3rd ACM/IEEE Workshop on Machine Learning for CAD (MLCAD)

August/30- September/3, 2021

Virtual

Sponsors & supporters

		Speaker	Title
Aug-30	Keynote 1	Ruchir Puri IBM Research, USA	Engineering the Future of AI for the Enterprises and the Chip Design Industry
	Session 1		ML for EDA
	Plenary 1	Ramesh Karri New York University, USA	Machine Learning Applications in Electronic Hardware Security CAD and Security Ramifications of Machine Learning for Electronic CAD
Aug-31	Keynote 2	Sachin Sapatnekar University of Minnesota, USA	Breaching the Last Bastion of the Expert Designer: Is ML the Answer to Automating Analog?
	Session 2		ML for Systems and FPGA
	Plenary 2	Elyse Rosenbaum University of Illinois, USA	Neural Networks for Transient Modeling of Circuits
Sop-01	Keynote 3	Azalia Mirhoseini & Anna Goldie Google	Learning to optimize systems and chips
	Session 3		ML for Modeling and Analysis
	Plenary 3	Janardhan Rao (Jana) Doppa Washington State University, USA	Harnessing Machine Learning to Accelerate Electronic Design Automation
Sep-02	Keynote 4	Shobha Vasudevan University of Illinois, USA Google Brain, USA	Making ML matter to us: How to tackle longstanding design verification challenges with ML
	Session 4		Special Session
	Plenary 4	Tal Kolan Intel, Israel	ML-Based Validation: a journey in the making
Sep-03	Plenary 5	Zhiru Zhang Cornell University, USA	Faster, Slimmer, Smarter: Machine Learning for Agile Hardware Specialization
	Session 5		ML for Circuit Design
	Keynote 5	Jeff Dyck Mentor, a Siemens Business, Canada	Getting the Right Answer with Machine Learning

Monday, August 30, 2021

Keynote 1:

Engineering the Future of AI for the Enterprises and the Chip Design Industry

7am PDT, 10am EDT, 4pm CEST, 7:30pm India, 10pm Asia, 11 pm Japan

Ruchir Puri
IBM Fellow; Chief Scientist
IBM Research, IBM Research Yorktown, USA

Abstract

Recent advances in AI are starting to transform every aspect of our society from healthcare, manufacturing, environment, and beyond. Future of AI for enterprises will be engineered with success along three foundational dimensions. We will dive deeper along these dimensions - Automation of AI; Trust of AI; and Scaling of AI - and conclude with the opportunities and challenges of AI for businesses and the Chip Design Industry.

Speaker Bio:

Dr. Ruchir Puri is the Chief Scientist of IBM Research and an IBM Fellow. He led IBM Watson as its CTO and Chief Architect from 2016-19 and has held various technical, research, and engineering leadership roles across IBM's AI and Research businesses. Dr. Puri is a Fellow of the IEEE, and has been an ACM Distinguished Speaker, an IEEE Distinguished Lecturer, and was awarded 2014 Asian American Engineer of the Year. Ruchir has been an adjunct professor at Columbia University, NY, and a visiting scientist at Stanford University, CA. He was honored with John Von-Neumann Chair at Institute of Discrete Mathematics at Bonn University, Germany. Dr. Puri is an inventor of over 75 United States patents and has authored over 120 scientific publications on software-hardware automation methods, microprocessor design, and optimization algorithms. He is the chair of AAAI-IAAI conference that focused on industrial applications of AI. Ruchir's technical opinions on the adoption of AI by society and businesses have been featured across New York Times, Wall Street Journal, Forbes, Fortune, IEEE spectrum among other.

Monday, August 30, 2021

Session 1: ML for EDA

8am PDT, 11am EDT, 5pm CEST, 8:30pm India, 11pm Asia, midnight Japan

Session Chair: Hai Helen Li – Duke University

1.1 Ensemble Learning Based Electric Components Footprint Analysis
Peng-Tai Huang, Xuan-Yi Lin, Tsung-Yi Ho – National Tsing Hua University, Taiwan
Yan-Jhih Wang - Footprintku Inc., Taiwan

1.2 ADAPT: An Adaptive Machine Learning Framework with Application to Lithography Hotspot Detection
Mohamed Baker Alawieh and David Z. Pan - The University of Texas at Austin, USA

1.3 Delving into Macro Placement with Reinforcement Learning
Zixuan Jiang, David Z. Pan - The University of Texas at Austin, USA
Ebrahim Songhori, Shen Wang, Anna Goldie, Azalia Mirhoseini, Joe Jiang, Young-Joon Lee - Google

1.4 ***Learning-Based Workload Phase Classification and Prediction using Performance Monitoring Counters [best paper candidate]***
Erika S. Alcorta and Andreas Gerstlauer – The University of Texas at Austin, USA

1.5 ***Using Deep Neural Networks and Derivative Free Optimization to Accelerate Coverage Closure***
Raviv Gal, **Marwa Mouallem**, Bilal Saleh and Avi Ziv – IBM Research, Haifa, Israel
Eldad Haber, Brian Irwin – The University of British Colombia, Vancouver

Monday, August 30, 2021

Plenary 1:
Machine Learning Applications in Electronic Hardware Security CAD and Security Ramifications of Machine Learning for Electronic CAD

9:15am PST, 12:15pm PST, 6:15pm CEST, 9:45pm India, 00:15am (Tue) Asia, 01:15am (Tue) Japan

Ramesh Karri
Electrical and Computer Engineering
New York University

Abstract
[TBD]

Speaker Bio:
Ramesh Karri is a Professor of Electrical and Computer Engineering at New York University. He co-directs the NYU Center for Cyber Security (http://cyber.nyu.edu). He co-founded the Trust-Hub (http://trust-hub.org) and organizes the Embedded Systems Challenge (https://csaw.engineering.nyu.edu/esc), the annual red team blue team event.

Ramesh Karri has a Ph.D. in Computer Science and Engineering, from the University of California at San Diego and a B.E in ECE from Andhra University. His research and education activities in hardware cybersecurity include trustworthy integrated circuits, processors and cyber-physical systems; security-aware computer-aided design, test, verification, validation, and reliability; nano meets security; hardware security competitions, benchmarks and metrics; biochip security; additive manufacturing security. He has published over 300 articles in leading journals and conference proceedings.

Ramesh Karri' work in trustworthy hardware received best paper award nominations (ICCD 2015 and DFTS 2015), awards (ACM TODAES 2017, ITC 2014, CCS 2013, DFTS 2013 and VLSI Design 2012, ACM Student Research Competition at DAC 2012, ICCAD 2013, DAC 2014, ACM Grand Finals 2013, Kaspersky Challenge and Embedded Security Challenge). He received the Humboldt Fellowship and the National Science Foundation CAREER Award. He is a Fellow of the IEEE for his contributions to and leadership in Trustworthy Hardware.

He is the Editor-in-Chief of ACM Journal of Emerging Technologies in Computing. Besides, he served/s as the Associate Editor of IEEE Transactions on Information Forensics and Security (2010-2014), IEEE Transactions on CAD (2014-), ACM Journal of Emerging Computing Technologies (2007-), ACM Transactions on Design Automation of Electronic Systems (2014-), IEEE Access (2015-), IEEE Transactions on Emerging Technologies in Computing (2015-), IEEE Design and Test (2015-) and IEEE Embedded Systems Letters (2016-). He served as an IEEE Computer Society Distinguished Visitor (2013-2015). He served on the Executive Committee of the IEEE/ACM Design Automation Conference leading the Security@DAC initiative (2014-2017). He has given keynotes, talks, and tutorials on Hardware Security and Trust.

Tuesday, August 31, 2021

Keynote 2:
Breaching the Last Bastion of the Expert Designer: Is ML the Answer to Automating Analog?

7am PDT, 10am EDT, 4pm CEST, 7:30pm India, 10pm Asia, 11 pm Japan

Sachin Sapatnekar
Department of Electrical and Computer Engineering
University of Minnesota, USA

Abstract

For decades, analog design has stubbornly resisted automation, even as significant parts of digital design flows have embraced it. The reasons for this resistance were embedded in the fact that analog designs were small and "easy" for an expert to comprehend. Despite the efforts of the brightest minds, using the best mathematical techniques of the time, analog EDA tools struggled in competition with the expert. Has ML changed anything? This talk is based on recent experience with developing ALIGN, an open-source analog layout automation flow that has been applied to a wide range of design types and technology nodes. The talk overviews the lessons learned -- the advantages, as well as the perils and pitfalls -- of applying ML to analog design to enhance designer productivity.

Speaker Bio:

Sachin S. Sapatnekar is the Henle Chair in ECE and Distinguished McKnight University Professor at the University of Minnesota. His current research interests include design automation methods for analog and digital circuits, circuit reliability, and algorithms and architectures for machine learning. He is a recipient of the NSF CAREER Award, the SRC Technical Excellence Award, the Semiconductor Industry Association's University Research Award, and 11 Best Paper awards. He has served as Editor-in-Chief of the IEEE Transactions on CAD and General Chair for the ACM/IEEE Design Automation Conference (DAC). He is a Fellow of the IEEE and the ACM.

Tuesday, August 31, 2021

Session 2: ML for Systems and FPGA

8am PDT, 11am EDT, 5pm CEST, 8:30pm India, 11pm Asia, midnight Japan

Session Chair: Avi Ziv – IBM Research, Haifa, Israel

2.1 Effective Machine-Learning Models for Predicting Routability During FPGA Placement [best paper candidate]
Timothy Martin, Shawki Areibi and Gary Grewal – University of Guelph, Canada

2.2 Domain-Adaptive Soft Real-Time Hybrid Application Mapping for MPSoCs [best paper candidate]
Jan Spieck, Stefan Wildermann and Juergen Teich - Chair for Hardware/Software Co-Design, FAU, Germany

2.3 On the Effectiveness of Quantization and Pruningon the Performance of FPGAs-based NNTemperature Estimation
Veera Venkata Ram Murali Krishna Rao Muvva, Marilyn Wolf - University of Nebraska-Lincoln, USA
Martin Rapp, Jorg Henke - Karlsruhe Institute of Technology, Germany
Hussam Amrouch - University of Stuttgart, Germany

2.4 ***Learning based Memory Interference Prediction for Co-running Applications on Multi-Cores***
Ahsan Saeed, Falk Rehm, Arne Hamann, Dirk Ziegenbein - Robert Bosch GmbH, Germany
Daniel Muller-Gritschneder, Ulf Schlichtmann - Technical University of Munich, Germany
Andreas Gerstlauer – The University of Texas at Austin, USA

2.5 ***Dynamic Transformer for Efficient Machine Translation on Embedded Devices***
Hishan Parry, Lei Xun, Amin Sabet, Jia Bi, Jonathon Hare and Geoff Merrett – University of Southampton, UK

Tuesday, August 31, 2021

Plenary 2: Neural Networks for Transient Modeling of Circuits

9:15am PST, 12:15pm PST, 6:15pm CEST, 9:45pm India, 00:15am (Wed) Asia, 01:15am (Wed) Japan

Elyse Rosenbaum
Electrical & Computer Engineering
University of Illinois Urbana-Champaign

Abstract

A circuit behavioral model obscures intellectual property (IP) and, ideally, is faster to simulate than a transistor-level netlist model. Theoretical analyses as well as case studies have established that behavioral models based on a recurrent neural network (RNN) are suitable for transient modeling of nonlinear circuits. After training, an RNN model can be implemented in Verilog-A and evaluated by a SPICE-type circuit simulator. Challenges associated with model training have impeded the wide-scale adoption of RNN circuit models. This talk will detail several of the reasons why large computational resources are needed to train a high-fidelity transient model. Recent advances will be described, including a new stability constraint that is demonstrated to guide model training and improve performance. The talk will conclude with a presentation of augmented RNNs that can accurately capture aging effects and represent process variations.

Speaker Bio:

Elyse Rosenbaum (Fellow, IEEE) is the Melvin and Anne Louise Hassebrock Professor in Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. She received the Ph.D. degree in electrical engineering from University of California, Berkeley. She is the director of the NSF-supported Center for Advanced Electronics through Machine Learning (CAEML), a joint project of the University of Illinois, Georgia Tech and North Carolina State University. Her current research interests include machine-learning aided behavioral modeling of microelectronic components and systems, compact models, circuit reliability simulation, component and system-level ESD reliability, and ESD-robust high-speed I/O circuit design.

Dr. Rosenbaum has authored or co-authored about 200 technical papers; she has been an editor for IEEE Transactions on Device and Materials Reliability and IEEE Transactions on Electron Devices. She was the recipient of a Best Student Paper Award from the IEDM, Outstanding and Best Paper Awards from the EOS/ESD Symposium, a Technical Excellence Award from the SRC, an NSF CAREER award, an IBM Faculty Award, and the ESD Association's Industry Pioneer Recognition Award.

Wednesday, September 01, 2021

Keynote 3: Learning to optimize systems and chips

7am PDT, 10am EDT, 4pm CEST, 7:30pm India, 10pm Asia, 11 pm Japan


Azalia Mirhoseini & Anna Goldie
Research Scientists, Google Brain
Google

Abstract

In the past decade, computer systems and chips have played a key role in the success of deep learning. Our vision in Google Brain's ML for Systems team is to use deep learning to transform the way in which computer systems and chips are designed. Many core problems in systems and hardware design are combinatorial optimization or decision making tasks with state and action spaces that are orders of magnitude larger than that of standard AI benchmarks in robotics and games. In this talk, we will describe some of our latest learning based approaches to tackling such large-scale optimization problems. First, we will talk about our work on deep reinforcement learning for computational resource allocation, a combinatorial optimization problem that repeatedly appears in systems. The proposed method is end-to-end and abstracts away the complexity of the underlying optimization space; the RL agent learns the implicit tradeoffs between computation and communication of the underlying resources and optimizes the allocation using only the true reward function (e.g., the runtime of the generated allocation). Next, we will discuss our work on a new domain-transferable reinforcement learning method for optimizing chip placement, a long pole in hardware design. Our approach is capable of learning from past experience and improving over time, resulting in more optimized placements on unseen chip blocks as the RL agent is exposed to a larger volume of data. Our objective is to minimize PPA (power, performance, and area), and we show that, in under 6 hours, our method can generate placements that are superhuman or comparable on modern accelerator chips, whereas existing baselines require human experts in the loop and can take several weeks.

Speaker Bio - Azalia Mirhoseini*:*

Azalia Mirhoseini is a Staff Research Scientist at Google Brain. She is the co-founder/tech-lead of the Machine Learning for Systems Team, a larger research effort focused on developing advanced learned optimization methods to design next generation computer systems and hardware. Azalia has published more than 40 peer-reviewed papers at scientific venues such as Nature, ICML, ICLR, NeurIPS, UAI, SIGMETRICS, DAC, DATE, and ICCAD. She has a Ph.D. in Electrical and Computer Engineering from Rice University. She has re ceived a number of awards, including the MIT Technology Review 35 under 35 award, the Best Ph.D. Thesis Award at Rice and a Gold Medal in the National Math Olympiad in Iran. Her work has been covered in various media outlets including MIT Technology Review and IEEE Spectrum.

Speaker Bio - Anna Goldie*:*

Anna Goldie is a Staff Research Scientist at Google Brain and co-founder/tech-lead of the Machine Learning for Systems Team, which focuses on deep reinforcement learning approaches to problems in computer systems. She is also a PhD student in the Stanford NLP Group, where she is advised by Prof. Chris Manning. At MIT, she earned a Masters of Computer Science, Bachelors of Computer Science, and Bachelors of Linguistics. She has published peer-reviewed articles in top scientific venues, such as Nature, NeurIPS, ICLR, and EMNLP. She was recently named one of MIT Technology Review's 35 Innovators Under 35, and her work has been covered in various media outlets, including CNBC, ABC News, MIT Technology Review and IEEE Spectrum.

Wednesday, September 01, 2021

Session 3: ML for Modeling and Analysis

8am PDT, 11am EDT, 5pm CEST, 8:30pm India, 11pm Asia, midnight Japan

Session Chair: Haoxing Ren (Mark) – Nvidia

3.1 An Efficient Timing Model of Flip-Flops Based on Artificial Neural Network
Madhvi Agarwal and **Sneh Saurabh** - IIIT Delhi, India

3.2 Connectivity-Based Machine Learning Compact Models for Interconnect Parasitic Capacitances
Mohamed Saleh Abouelyazid, Sherif Hammouda - Siemens EDA, Egypt
Yehea Ismail – American University in Cairo, Egypt

3.3 Fast and Accurate PPA Modeling with Transfer Learning
Luis Francisco, Paul Franzon and Rhett Davis - North Carolina State University, USA

3.4 ***Fast Electrostatic Analysis For VLSI Aging based on Generative Learning***
Subed Lamichhane, Shaoyi Peng, Wentian Jin and Sheldon X.-D. Tan – Department of Electrical and Computer Engineering, University of California- Riverside, USA

3.5 ***A Survey of Graph Neural Networks for Electronic Design Automation***
Daniela Sanchez Lopera, Lorenzo Servadei, Gamze Naz Kiprit, Souvik Hazra, Wolfgang Ecker - Infineon Technologies AG, Germany
Robert Wille - Johannes Kepler University Linz, Austria

Wednesday, September 01, 2021

Plenary 3: Harnessing Machine Learning to Accelerate Electronic Design Automation

9:15am PST, 12:15pm PST, 6:15pm CEST, 9:45pm India, 00:15am (Thu) Asia, 01:15am (Thu) Japan

Janardhan Rao (Jana) Doppa
Electrical Engineering and Computer Science
Washington State University, USA

Abstract

Advanced computing systems have long been enablers for breakthroughs in science and engineering applications including Artificial Intelligence (AI) either through sheer computational power or form-factor miniaturization. However, as algorithms become more complex and the size of datasets increase, existing computing platforms are no longer sufficient to bridge the gap between algorithmic innovation and hardware design. To address the computing needs of emerging applications from the edge to the cloud, we need high-performance, energy-efficient, and reliable computing systems targeted for these applications. Developing these application-specific hardware must become easy, inexpensive, and as seamless as developing application software.

In this talk, I will argue that synergistically combining the domain knowledge from hardware designers and machine learning (ML) algorithms will allow us to make faster progress towards this overarching goal. The key driving principle will be to learn appropriate models for performing intelligent design space exploration to significantly reduce the overall engineering cost and design time of application-specific hardware. I will describe a general framework for ML-guided design space exploration to optimize multiple design objectives and employ manycore system design as an example for illustration purposes. I will also discuss novel Bayesian optimization algorithms and principles for hardware design ranging from circuits to multi-core chips when designs can be evaluated through both single-fidelity (expensive and accurate) and multi-fidelity (vary in accuracy and resource cost) simulation experiments.

Speaker Bio:

Jana Doppa is the George and Joan Berry Distinguished Associate Professor in the School of Electrical Engineering and Computer Science at Washington State University, Pullman. He received his Ph.D. degree in Computer Science from Oregon State University and his M.Tech. degree from Indian Institute of Technology (IIT), Kanpur. His primary research focus is at the intersection of machine learning and electronic design automation. Specifically, he explores the synergies between these two mutually beneficial areas.

His research has been recognized with a number of awards, honors, and distinctions including the 2019 National Science Foundation CAREER Award; the 2021 Early Career Award in AI by the International Joint Conference on Artificial Intelligence for ML algorithms to accelerate design automation for science and engineering applications including electronic design automation; the 2021 Best Paper Award from ACM Transactions on Design Automation of Electronic Systems; the 2013 Outstanding Paper Award from the AAAI Conference on Artificial Intelligence; the 2018 Best Student Paper Award from the AAAI Conference on Artificial Intelligence; the 2015 Outstanding PhD Dissertation Award from Oregon State University and was nominated for ACM Doctoral Dissertation Award; a 2015 Google Faculty Research Award; and the 2020 Outstanding Junior Faculty Research Award and the 2018 Reid-Miller Teaching Excellence Award from the College of Engineering, Washington State University.

Thursday, September 02, 2021

Keynote 4:
Making ML matter to us: How to tackle longstanding design verification challenges with ML

7am PDT, 10am EDT, 4pm CEST, 7:30pm India, 10pm Asia, 11 pm Japan

Shobha Vasudevan
Electrical & Computer Engineering
University of Illinois Urbana-Champaign, USA
Google Brain, USA

Abstract

Whether in software, hardware or embedded systems, verification is widely considered to be the principal bottleneck occupying 70% of time and resources, while only 30% is spent in the actual system development itself. This is especially true in hardware design verification, with verification teams being 3 times the size of design teams. Verification is inherently complex even for small designs; the unprecedented size and complexity of modern hardware exacerbates this problem. In contemporary state of practice, there are no known good solutions for verification other than costly, multiple person-years worth of simulations. In state-of-the-art, static analysis and formal exhaustive analysis techniques have been researched.

In this talk, I will present a mode of thinking that addresses the verification problem using statistical techniques guided by static analysis. The insight behind these techniques is that static analysis is not scalable of its own, and statistical analysis lacks sufficient context. When combined, the intersection of these two solution spaces can be powerful. For more than a decade, in our research group, we have explored the possibilities at this intersection for hardware and software. I will talk about the "meta techniques" to use ML for a theoretically complex and practically infeasible problem like verification. I will outline ML-based solutions for test generation, assertion generation, diagnosis, root causing and debugging. I will also touch upon the use of deep learning for solving different aspects of the verification problem. I will provide real life case studies and examples of where many of these solutions have worked and been incorporated into practical industrial workflows.

Speaker Bio:

Shobha Vasudevan is an associate professor in the department of Electrical and Computer Engineering, and an affiliate in Computer Science at the University of Illinois at Urbana-Champaign. Her research interests span reliability of systems and machine learning algorithms. She has won several best paper awards including one at DAC 2014, one at VLSI Design 2014 and several best paper nominations. Her other honors include the NSF CAREER award, ACM SIGDA Outstanding New Faculty Award, IEEE CEDA early career award, IBM faculty award, Dean's award for research excellence in UIUC, and a YWCA/UIUC award for service to women in engineering. GoldMine, a verification software from her group has been developed into a commercial product since 2014 and has been licensed by multiple semiconductor and electronic design automation companies from UIUC. She conceptualized MyTri, a professional networking portal for women engineers in UIUC. She is a technical consultant for several companies. She enjoys mentoring young women engineers and scientists, and young women who can be future engineers and scientists.

Thursday, September 02, 2021

Session 4: Special Session

8am PDT, 11am EDT, 5pm CEST, 8:30pm India, 11pm Asia, midnight Japan

Session Chair: Hussam Amrouch – University of Stuttgart

Technical Invited Paper

4.1 Feeding Hungry Models Less: Deep Transfer Learning for Embedded Memory PPA Models

Felix Last - Technical University of Munich and Intel, Germany
Ulf Schlichtmann - Technical University of Munich, Germany

Abstract

Supervised machine learning requires large amounts of labeled data for training. In power, performance and area (PPA) estimation of embedded memories, every new memory compiler version is considered independently of previous versions. Since the data of different memory compilers originate from similar domains, transfer learning may reduce the amount of supervised data required by pre-training PPA estimation neural networks on related domains. We show that provisioning times of PPA models for new compiler versions can be reduced significantly by exploiting similarities across versions and technology nodes. Through transfer learning, we shorten the time to provision PPA models for new compiler versions by 50% to 90%, which speeds up time-critical periods of the design cycle. This is achieved by requiring less than 6,500 ground truth samples for the target compiler to achieve average estimation errors of 0.35% instead of 13,000 samples. Using only 1,300 samples is sufficient to achieve an almost worst-case (98th percentile) error of approximately 3% and allows us to shorten model provisioning times from over 40 days to less than one week.

Insights and Lessons from Negative Results

Published papers usually aim at presenting successful research work with positive results and achievements. However, from time to time, we end up with insufficient results, or even negative results, but many important lessons learned – about the complexity of the underline problem, and regarding the methods we used. In this part of the special session, we focus on works that did not reach CAD tools, but have interesting lesson learned.

4.2 Lessons from Machine Learning-Based Triage

Sam Ackerman, Raviv Gal, Wesam Ibraheem, Bilal Saleh and **Avi Ziv** – IBM Research, Haifa, Israel

Abstract

In hardware verification and software testing the term triage is used to describe the process of early assessment of (an often large) group of failures, such as the ones occurring in a nightly regression. The term triage comes from the medical domain, where it is used to describe the assignment of degrees of urgency to wounds or illnesses to decide the order of treatment of a large number of patients or casualties. The high-level goals of the triage process are simplifying the treatment (debug) process that follows it and improve its efficiency and quality. These goals can be easily translated to operational goals such as cluster together failures that have a common root cause, identify failures that already have a defect against them, and assign the right person to handle a group of failures. While many of the operational goals of triage can be easily mapped to classic machine learning techniques such as classification and clustering, machine

learning techniques are only sparsely used in large-scale triage and are not a dominant factor even in places where they are used. In this presentation we describe our attempts to build a machine learning based triage system that assists the human triage team in its daily work. We describe the techniques we used to achieve some of the triage goals, show the results from two high-end processor designs, and discuss some of the reasons that prevented our results from being good enough to be integrated into the current triage process. The main reasons for this unsuccessful outcome are the level of noise in the input data, the constant drift in results, the difficulty with getting accurately labeled data, the lack of common benchmarks to compare results with, and the level of expertise of the potential users. We conclude the work with some insights of what can make future machine learning-based triage-assist system feasible.

4.3 Reinforcement Learning-Driven Global Routing

Bei Yu - The Chinese University of Hong Kong

Abstract

Heuristics developed by human experts require heavy and recurring experimental design efforts, while may not be equally effective in unseen cases. Reinforcement learning has recently demonstrated the potential of exploring policies through experience, which points a promising direction to replace the manual workforce by GPU hours. As a case study, a learning-based global router planner is proposed to determine the routing order of nets. In particular, a neural agent, composed of simple convolutional layers and fully connected layers, is trained to generate a priority score for each batch of nets, and a differentiable sorting operation is used to enable end-to-end training of the whole framework. We will also discuss the lessons we learned from this project.

Thursday, September 02, 2021

Plenary 4: ML-Based Validation: a journey in the making

9:15am PST, 12:15pm PST, 6:15pm CEST, 9:45pm India, 00:15am (Fri) Asia, 01:15am (Fri) Japan

Tal Kolan
Senior Validation Engineer
Intel, Israel

Abstract

Modern CPU products' design size and complexity creates a real challenge to traditional validation techniques. Realizing this challenge, the Intel BigCore validation team started in 2015 a partnership with the Intel IT AI organization in order to incorporate Big Data and Machine Learning technologies into the mainstream pre-silicon functional validation processes. The journey of ML & pre-silicon validation has lasted ever since with great success. The ML-based solutions were enhanced, and more products were developed to solve various validation problems. On the other hand, the validation team, realizing the power of AI, has shifted its methodologies & tools to become more ML-oriented and enable more advanced ML-based solutions. In this presentation we will review this journey, its successes & its failures, sample in depth some of the solutions, explore the results and learnings and take a look into our ML-based HW validation future.

Speaker Bio:

Tal is a Principal-Engineer in Intel CPU development organization. Tal started working for Intel in 2006 under the validation team and starting 2017 he is leading the development of ML-based solution for his organization. Under this role, many solutions were identified, developed, and fully assimilated into the mainstream validation processes and were further scaled to other Intel validation organizations. Tal holds a BsC in computer science from the Technion - Israel Institute of Technology.

Friday, September 03, 2021

Plenary 5:
Faster, Slimmer, Smarter: Machine Learning for Agile Hardware Specialization

7am PDT, 10am EDT, 4pm CEST, 7:30pm India, 10pm Asia, 11 pm Japan

Zhiru Zhang
Computer Systems Laboratory, School of Electrical and Computer Engineering
College of Engineering, Cornell University, USA

Abstract

Targeted customization of functionality in hardware has become arguably the best means to achieving improved compute performance and energy efficiency for a broad range of emerging applications. However, building new application- or domain-specific hardware remains a highly unproductive practice even with more modern design methodologies such as high-level synthesis (HLS). For this reason, there is a strong appeal to leverage machine learning (ML) to further automate some of the critical steps of the design process to achieve agile hardware specialization.

This talk covers some of our recent progress on using ML to improve both digital design and design automation. In particular, we notice that the simulation traces generated by CAD tools provide a rich source of data for training design-specific ML models. We will first introduce a trace-based learning approach that automatically derives lightweight arbitration logic from cycle-level simulation to customize on-chip networks for high performance. We will also show that simulation traces are useful for enabling fast and accurate power estimation for reusable ASIC IPs such as RISC-V cores. Finally, we will discuss how we apply ML to improve various aspects of the HLS tools.

Speaker Bio:

Zhiru Zhang is an Associate Professor in the School of ECE at Cornell University. His current research investigates new algorithms, design methodologies, and automation tools for heterogeneous computing. His research has been recognized with a Facebook Research Award (2020), Google Faculty Research Award (2018), the DAC Under-40 Innovators Award (2018), the Rising Professional Achievement Award from the UCLA Henry Samueli School of Engineering and Applied Science (2018), a DARPA Young Faculty Award (2015), and the IEEE CEDA Ernest S. Kuh Early Career Award (2015), an NSF CAREER Award (2015), the Ross Freeman Award for Technical Innovation from Xilinx (2012), and multiple best paper awards and nominations. Prior to joining Cornell, he was a co-founder of AutoESL, a high-level synthesis start-up later acquired by Xilinx.

Friday, September 03, 2021

Session 5: ML for Circuit Design

7:45am PDT, 10:45am EDT, 4:45pm CEST, 8:15pm India, 10:45pm Asia, 23:45pm Japan

Session Chair: Paul Franzon – North Carolina State University

5.1 Massive Figure Extraction and Classification in Electronic Component Datasheets for Accelerating PCB Design Preparation

Kuan-Chun Chen, Chou-Chen Lee, **Mark Po-Hung Lin** – National Yang Ming Chiao Tung University, Taiwan

Yan-Jhih Wang, Yi-Ting Chen - Footprintku Inc., Taiwan

5.2 A Circuit Attention Network-Based Actor-Critic Learning Approach to Robust Analog Transistor Sizing

Yaguang Li, Yishuang Lin, Jiang Hu - Texas A&M University, USA

Meghna Madhusudan, Arvind Sharma, Sachin Sapatnekar, Ramesh Harjani - University of Minnesota, USA

5.3 Approximate Divider Design Based on Counting-Based Stochastic Computing Division

Shuyuan Yu, Yibo Liu and Sheldon Tan – University of California, Riverside, USA

5.4 Variation-aware Analog Circuit Sizing with Classifier Chains

Zhengfeng Wu and Ioannis Savidis – Drexel University, USA

Friday, September 03, 2021

Keynote 5: Getting the Right Answer with Machine Learning

9am PDT, 12pm EDT, 6pm CEST, 9:30pm India, midnight Asia, 1am (Sat) Japan

Jeff Dyck
Director, Engineering
Mentor, a Siemens Business, Saskatchewan, Canada

Abstract

Many engineering problems require a correct answer. For these cases, a machine learning technology that provides a decent guess, but can be wrong in some cases, is a non-starter. For example, in semiconductor verification, a machine learning model error can cause a catastrophic respin, costing millions. This talk reviews some of Siemens EDA's latest production-proven methods that are used to create consistently accurate machine learning answers for measuring chip variability and for generating timing models. These adaptive and self-verifying methods are general and can be applied widely in other engineering applications.

Speaker Bio:

Jeff Dyck is a Director of Engineering at Siemens EDA, responsible for R&D for three software product lines in the integrated circuit verification solutions (ICVS) division. Prior to joining Siemens, Jeff was VP of Engineering at Solido Design Automation, where he led Solido's R&D teams, managed Solido's product lines, and co-invented Solido's machine learning technologies. Solido was acquired by Siemens in 2017. Jeff is now working on evolving the active learning technology in Solido's products, as well as developing new disruptively differentiated tools within the Siemens EDA analog mixed signal product line.