given quality constraints. We also present how to apply the proposed LUT allocation solution for multiple computation kernels. Experimental results show the efficacy of our proposed methodology.

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:
Yun Long, Georgia Institute of Technology, US

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

Abstract

Modeling the dynamics of biophysical neural network (BNN) is essential to understand brain operation and design cognitive systems. Large-scale and biophysically plausible BNN modeling requires solving multiple-term, 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 11x speedup with insignificant loss of biological details in the region of interest.

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

3.3 Optimization Techniques for MPSoCs

Date: Tuesday, March 20, 2018

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
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 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.

Domain-specific computing is promising for high-performance low-power execution of applications with similar functionality. In particular, streaming applications with significant functional and structural similarities can tremendously benefit. However, current Design Space Exploration (DSE) focuses on individual applications in isolation. Hence, much of the domain optimization opportunities are missed. DSE methodologies need to broaden the scope from individual applications to optimizing across applications within a domain. This paper introduces a novel Domain-Specific DSE (DS-DSE) approach for domain-specific computing with a focus on streaming applications. Key contributions are: (1) a formalized method to extract the functional and structural similarities of domain applications, (2) a novel algorithm for hardware/software partitioning of a domain-specific platform to maximize throughput across domain applications (under certain constraints) and (3) a methodology to evaluate a domain platform. This paper demonstrates the benefits using 4 domains: OpenVX (vision processing), and 3 synthetic domains (with greater complexity). Our experiments demonstrate a performance improvement (average throughput) of 36.8% for OpenVX and 46.2% for synthetic domains of the DS-DSE generated platform compared to an application-specific platform.

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.

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 ASPM+T. 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.

This paper introduces a novel Domain-Specific DSE (DS-DSE) approach for domain-specific computing with a focus on streaming applications. Key contributions are: (1) a formalized method to extract the functional and structural similarities of domain applications, (2) a novel algorithm for hardware/software partitioning of a domain-specific platform to maximize throughput across domain applications (under certain constraints) and (3) a methodology to evaluate a domain platform. This paper demonstrates the benefits using 4 domains: OpenVX (vision processing), and 3 synthetic domains (with greater complexity). Our experiments demonstrate a performance improvement (average throughput) of 36.8% for OpenVX and 46.2% for synthetic domains of the DS-DSE generated platform compared to an application-specific platform.

In this paper, we focus on topological-aware virtual resource management for heterogeneous multicore systems. We introduce a runtime Topology-aware virtual Resource Management Scheme for heterogeneous NUMA multicore systems. vTRMS can improve application performance based on the comprehensive online monitor of the application resource access behaviors as well as an accurate and platform-independent detected NUMA topology metric. Experiment results show that, compared with state-of-art approaches, vTRMS can bring up an average throughput improvement of 28.3% and 36.2% on Intel and AMD NUMA machines respectively, when consolidating 32 VMs. At the same time, vTRMS only incurs a runtime overhead no more than 5%.

Virtualization technology consolidates multiple independent workloads on a single physical server with virtual resource management, which can result in a significant utilization improvement and energy saving. However, the management of virtual resource is becoming more and more challenging, due to the lack of accurate performance prediction model for the diverse applications’ irregular resource access behaviors as well as the complicated Non-Uniform Memory Access (NUMA) server architecture. These challenges drastically affect the overall consolidation performance. This paper proposes vTRMS, a runtime Topology-aware virtual Resource Management Scheme for heterogeneous NUMA multicore systems. vTRMS can improve application performance based on the comprehensive online monitor of the application resource access behaviors as well as an accurate and platform-independent detected NUMA topology metric. Experiment results show that, compared with state-of-art approaches, vTRMS can bring up an average throughput improvement of 28.3% and 36.2% on Intel and AMD NUMA machines respectively, when consolidating 32 VMs. At the same time, vTRMS only incurs a runtime overhead no more than 5%.

In this paper, we focus on topological-aware virtual resource management for heterogeneous multicore systems. We introduce a runtime Topology-aware virtual Resource Management Scheme for heterogeneous NUMA multicore systems. vTRMS can improve application performance based on the comprehensive online monitor of the application resource access behaviors as well as an accurate and platform-independent detected NUMA topology metric. Experiment results show that, compared with state-of-art approaches, vTRMS can bring up an average throughput improvement of 28.3% and 36.2% on Intel and AMD NUMA machines respectively, when consolidating 32 VMs. At the same time, vTRMS only incurs a runtime overhead no more than 5%.

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.

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 (ASPM+T), 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 ASPM+T. 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.

This paper introduces a novel Domain-Specific DSE (DS-DSE) approach for domain-specific computing with a focus on streaming applications. Key contributions are: (1) a formalized method to extract the functional and structural similarities of domain applications, (2) a novel algorithm for hardware/software partitioning of a domain-specific platform to maximize throughput across domain applications (under certain constraints) and (3) a methodology to evaluate a domain platform. This paper demonstrates the benefits using 4 domains: OpenVX (vision processing), and 3 synthetic domains (with greater complexity). Our experiments demonstrate a performance improvement (average throughput) of 36.8% for OpenVX and 46.2% for synthetic domains of the DS-DSE generated platform compared to an application-specific platform.

In this paper, we focus on topological-aware virtual resource management for heterogeneous multicore systems. We introduce a runtime Topology-aware virtual Resource Management Scheme for heterogeneous NUMA multicore systems. vTRMS can improve application performance based on the comprehensive online monitor of the application resource access behaviors as well as an accurate and platform-independent detected NUMA topology metric. Experiment results show that, compared with state-of-art approaches, vTRMS can bring up an average throughput improvement of 28.3% and 36.2% on Intel and AMD NUMA machines respectively, when consolidating 32 VMs. At the same time, vTRMS only incurs a runtime overhead no more than 5%.

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.

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 (ASPM+T), 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 ASPM+T. 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.

In this paper, we focus on topological-aware virtual resource management for heterogeneous multicore systems. We introduce a runtime Topology-aware virtual Resource Management Scheme for heterogeneous NUMA multicore systems. vTRMS can improve application performance based on the comprehensive online monitor of the application resource access behaviors as well as an accurate and platform-independent detected NUMA topology metric. Experiment results show that, compared with state-of-art approaches, vTRMS can bring up an average throughput improvement of 28.3% and 36.2% on Intel and AMD NUMA machines respectively, when consolidating 32 VMs. At the same time, vTRMS only incurs a runtime overhead no more than 5%.
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.

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

3.4 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–Aix, 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)

14:30
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, Wenyuan Lu, Shuhaos 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 proposed 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 ineffective ones. The ineffective convolutions in which either the neurons or synapses are all zeros 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 times 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 1.9x 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)

15:30 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
- 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.
### 3.5.1 GRADIENT IMPORTANCE SAMPLING: AN EFFICIENT STATISTICAL EXTRACTION METHODOLOGY OF HIGH-SIGMA SRAM DYNAMIC CHARACTERISTICS

**Speaker:**
Thom Haıne, Université catholique de Louvain, BE

**Authors:**
Thomas Haıne, Johan Segers, Denis Flandre and David BoP

1université catholique de louvain, BE; 2université catholique de Louvain, BE

**Abstract**
The impact of within-die transistor variability has increased with CMOS technology scaling up to the point where it has emerged as a systematic problem for the designer. Estimating extremely low failure rate, i.e. “high sigma” probabilities, by the conventional Monte Carlo (MC) approach requires millions of simulation runs, making it an impractical approach for circuit designers. To overcome this problem, alternative failure estimation methodologies, which require a smaller number of runs have been proposed. In this paper, we propose a novel methodology called “gradient importance sampling” (GIS) for fast statistical extraction of high-sigma circuit characteristics. It is based on conventional Importance Sampling combined with a gradient-based approach to find the most probable failure point (MPFP). By applying GIS to extract SRAM dynamic characteristics in 28nm FDSOI CMOS, we show that the proposed methodology is straightforward, computationally efficient and the results are in line with those obtained via standard MC. To the best of our knowledge, the GIS results are the best in their class for low failure rate estimation.

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

### 3.5.2 DEGRADATION ANALYSIS OF HIGH PERFORMANCE 14NM FINFET SRAM

**Speaker:**
Daniel Kraak, Delft University of Technology, NL

**Authors:**
Daniel Kraak, Innocent Agbo, Matteoalltao Toaullii, Said Hamdoull, Pieter Weck and Frandcy Catthoo

1Delft University of Technology, NL; 2imec vzw., BE

**Abstract**
Memory designs usually add design margins to compensate for chip aging; this may lead to yield and performance loss (in case of overestimation) or reduced reliability (in case of underestimation). This paper analyzes the impact of aging on cutting edge high performance 14nm FinFET SRAM using a calibrated aging model; it does not only analyze the impact of the SRAM's components individually, as it is the case in prior work, but it also investigates the contribution of the interaction of these components while considering different workloads; both the overall metric of the memory (i.e., the access time) as well as metrics of individual components (e.g., sensing delay for the sense amplifier) are examined. The results show that it is crucial to consider not only the aging of individual components, but also their interaction in order to provide accurate prediction of aging effects; considering only aging of single/individual components leads to either too optimistic or pessimistic results. For example, using our approach (which includes the components interaction) results approximately in 9.1% degradation of memory access time (for three years of aging), while using the traditional approach (based on adding the impact of individual components) results in 7.3% increase only; a relative difference of 25%, for which the timing and the address decoder components are the main contributors. With respect to individual components, the sense amplifier is the most fragile one (e.g., its offset voltage spec. degrades up to 58%).

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

### 3.5.3 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 Kairim 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)**

### 3.5.4 14NM FINFET SRAM

**Speaker:**
Sarath Mohananandhan Nair, Karlsruhe Institute of Technology, DE

**Authors:**
Sarath Mohananandhan 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 parameter 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 exacerbates 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 equips 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)**
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 10:50 - 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>
<tr>
<td></td>
<td></td>
<td></td>
<td>Authors: Jiankang Ren, Ran Bi, Xiaoyan Su, Qian Liu, Guowei Wu and Guozhen Tan, Dalian University of Technology, CN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>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 harmonicity 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)</td>
</tr>
<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>
<tr>
<td></td>
<td></td>
<td></td>
<td>Authors: Leandro Indrusiak&lt;sup&gt;1&lt;/sup&gt;, Alan Burns&lt;sup&gt;1&lt;/sup&gt; and Borislav Nikolic&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>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)</td>
</tr>
</tbody>
</table>
### 3.6.3 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 tasks for a broader design space exploration illustrates the efficacy of the proposed mechanism.

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

### 3.6.4 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; ²LJFL/University of Lille¹, 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)**

### 16:00 ONE-WAY SHARED MEMORY

**Speaker and Author:**
Martin Schoebel, 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 single 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)**

### 3.8 Innovative Products for Autonomous Driving (part 1)

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

**Location / Room:** Exhibition Theatre
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>Speakers: Christian Wolf&lt;sup&gt;1&lt;/sup&gt; and Hans-Jürgen Brand&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1IDT Europe GmbH, DE; 2IDT/ZMDI, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</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>Speakers: Stanislav Mudriievskyi and Vincent Latzko, Technical University Dresden, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</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>Speaker: Alexander Muffler, X-Fab Semiconductor Foundries AG, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</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 automotive 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 also develops NWIPs explicitly targeting automotive needs. These IPs 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: TopoPiano and MagCAD. Different technologies are supported, and new ones could be added thanks to their modular structure. TopoPiano 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 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:
Toongiai Markus1, Markus Mueller2 and Ulrich Brunnings1

1University of Heseldel, DE; 2Excell 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

DISGUISSING THE INTERCONNECTS: EFFICIENT PROTECTION OF DESIGN IP

Authors:
Johann Knechtle1, Satwik Patnala2, Mohammed Ashraf1 and Ozgur Sinanoglu2

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 to protect 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:
Ashraf Sangupta1, Muhammad Yasir2, Mohammed Nabeel3, Mohammed Ashraf1, Jayavijayan Rajendran3 and Ozgur Sinanoglu2


Abstract
With the globalization of integrated circuit (IC) supply chain, the semi-conductor industry is facing a number of threats, such as Intellectual Property (IP) piracy, hardware Trojans, and counterfeiting. To defend against such threats 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 microprocessor in 65nm technology.

More information ...

UB03.5

RECONFIGURABLE SELF-TIMED DATAFLOW ACCELERATOR

Authors:
Daniil Sokolov, Alessandro de Gennaro and Andrey Movkhov, 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 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 Hilde1, David Halliday1, Andy Tyrrell1, Jon Timmis1, Junxiu Liu2, Shivan Kastm2, Jim Hark1 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 ...
RECOM: An Efficient Resistive Accelerator for Compressed Deep Neural Networks

Authors:
Houxiang Ji, Shanghai Jiao Tong University, CN

Abstract
Deep Neural Networks (DNNs) play a key role in prevailing machine learning applications. Resilient-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 inefficient 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, where 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)
**IP1-2**  
**SPARSENN: 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, Jingbo Jiang, Xi Li, and Chi-Ying Tsui, Hong Kong University of Science and Technology, HK  
**Abstract:** Contemporary Deep Neural Network (DNN) contains millions of synaptic connections with tens to hundreds of layers. 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, SparseNN, is proposed to exploit both the input and output sparsity. SparseNN is a 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, SparseNN can achieve a 10%-70% improvement in throughput and a power reduction of around 50%.  
**Download Paper (PDF; Only available from the DATE venue WiFi)**

**IP1-3**  
**ACCLIB: ACCELERATORS AS LIBRARIES**  
**Speaker:** Jacob R. Stevens, Purdue University, US  
**Authors:** Jacob Stevens1, Yue Du1, Vivek Kozhikkottu2, and Anand Raghunathan1  
1Purdue University, US; 2IBM, US; 3Intel 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 identify a computational target for acceleration, design a hardware accelerator, verify the correctness of the accelerator, integrate the accelerator into the system, and rewrite 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, nor does it leverage the growing body of already existing hardware accelerators. In this work, we propose ACCLIB, a design 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.  
**Download Paper (PDF; Only available from the DATE venue WiFi)**

**IP1-4**  
**HPXA: A HIGHLY PARALLEL XML PARSER**  
**Speaker:** Smitul Sarangi, IIT Delhi, IN  
**Authors:** Issa Ahmad, Sanjog Patil and Smitul 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 inito, to ensure that they can process multiple XML tokens and tags 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 1.06 Gbps, which is roughly 6.5X faster than competing prior work.  
**Download Paper (PDF; Only available from the DATE venue WiFi)**

**IP1-5**  
**QOR-AWARE POWER CAPPING FOR APPROXIMATE BIG DATA PROCESSING**  
**Speaker:** Shervel Reda, Brown University, US  
**Authors:** Seyed Morteza Nabavinejad1, Xin Zhan2, Reza Azimi, Maziar Goudarzi2, and Shervel Reda2  
1Sharif University of Technology, Iran; 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 finish 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)**

**IP1-6**  
**EXACT MULTI-OBJECTIVE DESIGN SPACE EXPLORATION USING ASPMT**  
**Speaker:** Kai Neubauer, University of Rostock, DE  
**Authors:** Kai Neubauer1, Philipp Wanko2, Torsten Schaub2, 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)**
IP1-7

**HEP: HMC INSTRUCTION PREDICTION 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 Almeida3 and Marco Antonio Zanata Alves2

1Federal University of Paraíba, BR; 2UFRGS, BR; 3UFPF, 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 interblock register bank. In this work, we propose HMC Instruction Prediction Extensions (HEP), 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 HIVE 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)

IP1-8

**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 equips 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)

IP1-9

**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 single 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)

IP1-10

**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, Shuachui 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)

IP1-11

**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 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  
**OHEX: OS-AWARE HYBRIDIZATION TECHNIQUES FOR ACCELERATING MPSOC FULL-SYSTEM SIMULATION**  
**Speaker:** Róbert Lajos Bócs, Institute for Communication Technologies and Embedded Systems, RWTH Aachen University, DE  
**Authors:** Róbert Lajos Bócs1, Maximilian Frick2, Rainer Reuplers1, Gerd Ascheid1, Stephan Tobias1 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 (OSs). 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)

IP1-13  
**A HIGHLY EFFICIENT FULL-SYSTEM VIRTUAL PROTOTYPE BASED ON VIRTUALIZATION-ASSISTED APPROACH**  
**Speaker:** Hsin-I Wu, National Tsing Hua University, Department of Computer Science, Hsinchu, Taiwan, TW  
**Authors:** Hsin-I Wu, Chi-Kang Chen, Tsung-Ying Lu and Ren-Song Tsay, National Tsing Hua University, TW  
**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.  
Download Paper (PDF; Only available from the DATE venue WiFi)

IP1-14  
**INDUSTRIAL EVALUATION OF TRANSITION FAULT TESTING FOR COST EFFECTIVE OFFLINE ADAPTIVE VOLTAGE SCALING**  
**Speaker:** Mahroo Zandrahimi, TU Delft, NL  
**Authors:** Mahroo Zandrahimi1, Philippe Debaud2, Armand Castillejo2 and Zaid Al-Ars1  
1Delft University of Technology, NL; 2STMicroelectronics, FR  
**Abstract**  
Adaptive voltage scaling (AVS) has been used widely to compensate for process, voltage, and temperature variations as well as power optimization of integrated circuits. The current industrial state-of-the-art AVS approaches using Process Monitoring 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.  
Download Paper (PDF; Only available from the DATE venue WiFi)

IP1-15  
**AN ANALYSIS ON RETENTION ERROR BEHAVIOR AND POWER CONSUMPTION OF RECENT DDR4 DRAMS**  
**Speaker:** Deepak M. Mathew, University of Kaiserslautern, DE  
**Authors:** Deepak M. Mathew1, Martin Schulte1, Carl C. Rheinländer1, Chirag Sudarshan1, Matthias Jung2, Christian Weis1 and Norbert Wehn1  
1University of Kaiserslautern, DE; 2Fraunhofer IESE, DE  
**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).  
Download Paper (PDF; Only available from the DATE venue WiFi)

IP1-16  
**A BOOLEAN MODEL FOR DELAY FAULT TESTING OF EMERGING DIGITAL TECHNOLOGIES BASED ON AMBIPOlar DEVICES**  
**Speaker:** Davide Bertozzi, DE - University of Ferrara, IT  
**Authors:** Marcello Dalpasso1, Davide Bertozzi2 and Michele Favalli2  
1DEI - UNIV. of Padova, IT; 2DE - Univ. of Ferrara, IT  
**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.  
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>
</tr>
</thead>
<tbody>
<tr>
<td>17:00</td>
<td>4.1.1</td>
<td>IMPROVING CIRCUIT SIZE UPPER BOUNDS USING SAT-SOLVERS</td>
</tr>
</tbody>
</table>

Speaker and Author:
Alexander Kulikov, Steklov Mathematical Institute at St. Petersburg, RU

Abstract
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.

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

| 17:30 | 4.1.2 | PRACTICAL EXACT SYNTHESIS |

Speaker:
Winston Haaswijk, EPFL, CH

Authors:
Mathias Soeken1, Winston Haaswijk1, Eleonora Testa1, Alan Mishchenko2, Luca G. Amanu2, Robert K. Brayton2 and Giovanni De Micheli1

1EPFL, CH; 2University of California, Berkeley, US; 3Synopsys Inc., US

Abstract
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.

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

| 18:00 | 4.1.3 | SAT-BASED REDUNDANCY REMOVAL |

Speaker and Author:
Krishnuni Deb Nath, Synopsys, IN

Abstract
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.

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.

4.2 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, Karlsruhe 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. The session highlights the usefulness of modelling to improve the efficiency of system-level design and simulation. The first paper presents a framework to generate accurate representative software the tolerance to disturbances provoked by neighboring readings of read-intensive applications in NAND flash.

4.2.1 APPROXIMATE COMPUTING FOR BIOMETRIC SECURITY SYSTEMS: A CASE STUDY ON IRIS SCANNING
Speaker: Sherief Reda, Brown University, US
Authors: Soheil Hashemi, Hochkhay Tann, Francesco Buttafuoco and Sherief Reda, Brown University, US
Abstract
Exploiting the error resilience of emerging data-rich applications, approximate computing promotes the introduction of small amount of inaccuracy into computing systems to achieve significant reduction in computing resources such as power, design area, runtime or energy. Successful applications for approximate computing have been 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 its 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 its 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 48x on top of an already HW accelerated design, while meeting industry-standard accuracy requirements for iris scanning systems.

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

4.2.2 FLASH READ DISTURB MANAGEMENT USING ADAPTIVE CELL BIT-DENSITY WITH IN-PLACE REPROGRAMMING
Speaker: Sung Ming Wu, 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%.

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

4.2.3 HTF-MPR: A HETEROGENEOUS TENSORFLOW MAPPER TARGETING PERFORMANCE USING GENETIC ALGORITHMS AND GRADIENT BOOSTING REGRESSORS
Speaker: Nadir Bagherzadeh, University of California, Irvine, US
Authors: Ahmad Albaqami, Manya S. Hosseini and Nadir 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 the 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.

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

1.13, respectively.

4.3 System Modelling for Simulation and Optimisation
Date: Tuesday, March 20, 2018
Time: 17:30 - 18:30
Location / Room: Konst. 1
Chair: Frederic Mallet, Universite Nice Cote d’Azur, FR, Contact Frederic Mallet
Co-Chair: Gianluca Palermo, Politecnico di Milano, IT, Contact Gianluca Palermo

The session highlights the usefulness of modelling to improve the efficiency of system-level design and simulation. The first paper presents a framework to generate accurate representative workloads from big-data applications. The second paper minimises the energy consumption by reducing accesses to off-chip memory for convolutional neural networks (CNNs). The third paper introduces compile-time analysis to improve parallel SystemC simulation.
SMARTSHUTTLE: OPTIMIZING OFF-CHIP MEMORY ACCESSES FOR DEEP LEARNING ACCELERATORS

Speaker: Guihai Yan, Institute of Computing Technology, Chinese Academy of Sciences, CN
Authors: Jialun Li, Guihai 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 of 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)
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 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.

**Abstract**

Byung Su Kim, Samsung Electronics, Foundry, KR

**Speaker**

Byung Su Kim, Samsung Electronics, Foundry, KR

**Authors**

Byung Su Kim and Joon-Sung Yang

**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 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'99 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)**

**Time** 17:30  **Label** 4.4.2  **Presentation Title** BAYESIAN THEORY BASED SWITCHING PROBABILITY CALCULATION METHOD OF CRITICAL TIMING PATH FOR ON-CHIP TIMING SLACK MONITORING  **Authors** Byung Su Kim, Samsung Electronics, Foundry, KR; Sungkyunkwan University, KR

**Abstract**

For this reason, we propose a novel timing slack monitoring methodology based on switching probability and correlation on critical timing 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'99 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)**

---

**Time** 18:00  **Label** 4.4.3  **Presentation Title** PERFORMANCE BASED TUNING OF AN INDUCTIVE INTEGRATED VOLTAGE REGULATOR DRIVING A DIGITAL CORE AGAINST PROCESS AND PASSIVE VARIATIONS  **Authors** Venkata Chaitanya Krishna Chekuri, 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: Danilo Pau, STMicroelectronics, IT; Contact Danilo Pau

Co-Chair: Lukasz Rybak, Mentor Graphics Poland, PL; Contact Lukasz Rybak

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

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 17:00 | 4.5.1 | PRE-ASSEMBLY TESTING OF INTERCONNECTS IN EMBEDDED MULTI-DIE INTERCONNECT BRIDGE (EMIB) DIES | Speaker: Krishnendu Chakraborty, Duke University, US

Authors: Sudipta Mondal and Krishnendu Chakraborty, 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) |
| 17:30 | 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 Kuentzzer, 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 Becker¹
¹University of Freiburg, DE; ²Rhinence Technologies, DE

Abstract: We present a novel algorithm for characterizing the set of possibly detected faults, i.e., the set of faults that can be detected with a non-zero probability. This algorithm is based on an analysis of the circuit's fault coverage, which is computed using a statistical model of the circuit's power consumption. The results of this analysis can be used to design more efficient test patterns and to improve the overall reliability of the circuit.

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

A BOOLEAN MODEL FOR DELAY FAULT TESTING OF EMERGING DIGITAL TECHNOLOGIES BASED ON AMBIPOLAR DEVICES

Speaker: Davide Bertozzi, DE - University of Ferrara, IT
Authors: Marcello Balpasso¹, Davide Bertozzi² and Michele Favat²
¹DEI, University of Padova, IT; ²DE - University of Ferrara, IT

Abstract: We propose a boolean model for delay fault testing of emerging digital technologies based on ambipolar devices. This model allows for the testing of devices with complex and high-performance requirements, which cannot be tested using traditional techniques. The model is validated using experiments on benchmark circuits, which show that it can accurately predict the detection probability of faults.

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

ATPG POWER GUARDS: ON LIMITING THE TEST POWER BELOW THRESHOLD

Speaker: Virendra Singh, Indian Institute of Technology Bombay, IN
Authors: Rohini Gulve¹ and Virendra Singh²
¹Indian Institute of Technology Bombay, IN; ²IT Bombay, IN

Abstract: We propose a novel algorithm for limiting the power consumption of ATPG tools during the generation of test patterns. The algorithm takes into account the current consumption of the circuit under test (CUT) and adjusts the test pattern generation accordingly to avoid exceeding a predefined power limit. The algorithm is validated using experiments on benchmark circuits, which show that it can significantly reduce the power consumption of ATPG tools without compromising the fault coverage.

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

IN-GROWTH TEST FOR MONOLITHIC 3D INTEGRATED SRAM

Speaker: Yuxin Zhang, Shanghai Jiao Tong University, CN
Authors: Pu Pang¹, Yuxin Zhang¹, Tanjian Li¹, Sung Kyo Lim², Quan Chen¹, Xiaoyao Liang¹ and Li Jiang¹
¹Shanghai Jiao Tong University, CN; ²Georgia Tech, US

Abstract: We propose a novel in-growth test method for monolithic 3D SRAMs. The method involves the fabrication of a stack of 3D-S RAM cells and the testing of each layer independently. The results show that the method is able to detect a high percentage of faults, and that it is more efficient than traditional test methods.

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

4.6 Special Session: Securing Power-constrained System-on-Chips: Challenges and Opportunities

Date: Tuesday, March 20, 2018
Time: 17:00 - 18:30
Location / Room: Conf. 4
Chair: Mukhopadhyay Sabal, School of ECE, Georgia Institute of Technology, US; Contact Sabal Mukhopadhyay

The power is the key constraint for modern System-on-Chip (SoC) design. After decade long research and development, the low-power circuit techniques and run-time power-management schemes are maturing at circuit, logic, and system levels. Now the SoCs face a new but critical design challenge: how to keep them secure - and techniques are being investigated in software and hardware to achieve this goal. This session is dedicated to study the critical, but often complex, interplay between power and security, in different aspects of design-for-security practices in SoCs.
**4.6.1 ULTRA-LOW ENERGY CIRCUIT BUILDING BLOCKS FOR SECURITY TECHNOLOGIES**

**Speaker:**
Sanu Mathew, Intel Corporation, US

**Authors:**
Sanu Mathew\(^1\), Sudhir Satpathy\(^2\), Víkram Suresh\(^2\) and Ram Krishnamurthy\(^2\)

\(^1\)Intel Labs, US; \(^2\)Intel Corporation, Hillsboro, US

**Abstract**
Low-area energy-efficient security primitives are key building blocks for enabling end-to-end content protection, user authentication in IoT platforms. This paper describes 3 designs that employ energy-efficient circuit techniques with optimal hardware-friendly arithmetic. The paper presents power analysis attacks on a new hybrid system-on-a-chip (SoC) architecture. It demonstrates the feasibility of using integrated voltage regulator and dynamic voltage frequency scaling for power management, resulting in 100% stable encryption key. All-digital TRNG to achieve >0.99 min-entropy with 3pJ/bit energy-efficiency.

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

**4.6.2 EMBEDDED RANDOMNESS AND DATA DEPENDENCIES DESIGN PARADIGM: ADVANTAGES AND CHALLENGES**

**Speaker:**
Itamar Levi, Université catholique de Louvain (UCL), Belgium, BE

**Authors:**
Itamar Levi, Alexander Fish and Osnat Keren, Faculty of Engineering, Bar-Ilan University, IL

**Abstract**
Information leakage through physical channels is a major hurdle in embedded hardware security. This paper overviews the three key factors in the embedded hardware security space, focusing on gray-box (bounded resources) power analysis attacks: the adversary's knowledge and abilities, the security metrics used by adversaries' and security evaluators and gate level countermeasures. A new design paradigm, dubbed pAsynch, that utilizes internal signals and random signals to uniformly spread the information-carrying energy within the clock period in a specific way with a resolution below the band-width and noise-filtering abilities of advanced measurement equipment is introduced. The advantages and design challenges introduced by the pAsynch paradigm are discussed.

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

**4.6.3 EXPLOITING ON-CHIP POWER MANAGEMENT FOR SIDE-CHANNEL SECURITY**

**Speaker:**
Saibal Mukhopadhyay, Georgia Institute of Technology, US

**Authors:**
Avinash Singh\(^1\), Monodeep Kar\(^1\), Sanu Mathew\(^2\), Anand Rajan\(^2\), Vivek De\(^2\) and Saibal Mukhopadhyay\(^1\)

\(^1\)Georgia Institute of Technology, US; \(^2\)Intel Labs, US; \(^3\)Intel Corporation, US

**Abstract**
The high-performance and energy-efficient encryption engines have emerged as a key component for modern System-On-Chip (SoC) in various platforms including servers, desktops, mobile, and IoT edge devices. A key bottleneck to secure operation of encryption engines is leakage of information through various side-channels. For example, an adversary can extract the secret key by performing statistical analysis on measured power and electromagnetic (EM) emission signatures generated by the hardware during encryption. Countermeasures to such side-channel attacks often come at high power, area, or performance overheads. Therefore, design of side-channel secure encryption engines is a critical challenge for high-performance and/or power/energy efficient operations. This paper reviews that although low-power requirement imposes critical challenge for side-channel security, but circuit techniques traditionally developed for power management also present new opportunities for side-channel resistance. As a case study, we review the feasibility of using integrated voltage regulator and dynamic voltage frequency scaling normally used for efficient power management, for increasing power-side-channel resistance of AES engines. The hardware measurement results from test-chip fabricated in 130nm process are presented to demonstrate the impact of power management circuits on side-channel security.

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

**4.6.4 ENERGY-SECURE SWARM POWER MANAGEMENT**

**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)**

**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 heterogeneous hardware to fit application requirements.
4.7.1 RESCUING MEMRISTOR-BASED COMPUTING WITH NON-LINEAR RESISTANCE LEVELS

Speaker:
Jilan Lin, Tsinghua University, CN
Authors:
Jilan Lin1, Liuxue Xia1, Zhenhua Zhu1, Hanbo Sun1, Yi Cai1, Hui Gao1, Ming Cheng3, Xiaoming Chen2, Yu Wang3 and Huazhong Yang2
1Tsinghua University, Beijing, CN; 2University of Notre Dame, US

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 nonlinear resistance levels. We classify the resistance level distributions in RRAM devices 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 typical 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)

4.7.2 PX-CGRA: POLYMORPHIC APPROXIMATE COARSE-GRAINED RECONFIGURABLE ARCHITECTURE

Speaker:
Omid Akbari, University of Tehran, IR
Authors:
Omid Akbari1, Mehdi Kamali1, Ali Atzall-Kusha1, Massoud Pedram2 and Mohammad Shafique2
1University of Tehran, IR; 2University of Southern California, US; 3TU Wien, AT

Abstract
Coarse-Grained Reconfigurable Architectures (CGRAs) provide tradeoff between the energy-efficiency of Application Specific Integrated Circuits (ASICs) and the flexibility of General Purpose Processors (GPPs). State-of-the-art CGRAs only support exact architectures and precise application executions. However, a majority of the streaming applications such as multimedia and digital signal processing, which are amenable to CGRAs, are inherently error- resilient. Therefore, these applications can greatly benefit from the emerging trend of Approximate Computing that leverages this error-resiliency to provide higher energy efficiency proportional to the tolerable accuracy loss (can even be constrained). This paper, for the first time, introduces the novel concept of Polymorphic Approximate CGRA (PX-CGRA) that employs heterogeneous tiles of Polymorphic-Approximated ALU Clusters (PACs) connected in a 2-D mesh style connection. These PACs can implement different approximate modes as well as accurate modes depending upon their selected configuration as per the run-time requirements of executing applications. For designing an efficient PX-CGRA, we propose a bottom-up design flow. In addition, the flow of application mapping on PX-CGRA is discussed including accuracy-level mapping, scheduling, and binding steps. To comprehensively evaluate the efficacy of the proposed CGRA, the complete PX-CGRA architecture in different sizes as well as with different PACs configurations are synthesized using a 15-nm FinFET technology. Our results show up to 15%-45% energy efficiency improvement for 5%-35% output quality degradation, respectively, when compared to the state-of-the-art exact-mode CGRA. Our proposed architecture and design methodology enable a new era of accuracy-configurable CGRAs to provide significant energy gains.

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

4.7.3 MULTI-PRECISION CONVOLUTIONAL NEURAL NETWORKS ON HETEROGENEOUS HARDWARE

Speaker:
Mohammad Hosseinabady, University of Bristol, GB
Authors:
Moslem Amiri, Mohammad Hosseinabady, Simon McIntosh-Smith and Jose Nunez-Yanez, University of Bristol, GB

Abstract
Fully binarised convolutional neural networks (CNNs) deliver very high inference performance using single-bit weights and activations, together with XNOR type operators for the kernel convolutions. Current research shows that full binarisation results in a degradation of accuracy and different approaches to tackle this issue are being investigated such as using more complex models as accuracy reduces. This paper proposes an alternative based on a multi-precision CNN framework that combines a binarised and a floating point CNN in a pipeline configuration deployed on heterogeneous hardware. The binarised CNN is mapped onto an FPGA device and used to perform inference over the whole input set while the floating point network is mapped onto a CPU device and performs re-inference only when the classification confidence is low. A light-weight confidence mechanism enables a flexible trade-off between accuracy and throughput. To demonstrate the concept, we choose a Zynq 7020 device as the hardware target and show that the multi-precision network is able to increase the BNN accuracy from 78.5% to 82.5% and the CPU inference speed from 29.68 to 90.82 images/sec.

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

4.7.4 LOGIC SYNTHESIS AND DEFECT TOLERANCE FOR MEMRISTIVE CROSSBAR ARRAYS

Speaker:
Onur Tunali, Istanbul Technical University, TR
Authors:
Onur Tunali and Mustafa Altun, Istanbul Technical University, TR

Abstract
Contrary to abundant memory related studies of memristor based crossbar architectures, logic oriented applications are only gaining popularity in recent years. In this paper, we study logic synthesis, regarding both two-level and multi level designs, and defect aspects of memristor based crossbar architectures. First, we introduce our two-level and multi-level logic synthesis techniques. We elaborate on advantages and disadvantages of both approaches with experimental results regarding area cost. After that, we devise a defect model in alignment with the conventional stuck-at open and closed paradigm. In addition, we determine the effects of defects to the operational capacity of the crossbar. Furthermore, we propose a preliminary defect tolerant Boolean logic mapping approach. In order to evaluate our approach, we conduct extensive Monte Carlo simulations with industrial benchmarks. Finally, we discuss future directions concerning both existing two-level and prospective multi-level logic designs as well as defect tolerance with area redundancy.

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

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:** Tuesday, March 20, 2018

**4.8 Components for Secure IoT Systems**

**Time** | **Label** | **Presentation Title** | **Authors**
--- | --- | --- | ---
18:30 | IP2-2, 199 | A CO-DESIGN METHODOLOGY FOR SCALABLE QUANTUM PROCESSORS AND THEIR CLASSICAL ELECTRONIC INTERFACE | Jeroen van Dijk, Delft University of Technology, NL
**Authors:** Jeroen van Dijk$^1$, Andrei Vladimirescu$^2$, Masoud Babaei$^1$, Edoardo Chainton$^1$ and Fabio Sebastiani$^1$

$^1$Delft University of Technology, NL; $^2$University 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)**

18:31 | IP2-3, 757 | APPROXIMATE QUATERNARY ADDITION WITH THE FAST CARRY CHAINS OF FPGAS | Philip Brisk, University of California, Riverside, US
**Authors:**

Sina Boroumand$^1$, Hadi P. Ataene$^2$ and Philip Brisk$^3$

$^1$University of Tehran, IR; $^2$Qualcomm Research, US; $^3$University 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 | Seongmin Hong, Hongik University, KR
**Authors:**

Seongmin Hong$^1$, Inho Lee$^2$ and Yongjun Park$^2$

$^1$Hongik University, KR; $^2$Hanyang 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, CNTN, 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)**

18:38 | 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.

**4.8.1 SECURING THE INTERNET OF THINGS WITH TI SIMPLELINK PLATFORM**

**Time** | **Label** | **Presentation Title** | **Authors**
--- | --- | --- | ---
17:00 | 4.8.1 | SECURING THE INTERNET OF THINGS WITH TI SIMPLELINK PLATFORM | 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-optimized 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 Dobiasl, 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 for 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 RoMulus [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 Hald, 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 RoMulus. 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.

A SCHMITT-TRIGGER BASED SUB-THRESHOLD DIGITAL CMOS CIRCUIT DESIGN TECHNIQUE FOR ULTRA LOW-VOLTAGE AND ULTRA LOW-POWER APPLICATIONS

Speaker: Matthias Keller, University of Freiburg, 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 RoMulus, 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.

The presentation provides an overview of our research activities on a sub-threshold digital circuit design technique for ultra low-voltage and ultra low-power applications. At first, the fundamentals of the Schmitt-Trigger based design methodology are presented that allow reducing the supply voltage of digital circuits to a few tens of millivolts in a 130nm CMOS technology. Subsequently, the BMBF-funded project RoMulus is considered in which the Schmitt-Trigger based design methodology is applied to the design of a digital standard cell library in an X-FAB 180nm CMOS process. Measurement results as well as an outlook on the next steps in the project conclude the presentation.

END OF SESSION

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.

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

Authors: Shahzad Muzaffar1 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 data and to provide structured feedback to the user. 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.

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

Authors: Tobias Markus1, Markus Mueller2 and Ulrich Brunner1

1University of Heidelberg, DE; 2Ettol 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 design time, and a robust 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 the only functional part that 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.
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.
Abstract

I will share

Finally, I will argue that the biggest source of pain for creators of microfluidic chips isn't the chip itself, but the off-chip equipment needed to control the chip, and I will share recent work by my group and others aimed at reducing or eliminating off-chip equipment by integrating its control functions onto the chip itself using "pneumatic logic."

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)

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

Coffee Break 15:30 - 16:00

Coffee Break 09:00 - 10:00

Coffee Break 16:00 - 17:00

Coffee Break 13:30 - 14:20

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

Coffee Break 10:00 - 11:00

Coffee Break 08:30 - 09:30

Coffee Break 08:30 - 10:00

Coffee Break 09:00 - 10:00

Coffee Break 10:00 - 11:00
Lunch Break 12:30 - 14:00
Coffee Break 15:30 - 16:00
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><strong>Speaker:</strong> Alma Proebstl</td>
<td>Authors: Alma Proebstl1, Sangyoung Park1, Swaminnath Narayanaswamy2, Sebastian Steinhorst2 and Samavjti Chakraborty2</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Authors:</strong> Alma Proebstl1, Sangyoung Park1, Swaminnath Narayanaswamy2, Sebastian Steinhorst2 and Samavjti Chakraborty2</td>
<td><strong>Affiliations:</strong> 1Technical University of Munich, DE; 2TUM CREATE, SG</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract:</strong> Short drive range due to limited battery capacity and high battery depreciation 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 prefering 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>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
<tr>
<td>09:00</td>
<td>5.2.2</td>
<td>GIS-BASED OPTIMAL PHOTOVOLTAIC PANEL FLOORPLANNING FOR RESIDENTIAL INSTALLATIONS</td>
<td>Sara Vino, Politecnico di Torino, IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Speaker:</strong> Sara Vino, Lorenzo Bottacchi, Eduardo Patti, Andrea Acquaviva, Enrico Maci and Massimo Poncino, Politecnico di Torino, IT</td>
<td>Authors: Sara Vino, Lorenzo Bottacchi, Eduardo Patti, Andrea Acquaviva, Enrico Maci and Massimo Poncino, Politecnico di Torino, IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract:</strong> Shading is a crucial issue for the placement of PV installations, as it heavily impacts power production and the corresponding return of 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.</td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
<tr>
<td>09:30</td>
<td>5.2.3</td>
<td>CELL-BASED UPDATE ALGORITHM FOR OCCUPANCY GRID MAPS AND HYBRID MAP FOR ADAS ON EMBEDDED GPUs</td>
<td>Jörg Fickenscher, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Speaker:</strong> Jörg Fickenscher, Jens Schlumberger, Frank Hanning, Mohamed Essayed Bouzoua and Jürgen Teich</td>
<td>Authors: Jörg Fickenscher1, Jens Schlumberger1, Frank Hanning1, Mohamed Essayed Bouzoua2 and Jürgen Teich1</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Authors:</strong> Jörg Fickenscher1, Jens Schlumberger1, Frank Hanning1, Mohamed Essayed Bouzoua2 and Jürgen Teich1</td>
<td><strong>Affiliations:</strong> 1Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU); 2Concept Development Automated Driving, AUDI AG, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract:</strong> 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.</td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
<tr>
<td>10:00</td>
<td>203</td>
<td>IMPROVING FAST CHARGING EFFICIENCY OF RECONFIGURABLE BATTERY PACKS</td>
<td>Alexander Lamprecht, TUM CREATE, SG</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Speaker:</strong> Alexander Lamprecht1, Swaminnath Narayanaswamy1 and Sebastian Steinhorst2</td>
<td>Authors: Alexander Lamprecht1, Swaminnath Narayanaswamy1 and Sebastian Steinhorst2</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Authors:</strong> Alexander Lamprecht1, Swaminnath Narayanaswamy1 and Sebastian Steinhorst2</td>
<td><strong>Affiliations:</strong> 1TUM CREATE, SG; 2Technical University of Munich, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Abstract:</strong> 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.</td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
</tbody>
</table>
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)

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.

5.3 Heterogeneous multi-level caching

Date: Wednesday, March 21, 2018
Time: 08:30 - 10:00
Location / Room: Konf. 1

Chair:
Jeronimo Castrillon, Technische Universität Dresden, DE, Contact Jeronimo Castrillon

Co-Chair:
Lei Ju, Shandong University, CN, Contact Lei Ju

This session discusses different aspects and optimization for multi-level caching, involving various aspects of non-volatile memory technologies and embedded systems. The first paper proposes a novel management of last-level-cache for PCM-based main memory. The next paper presents a multi-level cache architecture with time-randomized behaviour, amenable for WCET analysis. The last paper explores the trade-off between retention time, performance, and energy for STT-RAM-based caches.
WALL: A WRITEBACK-AWARE LLC MANAGEMENT FOR PCM-BASED MAIN MEMORY SYSTEMS

Speaker:
Bahareh Pourshirazi, University of Illinois at Chicago, US

Authors:
Bahareh Pourshirazi1, Majed Valad Beigi2, Zhichun Zhu1 and Gokhan Memik2
1University of Illinois at Chicago, US; 2Northwestern University, US

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 1.25×, on average, on an 8-core system with a 4GB PCM main memory, running memory intensive applications.

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

DESIGN AND INTEGRATION OF HIERARCHICAL-PLACEMENT MULTI-LEVEL CACHES FOR REAL-TIME SYSTEMS

Speaker:
Pedro Benedito, Barcelona Supercomputing Center and Universitat Politecnica de Catalunya, ES

Authors:
Pedro Benedito1, Carles Hernandez2, Jaume Abella3 and Francisco Cazorla4
1Barcelona Supercomputing Center and Universitat Politecnica de Catalunya, ES; 2Barcelona Supercomputing Center, ES; 3Barcelona Supercomputing Center (BSC-CNS), ES; 4Barcelona Supercomputing Center and IIIA-CSIC, ES

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 analyzability 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.

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

LARS: LOGICALLY ADAPTABLE RETENTION TIME STT-RAM CACHE FOR EMBEDDED SYSTEMS

Speaker:
Tosiron Adegbija, University of Arizona, US

Authors:
Kyle Kuan and Tosiron Adegbija, University of Arizona, US

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 relies 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 overhead 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.

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

WALL: A WRITEBACK-AWARE LLC MANAGEMENT FOR PCM-BASED MAIN MEMORY SYSTEMS

Authors:
Bahareh Pourshirazi, University of Illinois at Chicago, US

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 1.25×, on average, on an 8-core system with a 4GB PCM main memory, running memory intensive applications.

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

DESIGN AND INTEGRATION OF HIERARCHICAL-PLACEMENT MULTI-LEVEL CACHES FOR REAL-TIME SYSTEMS

Authors:
Pedro Benedito1, Carles Hernandez2, Jaume Abella3 and Francisco Cazorla4
1Barcelona Supercomputing Center and Universitat Politecnica de Catalunya, ES; 2Barcelona Supercomputing Center, ES; 3Barcelona Supercomputing Center (BSC-CNS), ES; 4Barcelona Supercomputing Center and IIIA-CSIC, ES

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 analyzability 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.

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

LARS: LOGICALLY ADAPTABLE RETENTION TIME STT-RAM CACHE FOR EMBEDDED SYSTEMS

Authors:
Kyle Kuan and Tosiron Adegbija, University of Arizona, US

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 relies 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 overhead 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.

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 12:30 - 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

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 sessions 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 forth 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.

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 technique, 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)
Device Attestation: Past, Present, and Future

Speaker: Yier Jin, University of Florida, US
Authors: Orlando Aramoo1, Dean Sullivan1, Fahim Rahman2, Mark M. Teranipoor3 and Yier Jin2
1University of Central Florida, US; 2University of Florida, US

Abstract
In recent years we have seen a rise in popularity of networked devices. From traffic signals in a city’s busiest intersection and energy metering appliances, to internet-connected security cameras, these embedded devices have become entrenched in everyday life. As a consequence, a need to ensure secure and reliable operation of these devices has also risen. Device attestation is a promising solution to the operational demands of embedded devices, especially those widely used in Internet of Things and Cyber-Physical System. In this paper, we summarize the basics of device attestation. We then present a summary of attestation approaches by classifying them based on their functionality and reliability guarantees they provide to networked devices. Lastly, we discuss the limitations and potential issues current mechanisms exhibit and propose new research directions.

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

A Reconfigurable Scan Network Based IC Identification for Embedded Devices

Speaker: Omid Aramoo, University of Maryland, US
Authors: Omid Aramoo1, Xi Chen1 and Gang Qu2
1University of Maryland, US; 2 Univ. of Maryland, College Park, US

Abstract
Most of the Internet of Things (IoT) and embedded devices are resource constrained, making it impractical to secure them with the traditional computationally expensive crypto-based solutions. However, security and privacy are crucial in many IoT applications such as health monitoring. In this paper, we consider one of the most fundamental security problems: how to identify and authenticate an embedded device. We consider the fact that embedded devices are designed by reusing IP cores with reconfigurable scan network (RSN) as the standard testing facility and propose to generate unique integrated circuit (IC) identifiers (IDs) based on different configurations for the RSN. These circuit IDs not only solve the IC and device identification and authentication problems, they can also be considered as a lightweight security primitive in other applications such as IC metering and IP fingerprinting. We demonstrate through the ITC’02 benchmarks that the proposed approach can easily create from 10^7 to 10^186 unique IDs without any overhead. Finally, our method complies with the IEEE standards and thus has high practical value.

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

Early Detection of System-Level Anomalous Behaviour Using Hardware Performance Counters

Speaker: Mark Zwolinski, University of Southampton, GB
Authors: Lai Leng Woo1, Basel Halak2 and Mark Zwolinski1
1Electronics and Computer Science Department, Southampton University, GB; 2Southampton University, GB; 3University of Southampton, GB

Abstract
Embedded systems suffer from reliability issues such as variations in temperature and voltage, single event effects and component degradation, as well as being exposed to various security attacks such as control hijacking, malware, reverse engineering, eavesdropping and many others. Both reliability problems and security attacks can cause the system to behave anomalously. In this paper, we will present a detection technique that is able to detect a change in the system before the system encounters a failure, by using data from Hardware Performance Counters (HPCs). Previously, we have shown how HPC data can be used to create an execution profile of a system based on measured events and any deviation from this profile indicates an anomaly has occurred in the system. The first step in developing a detector is to analyse the HPC data and extract the features from the collected data to build a forecasting model. Anomalies are assumed to happen if the observed value falls outside a given confidence interval, which is calculated based on the forecast values and prediction confidence. The detector is designed to provide a warning to the user if anomalies that are detected occur consecutively for a certain number of times. We evaluate our detection algorithm on benchmarks that are affected by single bit flip faults. Our initial results show that the detection algorithm is suitable for use for this kind of univariate time series data and is able to correctly identify anomalous data from normal data.

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
5.5 Emerging Technologies for Future Computing

Date: Wednesday, March 21, 2018
Time: 08:30 - 10:00
Location / Room: Konf. 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>
<tbody>
<tr>
<td>08:30</td>
<td>5.5.1</td>
<td>COMPACT MODELING OF CARBON NANOTUBE THIN FILM TRANSISTORS FOR FLEXIBLE CIRCUIT DESIGN</td>
<td>Leilai Shao, University of California, Santa Barbara, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>Leilai Shao1, Tsung-Ching Huang2, Ting Le3, Zhenan Bao4 and Tim Cheng5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1University of California Santa Barbara, US; 2Revlet Packard Labs, US; 3Stanford University, US; 4HPE Labs, US; 5HKUST, HK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>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.</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>5.5.2</td>
<td>A HIGH-SPEED DESIGN METHODOLOGY FOR INDUCTIVE COUPLING LINKS IN 3D-ICS</td>
<td>Benjamin Fletcher, University of Southampton, GB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>Benjamin Fletcher1, Shidhartha Das2 and Terence Mak3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1University of Southampton, GB; 2ARM Ltd., GB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>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.</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>5.5.3</td>
<td>AN EXACT METHOD FOR DESIGN EXPLORATION OF QUANTUM-DOT CELLULAR AUTOMATA</td>
<td>Marcel Walter, University of Bremen, DE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>Marcel Walter1, Robert Wille2, Daniel Grosse3, Frank Sill Tomes4 and Rolf Drechsler2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1University of Bremen, DE; 2Johannes Kepler University Linz, AT; 3University of Bremen/DFKI GmbH, DE; 4Federal University of Minas Gerais, BR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>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 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 the proposed method.</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>5.5.4</td>
<td>ACCURATE MARGIN CALCULATION FOR SINGLE FLUX QUANTUM LOGIC CELLS</td>
<td>Soheil Nazar Shahsavani, Bo Zhang and Massoud Pedram, University of Southern California, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Authors:</td>
<td>Soheil Nazar Shahsavani1, Bo Zhang and Massoud Pedram, University of Southern California, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1University of Southern California, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>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 10:30 - 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
# 5.6 Reliability improvement and evaluation techniques

**Date:** Wednesday, March 21, 2018  
**Time:** 08:30 - 10:00  
**Location / Room:** Konf. 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>
</table>
| 08:30 | 5.6.1 | IMPROVING RELIABILITY FOR REAL-TIME SYSTEMS THROUGH DYNAMIC RECOVERY | Yue Ma, University of Notre Dame, US  
Authors: Yue Ma\(^1\), Tam Chantam\(^2\), Robert P. Dick\(^3\) and Xiaobo Sharon Hu\(^1\)  
\(^1\)University of Notre Dame, US; \(^2\)Virginia Tech, US; \(^3\)University of Michigan, US  
**Abstract**  
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. |

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

| 09:00 | 5.6.2 | OPTIMAL METASTABILITY-CONTAINING SORTING NETWORKS | Johannes Bund, Saarland University, DE  
Authors: Johannes Bund\(^1\), Christoph Lenzen\(^2\) and Mor Medina\(^2\)  
\(^1\)Saarland University, Saarland Informatics Campus, DE; \(^2\)Max Planck Institute for Informatics, Saarland Informatics Campus, DE; \(^3\)From 1/10/2017 in The Department of Electrical and Computer Engineering Ben-Gurion University, IL  
**Abstract**  
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 & 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. |

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

| 09:30 | 5.6.3 | MAUI: MAKING AGING USEFUL, INTENTIONALLY | Shou-Chun Li, Department of Computer Science, National Chiao Tung University, TW  
Authors: Kai-Chiang Wu\(^1\), Tien-Hung Tseng\(^2\) and Shou-Chun Li\(^3\)  
\(^1\)National Chiao Tung University, Taiwan, TW; \(^2\)National Chiao Tung University, Taiwan, Taiwan  
**Abstract**  
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 its 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. |

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

| 09:45 | 5.6.4 | EXPERT: EFFECTIVE AND FLEXIBLE ERROR PROTECTION BY REDUNDANT MULTITHREADING | HeeSoo So, Yonsei University, KR  
Authors: HeeSoo So\(^1\), Moslem Didehban\(^2\), Yohan Ko\(^1\), Aviral Shrivastava\(^2\) and Kyungwoo Lee\(^1\)  
\(^1\)Yonsei University, KR; \(^2\)Arizona State University, US  
**Abstract**  
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. |

**Download Paper (PDF; Only available from the DATE venue WiFi)**
5.7 Software-centric techniques for embedded systems

**Title:** ETISS-ML: A MULTI-LEVEL INSTRUCTION SET SIMULATOR WITH RTL-LEVEL FAULT INJECTION SUPPORT FOR THE EVALUATION OF CROSS-LAYER RESILIENCY TECHNIQUES

**Authors:**

- Martin Dittrich, Technical University of Munich, DE
- Daniel Mueller-Gritschneder, Martin Dittrich, Josef Weinzierl, Eric Cheng, Subhasish Mitra and Ulf Schlüchtmann

**Abstract**

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 100x 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 platform for superscalar processors, which has a smart checkpointing mechanism to accelerate injection time, attenuating the shortcomings imposed by the aforementioned levels of accuracy: low-level HW-based methods are accurate, but very expensive, need special equipment and the actual hardware, and lack controllability; while high-level simulation-based strategies are flexible, fast, easily accessible and have high controllability, but are not accurate since they are based on models that do not always reflect the low-level implementation, mainly when it comes to complex designs like out-of-order multiple-issue processors. In this work, we propose a cycle-accurate fault injection platform for superscalar processors, which has a smart checkpointing mechanism to accelerate injection time, attenuating the shortcomings imposed by the aforementioned fault injection methods while providing the same level of abstraction as detailed RTL models. Leveraging from this new platform, we evaluate a complex and parameterizable Out-of-Order processor (BOOM) by experimenting with different issue widths and analyzing the sensitivity of several hardware structures of the processor.

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

**Authors:**

- Rafael Tonetto, Gabriel Luca Nazar and Antonio Carlos Schneider Beck

**Speaker:**

Antonio Carlos Schneider Beck, Federal University of Rio Grande do Sul, BR

**Presentation Title:** PRECISE EVALUATION OF THE FAULT SENSITIVITY OF OOO SUPERSCALAR PROCESSORS

**Abstract**

Since superscalar processors lead the market, their resiliency evaluation by means of fault injection grows in importance. Fault injection strategies usually trade-off their levels of accuracy: low-level HW-based methods are accurate, but very expensive, need special equipment and the actual hardware, and lack controllability; while high-level simulation-based strategies are flexible, fast, easily accessible and have high controllability, but are not accurate since they are based on models that do not always reflect the low-level implementation, mainly when it comes to complex designs like out-of-order multiple-issue processors. In this work, we propose a cycle-accurate fault injection platform for superscalar processors, which has a smart checkpointing mechanism to accelerate injection time, attenuating the shortcomings imposed by the aforementioned fault injection methods while providing the same level of abstraction as detailed RTL models. Leveraging from this new platform, we evaluate a complex and parameterizable Out-of-Order processor (BOOM) by experimenting with different issue widths and analyzing the sensitivity of several hardware structures of the processor.

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

**Authors:**

- Martin Dittrich, Technical University of Munich, DE
- Antonio Carlos Schneider Beck

**Speaker:**

Martin Dittrich, Technical University of Munich, DE

**Presentation Title:** TECHNIQUES

**Abstract**

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 100x 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 platform for superscalar processors, which has a smart checkpointing mechanism to accelerate injection time, attenuating the shortcomings imposed by the aforementioned levels of accuracy: low-level HW-based methods are accurate, but very expensive, need special equipment and the actual hardware, and lack controllability; while high-level simulation-based strategies are flexible, fast, easily accessible and have high controllability, but are not accurate since they are based on models that do not always reflect the low-level implementation, mainly when it comes to complex designs like out-of-order multiple-issue processors. In this work, we propose a cycle-accurate fault injection platform for superscalar processors, which has a smart checkpointing mechanism to accelerate injection time, attenuating the shortcomings imposed by the aforementioned fault injection methods while providing the same level of abstraction as detailed RTL models. Leveraging from this new platform, we evaluate a complex and parameterizable Out-of-Order processor (BOOM) by experimenting with different issue widths and analyzing the sensitivity of several hardware structures of the processor.

**Download Paper (PDF; Only available from the DATE venue WiFi)**
9:00  5.7.2  CIRCUIT CARVING: A METHODOLOGY FOR THE DESIGN OF APPROXIMATE HARDWARE

Speaker: Ilaria Scarabottolo, ISI Lugano, CH
Authors: Ilaria Scarabottolo, Giovanni Ansaboni and Laura Pozzi, ISI Lugano, CH
Abstract: 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-off. 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 combinatorial 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.
Download Paper (PDF; Only available from the DATE venue WiFi)

9:30  5.7.4  TASK SCHEDULING FOR MANY-CORES WITH S-NUCA CACHES

Speaker: Anuj Pathania, Kartlsruhe Institute of Technology, IN
Authors: Anuj Pathania and Joerg Henkel, Kartlsruhe Institute of Technology, DE
Abstract: 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.
Download Paper (PDF; Only available from the DATE venue WiFi)

9:45  5.7.5  KVSSD: CLOSE INTEGRATION OF LSM TREES AND FLASH TRANSLATION LAYER FOR WRITE-EFFICIENT KV STORE

Speaker: Sung-Ming Wu, National Chiao-Tung University, TW
Authors: Sung-Ming Wu, Kai-Heng Lin and Li-Pin Chang, National Chiao-Tung University, TW
Abstract: 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%.
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: Hyujong 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.

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

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.

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 12:00 - 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

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

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 - 14:00

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.
IP2-1  IN-GROWTH TEST FOR MONOLITHIC 3D INTEGRATED SRAM

Speaker:  Yixin Zhang, Shanghai Jiao Tong University, CN
Authors:  Pu Pang1, Yixin Zhang1, Tianjian Li1, Sung Kyu Lim2, Quan Chen1, Xiaoyao Liang1 and Li Jiang1
1Shanghai Jiao Tong University, CN; 2Georgia Tech, US

Abstract:  Monolithic three-dimensional integration (3DI) 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 (3DSI). However, the Known-Good-Die problem haunting 3DI-SI-a faulty tier causes the failure of the entire stack-also occurs in 3DI. Lack of efficient test methodologies such as the pre-bond testing in 3DI-SI, 3DI 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 3DI 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 prosaic judgment to determine whether or not to stop the fabrication, in order to prevent from raising the cost of fabricating more tiers upon the irreparable 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)

IP2-2  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, Edoardo Charbon1 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 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)

IP2-3  APPROXIMATE QUATERNARY ADDITION WITH THE FAST CARRY CHAINS OF FPGAS

Speaker:  Philip Brisk, University of California, Riverside, US
Authors:  Sina Boroumand1, Hadi P. Atash2 and Philip Brisk1
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)

IP2-4  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)

IP2-5  IMPROVING FAST CHARGING EFFICIENCY OF RECONFIGURABLE BATTERY PACKS

Speaker:  Alexander Lamprecht, TUM CREATE, SG
Authors:  Alexander Lamprecht1, Swaminathan Narayanaswamy1 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)
IP2-6 CLOUD-ASSISTED CONTROL OF GROUND VEHICLES USING ADAPTIVE COMPUTATION OFFLOADING TECHNIQUES

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

Authors:
Arun Adithiyan1, Ramesh SP and Scheil Sami2
1City University of New York, US; 2General Motors R&D, US

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 crewtensed data for path planning.

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

IP2-7 FUSIONCACHE: USING LLC TAGS FOR DRAM CACHE

Speaker:
Evangelos Vasilakis, Chalmers University of Technology, SE

Authors:
Evangelos Vasilakis1, Vassilis Papaefstathiou1, Pedro Trancoso1 and Ioannis Soudris1
1Chalmers University of Technology, SE; 2FORTH-ICS, GR

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 off-chip SRAM caches. Moreover, their tag access latency and storage overheads are excessive. Storing tags for a large DRAM cache in SRAM 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 off-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.

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

IP2-8 IMPROVED SYNTHESIS OF CLIFFORD+T QUANTUM FUNCTIONALITY

Speaker:
Philipp Niemann, German Research Center for Artificial Intelligence (DFKI GmbH), DE

Authors:
Philipp Niemann1, Robert Wille2 and Rolf Drechsler2
1Cyber-Physical Systems, DFKI GmbH, DE; 2Technische Universität Darmstadt, DE

Abstract
The Clifford+T library provides robust and fault-tolerant realizations for quantum computations. Consequently, (logic) synthesis of Clifford+T quantum circuits became an important research problem. However, previously proposed solutions are either only applicable to very small quantum systems or lead to circuits that are far from being optimal—mainly caused by a local, i.e., column-wise, consideration of the underlying transformation matrix to be synthesized. In this paper, we suggest an improved approach that considers the matrix globally and, by this, overcomes many of these drawbacks. Preliminary evaluations show the promise of this direction.

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

IP2-9 ENERGY-EFFICIENT CHANNEL ALIGNMENT OF DWDM SILICON PHOTONIC TRANSCIEVERS

Speaker:
Yuyang Wang, University of California, Santa Barbara, US

Authors:
Yuyang Wang1, M. Ashkan Seyedi2, Rui Wu1, Jared Hulme2, Marco Fiorentino2, Raymond G. Beausoleil3 and Kwang-Ting Cheng3
1University of California, Santa Barbara, US; 2Philips Research, NY; 3Hong Kong University of Science and Technology, HK

Abstract
The comb laser-driven microring based dense wavelength division multiplexing silicon photonics is a promising candidate for next-generation optical interconnects. However, existing solutions for exploring the power-performance trade-off of such systems have been restricted to a limited design space, resulting from the unnecessary constraints of using an identical spacing for laser comb lines and microring channels, and of utilizing combinatorial generation techniques for data transmission. We propose an energy-efficient channel alignment scheme that aligns the microring channels to a subset of laser comb lines that are non-uniformly distributed in the free spectrum range of the micrings. Based on a well-established process variation model, our simulations show that the proposed scheme significantly reduces the microring tuning power in the presence of denser comb lines. The power saved from microring tuning can improve the overall system energy efficiency despite some power wasted in unused laser comb lines. We further conducted a case study for design space exploration using the proposed channel alignment scheme, seeking the most energy-efficient configuration in order to achieve a target aggregated data rate.

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

IP2-10 A PHYSICAL SYNTHESIS FLOW FOR EARLY DWDM ENGINEERING OF SILICON NANOWIRE BASED RECONFIGURABLE FETS

Speaker:
Shubham Rai, Chair For Processor Design, CFAED, Technische Universität Dresden, Dresden, DE

Authors:
Shubham Rai1, Ansh Rupani2, Dennis Walter1, Michael Raitza1, André Heinzig3, Christian May1, Walter Weber4 and Akash Kumar1
1Technische Universität Dresden, DE; 2Bíblica Institute of Technology and Science Pilani, Hyderabad Campus, IN; 3ÍmLab GmbH, DE; 4NalLab gGmbH and CIAED, DE

Abstract
Silicon Nanowire based reconfigurable transistors (RFETs) provide an open gate terminal called the program gate which gives the freedom of programming p-type or n-type functionality for the same device at runtime. This enables the circuit designer to pack more functionality per computational unit. This saves processing costs as only one device type is required. No doping and associated lithography steps are needed for this technology. In this paper, we present a complete design flow for circuits based on SiNW RFETs. We propose layouts of logic gates, Liberty and LEF (libray extension format) files for the physical synthesis flow and make these available under an open source license to enable further research in the domain of these novel, functionally enhanced transistors. We develop a table model based on a transistor cell with relaxed dimensions following an SOI-based 22 nm technology having a gate pitch of 110 nm and modeled our logic gates on dual-gate RFETs. For the sake of comparison, we use the same tool flow for CMOS. We show that in the first of its kind comparison, for these fully symmetrical reconfigurable transistors, the area after placement and routing for SiNW based circuits is 17% more than that of CMOS for MCNC benchmark. Further, we discuss areas of improvement for obtaining better area results from the silicon nanowire based RFETs from a fabrication and technology point of view. The future use of self-aligned techniques to structure two independent gates within a smaller pitch holds the promise of substantial area reduction.

Download Paper (PDF; Only available from the DATE venue WiFi)
2.1GHz, AIM can speed up the encryption process by 80 times for a 1GB NVM. 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 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 non-volatile main memory-based systems pose an opportunity for an attacker to readily access sensitive information on the memory because of its long retention time. While real-time encryption memory on 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 low bandwidth intensive encryption application. We take advantage of NVMMs 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.
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 interferers 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)
More information ...

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

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

optimization algorithms, do not satisfy industrial requirements. A promising alternative is given by procedural approaches (also known as "generators"); They (a) emulate experts'

While digital design automation is highly developed, analog design automation still remains behind the demands. Previous circuit synthesis approaches, which are usually based on

Our contribution deals with a fully automated functional approximation methodology for combinational digital circuits. We present 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.fit.vutbr.cz/research/groups/ehw/approxlib/


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,

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.

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

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. Reconfiguration 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 ...

The flexibility of on-chip instrument access enabled by IEEE 1687 (IUTAG) has shown tremendous improvements in modern industrial designs. Due to a constantly increasing

spektrum 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 IUTAG 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 IUTAG test infrastructures.

More information ...

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 (Poell, 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 ...

12:00 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
- 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>Author: Michael Hoffmann, NaMLab gGmbH, DE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: A transistor looking from the gate essentially acts as a series combination of two capacitors: the gate oxide capacitor and the channel capacitor. When the gate oxide is an appropriate ferroelectric, this series combination can stabilize the ferroelectric material at a state of negative capacitance. At this state, the total capacitance of the series combination is enhanced, leading to more charge at the channel at the same voltage. This boost of charge, in turn, leads to larger current at the same voltage. In fact, this boost makes it possible to reduce supply voltage of transistors below the traditional Boltzmann limit --- often termed as the Boltzmann tyranny. In the recent years, many groups around the world, both in academy and in the industry, have demonstrated the fundamental effect and the Negative Capacitance Transistors. In this work, we shall describe the physical origin of the negative capacitance effect and our current understanding of the recent experimental work. We shall also discuss potential ways to optimize devices that could lead to significant improvement in energy 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>Author: Lian-Mao Peng, Peking University, CN</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Carbon nanotube (CNT)-based electronics has been considered one of the most promising candidates to replace Si complementary metal-oxide-semiconductor (CMOS) technology, which will soon meet its performance limit. Prototype device studies on individual CNTs revealed that CNT based devices have the potential to outperform Si CMOS technology in both performance and power consumption, especially at sub-10 nm technology nodes, which are close to the theoretical limits; and various optoelectronic device such as light-emitting diodes, photodetectors and photovoltaic (PV) cells have been demonstrated. In this talk, I will discuss the use of randomly oriented CNT film to build CNT CMOS and optoelectronic devices, and show that the performance of CNT film devices and systems can be dramatically improved by optimizing the material purity, device structure and fabrication processes, thus yie tasting CNT devices with outstanding performance comparable to that of Si CMOS and ICs working in the GHz regime, and integrated electronic and optoelectronic systems for communications between nanoelectronic circuits using 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. Resta\(^1\), Jorge Romero Gonzalez\(^2\), Tarun Kumar Agarwal\(^3\), Dennis Lin\(^3\), Francky Caethoven\(^3\), Iuliana P. Radu\(^3\), Giovanni De Micheli\(^1\) and Pierre-Emmanuel Gaillardon\(^4\)

\(^1\)Integrated System Laboratory – EPFL, CH; \(^2\)Laboratory of NanoIntegrated Systems (LNIS), Department of Electrical and Computer Engineering, University of Utah, US; \(^3\)IMEC, BE; \(^4\)University 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 MoS\(_2\) and WSe\(_2\), 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-WSe\(_2\) 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
- Keynote Lecture in “Saal 2” 13:20 - 13:50
- 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](mailto:)

**Co-Chair:**  
Todd Austin, University of Michigan, US, [Contact Todd Austin](mailto:)

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.
### 11:00 6.2.1 DYNAMIC SKEWED TREE FOR FAST MEMORY INTEGRITY VERIFICATION

**Speaker:** Saru Vig, Nanyang Technological University, SG

**Authors:** Saru Vig, Jiayuan Liu, and Lam Siow 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 well-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 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

**Speaker:** Cezar Rodolfo W. Reinbrecht, UFRGS, BR

**Authors:** Cezar Rodolfo Wedig Reinbrecht, Bruno Endres Forlin, Andreas Zankl, and Johanna Sepulveda

**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

**Speaker:** Jakob Stangl, Austrian Institute of Technology (AIT), AT

**Authors:** Jakob Stangl, Thomas Lorenz, and Sai Dinakar

**Abstract**

Outsourcing data into the cloud gives wide benefits and opportunities to customers. Beside 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 parametric 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

**Speaker:** Jingtong Hu, University of Pittsburgh, US

**Authors:** Ming Hu, University of Pittsburgh, US; University of California, Santa Barbara, US; Huawei Technologies, China, CN

**Abstract**

Non-volatile main memory-based systems pose an opportunity for an attacker to readily 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)

### 12:31 IP2-16, 748 SAT-BASED BIT-FLIPPING ATTACK ON LOGIC ENCRYPTIONS

**Speaker:** Hai Zhou, Northwestern University, US

**Authors:** Yuang Qi Shen, Amin Rezaei, and Hai Zhou, Northwestern University, US

**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 SARKock 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-flipping attacks 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)
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

Date: Wednesday, March 21, 2018
Time: 11:00 - 12:30
Location / Room: Konf. 1

Chair: Marie-Minerve Louerat, LIP6, FR, Contact Marie-Minerve Louerat

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.

11:00 6.3.1 ENHANCED ANALOG AND RF IC SIZING METHODOLOGY USING PCA AND NSGA-II OPTIMIZATION KERNEL
Speaker: Nuno Lourenço, Instituto de Telecomunicações, PT
Authors: Tiago Pessoa, Nuno Lourenço, Ricardo Martins, Ricardo Povo and Nuno Horta, Instituto de Telecomunicações /Instituto Superior Técnico – Universidade de Lisboa, PT
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) using 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.

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

11:30 6.3.2 A SYSTEMC-BASED SIMULATOR FOR DESIGN SPACE EXPLORATION OF SMART WIRELESS SYSTEMS
Speaker: Gabriele Morandi, University of Verona, IT
Authors: Gabriele Morandi1, Francesco Stefanni2, Federico Fraccaroli3 and Davide Quaglia1
1University of Verona, IT; 2EDA Lab s.r.l., IT; 3Wagoo Italia s.r.l.s., IT
Abstract
Smart wireless techniques are at the core of many today's telecommunication and networked embedded systems. Performance improvements in software-defined radio frequency (RF) and digital aspects are therefore highly desired by both the 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, cost-performance is limited by the computational time-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.

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 Aghiaie, Ralf Wunderlich and Stefan Heinrich, 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 interferers 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 the 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-continuous transmitter to detect the effects on system level.

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

### Testbench Qualification for SystemC-AMS Timed Data Flow Models

**Speaker:** Muhammad Hassan, DFKI GmbH, DE

**Authors:** Muhammad Hassan¹, Daniel Grosse², Hoang M. Le², Thilo Voertler², Karsten Einwich¹ and Rolf Drechsler²

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.

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

### 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)**
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” 13:50 - 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

6.4 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.

STABILITY-AWARE INTEGRATED ROUTING AND SCHEDULING FOR CONTROL APPLICATIONS IN ETHERNET NETWORKS

Rouhollah Mahfouzi, Linköping University, SE; Amir Aminifar, Swiss Federal Institute of Technology in Lausanne (EPFL), CH; General Motors Research & Development, US

Authors: Rouhollah Mahfouzi1, Amir Aminifar2, Soheil Samii1, Ahmed Rezine1, Petru Elea1 and Zebo Peng1
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 jitter. 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 Cong1, Kun Cao1, Junlong Zhuo2, Tongquan Wei1, Mingsong Chen1 and Xiaobo Sharan Hu2
1East China Normal University, CN; 2Nanjing University of Science and Technology, CN; 3University of 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 Chakraborty2
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 IP9-3, 135  TTW: A Time-Triggered Wireless Design for CPS
Speaker: Romain Jacob, ETH Zurich, CH
Authors: Romain Jacob1, Leciong Zhang2, Marco Zimmerling3, Jan Beutel1, Samarjit Chakraborty2 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 DPP, the closest related work, validating the suitability of TTW for new emerging wireless CPS applications.
Download Paper (PDF; Only available from the DATE venue WiFi)

12:31 IP9-4, 429  Phylax: Snapshot-Based Profiling of Real-Time Embedded Devices via JTAG Interface
Speaker: Eduardo Chielle, New York University Abu Dhabi, UBR
Authors: Charalambos 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 applied 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 IP9-5, 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: 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 toaperiodically 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).

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.

**Abstract**

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 dataset. 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, 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 Tiny YOLO 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 2.08 Wh, 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).

Coffee Break 10:30 - 11:30
Lunch Break 13:00 - 14:30
Coffee Break 16:00 - 17:00

Coffee Break 10:00 - 11:00
Lunch Break 12:30 - 14:00
Coffee Break 16:00 - 17:00

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

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

6.8 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.
**22FDX Ultra-Low-Voltage Design Based on Adaptive Body Bias**

**Speaker:** Holger Eisenreich, Racyics GmbH, DE

**Abstract**

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.

---

**A New ADAS Chip Design in 22 NM FDSOI Technology for Automotive Computer Vision Applications**

**Speaker:** Jens Benndorf, Dream Chip Technologies, DE

**Abstract**

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:

- Short company introduction
- Project setup and target applications
- Chip architecture and performance requirements
- Hardware Demonstration System
- Outlook
- Summary

---

**Accelerating Physical Signoff for Leading Edge Chip Designs**

**Speaker:** David DeMarcos, Synopsys, DE

**Abstract**

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 fixing within the IC Compiler and IC Compiler II environments.

---

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

**UB06 Session 6**

**Date:** Wednesday, March 21, 2018

**Time:** 12:00 - 14:00

**Location / Room:** Booth 1, Exhibition Area
UB06.1 SYSTEM-LEVEL OPERATING CONDITION CHECKS: AUTOMATED AUGMENTATION OF VERILOGAMS MODELS
Authors: Georg Gläser¹, Martin Grabmann¹, Gent Kropp¹ and Andreas Fürtig¹
¹Institut für Mikroelektronik- und Mechatronik-Systeme gemeinnützige GmbH, DE; ²Goethe University Frankfurt, DE
Abstract
System-Level Operating Condition Checks are a critical step in hardware implementation flow. Applying these checks manually, however, can be a time-consuming and error-prone process. We present a method for automatically augmenting VerilogAMS models with operating condition checks.

More information ...

UB06.2 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 present TopolinoNano, a design and simulation framework for emerging technologies. The framework is based on a modular design flow that allows for the exploration of new technologies and the integration of existing ones.

More information ...

UB06.3 ADVANCED SIMULATION OF QUANTUM COMPUTATIONS
Authors: Zulehner Albin and Robert Wille, Johannes Kepler University Linz, AT
Abstract
Quantum computing is a rapidly growing field, and computer simulation is a crucial tool for its development. In this contribution, we propose a novel simulation methodology that allows for the exploration of quantum algorithms and their implementation on emerging quantum computing hardware.

More information ...

UB06.4 ROS X FPGA FOR ROBOT-CLOUD SYSTEM: ROBOT-CLOUD COOPERATIVE VISUAL SLAM PROCESSING USING ROS-COMPATIBLE FPGA COMPONENT
Authors: Takanori Ohkawa, Yuhei Sugata, Aoi Soya, Kanemitsu Ootsu and Takashi Yokota, Utsunomiya University, JP
Abstract
We present a novel system for robot-cloud visual SLAM processing using ROS-compatible FPGA components. The system allows for efficient and real-time processing of large-scale environments.

More information ...

UB06.5 WIRELESS SENSOR SYSTEM WITH ELECTROMAGNETIC ENERGY HARVESTER FOR INDUSTRY 4.0 APPLICATIONS
Authors: Bianca Leistritz, Elena Chesvokova, 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, the system can adapt to the changing conditions of the environment.

More information ...

UB06.6 TTOOL/OMC: OPTIMIZED COMPILED COMPILATION OF EXECUTABLE UML/SYSML DIAGRAMS FOR THE DESIGN OF DATA-FLOW APPLICATIONS
Authors: Andrea Emili¹, Julien Lallet¹, Renaud Pacement² and Ludovic Avrile²
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., FPGAs). Efficiently programming these architectures is not trivial due to the intrinsic 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 ...

UB06.7 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 modeling and predictive thermal management, operating across multiple applications. Platform-independent runtime management is demonstrated using two heterogeneous platforms.

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 too often inhibited by the inherent differences between IP-XACT and the traditional hardware description languages. In this work, we take 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...

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...

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 10:00 - 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

7.0 LUNCH TIME KEYNOTE SESSION: From Inverse Design to Implementation of Robust and Efficient Photonics for Computing

Date: Wednesday, March 21, 2018
Time: 13:50 - 14:20
Location / Room: Saal 2

Chair:
Ayse Coskun, Boston University, US, Contact Ayse Coskun

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 is 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.
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 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

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

GENERATING FULL-CUSTOM SCHEMATICS IN A MIXED-SIGNAL TOP-DOWN DESIGN FLOW
Authors:
Tobias Markus, Markus Mueller and Ulrich Bruening
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.
More information ...

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

Authors:  
Takeshi Ohkawa, Yuhei Sugata, Aoi Soya, Kenanitsu Ootsu and Takashi Yokota, Utsumoninya University, JP  

Abstract  
Distributed processing in robot-cloud cooperative 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-hardwired 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 50W. (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 Chervakova, 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 benefits 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 managed 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-ISAS 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 Dameshkal2, Mikael Ekstrom2 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 GIMM2E2 [1], a high-throughput, and cost efficient FPGA-based stereo vision system. In the next step, we present a novel accelerator which is also compatible with GIMM2E2. Our accelerator tries to map neural network (NN) based image classification algorithms to FPGA by using DeepMarker which is an evolutionary based module embed in our accelerator that regenerates a near-optimal NN in term of accuracy. Then, the back-end side of DeepMarker maps the generated NN to FPGA. We will demo a GIMM2E2-based accelerator for image classification applications.

More information ...

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

Authors:  
Martin Schoeberl, 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 Aglo 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 processor 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. 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 IPs, and 3) automated design porting due to highly technology-independent blocks down to 22nm.

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...

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 (Pcell, Pycell). 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 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 these 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).

Time Label Presentation Title Authors
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
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>Trinayan Baruah, Northeastern University, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Trinayan Baruah, Northeastern University, US</td>
<td>Authors: Trinayan Baruah1, Yitan Sun1, Shi Dong1, David Kaeli1 and Norm Rubin2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1Northeastern University, Boston, US; 2RV/DIA, US</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, along with this increased gain in performance, power and energy management is a larger challenge. In this paper, we address the issues 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. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
</tr>
<tr>
<td>15:00</td>
<td>7.2.2</td>
<td>ALL-DIGITAL EMBEDDED METERS FOR ON-LINE POWER ESTIMATION</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 William Fornaciari, Politecnico di Milano, IT</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 prediction error smaller than 3%, with limited area and power overheads. Download Paper (PDF; Only available from the DATE venue WiFi)</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>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 Samaeit Chakraborty, Technical University of Munich, DE</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. Download Paper (PDF; Only available from the DATE venue WiFi)</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>Hanchen Yang1, Feiyang Kang2, Guowei Ding3, Ji Li4, Jaemin Kim5, Dongyu Baek6, Shahin Nazarian1, Xue Lin7, Paul Bogdan4 and Naehyuck Chang8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Hanchen Yang1, Feiyang Kang2, Guowei Ding3, Ji Li4, Jaemin Kim5, Dongyu Baek6, Shahin Nazarian1, Xue Lin7, Paul Bogdan4 and Naehyuck Chang8</td>
<td>Authors: 1Beijing University of Posts and Telecommunications, CN; 2Zhejiang University, CN; 3Yracuse University, US; 4University of Southern California, US; 5Seoul National University, KR; 6Korea Advanced Institute of Science and Technology, KR; 7Northeastern University, US; 8RAIST, KR</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. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
<tr>
<td>16:00</td>
<td>IP3-6, 554</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: Xue Lin, Northeastern University, US; Feiyang Kang2, Guowei Ding3, Ji Li4, Jaemin Kim5, Dongyu Baek6, Shahin Nazarian1, Paul Bogdan4 and Naehyuck Chang8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract: Thermoelectric generator has increasingly drawn attention for being environmentally friendly. However, only a few of the prior researches on thermoelectric generators 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. Moreover, the prediction 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. Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
</tbody>
</table>
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 Zhao⁰ and Zhuofan 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)
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.

7.3 Advances in Logic Synthesis and Technology Mapping

Presentation Title
Authors

<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>7.3.1</td>
<td>IMPROVEMENTS TO BOOLEAN RESYNTHESIS</td>
<td></td>
</tr>
</tbody>
</table>

Speaker:
Mathias Soeken, EPFL, CH

Authors:
Luca Amaru¹, Mathias Soeken², Patrick Vuillod³, Jiong Luo⁴, Alan Mishchenko⁴, Janet Olson⁴, Robert Brayton⁴ and Giovanni De Micheli⁵

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)
15:00 7.3.2 LOGIC OPTIMIZATION WITH CONSIDERING BOOLEAN RELATIONS
Speaker:
Chia-Cheng Wu, National Tsing Hua University, TW
Authors:
Tung-Yuan Lee¹, Chia-Cheng Wu¹, Chia-Chun Lin², Yung-Chih Chen² and Chun-Yao Wang⁴
¹National Tsing Hua University, TW; ²Yuan Ze University, TW; ³Dept. CS, National Tsing Hua University, TW
Abstract
Logic optimization considering BR can exploit the potential flexibility existed in logic networks to minimize the circuits. In this paper, we present a logic optimization approach considering BR. The approach identifies a proper sub-circuit and locally changes its functionality by solving the corresponding BR in the sub-circuit without altering the overall functionality of the circuit. We conducted experiments on a set of MCNC benchmarks that cannot be further optimized by resyn2 script in ABC. The experimental results show that the node counts of these benchmarks can be further reduced.
Additionally, when we apply our approach followed by the resyn2 script repeatedly, we can obtain 6.11% improvements in average.
Download Paper (PDF; Only available from the DATE venue WiFi)

16:00 8.4.2 APPROXIMATE HARDWARE GENERATION USING SYMBOLIC COMPUTER ALGEBRA EMPLOYING GRÖBNER BASIS
Speaker:
Saman Froehlich, DFKI GmbH, DE
Authors:
Saman Froehlich¹, Daniel Grosse² and Rolf Drechsler²
¹Cyber-Physical Systems, DFKI GmbH, DE; ²University of Bremen/DFKI GmbH, DE
Abstract
Many applications are inherently error tolerant. Approximate Computing is an emerging design paradigm, which gives the opportunity to make use of this error tolerance, by trading off 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 wrt. applicationspecific input probabilities. To the best of our knowledge taking input probabilities, which are known for many practical applications, into account has not been considered before. We employ the proposed approach to generate 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)

15:30 7.3.3 TECHNOLOGY MAPPING FLOW FOR EMERGING RECONFIGURABLE SILICON NANOWIRE TRANSISTORS
Speaker:
Shubham Rai, Chair For Processor Design, CFAED, Technische Universität Dresden, Dresden, DE
Authors:
Shubham Rai, Michael Raitza and Akash Kumar, Technische Universität Dresden, DE
Abstract
Efficient circuit designs can make use of ambipolar nature of silicon nanowire (SiNW) over CMOS. Conventional circuit Design-Flow fails to use this inherent functional flexibility as CMOS based mapping considers a single logical output from logic gates. To address this, we propose an area optimized technology mapping which uses this innate reconfigurability, offered by SiNW transistors for efficient circuit designs. To enable this objective, we use higher order functions (HOF) to encapsulate this extended functionality. Additionally, the electrical properties of SiNW allow us to take advantage of the available inverted forms of fan-ins for additional savings of area for XOR logic family. Experimental results using our technology mapping shows that area of SiNW based logic design is less by an average of 18.36% as compared to CMOS flow for complete mcnc benchmarks suite. Further, we evaluate our flow for both reconfigurability-aware and static layout for SiNW based logic gates. The whole flow including the new SiNW based genlib and the modified ABC tool is made available under open source license to enable further research for any kind of emerging ambipolar transistors.
Download Paper (PDF; Only available from the DATE venue WiFi)

16:01 9.6.4 RECONFIGURABLE IMPLEMENTATION OF $GF(2^m)$ BIT-PARALLEL MULTIPLIERS
Speaker and Author:
José L. Imaña, Complutense University of Madrid, ES
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 $\{y^i + y^{i+2} + y^{i+1} + y^n + 1\}$ 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$-$times$@$time parameter when compared with similar multipliers found in the literature.
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

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)
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.

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 10:00 - 11:00
Coffee Break 12:30 - 14:00
Coffee Break 16:00 - 17:00
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 provides aging mitigation schemes for voltage regulators. The third paper addresses program vulnerability in GPU applications.

### Time Table

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Speaker</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 14:30  | 7.5.1 | LOW-COST HIGH-ACCURACY VARIATION CHARACTERIZATION FOR NANOSCALE IC TECHNOLOGIES VIA NOVEL LEARNING-BASED TECHNIQUES | Zhijian Pan | Tsinghua University, CN  
Zhijian Pan¹, Mao Li², Jian Yao², Hong Lu², Zuchang Ye¹, Yanfeng Li² and Yan Wang¹  
¹Tsinghua University, CN; ²Platform Design Automation, Inc., CN |
|        |       | **Abstract**                                                                        |         | Faster and more accurate variation characterizations of semiconductor devices/circuits are in great demand as process technologies scale down to Fin-FET era. Traditional methods with intensive data testing are extremely costly. In this paper, we propose a novel learning-based high-accuracy data pre-diction framework inspired by learning methods from computer vision to efficiently characterize variations of device/circuit behaviors induced by manufacturing process variations. The key idea is to adaptively learn the underlying data pattern among data with variations from a small set of already obtained data and utilize it to accurately predict the unmeasured data with minimum physical measurement cost. To realize this idea, novel regression modeling techniques based on Gaussian process regression and partial least squares regression with feature extration and matching are developed. We applied our approach to real-time variation characterization for transistors with multiple geometries from a foundry 28nm CMOS process. The results show that the framework achieves about 14x time speed-up with on average 0.1% error for variation data prediction and under 0.3% error for statistical estimation compared to traditional physical measurements, which demonstrates the efficacy of the framework for accurate and fast variation analysis and statistical modeling. |

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

| 15:00  | 7.5.2 | MITIGATION OF NBTI INDUCED PERFORMANCE DEGRADATION IN ON-CHIP DIGITAL LDOS           | Longfei Wang | University of South Florida, US  
Longfei Wang¹, S. Karen Khatamfard², Ulya Kaspuzu² and Selauck Kose³  
¹University of South Florida, US; ²University of Minnesota, US; ³University of South Florida, US |
|        |       | **Abstract**                                                                        |         | On-chip digital low-dropout voltage regulators (LDOs) have recently gained impetus and drawn significant attention for integration within both mobile devices and microprocessors. Although the benefits of easy integration and fast response speed surpass analog LDOs and other voltage regulator types, NBTI induced performance degradation is typically overlooked. The conventional bi-directional shift register based controller can even exacerbate the degradation, which has been demonstrated theoretically and through practical applications. In this paper, a novel uni-directional shift register is proposed to evenly distribute the electrical stress and mitigate the NBTI effects under arbitrary load conditions with nearly no extra power and area overhead. The benefits of the proposed design as well as reliability aware design considerations are explored and highlighted through simulation of an IBM POWER8 like processor under several benchmark applications. It is demonstrated that the proposed NBTI-aware design can achieve up to 43.2% performance improvement as compared to a conventional one. |

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

| 15:30  | 7.5.3 | EVALUATING THE IMPACT OF EXECUTION PARAMETERS ON PROGRAM VULNERABILITY IN GPU APPLICATIONS | Fritz Previlon | Northeastern University, US  
Fritz Previlon¹, Charu Kalra¹, Paolo Rech² and David Kaeli³  
¹Northeastern University, US; ²Universidade Federal do Rio Grande do Sul, BR; ³University of Minnesota, US |
|        |       | **Abstract**                                                                        |         | While transient faults continue to be a major concern for the High Performance Computing (HPC) community, we still lack a clear understanding of how these faults propagate in applications. This paper addresses two particular aspects of the vulnerabilities of HPC applications as run on Graphics Processing Units (GPUs): their dependence on input data and on thread-block size. To characterize fault propagation as a function of input parameters, we leverage an ISA-level fault injection framework and carry out an extensive fault injection campaign to characterize the vulnerability of a suite of GPU applications. Our results show that the vulnerability of most of the programs studied is insensitive to changes in input values, except in less common cases when input values were highly-biased, i.e., values that exhibit a special vulnerability behavior. For example, the multiplication property of any value with a zero value (zero times any number is equal to zero) makes it a biased input for multiplication operations. Our study also examines the effects of changing the GPU thread-block size and its impact on vulnerability. We found that, similar to performance, the vulnerability of an application can depend on the block size of the kernels in the application. In some applications, we found that the silent data corruption rate can vary by as much as 8% when changing the block size of a kernel. |

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

| 16:00  |       | AN EFFICIENT NBTI-AWARE WAKE-UP STRATEGY FOR POWER-GATED DESIGNS                     | Yu-Guang Chen | Yuan Ze University, TW  
Yu-Guang Chen¹, Yu-Guang Chen² and Ing-Chao Lin³  
¹National Cheng Kung University, TW; ²Yuan Ze University, TW; ³National Cheng Kung 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

Speaker: Nacer-Eddine Zergainoh, TIMA, FR
Authors: Thierry Bonnot, Faidy 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)

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

7.6 Special Session: Next Generation Processors and Architectures for Deep Learning

Date: Wednesday, March 21, 2018
Time: 14:30 - 16:00
Location / Room: Konf. 4

Chair: Theocharis Theocharides, University of Cyprus, CY, Contact Theocharis Theocharides

Co-Chair: Shatique Muhammad, TU Wien, AT, Contact Shatique Muhammad

Machine Learning is nowadays embedded in several computing devices, consumer electronics and cyber-physical systems. Smart sensors are deployed everywhere, in applications such as wearables and perceptual computing devices, and intelligent algorithms power the so-called “Internet of Things”. Similarly, smart cyber-physical systems emerge as a vital computation paradigm in a vast application spectrum ranging from consumer electronics to large-scale complex critical infrastructures. The need for smartification of such systems, and intelligent data analytics (especially in the era of Big Data), emphasizes the need of revolutionizing the way we build processors and systems geared towards machine learning (and deep learning in particular). Issues related from memory, to interconnect, and spanning across the hardware and software spectrum, need to be addressed typically by advances in technology, design methodologies, and new programming paradigms among others. The emergence of powerful embedded devices and ultra-low-power hardware has enabled us to transfer the paradigm of deep learning architectures and systems, from high-end costly clusters/supercomputers, to affordable systems and even mobile devices. Such systems and devices have not received the required attention so far in research and development, until the last few years, when such systems where initially proposed to accelerate deep learning, providing previously unattainable levels of performance of such algorithms, whilst maintaining the power and reliability constraints imposed by the nature of these embedded applications. This special session aims to present a holistic overview of emerging works in such architecture and systems, using all available technology spectrums, and bring together views from academia and industry in order to exchange information and explain how we can take advantage of existing and emerging hardware technologies in addressing the associated challenges.
7.6.1 RERAM-BASED ACCELERATOR FOR DEEP LEARNING

Speaker: Hai Li, Duke university, US
Authors: Li Bing1, Linghao Song2, Fan Chen2, Xuehai Qian2, Yiran Chen2 and Hai (Helen) Li4

Abstract
Big data computing applications such as deep learning and graph analytic usually incur a large amount of data movements. Deploying such applications on conventional von Neumann architecture that separates the processing units and memory components likely leads to performance bottleneck due to the limited memory bandwidth. A common approach is to develop architecture and memory co-design methodologies to overcome the challenge. Our research follows the same strategy by leveraging resistive memory (ReRAM) to further enhance the performance and energy efficiency. Specifically, we employ the general principles behind processing-in-memory to design efficient ReRAM based accelerators that support both testing and training operations. Related circuit and architecture optimization will be discussed too.

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

7.6.2 EXPLOITING APPROXIMATE COMPUTING FOR DEEP LEARNING ACCELERATION

Speaker: Jungook Choi, IBM Research, US
Authors: Cha-Yu Chen, Jungwook Choi, Kalilash Gopalakrishnan, Viji Srinivasan and Swagath Venkataramani, IBM T. J. Watson Research Center, US

Abstract
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 (jemAdaComp), and computation cost for deep learning inference via Prameterized clipping (jem 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.

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

7.6.3 AN OVERVIEW OF NEXT-GENERATION ARCHITECTURES FOR MACHINE LEARNING: ROADMAP, OPPORTUNITIES AND CHALLENGES IN THE IOT ERA

Speaker: Muhammad Shafique, Vienna University of Technology (TU Wien), AT
Authors: Muhammad Shafique1, Theodoris Theocharides2, Christos Bouganis2, Muhammad Abdullah Hanif3, Faiz Khalid Loth4, Rehan Hafiz2 and Sameen Rehman3
1TU Wien, AT; 2University of Cyprus, CY; 3Imperial College London, GB; 4Department of Computer Engineering, Vienna University of Technology, AT; 5TU, PK

Abstract
The number of connected Internet of Things (IoT) devices are expected to reach over 20 billion by 2020. These range from basic sensor nodes that log and report the data to the ones that are capable of processing the incoming information and taking an action accordingly. Machine learning, and in particular deep learning, is the de facto processing paradigm for intelligently processing these immense volumes of data. However, the resource inhibited environment of IoT devices, owing to their limited energy budget and lower compute capabilities, render them a challenging platform for deployment of desired data analytics. This paper provides an overview of the current and emerging trends in designing highly efficient, reliable, secure and scalable machine learning architectures for such devices. The paper highlights the focal challenges and obstacles being faced by the community in achieving its desired goals. The paper further presents a research roadmap that can help in addressing the highlighted challenges and thereby designing scalable, high-performance, and energy efficient architectures for performing machine learning on the edge.

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

7.6.4 INFERENCe OF QUANTIZED NEURAL NETWORKS ON HETEROGENEOUS ALL-PROGRAMMABLE DEVICES

Speaker: Thomas Preußer, Xilinx Inc., IE
Authors: Thomas Preußer1, Giulio Gambardella2, Nicholas Fraser2 and Michaela Blott3
1Technische Universität Dresden, DE; 2Xilinx Research Labs, IE; 3Xilinx, IE

Abstract
Neural networks have established as a generic and powerful means to approach challenging problems such as image classification, object detection or decision making. Their successful deployment rests on an enormous demand of compute. The quantity of network parameters and the processed data has proven a valuable measure to reduce the challenges of network inference so effectively that the feasible scope of applications is expanded even into the embedded domain. This paper describes the making of a real-time object detection in a live video stream processed on an embedded all-programmable device. The presented case illustrates how the required processing is tamed and parallelized across both the CPU cores and the programmable logic and how the most suitable resources and powerful extensions, such as NEON vectorization, are leveraged for the individual processing steps. The crafted result is an extended Darknet framework implementing a fully integrated, end-to-end solution from video capture over object annotation to video output applying neural network inference at different quantization levels running at 16-frames per second on an embedded Zynq UltraScale+ (XCGZU3EG) platform.

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

7.7 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.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00</td>
<td>Coffee Break in Exhibition Area</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30</td>
<td>7.7.1</td>
<td>CHASE: CONTRACT-BASED REQUIREMENT ENGINEERING FOR CYBER-PHYSICAL SYSTEM DESIGN</td>
</tr>
<tr>
<td></td>
<td>Speaker:</td>
<td>Pierlugi Nuzzo, University of Southern California, US</td>
</tr>
<tr>
<td></td>
<td>Authors:</td>
<td>Pierlugi Nuzzo1, Michele Lora2, Yishai Feldman3 and Alberto Sangiovanni-Vincentelli4</td>
</tr>
<tr>
<td></td>
<td>1University of Southern California, US; 2University of Verona, IT; 3IBM Research, Haifa, IL; 4University of California at Berkeley, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abstract</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:00</td>
<td>7.7.2</td>
<td>RESILIENCE EVALUATION VIA SYMBOLIC FAULT INJECTION ON INTERMEDIATE CODE</td>
</tr>
<tr>
<td></td>
<td>Speaker:</td>
<td>Hoang M. Le, University of Bremen, DE</td>
</tr>
<tr>
<td></td>
<td>Authors:</td>
<td>Hoang M. Le1, Vladimir Herdt2, Daniel Grosse2 and Rolf Drechsler2</td>
</tr>
<tr>
<td></td>
<td>1University of Bremen, DE; 2University of Bremen/DFKI GmbH, DE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abstract</td>
<td>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 (SWFI) techniques on compiler's intermediate code, this paper proposes a novel resilience evaluation framework combining LLVM-based SWFI and SMT-based symbolic execution. This novel combination offers significant advantages over state-of-the-art approaches with respect to accuracy and coverage.</td>
</tr>
<tr>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
</tbody>
</table>
ONLINE ANALYSIS OF DEBUG TRACE DATA FOR EMBEDDED SYSTEMS

Speaker:
Philip Gottschling, TU Darmstadt, DE

Authors:
Normann Decker¹, Boris Dreyer², Philip Gottschling², Christian Hochberger³, Alexander Lange³, Martin Leucker³, Simon Wegener⁴ and Alexander Weiss⁵
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

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 16:30 - 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.

Time | Label | Presentation Title
---|---|---
14:30 | 7.8.1 | 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.
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**

**Authors:**
Muhammad Hassan, DFKI GmbH, DE
José Medeiros, University of Brasilia, BR
Muhammad Hassan1, Daniel Grosse2, Hoang M. Le3, Thilo Voelter4, Karsten Einwich5 and Roland Drechsle6
1Cyber Physical Systems, DFKI, DE; 2University of Bremen/DFKI GmbH, DE; 3University of Bremen, DE; 4COSEDA Technologies GmbH, DE

**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)**

---

**IP3-2**
**An Algebra for Modeling Continuous Time Systems**

**Authors:**
José Medeiros, University of Brasilia, BR
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**

**Authors:**
Romain Jacob, ETH Zurich, CH
Romain Jacob1, Licong Zhang2, Marco Zimmerling3, Jan Beutel4, Samarjit Chakraborty5 and Lothar Thiele1
1ETH Zurich, CH; 2Technical University of Munich, DE; 3Technische 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 still cannot 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**

**Authors:**
Eduardo Chielle, New York University Abu Dhabi, BR
Charalambos Konstantinou1, Eduardo Chielle1 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 relays 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**

**Authors:**
Chung-Ta King, National Tsing Hua University, TW
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)**
IP3-6  PREDICTION-BASED FAST THERMOELECTRIC GENERATOR RECONFIGURATION FOR ENERGY HARVESTING FROM VEHICLE RADIATORS

Speaker: Xue Lin, Northeastern University, US
Authors: Hanchen Yang1, Feiyang Kang2, Caiwen Feng3, Daheng Yue1, Zhenyu Zhao1 and Zhuofan Liao2
1Beijing University of Posts and Telecommunications, CN; 2National University of Defense Technology, CN; 3National University of Science and Technology, CN
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. More precisely, with O(N) time complexity, 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.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP3-7  A PARAMETERIZED TIMING-AWARE FLIP-FLOP MERGING ALGORITHM FOR CLOCK POWER REDUCTION

Speaker: Chaocao Feng, National University of Defense Technology, CN
Authors: Chaocao Feng1, Daheng Yue1, Zhenyu Zhao1 and Zhuofan 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, both 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 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 better 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)

IP3-8  FAST CHIP-PACKAGE-PCB COANALYSIS METHODOLOGY FOR POWER INTEGRITY OF MULTI-DOMAIN HIGH-SPEED MEMORY: A CASE STUDY

Speaker: Seungwon Kim, Ulsan National Institute of Science and Technology, KR
Authors: Seungwon Kim1, Ki Jin Han2, Youngmin Kim3 and Seokkyeong 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 LLPDDR4 environment.
Download Paper (PDF; Only available from the DATE venue WiFi)

IP3-9  APPROXIMATE HARDWARE GENERATION USING SYMBOLIC COMPUTER ALGEBRA EMPLOYING GRÖBNER BASIS

Speaker: Saman Froehlich, DFKI GmbH, DE
Authors: Saman Froehlich1, Daniel Grosse2 and Rolf Drechsler2
1Cyber-Physical Systems, DFKI GmbH, DE; 2University of Bremen/DFKI GmbH, DE
Abstract
Many applications are inherently error tolerant. Approximate Computing is an emerging design paradigm, which gives the opportunity to make use of this error tolerance, by trading off 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 algorithm 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 w.r.t. application-specific input probabilities. To the best of our knowledge taking input probabilities, which are known for many practical applications, into account has not been considered before. We employ the proposed approach to generate 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)

IP3-10  RECONFIGURABLE IMPLEMENTATION OF $GF(2^m)\times$ 8 BIT-PARALLEL MULTIPLIERS

Speaker and Author: José L. Imaña, Complutense University of Madrid, ES
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 $f(y) = y^m + y^{n+2} + y^{n+1} + y^n + 1$ $y^n + 1$ are presented. These pentanomials are important because all five binary fields recommended by NIST for ECDSA can be constructed using such pentanomials. In this work, a new approach for $GF(2^m)\times$ 8 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 area8 time8 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: Muhammad Abdullah Hanif, Vienna University of Technology, Vienna, AT
Authors: Muhammad Abdullah Hanif1, Rehan Hafiz2 and Muhammad Shafique1
TU Wien, AT; ITU, PK
Abstract
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 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: Kun-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 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)

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 Nicolaides, 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 Schoeberl, 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 Hanif1, Vienna University of Technology, Vienna, AT
Authors: Muhammad Abdullah Hanif1, Rehan Hafiz2 and Muhammad Shafique1
TU Wien, AT; ITU, 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:
Semreen Rehan, Vienna University of Technology (TU Wien), AT

Authors:
Bharath Srinivas Prabakaran1, Semreen Rehan1, Muhammad Abdullah Hanif1, Saim Ullah2, Ghazali 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 area, 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 approximate 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, synthesized 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%, 38%, and 53%, respectively. We also compare our approximate adders to that of the art approximate adders specialized for ASIC and FPGA fabrics and demonstrate the benefits of our approach. We will make the RTL and behavioral models of our approach available.

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 Moazzemi 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)
Projected energy efficiency benefits are significant. Hardware prototypes demonstrate the feasibility of such nanosystems.

This session presents end-to-end approaches for building nanosystems that enable new applications by connecting nanotechnologies for logic, memory and sensing with new architecture.

Mohamed M. Sabry Aly, Nanyang Technological University, SG, 8.1 Special Day Session on Future and Emerging Technologies: NanoSystems: Connecting Devices, Architectures, and Applications

Wednesday, March 21, 2018
17:00 - 18:30
Saal 2

18:00 End of session

3D NANOSYSTEMS: THE PATH TO 1,000X ENERGY EFFICIENCY

Max Shulaker, MIT, US

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.

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 ºC), thermal stability (>500 ºC) 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 biomimicry.

RESISTIVE RAM FOR NEW COMPUTING SYSTEMS: FROM DEEP LEARNING TO BIOMIMICRY

Elisa Vianello, CEA LETI, FR

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 ºC), thermal stability (>500 ºC) 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 biomimicry.

EXPRESS BASED AUTOMATION OF ANALOG IC DESIGN

Florian Leber and Juergen Scheible, Reutlingen University, DE

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".
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 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 are both more data efficient, in terms of learning, and specifically well suited for energy-aggressive hardware architectures. Illustrations will be presented based on integrated LAE-CMOS sensory-compute system prototypes.
### 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, Dpto 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.

### 17:00 ONLINE EFFICIENT BIO-MEDICAL VIDEO TRANSCODING ON MPSOCs THROUGH CONTENT-AWARE WORKLOAD ALLOCATION

**Speaker:** Aman Irandar, Embedded Systems Lab (ESL), EPFL, CH  
**Authors:** Aman Irandar1, Ali Pahlevan1, Marina Zapater1, Martin Žagar2, 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 anundenable 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 (MPSoC)s 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 bi-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)**

### 17:30 HIGHLY EFFICIENT AND ACCURATE SEIZURE PREDICTION ON CONSTRAINED IOT DEVICES

**Speaker:** Farzad Same, Karlsruhe Institute of Technology (KIT), DE  
**Authors:** Farzad Same, 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 make 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 processing capability, 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

**Date:** Wednesday, March 21, 2018  
**Time:** 17:00 - 18:30  
**Location / Room:** Konf. 2

**Chair:** Göhringer Diana, Technische Universität Dresden, DE, Contact Diana Goehringer  
**Co-Chair:** Jie Han, University of Alberta, CA, Contact Jie Han

This session covers the exploitation of techniques to improve energy efficiency and resilience in memory and computing architectures. The first paper proposes a method using hybrid vertex-edge memory hierarchy to reduce the energy consumption of resistive random-access memory (ReRAM) systems. This method significantly improves the energy efficiency compared to conventional DRAM-based systems. The second paper examines the use of neural networks to improve resilience. The third paper addresses the performance bottleneck in von Neumann architectures by proposing an efficient algorithm for matrix multiplication in the memory of resistive associative processors (ReAPs). Finally, the last paper covers run-time application mapping on manycore DRAM-based systems.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 18:00 | 8.3.3 | A WEARABLE LONG-TERM SINGLE-LEAD ECG PROCESSOR FOR EARLY DETECTION OF CARDIAC ARRHYTHMIA | Muhammad Awais Bin Altaf, Lahore University of Management Sciences (LUMS), PK  
Syed Muhammad Abubakar, Wala Saadah and Muhammad Awais Bin Altaf, Lahore University of Management Sciences (LUMS), PK |
|      |       | **Abstract** | Cardiac arrhythmia (CA) is one of the most serious heart diseases that lead to a very large number of annual casualties around the world. The traditional electrocardiography (ECG) devices usually fail to capture arrhythmia symptoms during patients’ hospital visits due to their recurrent nature. This paper presents a wearable long term single-lead ECG processor for the CA detection at an early stage. To achieve on-sensor integration and long-term continuous monitoring, an ultra-low complexity feature extraction engine using reduced feature set of four (RFS4) is proposed. It reduces the area by >25% compared to the conventional QRS complex detection algorithms without compromising the accuracy. Moreover, RFS4 eliminates the need for complex machine-learning decision logics for the detection of premature ventricular contraction (PVC) and nonsustained ventricular tachycardia (NVT). To ensure core functional verification, the proposed system is implemented on FPGA and tested using the MIT-BIH ECG arrhythmia database. It achieves a sensitivity and specificity of 94.64% and 99.41%, respectively. The proposed processor is also synthesized using 0.18um CMOS technology with an overall energy efficiency of 1.39 nJ/detection. |
| 18:15 | 8.3.4 | DRONET: EFFICIENT CONVOLUTIONAL NEURAL NETWORK DETECTOR FOR REAL-TIME UAV APPLICATIONS | Christos Kyrkou, University of Cyprus, KIOS CoE, CY  
Muhammad Abdullah Hanif, Vienna University of Technology, Vienna, AT |
|      |       | **Abstract** | 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 exploring their resilience characteristics in a systematic way. |
| 18:30 | IP3-889 | ERROR RESILIENCE ANALYSIS FOR SYSTEMATICALLY EMPLOYING APPROXIMATE COMPUTING IN CONVOLUTIONAL NEURAL NETWORKS | Muhammad Abdullah Hanif1, Rehan Hefzi2 and Muhammad Shafique1  
1TU Wien, AT; 2TU, 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 exploring their resilience characteristics in a systematic way. |
| 18:30 | End of session | | |
18:00 8.4.4 HIMAP: A HIERARCHICAL MAPPING APPROACH FOR ENHANCING LIFETIME RELIABILITY OF DARK SILICON MANYCORE SYSTEMS

Speaker: Vivek Chaturvedi, Nanyang Technological University, SG

Authors: Naveen Rathore1, Vivek Chaturvedi1, Amit Kumar Singh2, Thambipillai Srikantian3, Rohit R1, Siew Kei Lam4 and Muhammad Shafique2

1Nanyang Technological University, SG; 2University of Essex, GB; 3TU Wien, AT

Abstract

Technology scaling into the nano-scale CMOS regime has resulted in increased leakage and readback on voltage scaling, which has led to several issues like high power density and elevated on-chip temperature. This consequently aggravates device aging, compromising lifetime reliability of the manycore systems. The proposed approach combines two levels: (1) it identifies a region of cores suitable for mapping, and (2) it maps threads in the region and interconnects cores for thermal mitigation while considering the current heat of the cores. Both the levels strive to reduce aging variance across the chip. We evaluated HIMAP for 64-core and 256-core systems. Results demonstrate an improved system lifetime reliability by up to 2 years at the end of 3.25 years of use, as compared to the state-of-the-art.

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

18:15 8.4.4 RAPID IN-MEMORY MATRIX MULTIPLICATION USING ASSOCIATIVE PROCESSOR

Speaker: Hasan Erdem Yantir, University of California Irvine, US

Authors: Naveen Rathore1, Vivek Chaturvedi1, Amit Kumar Singh2, Thambipillai Srikantian3, Rohit R1, Siew Kei Lam4 and Muhammad Shafique2

1Nanyang Technological University, SG; 2University of Essex, GB; 3TU Wien, AT

Abstract

Memory hierarchy latency is one of the main problems that prevents processors from achieving high performance. To eliminate the need of loading/storing large sets of data, Resistive Associative Processors (ReAP) have been proposed as a solution to the von Neumann bottleneck. In ReAPs, logic and memory structures are combined together to allow in-memory computations. In this paper, we propose a new algorithm to compute the matrix multiplication inside the memory that exploits the benefits of ReAP. The proposed approach is based on the Cannon algorithm and uses a series of rotations without duplicating the data. It runs in O(n), where n is the dimension of the matrix. The method also applies to a large set of row by column matrix-based applications. Experimental results show several orders of magnitude increase in performance and reduction in energy and area when compared to the latest FPGA and CPU implementations.

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

18:30 8.3 IP-16 DEMAS: AN EFFICIENT DESIGN METHODOLOGY FOR BUILDING APPROXIMATE ADDERS FOR FPGA-BASED SYSTEMS

Speaker: Sumeen Rehman, Vienna University of Technology (TU Wien), AT

Authors: Bharath Srinivas Prabakaran1, Sumeen Rehman1, Muhammad Abdullah Hanifi2, Salim Ulha2, Ghazal Mazaheer2, Akash Kumar2 and Muhammad Shafique1

1TU Wien, AT; 2Technische Universität Dresden, DE; 3TU 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 area, 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 generic 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 18-bit accurate adder, our designs are successful in achieving area, latency and power/energy product gains of 50%, 38%, and 53%, respectively. We also compare our approximate adders to state-of-the-art approximate 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)
### 8.3 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.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 17:00 | IP3-17, 515 | GAIN SCHEDULED CONTROL FOR NONLINEAR POWER MANAGEMENT IN CMPS | Niki Dutt, University of California, Irvine, US  
**Authors:** Bryan Donyanavard, Amir M. Rahmani, Tiago Muck, Kasra Moazzeni and Niki 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. |

---

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 17:15 | IP4-1, 376 | EFFICIENT MAPPING OF QUANTUM CIRCUITS TO THE IBM QX ARCHITECTURES | Alwin Zulehner, Johannes Kepler University Linz, AT  
**Authors:** Alwin Zulehner, Alexandru Paler and Robert Wille, Johannes Kepler University Linz, AT  
**Abstract:** 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 be properly configured 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. |

---

<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>End of session</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Amit Singh, University of Essex, GB,

Co-Chair: Oliver Bringmann, Universität Tübingen, DE,

Chair: Thomas Maurin, CEA, Leti, Univ. Grenoble Alpes, FR

Location / Room: Konf. 4

Time: 17:00 - 18:30

Date: Wednesday, March 21, 2018

17:30 8.5.3 A CASE STUDY FOR USING DYNAMIC PARTITIONING BASED SOLUTION IN VOLUME DIAGNOSIS

Speaker: Tao Wang, Mentor, A Siemens Business, US

Authors: Tao Wang1, Zhangzhun Shi1, Junlin Huang1, Huaxing Tang2, Wu Yang2 and Junna Zhong2

1Helicon Technologies Co., Ltd, CN; 2Mentor, A Siemens Business, US; 2Mentor, 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 ONLINE 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, Yassin Al Hamza3, Gero Hinz4, Georg Nüenemberg5, Markus Rickert6 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 Ducruex1, George Caraiman2 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
18:30 8.6.1 SUPPORTING RUNTIME RECONFIGURABLE VLIW CORES THROUGH DYNAMIC BINARY TRANSLATION

Speaker: Simon Rokicki, Univ Rennes, INRIA, CNRS, IRISA, FR
Authors: Simon Rokicki1, Erven Rohou2 and Steven Derrien3
1IRISA, FR; 2Inria, FR; 3University of Rennes 1/IRISA, FR
Abstract
Single ISA Heterogeneous multicore such as the ARM big.LITTLE have proven to be an attractive solution to explore different energy/performance trade-offs. Such architectures combine Open Order 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 programing model, we resort to Dynamic Binary Translation, and use this technique to enable dynamic code specialization for runtime reconfigurable VLIW cores. Our proposed DRT framework targets the RISC-V ISA, for which both OoO and in-order implementations exist. Our experimental results show that this 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)

17:30 8.6.2 USFI: ULTRA-LIGHTWEIGHT SOFTWARE FAULT ISOLATION FOR IOT-CLASS DEVICES

Speaker: Zelalem Abeke, University of Michigan, US
Authors: Zelalem Beyhanu Abeke and Todd Austin, University of Michigan, US
Abstract
Embedded device security is a particularly difficult challenge, as the quantity of devices makes them attractive targets, 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 sandboxing 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 rudimentary memory protection support found in modern IoT-class microcontrollers to build a low-profile, low-overhead, flexible sandboxing mechanism that can provide isolation between tightly-coupled software modules.
With our approach, named USFI, only the trust management code must be trusted. Through the use of a static verifier and monitored inter-module transitions, code at all privilege levels (including the kernel) is able to run uninstrumented and untrusted code. We implemented USFI on an ARMv7-M based processor, both bare metal and 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)

18:00 8.6.3 CONVERGING SAFETY AND HIGH-PERFORMANCE DOMAINS: INTEGRATING OPENMP INTO ADA

Speaker: Sara Royuela, Barcelona Supercomputing Center, ES
Authors: Sara Royuela1, Eduardo Quinones2 and Luis Miguel Pinho2
1Barcelona Supercomputing Center, ES; 2Polytechnic 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)

18:15 8.6.4 COMPILER-DRIVEN ERROR ANALYSIS FOR DESIGNING APPROXIMATE ACCELERATORS

Speaker: Jorge Castro-Godinez, Chair for Embedded Systems (CES), Kaisruhe Institute of Technology (KIT), DE
Authors: Jorge Castro-Godinez1, Sven Esser1, Muhammad Shafique2, Santiago Pagani3 and Joerg Henkel3
1Karlsruhe Institute of Technology, DE; 2TU Wien, AT
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) by making approximations. 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 SEAD, 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)

18:31 IP4-2, 105 PARALLEL CODE GENERATION OF SYNCHRONOUS PROGRAMS FOR A MANY-CORE ARCHITECTURE

Speaker: Amaury Graillat, Verimag - Univ. Grenoble Alpes, FR
Authors: Amaury Graillat1, Matthieu Myo2, Pascal Raymond2 and Benoît Dupont de Dinechin4
1Verimag - Univ. Grenoble Alpes, FR; 2Univ. Grenoble Alpes, Verimag, FR; 3Verimag/CNRS, FR; 4Kalray, FR
Abstract
Embedded 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)
### BEST-IN-CLASS RF INTEGRATED CIRCUITS FOR MULTI-GBPS COMMUNICATION IN 22FDX

**Speaker:** Corrado Carta, Technical University Dresden, DE

**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 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 50 \( \Omega \). Error-free (BER < 10^{-12}) transmission at 30 Gb/s is demonstrated. By using a switched output stage, only a small static dc-power is consumed, resulting in the most energy-efficient MZM driver with only 2.2 pJ/bit.

LARA has been used to hide the glue code insertion, thus separating the pure functional application description from extra-functional requirements. The MARGOT has been used for the automatic selection of the best configuration according to the runtime evaluation of the application. To demonstrated the effectiveness of the proposed approach, we evaluated LARA 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)
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

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

Ahmed Jerraya, CEA, FR,

Location / Room:

Time:

Date:

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
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 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 datathroughput advances 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.

### Time  Presentation Title Authors

**9:00** | **9.2.2**  
**A TRANSPRECISION FLOATING-POINT PLATFORM FOR ULTRA-LOW POWER COMPUTING**

**Speaker:** Giuseppe Taglavini, Università di Bologna, IT  
**Authors:** Giuseppe Taglavini1, Stefano Mach2, Davide Rossetti2, Andrea Marongiu2 and Luca Benini1  
**Abstract:**
In modern low-power embedded platforms, the execution of floating-point (FP) operations engenders as a major contributor to the energy consumption of compute-intensive applications with large dynamic range. Experimental evidence shows that 50% of the energy consumed by a core and its data memory is related to FP computations. The adoption of FP formats requiring a lower number of bits is an interesting opportunity to reduce energy consumption, since it allows to simplify the arithmetic circuitry and to reduce the memory bandwidth required to transfer data between memory and registers by enabling vectorization. From a theoretical look, the adoption of multiple FP types perfectly fits with the principle of transprecision computing, allowing fine-grained control of approximation while meeting specified constraints on the precision of final results. In this paper we propose an extended FP type system with complete hardware support to enable transprecision computing on low-power embedded processors, including two standard formats (binary32 and binary16) and two new formats (binary8 and binary16alt). First, we introduce a software library that enables exploration of FP types by tuning both precision and dynamic range of program variables. Then, we present a methodology to integrate our library with an external tool for precision tuning, and experimental results that highlight the clear benefits of introducing the new formats. Finally, we present the design of a transprecision FP unit capable of handling 8-bit and 16-bit operations in addition to standard 32-bit operations. Experimental results on FP-intensive benchmarks show that up to 90% of FP operations can be safely scaled down to 8-bit or 16-bit formats. Thanks to precision tuning and vectorization, execution time is decreased by 12% and memory accesses are reduced by 27% on average, leading to a reduction of energy consumption up to 30%.

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

**9:30** | **9.2.3**  
**A PERIPHERAL CIRCUIT REUSE STRUCTURE INTEGRATED WITH A RETIMED DATA FLOW FOR LOW POWER RRAM CROSSBAR-BASED CNN**

**Speaker:** Keni Qiu, Capital Normal University, CN  
**Authors:** Keni Qiu1, Weiwen Chen1, Yuanchao Xu1, Lixue Xia2, Yu Wang3 and Zili Shao3  
1Capital Normal University, China; 1Tsinghua University, China; 2The Hong Kong Polytechnic University, HK  
**Abstract:**
Convolutional computations implemented in RRAM crossbar-based Computing System (RCS) demonstrate the outstanding advantages of high performance and low power. However, current designs are energy-unbalanced among the three parts of RRAM crossbar computation, peripheral circuits and memory accesses, and 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 proposes a Peripheral Circuit Unit (PeriCU)-Reuse scheme to meet power budgets in energy constrained embedded systems. The underlying idea is to put the expensive ADCs/DACs onto spotlight and arrange multiple convolution layers to be sequentially served by the same PeriCU. In the solution, the first step is to determine the number of PeriCUs which are organized by cycle frames. Inside a cycle frame, the layers are computed in parallel inter-PeriCUs while sequentially intra-PeriCU. Furthermore, a layer retiming technique is exploited to significantly limit the potential gains of RCS. Addressing the problem of high power overhead of peripheral circuits in RCS, this paper proposes a Peripheral Circuit Unit (PeriCU)-Reuse scheme integrated with the retiming technique can efficiently meet variable power budgets, and further reduce energy consumption efficiently.

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

**9:45** | **9.2.4**  
**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)
<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>IP4-5</td>
<td>ENERGY-PERFORMANCE DESIGN EXPLORATION OF A LOW-POWER MICROPROGRAMMED DEEP-LEARNING ACCELERATOR</td>
<td>Andrea Calimera, Politecnico di Torino, IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Andrea Calimera, Politecnico di Torino, IT</td>
<td>Authors: Andrea Calimera¹, Mario R. Casu², Giulia Santoro¹, Valentino Peluso¹ and Massimo Aiolo³</td>
</tr>
<tr>
<td></td>
<td></td>
<td>¹Politecnico di Torino, IT; ²Department of Electronics and Telecommunications, IT; ³National University of Singapore, SG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>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)</td>
<td></td>
</tr>
<tr>
<td>10:01</td>
<td>IP4-6</td>
<td>GENPIM: GENERALIZED PROCESSING IN-MEMORY TO ACCELERATE DATA INTENSIVE APPLICATIONS</td>
<td>Tajana Rosing, UC San Diego, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speaker: Mohsen Imani, Saransh Gupta and Tajana Rosing, University of California, San Diego, US</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>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 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)</td>
<td></td>
</tr>
<tr>
<td>10:00</td>
<td></td>
<td>End of session Coffee Break in Exhibition Area</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

### 9.3 Advances in Reconfigurable Computing

**Date:** Thursday, March 22, 2018  
**Time:** 09:30 - 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 convolution neural networks.

**Time   | Label | Presentation Title                                                                 | Authors                                                                                     |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Advances in Reconfigurable Computing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Date: Thursday, March 22, 2018</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time: 09:30 - 10:00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Location / Room: Konf. 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chair: Jürgen Teich, Friedrich-Alexander Universität, DE, Contact Jürgen Teich</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Co-Chair: Florent de Dinechin, INSA-Lyon, FR, Contact Florent de Dinechin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>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 convolution neural networks.</td>
<td></td>
</tr>
</tbody>
</table>
Abstract

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 (conditional). 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 CRA, 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/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 12.19fps 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: Flavio Guiani, Lund University, SE, Contact Flavio Guiani  

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.

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, and this leads to the following limitations: (a) resource proportionality of a multi-tier system is bounded by the basic building block (mainboard), (b) resource allocation to processes or virtual machines (VM) is bounded by the available resources within the boundary of the mainboard, leading to sparse resource fragmentation and inefficiencies, and (c) upgrades 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.

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" 13:30 - 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 Elif, Infineon Technologies, DE, Contact Bilge Kavun Elif

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.

9.5.1 AN INSIDE JOB: REMOTE POWER ANALYSIS ATTACKS ON FPGAS

Speaker:
Falk Schellenberg, Ruhr-Universität Bochum, DE

Authors:
Falk Schellenberg1, Dennis GnadF, Amir Moradi1 and Mehdi Tahoori2
1Ruhr University Bochum, DE; 2Karlsruhe Institute of Technology, DE

Abstract
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.

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

9.5.2 CONFIDENT LEAKAGE DETECTION - A SIDE-CHANNEL EVALUATION FRAMEWORK BASED ON CONFIDENCE INTERVALS

Authors:
Florian Bach1, Christina Plump1 and Tim Güneysu2
1University of Bremen, DE; 2University of Bremen & DFKI, DE

Abstract
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 not only enables 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.

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</td>
</tr>
<tr>
<td></td>
<td>Speaker:</td>
<td>Alwin Zulehner, Alexandru Paler and Robert Wille, Johannes Kepler University Linz, AT</td>
</tr>
<tr>
<td></td>
<td>Authors:</td>
<td>Alwin Zulehner, Johannes Kepler University Linz, AT</td>
</tr>
<tr>
<td></td>
<td>Abstract:</td>
<td>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.</td>
</tr>
<tr>
<td></td>
<td>Download Paper (PDF; Only available from the DATE venue WiFi)</td>
<td></td>
</tr>
</tbody>
</table>

| IP4-2 | PARALLEL CODE GENERATION OF SYNCHRONOUS PROGRAMS FOR A MANY-CORE ARCHITECTURE | Amaury Graillat, Verimag - Univ. Grenoble Alpes, FR |
|       | Speaker:           | Amaury Graillat¹, Matthieu Moy², Pascal Raymond³ and Benoît Dupont de Dinechin⁴ |
|       | Authors:           | Amaury Graillat¹, Matthieu Moy², Pascal Raymond³ and Benoît Dupont de Dinechin⁴ |
|       | ¹Verimag - Univ. Grenoble Alpes, FR; ²Univ. Grenoble Alpes, Verimag, FR; ³VERIMAG/CNRS, FR; ⁴Kalray, FR |
|       | Abstract:          | Embedded 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) |
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 generic hardware generator. It is focused on basic posit arithmetic (floating-point to posit conversion, posit to floating point conversion, addition/subtraction and multiplication).

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 to ES 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.

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

Abstract

Authors

Hayden K.-H. So, The University of Hong Kong, HK

Abstract

Post 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 ES 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 posit system.
Boolean components for quantum logic. In contrast to the current state-of-the-art, we dedicate our efforts to address the main reasons causing the additional required qubits (namely the 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 these components. Since quantum computations typically contain large Boolean components, design automation techniques are required to realize the respective Boolean functions in a computationally efficient manner.

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

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). However, 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 12.19fps under 150MHz working frequency, with all intermediate data staying on-chip.

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

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 designs 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)

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 loss of accuracy. In the second part of the work, we develop a routing mechanism, which tackles the 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)

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 loss of accuracy. In the second part of the work, we develop a routing mechanism, which tackles the 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)

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 12.19fps under 150MHz working frequency, with all intermediate data staying on-chip.

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

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 loss of accuracy. In the second part of the work, we develop a routing mechanism, which tackles the 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)
IP4-13

POWER OPTIMIZATION THROUGH PERIPHERAL CIRCUIT REUSING INTEGRATED WITH LOOP TILING FOR RRAM CROSSBAR-BASED CNN

Authors:
Yuanzhui Ni, Weiwen Chen and Keni Qiu, Capital Normal University, CN

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 CNN.

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 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)

IP4-14

ORIENT: ORGANIZED INTERLEAVED ECCS FOR NEW STT-RAM CACHES

Speaker:
Hamed Farbeh, School of Computer Science, Institute for Research in Fundamental Sciences (IPM), IR

Authors:
Zahra Azadi1, Hamed Farbeh2 and Amir Mahdi Hosseini Monazzah3
1Sharif University of Technology, IR; 2School of Computer Science, Institute for Research in Fundamental Sciences (IPM), IR; 3Department of Computer Engineering, Sharif University of Technology, Tehran, IR

Abstract
Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM) is a promising alternative to SRAM in cache memories. However, STT-MRAMs face with high probability of write errors due to its stochastic switching behavior. To correct the write errors, Error-Correcting Codes (ECCs) used in SRAM caches are conventionally employed. A cache line consists of several codewords and the data bits are selected in such a way that the maximum correction capability is provided based on the error patterns in SRAMs. However, the different write error patterns in STT-MRAMs leads to inefficiency of conventional ECC configurations. In this paper, first we investigate the efficiency of ECC configurations and demonstrate that the vulnerability of codewords in a cache line varies by up to 17x. This variation means that, while some words are overprotected, some others are highly probable to experience uncorrectable errors. Then, we propose an ECC bit selection scheme, so-called ORIENT, to reduce the vulnerability variation of codewords to 1.4x. The simulation results show that conventional ECC configuration increases the write error rate by up to about 64.4% compared with the optimum ECC bit selection, whereas this value for ORIENT is only 4.5%.

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

IP4-15

ERASMUS: EFFICIENT REMOTE ATTESTATION VIA SELF-MEASUREMENT FOR UNATTENDED SETTINGS

Speaker:
Norrathep Rattanavipanon, University of California, Irvine, TH

Authors:
Xavier Carpent1, Norrathep Rattanavipanon2 and Gene Tsudik2
1UC Irvine, US; 2UCI, US

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 engage in a potentially expensive computation, which can negatively impact safety-critical or real-time devices. To this end, we introduce the concept of self-measurement whereby a prover periodically (and securely) measures and records its own software state. A verifier then collects and verifies these measurements. We demonstrate a concrete technique called ERASMUS, justify its features and evaluate its performance. We show that ERASMUS is well suited for safety-critical applications. We also define a new metric – Quality of Attestation (QoA).

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

IP4-16

END-TO-END LATENCY ANALYSIS OF CAUSE-EFFECT CHAINS IN AN ENGINE MANAGEMENT SYSTEM

Speaker:
Junchul Choi, Seoul National University, KR

Authors:
Junchul Choi, Donghyun Kang and Soochoi Ha, Seoul National University, KR

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)

IP4-17

GENERAL FLOORPLANNING METHODOLOGY FOR 3D ICS WITH AN ARBITRARY BONDING STYLE

Speaker:
Chien-Yu Huang, Department of Electrical Engineering, National Cheng Kung University, TW

Authors:
Jialing Lin and Chien-Yu Huang, Department of Electrical Engineering, National Cheng 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 utilizations of bins required by the 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 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)
UB09.1 
CCF: A CGRA COMPILATION FRAMEWORK
Authors: Shail Dave and Avrati 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&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 libraries. The circuit 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 ...

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 SoCs 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 MPSoCs
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 aimed 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: 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 enforce the overall system behavior. To demonstrate my methodologies a mixed-critical multi-processor architecture and its corresponding virtual prototype is used. The multi-processor 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 ...

UB09.6 
WARE: WEARABLE ELECTRONICS DIRECTIONAL AUGMENTED REALITY
Authors: Gabriele Morandi1, Walter Vendraminotto2, Federico Fraccaroli3, Davide Quaglia1 and Gianluca Benedetti4
1University of Verona, IT; 2REDALab Srl, IT; 3Wagoo LLC, IT; 4Wagoo 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 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 Chupliko, Alexander Kamkin and Andrei Tatarmkov, 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 RML/mmML 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 ...
A 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. Furthermore, processes, methods, and tools need to speed up to cope with all the consequences in validation and verification.

Autonomous systems are an important part of today's 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, and 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.
11:00 10.1.1 AUTO IN MOTION - TRENDS IN AUTOMOTIVE ENGINEERING

Author:
Eric Sax, Karlsruhe Institute of Technology, DE

Abstract
Technology is on the rise and enables new functions in modern cars. Automated vehicles, connected functions and alternative drives open new levels of safety, comfort and business models. But the processes, methods and tools lack behind and need to speed up to cope with all the consequences in validation and verification.

11:25 10.1.2 DRIVING AND BEING DRIVEN: FUTURE MOBILITY AND TECHNOLOGICAL ENABLERS

Author:
Hans-Jörg Vögel, BMW Group, DE

Abstract
Upcoming architectures for vehicular electrics, electronics, communication and software will face - among others - technological challenges driven by strong strategic trends. Those trends - Automated Driving, Artificial Intelligence, Drone/Drone Electrification and Connected Systems and Services - are driving increasing system complexity, demanding formal verifiability and semantically rich system integration; and they are driving innovation, requiring new technologies, subsystems and components to be integrated as well as new novel approaches in system integration and operation. Automation, in particular, not only introduces a plethora of additional sensors, compute-heavy algorithms, and the corresponding load on the vehicle's physical network, but also requires formalized functional safety engineering to a much greater extent than previous vehicle generations. Moreover, intelligent algorithms not only help automated vehicles drive safely, but artificial intelligence will be making its way into quite every aspect of functional vehicle design with the advent of intelligent personal assistants, affective multi-modal natural user interaction, situational awareness and augmented reality. The vehicle's self-awareness required is further driving architectural design challenges.

11:50 10.1.3 DIGITALIZATION IN THE INDUSTRY AUTOMATION - FROM THE COMPONENT TO THE CLOUD

Author:
Thilo Streichert, Festo AG & Co. KG, DE

Abstract
Products from Festo are becoming increasingly intelligent and incorporate more and more software-based applications as well as the ability to integrate embedded functions. Smart products optimize themselves, adapt to external influences, identify themselves and have a digital map in the form of a product key. Integrated sensor technology ensures that processes are transparent and self-diagnosing, thus enabling preventive maintenance. Universal standards and interfaces such as TSN and OPC UA fulfill the requirements for smart products suitable for plug&produce applications and Industry 4.0. In this contribution, the transformation from mechanical actuators to software-defined motion is presented. We will show how these automation components will be integrated in a networked machine cell and result in a higher overall equipment efficiency.

12:10 10.1.4 5G CHALLENGES FOR CONNECTED, COOPERATIVE AND AUTOMATED TRANSPORT SYSTEMS

Author:
Jérôme Härri, Eurecom - Communication systems, FR

Abstract
The WiFi-based V2X Communication technology ITS-G5 (a.k.a DSRC in the US) is the currently only market-ready solution to provide safety-critical V2X communications for future C-ITS applications, such as road hazard warning, lane-change warning or intersection-collision warning. Yet, future Connected Cooperative Automated Transport Systems (CATS) are expected to require ultra-reliable and low latency connectivity that neither the current WiFi technology nor even the new Cellular V2X technology can provide. This talk will first provide a rapid overview of the benefit of these future CATS as well as their expected communication requirements, then briefly review the state of the art, and finally describe the 5G challenges and roadmap to meet the communication, networking and services required by these future Connected Cooperative Automated Transport Systems.

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).

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" 13:50 - 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

10.2 Neural Networks and Neurotechnology
Neurotechnology applications and an optimisation strategy for efficiently mapping spiking neural networks to neuromorphic hardware are also presented.

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 Robert Wille, University of Linz, AT,

Co-Chair: Jim Harkin, University of Ulster, GB, Contact Jim Harkin

11:00 10.2.1 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 Niemer 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)

11:30 10.2.2 LOW-POWER 3D INTEGRATION USING INDUCTIVE COUPLING LINKS FOR NEUROTECHNOLOGY APPLICATIONS
Speaker: Benjamin Fletcher1, Shridhartha Das2, Chi-Sang Poon3 and Terrence Mak1
Authors: Benjamin Fletcher1, Shridhartha 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 SPIICE 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)

12:00 10.2.3 MAPPING OF LOCAL AND GLOBAL SYNAPSES ON SPIKING NEUROMORPHIC HARDWARE
Speaker: Franky Catthoor, IMEC Fellow, BE
Authors: Anup Das1, Yuefeng Wu1, Khanh Huynh1, Francesco Dell Anna2, Franky 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 partitions 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)

12:15 10.2.4 ENERGY-EFFICIENT NEURAL NETWORKS USING APPROXIMATE COMPUTATION REUSE
Speaker: Xun Jiao, University of California San Diego, US
Authors: Xun Jiao1, Vahideh 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)
12:30  IP4-11, 428

**AN ENERGY-EFFICIENT STOCHASTIC COMPUTATIONAL DEEP BELIEF NETWORK**

**Authors:**
Yidong Liu, University of Alberta, CA
Yanzhi Wang, Syracuse University, US
Fabrizio Lombardi, Northeastern University, US
Jie Han, University of Alberta, CA

**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)

12:31  IP4-12, 375

**PUSHING THE NUMBER OF QUBITS BELOW THE "MINIMUM": REALIZING COMPACT BOOLEAN COMPONENTS FOR QUANTUM LOGIC**

**Authors:**
Alwin Zulehner, 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**

As leakage increases proportionally with the technology downsizing, it becomes extremely challenging to manage to meet the total power budget. This is because, CMOS-based logic blocks cannot be completely powered-gated as their flip-flops always require a retention supply. Alternatively, their data can be stored in a separate memory during the standby mode, however, that result in a huge area and energy overhead. Spin Transfer Torque (STT) based non-volatile flip-flops can offer normally-off, instant-on computing features to reduce leakage by complete power-shut-down without the need to transfer and restore system states before and after the power-down phases. The non-volatile component of such flip-flops can be easily shared for the overall design optimizations. In this paper, we design a unique multi-bit non-volatile flip-flop architecture using STT devices to reduce the area and energy costs associated with non-volatile components. This 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-flops 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-flops designs.

Download Paper (PDF; Only available from the DATE venue WiFi)
### POWER OPTIMIZATION THROUGH PERIPHERAL CIRCUIT REUSING INTEGRATED WITH LOOP TILING FOR RRAM CROSSBAR-BASED CNN

**Authors:** Yuanhui Nü, Weilwen Chen and Keni Qi, Capital Normal University, CN

**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 CNN. 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 (PerCU)-Reuse scheme to meet a certain power budget. The underlying idea is to put the expensive AD/DA to spotlights and arrange multiple convolution layers to be sequentially served by the same PerCU. Furthermore, it is observed that memory accesses can be bypassed if two adjacent layers are assigned in the different PerCU. 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 PerCU-Reuse scheme integrated with the loop tiling technique can efficiently meet power requirement, and further reduce energy consumption by 61.7%.

**Download Paper (PDF; Only available from the DATE venue Wi-Fi)**

### ORIENT: ORGANIZED INTERLEAVED ECCS FOR NEW STT-RAM CACHES

**Speaker:** Hamed Farbeh, School of Computer Science, Institute for Research in Fundamental Sciences (IPM), IR

**Authors:**

1Sharif University of Technology, IR; 2School of Computer Science, Institute for Research in Fundamental Sciences (IPM), IR; 3Department of Computer Engineering, Sharif University of Technology, Tehran, Iran, IR

**Abstract**

Spin-Transfer Torque Magnetic Random Access Memory (STT-MRAM) is a promising alternative to SRAM in cache memories. However, STT-MRAM faces with high probability of write errors due to its stochastic switching behavior. To correct the write errors, Error-Correcting Codes (ECCs) used in SRAM caches are conventionally employed. A cache line consists of several codewords and the data bits are selected in such a way that the maximum correction capability is provided based on the error patterns in SRAMs. However, the different write error patterns in STT-MRAM caches leads to inefficiency of conventional ECC configurations. In this paper, first we investigate the efficiency of ECC configurations and demonstrate that the vulnerability of codewords in a cache line varies by up to 17x. This variation means that, while some words are overprotected, some others are highly probable to experience uncorrectable errors. Then, we propose an ECC bit selection scheme, so-called ORIENT, to reduce the vulnerability variation of codewords to 1.4x. The simulation results show that conventional ECC configuration increases the write error rate by up to about 64.4% compared with the optimum ECC bit selection, whereas this value for ORIENT is only 4.5%.

**Download Paper (PDF; Only available from the DATE venue Wi-Fi)**

### EFFICIENT WEARING LEVEL OF INODES FOR FILE SYSTEMS ON PERSISTENT MEMORIES

**Speaker:** Xianzhang Chen, Chongqing University, CN

**Authors:**

1Chongqing University, CN; 2East China Normal University, CN; 3Capital Normal University, CN; 4School of Computer Science, Institute for Research in Fundamental Sciences (IPM), IR

**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 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 with common file operations.

**Download Paper (PDF; Only available from the DATE venue Wi-Fi)**

### 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
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.

### Time | Label | Presentation Title | Authors |
--- | --- | --- | --- |
11:00 | 10.4.1 | **BINARY RING-LWE HARDWARE WITH POWER SIDE-CHANNEL COUNTERMEASURES** | Ye Wang, The University of Texas at Austin, US |
| | | Speaker: Ye Wang, The University of Texas at Austin, US | Authors: Aydin Ayşu, Mohit Tiwari and Michael Orshansky, University of Texas at Austin, US |
| | | 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. **Download Paper** (PDF; Only available from the DATE venue WiFi) | |
11:30 | 10.4.2 | **HIGH SPEED ASIC IMPLEMENTATIONS OF LEAKAGE-RESISTANT CRYPTOGRAPHY** | Thomas Unterluggauers, Graz University of Technology, AT |
| | | Speaker: Thomas Unterluggauer, Graz University of Technology, AT | Authors: Robert Schilling¹, Thomas Unterluggauer², Stefan Mangard³, Frank Gürkanayi⁴, Michael Muehberghuber⁴ and Luca Benini⁵ |
| | | 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 | 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-resistant 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 Gbit/s and a best-case energy efficiency of 137 Gbit/s/W making this accelerator suitable for high-speed secure IoT applications. **Download Paper** (PDF; Only available from the DATE venue WiFi) | |
12:00 | 10.4.3 | **OPTIMIZATION OF THE PLL CONFIGURATION IN A PLL-BASED TRNG DESIGN** | Elie Noumon Allini, Hubert Curien Laboratory, Jean Monnet University, FR |
| | | Speaker: Elie Noumon Allini, Laboratoire Hubert Curien, University of Saint-Etienne, FR | Authors: Elie Noumon Allini, Oto Petura, Viktor Fischer and Florent Bernard, Hubert Curien Laboratory, Jean Monnet University, FR |
| | | Abstract: Several recent designs show that the phase locked loops (PLLs) are well suited for building true random number generators (TRNG) 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 hardwired 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. **Download Paper** (PDF; Only available from the DATE venue WiFi) | |
12:30 | IP4-15.187 | **ERASMUS: EFFICIENT REMOTE ATTESTATION VIA SELF-MEASUREMENT FOR UNATTENDED SETTINGS** | Norrathep Rattanavipanon, University of California, Irvine, CA |
| | | Speaker: Norrathep Rattanavipanon and Gene Tsudik | Authors: Xavier Casper¹, Norrathep Rattanavipanon² and Gene Tsudik² |
| | | ¹UC Irvine, US; ²UCI, US | 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 engage in a potentially expensive computation, which can negatively impact safety-critical or real-time devices. To this end, we introduce the concept of self-measurement whereby a prover periodically (and securely) measures and records its own software state. A verifier then collects and verifies these measurements. We demonstrate a concrete technique called ERASMUS, justify its features and evaluate its performance. We show that ERASMUS is well-suited for safety-critical applications. We also define a new metric -- Quality of Attestation (QoA). **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/LIF, FR

Authors:
Leonel Acunha Guimarães, Thiago Fereira 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)

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.

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 Pimentel

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.

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)
MIXED-CRITICALITY SCHEDULING WITH MEMORY BANDWIDTH REGULATION

Speaker:
Muhammad Ali Awan, CISTER/INESC-TEC and ISEP/IPP, Porto, Portugal, PT

Authors:
Muhammad Ali Awan1, Pedro Souto2, Konstantinos Bletsas1, Benny Akesson1 and Eduardo Tovar3
1CISTER/INESC-TEC, ISEP, PT; 2Faculty of Engineering of the University of Porto, PT

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) Vesta 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)

DESIGN AND VALIDATION OF FAULT-TOLERANT EMBEDDED CONTROLLERS

Speaker:
Soumyajit Dey, IIT Kharagpur, IN

Authors:
Saurav Kumar Ghosh1, Soumyajit Dey2, Dip Goswami3, Daniel Mueller-Gritschneder4 and Samarjit Chakraborty4
1Dept. of CSE, IIT Kharagpur, IN; 2Indian Institute of Technology Kharagpur, IN; 3Eindhoven University of Technology, NL; 4Technical University of Munich, DE

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)

END-TO-END LATENCY ANALYSIS OF CAUSE-EFFECT CHAINS IN AN ENGINE MANAGEMENT SYSTEM

Speaker:
Junchul Choi, Seoul National University, KR

Authors:
Junchul Choi, Donghyun Kang and Soonhoi Ha, Seoul National University, KR

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)

TOWARDS FULLY AUTOMATED TLM-TO-RTL PROPERTY REFINEMENT

Speaker:
Vladimir Herdt, University of Bremen, DE

Authors:
Vladimir Herdt1, Hoang M. Le1, Daniel Gross2 and Rolf Drechsler2
1University of Bremen, DE; 2University 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.
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
- 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) are 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>
</tr>
</thead>
</table>
| 11:00 | 10.6.1 | OUTLOOK FOR LOW-POWER BEYOND-CMOS DEVICES  
**Author:** An Chen, Semiconductor Research Corporation, US  
**Abstract** tbd |
| 11:30 | 10.6.2 | EXPLOITING FERROELECTRIC FETS: FROM LOGIC-IN-MEMORY TO NEURAL NETWORKS AND BEYOND  
**Speaker and Author:** Xiaobo Sharon Hu, University of Notre Dame, US  
**Abstract** tbd |
| 12:00 | 10.6.3 | FEFETS: FROM NON-VOLATILE MEMORY TO NON-VOLATILE COMPUTING  
**Author:** Stefan Slesazeck, NaMLab gGmbH, DE  
**Abstract** tbd |
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 Schreiber, The MathWorks, DE, Contact Alexander Schreiber

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 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
UB10.4 RISC-V PROCESSOR MODELING IN IP-XACT USING KACTUS2
Authors:
Esko Pekkarinen 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 take 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:
Daniil 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 Bitwurture 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 Enríquez, Julien Lallet, Renaud Pacalet and Ludovic Aprile
1Nokia Bell Labs, FR; Télécom ParisTech, FR
Abstract
Future 5G networks are going 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., FPGAs). 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 will 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 Eichter 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 Kammar, 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 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 Sat-X features.1 Abhaya C. Kammar, 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 ...

14:30 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
- 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, 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 &quot;something everywhere&quot;, the revolutionary approach following the strategy &quot;everything somewhere&quot; 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 &quot;how good is good enough&quot; 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 10:30 - 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

11.1 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 task, and for some applications it is a near-impossible 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.

11.1.1 THE CAMEL APPROACH TO STACKED SENSOR SMART CAMERAS

Speaker:
Marilyn Wolf, Georgia Institute of Technology, US

Authors:
Saibal Mukhopadhyay1, Marilyn Wolf1 and Evan Gebhardt2
1Georgia Institute of Technology, US; 2School of ECE, Georgia Institute of Technology, US

Abstract
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.

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

11.1.2 A DESIGN TOOL FOR HIGH PERFORMANCE IMAGE PROCESSING ON MULTICORE PLATFORMS

Speaker:
Shuvra Bhattacharya, University of Maryland, College Park, MD, USA and Tampere University of Technology, Finland, US

Authors:
Jiahao Wu1, Timothy Blattner2, Wald Keyrouz2 and Shuvra S. Bhattacharyya1
1University of Maryland, US; 2National Institute of Standards and Technology, US

Abstract
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.

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>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>Speaker: Bart Goossens, Ghent University - imec, BE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Authors: Bart Goossens, Hép Luong, Jan Aelterman and Wilfried Philips, Ghent University, Dept. of Telecommunications and Information Processing, BE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Abstract: We present Quasar, a new programming framework that handles many complex aspects in the design of smart vision systems on embedded platforms, such as parallelization, data flow management, scheduling and load balancing. Quasar, as a highlevel programming language, is nearly hardware-agnostic, 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>
</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>Speaker: Ricardo Camarona-Galan, Instituto de Microelectronica de Sevilla (CSIC-Univ. de Sevilla), ES</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Authors: Manu Trevisi1, Héctor C Bandala2, Jorge Fernández-Bení3, Ricardo Camarona-Galan1 and Ángel Rodríguez-Vázquez1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Abstract: Compressive sampling allows wrapping the relevant content of an image in a reduced set of data. It 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 composing 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>
</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>Speaker: Axel Weissenfeld, AIT Austrian Institute of Technology GmbH, AT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Authors: Axel Weissenfeld and Bernhard Strobl, Austrian Institute of Technology, AT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Abstract: 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>
</tr>
<tr>
<td>15:30</td>
<td></td>
<td></td>
<td>Coffee Break in Exhibition Area</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
A FAITHFUL BINARY CIRCUIT MODEL WITH ADVERSARIAL NOISE
Speaker: Jürgen Maier, TU Wien, AT
Authors: Matthias Függer¹, Jürgen Maier², Robert Najvirt², Thomas Nowak² and Ulrich Schmid²
¹LV, 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. Although 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 Ganytiou, Nestor Ermopoulou 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)

A FAITHFUL BINARY CIRCUIT MODEL WITH ADVERSARIAL NOISE
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 computation of wirelength and density 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 CPU method achieves a better performance over both the GPU 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 Cheng Kung University, TW
Authors: Jai-Ming Lin and Chien-Yu Huang, Department of Electrical Engineering, National Cheng Kung University, TW
Abstract

This paper proposes a general floorplanning methodology which can be applied to 3D ICs with an arbitrary bonding style. Some researchers 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 utilizations of bins required by the analytical-based approach. Our experimental results show the cosine-shaped function can obtain a little better result than the bell-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)
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

Routing algorithms were extensively studied first in multi-computer systems, then in multi- and many-core architectures. Among the commonly used routing techniques, the 3D Network-on-Chip (3D-NoC) architectures are capable of achieving 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)

15:45 11.3.4

RSA: AN INTER/INTRA-CHIP SILICON PHOTONIC NETWORK FOR RACK-SCALE COMPUTING SYSTEMS

Speaker: Biresh Joardar, Washington State University, US
Authors: Biresh Joardar, Min Xie1, Luan H.K. Duong1, Zhifei Wang1, Zhehui Wang1, Jiang Xu1, Peng Yang1, Zhengbin Pang1, Zhiwei Wang1, Zehuai Wang1, Min Xie2, Xuangui Chen1, Luan H.K. Duong1 and Jiang Xu1
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.

Experimental results obtained from a partially and vertically connected 3D-NoC indicate that our solution can provide a high level of reliability in the presence of errors, at the expense of an area and power overhead of 4.1% and 6.8% respectively.

Download Paper (PDF; Only available from the DATE venue WiFi)
The session presents new solutions to evaluate and optimize data location and application timing. The first paper presents a hybrid memory emulator to analyze the performance characteristics of non-volatile memories. The second paper proposes a tool and a library to synthesize distributed protocols on concurrent systems. The third paper presents an optimization framework to find the best data partitioning schemes for processing-in-memory architectures. Finally, the last paper uses loop acceleration to advance source level timing simulation.

**11.4 Evaluating and optimizing memory and timing across HW and SW boundaries**

**Date:** Thursday, March 22, 2018  
**Time:** 14:00 - 15:30  
**Location / Room:** Konf. 2

**Chair:**  
Sara Vinco, Politecnico di Torino, IT, Contact Sara Vinco

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

The session presents new solutions to evaluate and optimize data location and application timing. The first paper presents a hybrid memory emulator to analyze the performance characteristics of non-volatile memories. The second paper proposes a tool and a library to synthesize distributed protocols on concurrent systems. The third paper presents an optimization framework to find the best data partitioning schemes for processing-in-memory architectures. Finally, the last paper uses loop acceleration to advance source level timing simulation.

**11.4.1 HME: A LIGHTWEIGHT EMULATOR FOR HYBRID MEMORY**

**Speaker:**  
Jie Xu, Huazhong University of Science and Technology, CN

**Authors:**  
Zhuchui Duan, Hakun Liu, Xiaofei Liao and Hai Jin, Huazhong University of Science and Technology, CN

**Abstract**  
Emerging non-volatile memory (NVM) technologies have been widely studied in recent years. Those studies mainly rely on cycle-accurate architecture simulators because the commercial NVM hardware is still unavailable. However, current simulation approaches are either too slow, or cannot simulate complex and large-scale workloads. In this paper, we propose a DRAM-based hybrid memory emulator, called HME, to emulate the performance characteristics of NVM devices. HME exploits hardware features available in commodity Non-Uniform Memory Access (NUMA) architectures to emulate two kinds of memories: fast, local DRAM, and slower, remote NVM on other NUMA nodes. HME can emulate a wide range of NVM latencies by injecting software-created memory access delays on the remote NUMA nodes. To evaluate the impact of hybrid memories on the application performance, we also provide application programming interfaces to allocate memory from NVM or DRAM regions. We evaluate the accuracy of the read/write delay injection models by using SPEC CPU2006 and compare the results with a state-of-the-art NVM emulator Quartz. Experimental results demonstrate that the average emulated errors of NVM read and write latencies are less than 5% in HME, which is much lower than Quartz. Moreover, the application performance overhead in HME is one order of magnitude lower than Quartz.

**Download Paper (PDF; Only available from the DATE venue WiFi)**
IN-MEMORY COMPUTING USING PATHS-BASED LOGIC AND HETEROGENEOUS COMPONENTS

Authors:
Alvaro Velasquez and Sumit Kumar Jha, University of Central Florida, US

Abstract
The memory-processor bottleneck and scaling difficulties of the CMOS transistor have given rise to a plethora of research initiatives to overcome these challenges. Popular among these is 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 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

11.5 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.

### Time | Label | Presentation Title | Authors
--- | --- | --- | ---
14:00 | 11.5.1 | STORAGE-AWARE SAMPLE PREPARATION USING FLOW-BASED MICROFLUIDIC LAB-ON-CHIP | Robert Wille, Institute for Integrated Circuits, Johannes Kepler University Linz, 4040 Linz, Austria, AT Authors: Sukanta Bhattacharjee1, Robert Wille2, Juinn-Dar Huang3 and Bhargab Bhattacharya3 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 biochips 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)
14:30 11.5.2 PUMP-AWARE FLOW ROUTING ALGORITHM FOR PROGRAMMABLE MICROFLUIDIC DEVICES
Speaker: Tsung-Yi Ho, National Tsing Hua University, TW
Authors: Tsung-Yi Ho1, National Tsing Hua University, TW
Guang-Ru Lai1, National Tsing Hua University, TW
Abstract: In this work, we propose a pump-aware flow routing algorithm for programmable microfluidic devices (PMDs) that addresses the complex routing congestion while minimizing the assay completion time within a given upper limit. The proposed algorithm utilizes a two-level routing strategy that first assigns pumps to virtual channels and then routes droplets within each channel. Simulation results demonstrate that the proposed algorithm achieves a significant reduction in assay completion time compared to existing approaches.

15:00 11.5.3 ADAPTIVE APPROXIMATION IN ARITHMETIC CIRCUITS: A LOW-POWER UNSIGNED DIVIDER DESIGN
Speaker: Honglai Jiang, University of Alberta, CA
Authors: Honglai Jiang1, Leibo Liu1, Fabrizio Lombardi2, and Je Han1
1University of Alberta, CA; 2Tsinghua University, CN; 3Rutgers University, US
Abstract: We propose a low-power unsigned divider design that adapts its accuracy during computation. The proposed design uses a single optimal level that minimizes the mean error distance, resulting in very low power consumption.

15:15 11.5.4 CORRELATION MANIPULATING CIRCUITS FOR STOCHASTIC COMPUTING
Speaker: Vincent T. Lee, University of Washington, US
Authors: Vincent T. Lee, University of Washington, US; Amin Alaghi, UC Berkeley, US; Mohamed Ibrahim1,2, and Ulf Schlichtmann1
1Technical University of Munich, DE; 2Duke University, US
Abstract: We present a novel set of correlation manipulating circuits for stochastic computing. These circuits enable efficient manipulation of correlation in stochastic computations, improving accuracy and energy efficiency.

15:30 11.5.5 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: We propose a fault-tolerant valve-based microfluidic routing fabric for droplet barcoding in single-cell analysis. The proposed design is capable of handling faults in the valve-based routing fabric, ensuring accurate and efficient single-cell analysis.

15:31 11.5.6 OPTIMIZING POWER-ACCURACY TRADE-OFF IN APPROXIMATE ADDERS
Speaker: Celia Dharmaraj, Indian Institute of Technology Madras, IN
Authors: Celia Dharmaraj, IN; Vinita Vasudevan and Nitin Chandrachoodan, Indian Institute of Technology Madras, IN
Abstract: We propose an approximate adder design that balances power and accuracy. The proposed adder is shown to be 17% to 55% more power efficient than existing approximate adders over a wide range of accuracy values.

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

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](mailto:Contact Carles Hernandez)

**Co-Chair:**  
Shahar Kvatinsky, Technion, IL,  
[Contact Shahar Kvatinsky](mailto:Contact Shahar Kvatinsky)

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>
</table>
| 14:00  | 11.6.1    | **XNOR-RRAM: A SCALABLE AND PARALLEL RESISTIVE SYNAPTIC ARCHITECTURE FOR BINARY NEURAL NETWORKS** | Shimeng Yu, Arizona State University, CN  
Authors: Xiaoyu Sun, Shihui Yin, Xiaochen Peng, Rui Liu, Jae-sun Seo and Shimeng Yu, Arizona State University, US  
**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)](mailto:Download Paper (PDF; Only available from the DATE venue WiFi))
15:30 11.6.2 A NOVEL FAULT TOLERANT CACHE ARCHITECTURE BASED ON ORTHOGONAL LATIN SQUARES THEORY
Speaker: Georgios Keramidas, Think Silicon S.A., GR
Authors: Filippos Filippou1, Georgios Keramidas2, Michail Mavropoulos1 and Dimitris Niklos1
1University of Patras, GR; 2Think Silicon S.A./Technological Educational Institute of Western Greece, GR
Abstract
Aggressive dynamic voltage and frequency scaling is widely used to reduce the power consumption of microprocessors. Unfortunately, voltage scaling increases the impact of process variations on memory cells resulting in an exponential increase in the number of malfunctioning memory cells. As a result, various cache fault-tolerant (CFT) techniques have been proposed. In this work, we propose a new CFT technique which applies a systematic redistribution (permutation) of the cache blocks (assuming various block granularity levels) within the cache structure using the orthogonal Latin Square concept and taking as input the location of the malfunctioning cells in the cache array. The aim of the redistribution is twofold. First, to uniformly distribute the faulty blocks to sets and second, to gather the faulty subblocks to a minimum number of blocks, so as the fault free blocks are maximized. Our evaluation results using the benchmarks of SPEC2006 suite, 100 memory fault maps, and four percentages of malfunctioning cells show that our proposal exhibits strong capability to reduce cache performance degradation especially in situations with high percentages of faulty cells and compares favorably to already known techniques.
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; 2Synopsys, 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, for the first time, 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 Gorkana Ghasemi1, Imam Ahmadpour2, Mehdi Ardebili3 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 in 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 Similarty-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 15:31 IPS-8, IPS-9
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 Sudarshan3, Deepak M. Mathew1, Christian Weis1 and Norbert Wehn1
1University of Kaiserslautern, DE; 2Fraunhofer IIESE, DE
Abstract
In this paper, we present a new communication theoretic channel model for Dynamic Random Access Memory (DRAM) retention errors, that relies on the fully asymmetric retention error behavior of DRAM cells. This new model shows that the traditional approach is over pessimistic and we confirm this with real measurements of DDR3 and DDR4 DRAM devices. Together with an exploitation of the vendor specific true- and anti-cell structure, a low complexity bit-flipping approach is presented, that can largely increase DRAMs reliability with minimum overhead.
Download Paper (PDF; Only available from the DATE venue WiFi)

ARCHITECTURE AND OPTIMIZATION OF ASSOCIATIVE MEMORIES USED FOR THE IMPLEMENTATION OF LOGIC FUNCTIONS BASED ON NANOELECTRONIC 1S1R CELLS
Speaker: Ame Heitlman, RWTH-Aachen University, DE
Authors: Ame Heitlman 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 any 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 compactation.
Download Paper (PDF; Only available from the DATE venue WiFi)
15:32  IP5-17, 62  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)  

15:30  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).  

Coffee Breaks after the End of Session  

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  

Lunch Breaks in "Saal 2"  

Coffee Breaks are also available in "Saal 2" during the following times:  

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  

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 sides 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: Furkan Eris1, Ajay Joshi1, Andrew B. Kahng2, Yenai Ma1, Safiul Mojumder1 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 silicon’ 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 as 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 sequences 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 COMPRESSION
Speaker: Ye Wang, University of Texas at Austin, US
Authors: Ye Wang and Michael Orlansky, 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 Mark3
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 logical value, 0 or 1, modules the thermal signal’s duty cycle. Next, we show in this study that proper selection of transmission frequency can significantly minimize thermal interference. In addition, we propose a robust end-to-end communication protocol for reliable communications. Our experiments have confirmed that, compared to an existing thermal covert channel attack [1][2], a thermal covert channel enhanced with all the improvements proposed in this study is seeing significant BER reduction (by as much as 75%), and transmission rate boost (by more than threefold).
Download Paper (PDF; Only available from the DATE venue WiFi)

15:30 IPS-10, 77  ACCURATE PREDICTION OF SMARTPHONES’ SKIN TEMPERATURE BY CONSIDERING EXOTHERMIC COMPONENTS
Speaker: Jihoon Park and Hojung Cho, Dept. of Computer Science, Yonsei University, KR
Authors: Jihoon Park and Hojung Cho, Dept. of Computer Science, Yonsei University, KR
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)
15:31 | IP5-11, 735 | **TRUSTWORTHY PROOFS FOR SENSOR DATA USING FPGA BASED PHYSICALLY UNCLONABLE FUNCTIONS**  
Speaker: Ubir Chatterjee, Indian Institute of Technology Kharagpur, IN  
Authors: Ubir Chatterjee, Durga Prasad Sahoo, Debdeep Mukhopadhyay and Rajat Subhra Chakraborty 
1Indian Institute of Technology Kharagpur, IN; 2Bosch India (RBEI/ETI), IN; 3Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, IN  
Abstract: 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. 
Download Paper (PDF; Only available from the DATE venue WiFi)

15:32 | IP5-15, 878 | **TOWARDS INTER-VENDOR COMPATIBILITY OF TRUE RANDOM NUMBER GENERATORS FOR FPGAS**  
Speaker: Miloš Grujić, imec-COSIC, KU Leuven, BE  
Authors: Miloš Grujić, 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)

15:30 | Coffee Break in Exhibition Area
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 10:30 - 11:30
Lunch Break 12:30 - 14:30
Awards Presentation and Keynote Lecture in “Saal 2” 13:00 - 14:20
Coffee Break 16:00 - 17:00

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

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
UB11.1 ARCHON: AN ARCHITECTURE-OPEN RESOURCE-DRIVEN CROSS-LAYER MODELLING FRAMEWORK

Authors: Fei Xia1, Ashur Rafay1, Mohammad Al-Hayam2, Alexei Ilison1, Rishad Shafik1, 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 ...

UB11.3 OSC MULTICORE STENCIL PROCESSOR: ONE INSTRUCTION-SET COMPUTER-BASED MULTICORE PROCESSOR FOR STENCIL COMPUTING

Authors: Kaoru Saso, Jing Yuan Zhao and Yuku Haru-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 1/20x circuit area and 1/10x power consumption compared with 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 ...

UB11.6 SPANNER: SELF-REPAIRING SPIKING NEURAL NETWORK CONTROLLER FOR AN AUTONOMOUS ROBOT

Authors: Alan Millard1, Anju Johnson1, James Hider1, David Halliday1, Andy Tymel1, Jon Timmers1, Junxu Liu2, Shvan Katim2, Jim Harkin2 and Liam McDaid2
1University of York, GB; 2Lister 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: Mohammadadegh Dalvandi, Michael Butler and Asesh Salehi Fatubadi, 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 ...

UB11.9 ABSYNTH: A COMPREHENSIVE APPROACH TO FRONT TO BACK ANALOG BLOCK DESIGN AUTOMATION

Authors: Abhaya Chandra Kammara S1,2, Soheil Nazar Shahsavani, Alireza Shafaei Bejestan and Massoud Pedram, University of Southern California, US
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 cells in a research project. Here, we present the extension to our nested optimization approach that creates a symmetric and well matched AB SYNTH 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

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: Scheil Nazar Shahsavani, Arefez Shetaeli 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 w榆林的信号 net 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 w榆林的信号 net and area by 8% compared with state-of-the-art techniques.

Download Paper (PDF; Only available from the DATE venue WiFi)
Optimizing Power-Accuracy Trade-Off in Approximate Adders

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 approximated 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)
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
The memory-processor bottleneck and scaling difficulties of the CMOS transistor have given rise to a plethora of research initiatives to overcome these challenges. Popular among these is 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.
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)

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 Yang1, Weiwen Jiang1 and Qingfeng Zhuge2
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)

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)

NON-INTRUSIVE TESTING TECHNIQUE FOR DETECTION OF TROJANS IN ASYNCHRONOUS CIRCUITS

Speaker: Rodrigo Possamai Bastos, TIMA Laboratory, CNRS/Grenoble INP/UJF, FR
Authors: Leonel Acunha Guimarães, Thiago Ferreira de Paiva Leite, Rodrigo Possamai Bastos and Laurent Fesquet, TIMA - Grenoble Institute of Technology, FR
Abstract
Asyncronous circuits are not only 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 Gujic, imec-COSIC, KU Leuven, BE
Authors: Milos Gujic, 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)

EXPLORING NON-VOLATILE MAIN MEMORY ARCHITECTURES FOR HANDHELD DEVICES

Speaker: Sneha Ved and Manu Awasthi, Indian Institute of Technology Gandhinagar, 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 interferences, 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)
12.2 Searching for corner cases, bugs and security vulnerabilities

Date: Thursday, March 22, 2018
Time: 16:00 - 17:30
Location/Room: Konf. 6

Chair: Daniel Grosse, University of Bremen / DFKI, DE; Contact Daniel Grosse
Co-Chair: Jaan Raik, Tallinn University of Technology, EE; Contact Jaan Raik

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 on RTL models. The second work uses historical debugging data to predict future bug locations. The last presentation automatically mines formal assertions to highlight security vulnerabilities on IP firmware.

16:00 12.2.1 DIRECTED TEST GENERATION USING CONCOLIC TESTING ON RTL MODELS

Speaker: Jonathan Cruz, University of Florida, US
Authors: Ali Ahmed, Farimah Farahmandi and Prabhat Mishra, University of Florida, US

Abstract:
Functional validation is one of the most time-consuming steps in System-on-Chip (SoC) design methodology. In today’s industrial practice, simulating designs using billions of random or constrained-random tests can lead to high functional coverage. However, it is hard to cover the remaining small fraction of corner cases and rare functional 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 for generating directed tests using concolic testing of RTL models. While application of concolic testing on hardware designs has shown some promising results, existing approaches are tuned for improving overall coverage, rather than covering a specific target. We developed a Control Flow Graph (CFG) assisted directed 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 compared to the state-of-the-art test generation methods.

Download Paper (PDF; Only available from the DATE venue WiFi)
### 12.3 Verification and Formal Synthesis

**Date:** Thursday, March 22, 2018  
**Time:** 16:00 - 17:30  
**Location / Room:** [Konf. 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.

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 16:30 | 12.2.2 | IMPROVING AND EXTENDING THE ALGEBRAIC APPROACH FOR VERIFYING GATE-LEVEL MULTIPLIERS | Armin Biere, Johannes Kepler Universität Linz, AT  
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)](#)

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 16:30 | 12.2.2 | RECONFIGURABLE ASYNCHRONOUS PIPELINES: FROM FORMAL MODELS TO SILICON | Danil Sokolov, Newcastle University, GB  
Alessandro de Gennaro and Andrey Mokhov, Newcastle University, GB |

**Abstract**  
Dataflow 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)](#)

<table>
<thead>
<tr>
<th>Time</th>
<th>Label</th>
<th>Presentation Title</th>
<th>Authors</th>
</tr>
</thead>
</table>
| 17:00 | 12.2.3 | SYMBOLIC ASSERTION MINING FOR SECURITY VALIDATION | Alessandro Danese, University of Verona, IT  
Alessandro Danese1, Valeria Bertacco2 and Graziano Pravadelli3 |

**Abstract**  
This paper presents DUVE, 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, DUVE 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)](#)

### Summary

- **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](#)

- **12.2.2** Reconfigurable Asynchronous Pipelines: From Formal Models to Silicon  
  - Speaker: Danil Sokolov, Newcastle University, GB  
  - Authors: Danil Sokolov, Alessandro de Gennaro and Andrey Mokhov, Newcastle University, GB  
  - Abstract: Dataflow 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](#)

- **12.2.3** Symbolic Assertion Mining for Security Validation  
  - Speaker: Alessandro Danese, University of Verona, IT  
  - Authors: Alessandro Danese1, Valeria Bertacco2 and Graziiano Pravadelli3  
  - Abstract: This paper presents DUVE, 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, DUVE 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](#)
12.4 Hardware-assisted Security

Date: Thursday, March 22, 2018
Time: 16:00 - 17:30
Location/Room: Kont. 2
Chair: Iliya Polian, University of Stuttgart, DE, Contact Iliya 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.

12.4.1 HARDWARE-ASSISTED ROOTKIT DETECTION VIA ON-LINE STATISTICAL FINGERPRINTING OF PROCESS EXECUTION

Speaker: Yiyong Makris, University of Texas at Dallas, US
Authors: Liwei Zhou and Yiyong Makris, The University of Texas at Dallas, US

Abstract
Kemel 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)

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 Mangard³

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)

12.4.3 TOWARDS PROVABLY-SECURE PERFORMANCE LOCKING

Speaker: Abhrajit Sengupta, Texas A&M University, US
Authors: Monir Zaman¹, Abhrajit Sengupta², Danguo Liu³, Ozgur Sinanoglu⁴, Yiyong Makris⁵ and Jayavijayan Rajendran⁶

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) thwarts 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 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.

adverse impact of writes from prospectives of cache replacement policy, frequent pattern compression, SLC/MLC hybrid organization, inode virtualization, as well as Android application specific

Chengmo Yang, University of Delaware, US,
Arne Heittman, RWTH, DE,
Chair:
Location / Room:
Date:
12.5 Lifetime Improvement for Persistent Memory

12:00 - 12:30

Chair: Ane Heitman, RWTH, DE, Contact Ane Heitman
Co-Chair: Chengmo Yang, University of Delaware, US, Contact Chengmo Yang

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 perspectives of cache replacement policy, frequent pattern compression, SLC/MLC hybrid organization, inode virtualization, as well as Android application specific address mapping.

12:30 End of session

12:50 12.5.1 EXTENDING THE LIFETIME OF NVMS WITH COMPRESSION

Speaker: Jie Xu, Huazhong University of Science and Technology, CN
Authors: Jie Xu, Dan Feng, Yu Hua, Wei Tong, Jingning Liu and Chunyan Li, Wuhan National Lab for Optoelectronics, School of Computer Science and Technology Huazhong University of Science and Technology, Wuhan, China, CN

Abstract Emerging Non-Volatile Memories (NVMs) such as Phase Change Memory (PCM) and Resistive RAM (RRAM) are promising to replace traditional DRAM technology. However, they suffer from limited write endurance and high write energy consumption. Encoding methods such as Flip-N-Write, FlipMin and CAFO can reduce the bit flips of NVMs by exploiting additional capacity to store the tag bits of encoding methods. The effects of encoding methods are limited by the capacity overhead of the tag bits. In this paper, we propose COE to CoPress cache line for Extending the lifetime of NVMs. COE exploits the space saved by compression to store the tag bits of data encoding methods. Through combining data compression techniques with data encoding methods, COE can reduce the bit flips with negligible capacity overhead. We further observe that the saved space size of each compressed cache line varies, and different encoding methods have different tradeoffs between capacity overhead and effects. To fully exploit the space saved by compression for improving lifetime, we select the proper encoding methods according to the saved space size. Experimental results show that our scheme can reduce the bit flips by 14.2%, decrease the energy consumption by 11.8% and improve the lifetime by 27.5% with only 0.2% capacity overhead.

12:30 12.5.2 HETEROGENEOUS PCM ARRAY ARCHITECTURE FOR RELIABILITY, PERFORMANCE AND LIFETIME ENHANCEMENT

Speaker: Taehyun Kwon, Sungkyunkwan University, KR
Authors: Taehyun Kwon1, Muhammad Imran1, Jung Min You2 and Joon-Sung Yang1
1Sungkyunkwan University, KR
2Sungkyunkwan University, KR

Abstract Conventional DRAM and flash memory are reaching their scaling limits thus motivating research in various emerging memory technologies as a potential replacement. Among these, phase change memory (PCM) has received considerable attention owing to its high scalability and multi-level cell (MLC) operation for high storage density. However, due to the resistance drift over time, the soft error rate in MLC PCM is high. Additionally, the iterative programming in MLC negatively impacts performance and cell endurance. The conventional methods to overcome the drift problem incur large overheads, impact memory lifetime and are inadequate in terms of acceptable soft error rate (SER). 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)
general sense of computing information locally within large data storage. To mitigate the write overhead of PCM-based main memory, we propose a Fine-grained Dritiness 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, evicting 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.

12.5.4 DFPC: A DYNAMIC FREQUENT PATTERN COMPRESSION SCHEME IN NVM-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 PPC 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)
<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 | Speaker: Maha Kooli, CEA-Leti, FR  
Authors: Maha Kooli¹, Henri-Pierre CHARLES², Bastien Giraud³ and Jean-Philippe Noel⁴  
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                                   | Speaker: Said Hamdioui, Delft University of Technology, NL  
Authors: Jintao Yu, HoangAnh DuNguyen, Mottaqiallah Taouil 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                                      | Speaker: Shubham Jain, Purdue University, US  
Authors: Shubham Jain¹, Sachin Sapatnekar², Jian-Ping Wang³, Kaushik Roy⁴ and Anand Raghunathan⁵  
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