



Published on DATE 2019 (<https://past.date-conference.com>)

[Home](#) > [Printer-friendly PDF](#) > [Printer-friendly PDF](#)

## IP5 Interactive Presentations

**Date:** Thursday, March 28, 2019

**Time:** 15:30 - 16:00

**Location / Room:** Poster Area

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	<b>THERMAL-AWARENESS IN A SOFT ERROR TOLERANT ARCHITECTURE</b> <b>Speaker:</b> Sajjad Hussain, Chair for Embedded Systems, KIT, DE <b>Authors:</b> Sajjad Hussain <sup>1</sup> , Muhammad Shafique <sup>2</sup> and Joerg Henkel <sup>1</sup> <sup>1</sup> Karlsruhe Institute of Technology, DE; <sup>2</sup> Vienna University of Technology (TU Wien), AT <b>Abstract</b> <i>It is crucial to provide soft error reliability in a power-efficient manner such that the maximum chip temperature remains within the safe operating limits. Different execution phases of an application have diverse performance, power, temperature and vulnerability behavior that can be leveraged to fulfill the resiliency requirements within the allowed thermal constraints. We propose a soft error tolerant architecture with fine-grained redundancy for different architectural components, such that their reliable operations can be activated selectively at fine-granularity to maximize the reliability under a given thermal constraint. When compared with state-of-the-art, our temperature-aware fine-grained reliability manager provides up to 30% reliability within the thermal budget.</i> <a href="#">Download Paper (PDF; Only available from the DATE venue WiFi)</a>
IP5-2	<b>A SOFTWARE-LEVEL REDUNDANT MULTITHREADING FOR SOFT/HARD ERROR DETECTION AND RECOVERY</b> <b>Speaker:</b> Hwisoo So, Yonsei University, KR <b>Authors:</b> Moslem Didehban <sup>1</sup> , HwiSoo So <sup>2</sup> , Aviral Shrivastava <sup>1</sup> and Kyoungwoo Lee <sup>2</sup> <sup>1</sup> Arizona State University, US; <sup>2</sup> Yonsei University, KR <b>Abstract</b> <i>Advances in semiconductor technology have enabled unprecedented growth in safety-critical applications. In such environments, error resiliency is one of the main design concerns. Software level Redundant MultiThreading is one of the most promising error resilience strategies because they can potentially serve as inexpensive and flexible solutions for hardware unreliability issues i.e. soft and hard errors. However, the error coverage of the existing software level RMT solutions is limited to soft error detection and they rely on external schemes for error recovery. In this paper, we investigate the potential of software-level RMT schemes for complete soft and hard error detection and recovery. First, we pinpoint the main reasons behind ineffectiveness of basic software level triple redundant multithreading (STRMT) in protection against soft and hard errors. Then we introduce FISHER (Flexible Soft and Hard Error Resiliency) as a software-only RMT scheme which can achieve comprehensive error resiliency against both soft and hard errors. Rather than performing centralized voting operations for critical instructions operands, FISHER distributes and intertwines error detection and recovery operations between redundant threads. To evaluate the effectiveness of the proposed solution, we performed more than 135,000 soft and hard error injection experiments on different hardware components of an ARM cortex53-like <math>\mu</math>-architecturally simulated microprocessor. The results demonstrate that FISHER can reduce programs failure rate by around 261x and 162x compared to original and basic STRMTprotected versions of programs, respectively.</i> <a href="#">Download Paper (PDF; Only available from the DATE venue WiFi)</a>
IP5-3	<b>COMMON-MODE FAILURE MITIGATION:INCREASING DIVERSITY THROUGH HIGH-LEVEL SYNTHESIS</b> <b>Speaker:</b> Farah Naz Taher, University of Texas at Dallas, US <b>Authors:</b> Farah Naz Taher <sup>1</sup> , Matthew Joslin <sup>1</sup> , Anjana Balachandran <sup>2</sup> , Zhiqi Zhu <sup>1</sup> and Benjamin Carrion Schaefer <sup>1</sup> <sup>1</sup> The University of Texas at Dallas, US; <sup>2</sup> The Hong Kong Polytechnic University, HK <b>Abstract</b> <i>Fault tolerance is vital in many domains. One popular way to increase fault-tolerance is through hardware redundancy. However, basic redundancy cannot cope with Common Mode Failures (CMFs). One way to address CMF is through the use of diversity in combination with traditional hardware redundancy. This work proposes an automatic design space exploration (DSE) method to generate optimized redundant hardware accelerators with maximum diversity to protect against CMFs given as a single behavioral description for High-Level Synthesis (HLS). For this purpose, this work exploits one of the main advantages of C-based VLSI design over the traditional RT-level design based on low-level Hardware Description Languages (HDLs): The ability to generate micro-architectures with unique characteristics from the same behavioral description. Experimental results show that the proposed method provides a significant diversity increment compared to using traditional RTL-based exploration to generate diverse designs.</i> <a href="#">Download Paper (PDF; Only available from the DATE venue WiFi)</a>
IP5-4	<b>EXPLOITING WAVELENGTH DIVISION MULTIPLEXING FOR OPTICAL LOGIC SYNTHESIS</b> <b>Speaker:</b> David Z. Pan, University of Texas, Austin, US <b>Authors:</b> Zheng Zhao <sup>1</sup> , Derong Liu <sup>2</sup> , Zhoufeng Ying <sup>1</sup> , Biying Xu <sup>1</sup> , Chenghao Feng <sup>1</sup> , Ray T. Chen <sup>1</sup> and David Z. Pan <sup>1</sup> <sup>1</sup> University of Texas, Austin, US; <sup>2</sup> Cadence Design Systems, US <b>Abstract</b> <i>Photonic integrated circuit (PIC), as a promising alternative to traditional CMOS circuit, has demonstrated the potential to accomplish on-chip optical interconnects and computations in ultra-high speed and/or low power consumption. Wavelength division multiplexing (WDM) is widely used in optical communication for enabling multiple signals being processed and transferred independently. In this work, we apply WDM to optical logic PIC synthesis to reduce the PIC area.</i> <a href="#">Download Paper (PDF; Only available from the DATE venue WiFi)</a>

IP5-5 **IGNORETM: OPPORTUNISTICALLY IGNORING TIMING VIOLATIONS FOR ENERGY SAVINGS USING HTM**

**Speaker:**

Dimitra Papagiannopoulou, University of Massachusetts Lowell, US

**Authors:**

Dimitra Papagiannopoulou<sup>1</sup>, Sungseob Whang<sup>2</sup>, Tali Moreshet<sup>3</sup> and Iris Bahar<sup>4</sup>

<sup>1</sup>University of Massachusetts Lowell, US; <sup>2</sup>CloudHealth Technologies, US; <sup>3</sup>Boston University, US; <sup>4</sup>Brown University, US

**Abstract**

Energy consumption is the dominant factor in many computing systems. Voltage scaling is a widely used technique to lower energy consumption, which exploits supply voltage margins to ensure reliable circuit operation. Aggressive voltage scaling will slow signal propagation; without coherent frequency relaxation, timing violations may be generated. Hardware Transactional Memory (HTM) offers an error recovery mechanism that allows reliable execution and power savings with modest overhead. We propose IgnoreTM, an adaptive error management framework, that tolerates (i.e., opportunistically ignores) timing violations, allowing for more aggressive voltage scaling. Our experimental results show that IgnoreTM allows up to 47% total energy savings with negligible impact on runtime.

[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)

IP5-6 **USING MACHINE LEARNING FOR QUALITY CONFIGURABLE APPROXIMATE COMPUTING**

**Speaker:**

Mahmoud Masadeh, Concordia University, CA

**Authors:**

Mahmoud Masadeh, Osman Hasan and Sofiene Tahar, Department of Electrical and Computer Engineering, Concordia University, Montreal, Quebec, CA

**Abstract**

Approximate computing (AC) is a nascent energy-efficient computing paradigm for error-resilient applications. However, the quality control of AC is quite challenging due to its input-dependent nature. Existing solutions fail to address fine-grained input-dependent controlled approximation. In this paper, we propose an input-aware machine learning based approach for the quality control of AC. For illustration purposes, we use 20 configurations of 8-bit approximate multipliers. We evaluate these designs for all combinations of possible input data. Then, we use machine learning algorithms to efficiently make predictive decisions for the quality control of the target approximate application, based on experimentally collected training data. The key benefits of the proposed approach include: (1) fine-grained input-dependent approximation, (2) no missed approximation opportunities, (3) no rollback recovery overhead, (4) applicable to any approximate computation with error-tolerant components, and (5) flexibility in adapting various error metrics.

[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)

IP5-7 **PREDICTION-BASED TASK MIGRATION ON S-NUCA MANY-CORES**

**Speaker:**

Martin Rapp, Karlsruhe Institute of Technology, DE

**Authors:**

Martin Rapp<sup>1</sup>, Anuj Pathania<sup>1</sup>, Tulika Mitra<sup>2</sup> and Joerg Henkel<sup>1</sup>

<sup>1</sup>Karlsruhe Institute of Technology, DE; <sup>2</sup>National University of Singapore, SG

**Abstract**

Performance of a task running on a many-core with distributed shared Last-Level Cache (LLC) strongly depends on two factors: the power budget needed to guarantee thermally safe operation and the LLC latency. The task's thread-to-core mapping determines both the factors. Arrival and departure of tasks on a many-core deployed in an open system can change its state significantly in terms of available cores and power budget. Task migrations can thereupon be used as a tool to keep the many-core operating at the peak performance. Furthermore, the relative impacts of power budget and LLC latency on a task's performance can change with its different execution phases mandating its migration on-the-fly. We propose the first run-time algorithm PCMig that increases the performance of a many-core with distributed shared LLC by migrating tasks based on their phases and the many-core's state. PCMig is based on a performance-prediction model that predicts the performance impact of migrations. PCMig results in up to 16% reduction in the average response time compared to the state-of-the-art.

[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)

IP5-8 **DESIGN OF HARDWARE-FRIENDLY MEMORY ENHANCED NEURAL NETWORKS**

**Speaker:**

Ann Franchesca Laguna, University of Notre Dame, US

**Authors:**

Ann Franchesca Laguna, Michael Niemier and X, Sharon Hu, University of Notre Dame, US

**Abstract**

Neural networks with external memories have been proven to minimize catastrophic forgetting, a major problem in applications such as lifelong and few-shot learning. However, such memory enhanced neural networks (MENNs) typically often require a large number of floating point-based cosine distance metric calculations to perform necessary attentional operations, which greatly increases energy consumption and hardware cost. This paper investigates other distance metrics in such neural networks in order to achieve more efficient hardware implementations in MENNs. We propose using content addressable memories (CAMs) to accelerate and simplify attentional operations. We focus on reducing the bit precision, memory size (MxD) and using alternative distance metric calculations such as L1, L2, and L $\infty$  to perform attentional mechanism computations for MENNs. Our hardware friendly approach implements fixed point L $\infty$  distance calculations via ternary content addressable memories (TCAM) and fixed point L1 and L2 distance calculations on a general purpose graphical processing unit (GPGPU) (Computing-in-memory arrays (CIM) might also be used). As a representative example, a 32-bit floating point-based cosine distance MENN with MD multiplications has a 99.06% accuracy for the Omniglot 5-way 5-shot classification task. Based on our approach with just 4-bit fixed point precision, a L $\infty$ -L1 distance hardware accuracy of 90.35% can be achieved with just 16 TCAM lookups and 16D addition and subtraction operations. With 4-bit precision and a L $\infty$ -L2 distance, hardware classification accuracies of 96.00% are possible. Hence, 16 TCAM lookups and 16D multiplication operations are needed. Assuming the hardware memory has 512 entries, the number of multiplication operations is reduced by 32x versus the cosine distance approach.

[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)

IP5-9 **ENERGY-EFFICIENT INFERENCE ACCELERATOR FOR MEMORY-AUGMENTED NEURAL NETWORKS ON AN FPGA**

**Speaker:**

Seongsik Park, Seoul National University, KR

**Authors:**

Seongsik Park, Jaehee Jang, Seijoon Kim and Sungho Yoon, Seoul National University, KR

**Abstract**

Memory-augmented neural networks (MANNs) are designed for question-answering tasks. It is difficult to run a MANN effectively on accelerators designed for other neural networks (NNs), in particular on mobile devices, because MANNs require recurrent data paths and various types of operations related to external memory access. We implement an accelerator for MANNs on a field-programmable gate array (FPGA) based on a data flow architecture. Inference times are also reduced by inference thresholding, which is a data-based maximum inner-product search specialized for natural language tasks. Measurements on the bAbI data show that the energy efficiency of the accelerator (FLOPS/kJ) was higher than that of an NVIDIA TITAN V GPU by a factor of about 125, increasing to 140 with inference thresholding.

[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)

- IP5-10 **HDCLUSTER: AN ACCURATE CLUSTERING USING BRAIN-INSPIRED HIGH-DIMENSIONAL COMPUTING**  
**Speaker:**  
Mohsen Imani, University of California, San Diego, US  
**Authors:**  
Mohsen Imani, Yeseong Kim, Thomas Worley, Saransh Gupta and Tajana Rosing, University of California San Diego, US  
**Abstract**  
*Internet of things has increased the rate of data generation. Clustering is one of the most important tasks in this domain to find the latent correlation between data. However, performing today's clustering tasks is often inefficient due to the data movement cost between cores and memory. We propose HDCluster, a brain-inspired unsupervised learning algorithm which clusters input data in a high-dimensional space by fully mapping and processing in memory. Instead of clustering input data in either fixed-point or floating-point representation, HDCluster maps data to vectors with dimension in thousands, called hypervectors, to cluster them. Our evaluation shows that HDCluster provides better clustering quality for the tasks that involve a large amount of data while providing a potential for accelerating in a memory-centric architecture.*  
[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)
- IP5-11 **FINDING ALL DC OPERATING POINTS USING INTERVAL-ARITHMETIC BASED VERIFICATION ALGORITHMS**  
**Speaker:**  
Itrat A. Akhter, University of British Columbia, CA  
**Authors:**  
Itrat Akhter, Justin Reiher and Mark Greenstreet, University of British Columbia, CA  
**Abstract**  
*This paper applies interval-arithmetic based verification algorithms to circuit verification problems. In particular, we use Krawczyk's operator to find all DC operating points of CMOS circuits. We present what we believe to be the first, completely automatic verification of the Rambus ring-oscillator start-up problem. Comparisons with the dReal and Z3 SMT shows large performance and scalability advantages to the interval verification approach. We provide an open-source implementation that supports state-of-the-art short-channel device models.*  
[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)
- IP5-12 **GENIE: QoS-GUIDED DYNAMIC SCHEDULING FOR CNN-BASED TASKS ON SME CLUSTERS**  
**Speaker:**  
Zhaoyun Chen, National University of Defense Technology, CN  
**Authors:**  
Zhaoyun Chen, Lei Luo, Haoduo Yang, Jie Yu, Mei Wen and Chunyuan Zhang, National University of Defense Technology, CN  
**Abstract**  
*Convolutional Neural Network (CNN) has achieved dramatic developments in emerging Machine Learning (ML) services. Compared to online ML services, offline ML services that are full of diverse CNN workloads are common in small and medium-sized Enterprises (SMEs), research institutes and universities. Efficient scheduling and processing of multiple CNN-based tasks on SME clusters is both significant and challenging. Existing schedulers cannot predict the resource requirements of CNN-based tasks. In this paper, we propose GENIE, a QoS-guided dynamic scheduling framework for SME clusters that achieves users' QoS guarantee and high system utilization. Based on a prediction model derived from lightweight profiling, a QoS-guided scheduling strategy is proposed to identify the best placements for CNN-based tasks. We implement GENIE as a plugin of Tensorflow and experiment with real SME clusters and large-scale simulations. The results of the experiments demonstrate that the QoS-guided strategy outperforms other baseline schedulers by up to 67.4% and 28.2% in terms of QoS-guarantee percentage and makespan.*  
[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)
- IP5-13 **ADIABATIC IMPLEMENTATION OF MANCHESTER ENCODING FOR PASSIVE NFC SYSTEM**  
**Speaker:**  
Sachin Maheshwari, University of Westminster, GB  
**Authors:**  
Sachin Maheshwari<sup>1</sup> and Izzet Kale<sup>2</sup>  
<sup>1</sup>University of Westminster, GB; <sup>2</sup>University of Westminster, GB  
**Abstract**  
*Energy plays an important role in NFC passive tags as they are powered by radio waves from the reader. Hence reducing the energy consumption of the tag can bring large interrogation range, increase security and maximizes the reader's battery life. The ISO 14443 standard utilizes Manchester coding for the data transmission from passive tag to the reader in the majority of the cases for NFC passive communications. This paper proposes a novel method of Manchester encoding using the adiabatic logic technique for energy minimization. The design is implemented by generating replica bits of the actual transmitted bits and then flipping the replica bits, for generating the Manchester coded bits. The proposed design was implemented using two adiabatic logic families namely; Positive Feedback Adiabatic Logic (PFAL) and Improved Efficient Charge Recovery Logic (IECRL) which are compared in terms of energy for the range of frequency variations. The energy comparison was also made including the power-clock generator designed using 2-stepwise charging circuit (SWC) with FSM controller. The simulation results presented for 180nm CMOS technology at 1.8V power supply shows that IECRL shows approximately 40% less system energy compared to PFAL family.*  
[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)
- IP5-14 **A PULSE WIDTH MODULATION BASED POWER-ELASTIC AND ROBUST MIXED-SIGNAL PERCEPTRON DESIGN**  
**Speaker:**  
Sergey Mileiko, MR, GB  
**Authors:**  
Sergey Mileiko<sup>1</sup>, Rishad Shafik<sup>1</sup>, Alex Yakovlev<sup>1</sup> and Jonathan Edwards<sup>2</sup>  
<sup>1</sup>Newcastle University, GB; <sup>2</sup>Temporal Computing, GB  
**Abstract**  
*Neural networks are exerting burgeoning influence in emerging artificial intelligence applications at the micro-edge, such as sensing systems. As many of these systems are typically self-powered, their circuits are expected to be resilient and efficient to continuous power variations imposed by the harvesters. In this paper, we propose a novel mixed-signal (i.e. analogue/digital) approach of designing a power-elastic perceptron using the principle of pulse width modulation (PWM). Fundamental to the design are a number of parallel inverters that transcode the input-weight pairs based on the principle of PWM duty cycle. Since PWM-based inverters are typically resilient to amplitude and frequency variations, the perceptron shows a high degree of power elasticity and robustness in the presence of these variations. Our extensive design analysis also demonstrates significant power and area efficiency, leading to significant reduction in dynamic and leakage energy when compared with a purely digital equivalent.*  
[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)
- IP5-15 **FAULT LOCALIZATION IN PROGRAMMABLE MICROFLUIDIC DEVICES**  
**Speaker:**  
Ulf Schlichtmann, TUM, DE  
**Authors:**  
Alessandro Bernardini, Chunfeng Liu, Bing Li and Ulf Schlichtmann, TUM, DE  
**Abstract**  
*Programmable Microfluidic Devices (PMDs) have revolutionized the traditional biochemical experiment flow. Test algorithms for PMDs have recently been proposed. Test patterns can be generated algorithmically. But an algorithm for fault localization once some faults have been identified is not yet available. When testing a PMD, once a test pattern fails it is unknown where the stuck valve is located. The stuck valve can be any one valve out of many valves forming the test pattern. In this paper, we propose an effective algorithm for the localization of stuck-at-0 faults and stuck-at-1 faults in a PMD. The stuck valve is localized either exactly or within a very small set of candidate valves. Once the locations of faulty valves are known, it becomes possible to continue to use the PMD by resynthesizing the application*  
[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)

**Speaker:**

Mengquan Li, Chongqing University, CN

**Authors:**

Weichen Liu<sup>1</sup>, Mengquan Li<sup>2</sup>, Wanli Chang<sup>3</sup>, Chunhua Xiao<sup>2</sup>, Yiyuan Xie<sup>4</sup>, Nan Guan<sup>5</sup> and Lei Jiang<sup>6</sup>

<sup>1</sup>Nanyang Technological University, SG; <sup>2</sup>Chongqing University, CN; <sup>3</sup>University of York, GB; <sup>4</sup>Southwest University, CN; <sup>5</sup>Hong Kong Polytechnic University, HK; <sup>6</sup>Indiana University Bloomington, US

**Abstract**

*In this paper, we for the first time utilize the micro-ring resonators (MRs) in optical networks-on-chip (ONoCs) to implement thermal sensing without requiring additional hardware or chip area. The challenges in accuracy and reliability that arise from fabrication-induced process variations (PVs) and device-level wavelength tuning mechanism are resolved. We quantitatively model the intrinsic thermal sensitivity of MRs with fine-grained consideration of wavelength tuning mechanism. Based on it, a novel PV-tolerant thermal sensor design is proposed. By exploiting the hidden 'redundancy' in wavelength division multiplexing (WDM) technique, our sensor achieves accurate and efficient temperature measurement with the capability of PV tolerance. Evaluation results based on professional photonic component and circuit simulations show an average of 86.49% improvement in measurement accuracy compared to the state-of-the-art on-chip thermal sensing approach using MRs. Our thermal sensor achieves stable performance in the ONoCs employing dense WDM with an inaccuracy of only 0.8650 K.*

[Download Paper \(PDF; Only available from the DATE venue WiFi\)](#)

---

Source URL: <https://past.date-conference.com/conference/session/IP5>