FINAL PROGRAM

SLIP 2015 Technical Program
June 6, 2015 Saturday
Co-Located with DAC in San Francisco
www.sliponline.org

Opening Remarks (8:00am-8:15am)
Rasit Topaloglu, IBM, General Chair
Baris Taskin, Drexel University, Technical Program Chair

Keynote (8:15am-9:15am)
"Asynchronous and GALS NoC's for Scalable High-Performance Computer Systems"
Steven Nowick, Columbia University

Break (9:15am-9:30am)

Session I: Interconnects (9:30am-10:45am)
Session Chair: Yanzhi Wang, USC

09:30am-09:50am
"SI for Free: Machine Learning of Interconnect Coupling Delay and Transition Effects"
Andrew Kahng, Mulong Luo, Siddhartha Nath
UCSD

09:50am-10:10am
"Smart I/Os: A Data-pattern Aware 2.5D Interconnect with Space-Time Multiplexing"
Sai Dinakarrao, Kanwen Wang, Hantao Huang and Hao Yu
Nanyang Technological University

10:10am-10:30am
"Compact Modeling and System Implications of Microring Modulators in Nanophotonic Interconnects"
Rui Wu[1], Chin-Hui Chen[2], Jean-Marc Fedeli[3], Maryse Fournier[3], Raymond G. Beausoleil[2], Kwang-Ting Cheng[1]
[1]UCSB [2]HP [3]CEA LETI

10:30am-10:45am Post-session review

Break (10:45am-11:00am)

Session II: Timing and Clocking (11:00am-12:25pm)
Session Chair: Ioannis Savidis, Drexel University

11:00am-11:20am
"Clock Clustering and IO Optimization for 3D Integration"

Samyoung Bang[1], Kwangsoo Han[2], Andrew B. Kahng[2], Vaishnav Srinivas[2]
[1]Samsung [2]UCSD

11:20am-11:40am
"On Fast Timing Closure: Speeding up Incremental Path-Based Timing Analysis with MapReduce"
Tsung-Wei Huang and Martin D. F. Wong
UIUC

11:40am-12:10pm
(Invited) "An Analytic Evaluation of the Effects of Power Supply Noise on the Clock Distribution Network"
Peter Feldmann, J. Adam Butts, Timothy J. Correia, David E. Shaw
D.E. Shaw Research

12:10pm-12:25pm Post-session review

Lunch (12:30pm-1:30pm)

Invited Talk 1 (1:30pm-2:00pm)
"Signal-Integrity-Aware Power Delivery Modeling for Future-Generation I/O's"
Wei Kai Shih
Intel

Invited Talk 2 (2:00pm-2:30pm)
Modeling the Performance of Advanced Architectures
Kevin J. Barker
Pacific Northwest National Lab (PNNL)

Panel (2:30pm-3:45pm)
When Interconnect Meets Architecture: Cross-Layer Design and Optimization
Chair: Cheng Zhuo, Intel

Jing-Jia Liou, National Tsing Hua University, Taiwan
Wei-Kai Shih, Intel
Mircea Stan, University of Virginia
Sheldon Tan, UC Riverside

Break (3:45pm-4:00pm)

Session III: System Design (4:00pm-5:15pm)
Session Chair: Mircea Stan, University of Virginia

4:00pm-4:20pm
"Power Line Communication for Hybrid Power/Signal Pin SoC Design"
Xiang Zhang[1], Yang Liu[2], Ryan Coutts[1] and Chung-Kuan Cheng[1]

[1]UCSD, [2]Xidian University

4:20pm-4:40pm
"Lynx: A Self-Organizing Wireless Sensor Network with Commodity Palmtop Computers"
Haifeng Xu, Melissa Bilec, William Collinge, Laura Schaefer, Amy Landis, Alex Jones
University of Pittsburgh

4:40pm-5:00pm
"Multi-Product Floorplan and Uncore Design Framework for Chip Multiprocessors"
Marco Escalante[1], Andrew Kahng[2], Michael Kishinevsky[1], Umit Ogras[3], Kambiz Samadi[4]
[1]Intel [2]UCSD [3]ASU [4]Qualcomm

5:00pm-5:15pm Post-session review

Closing Remarks and Awards Presentation (5:15pm-5:30pm)

ABSTRACTS

Keynote (8:15am-9:15am)
"Asynchronous and GALS NoC's for Scalable High-Performance Computer Systems"
Steven M. Nowick, Columbia University

Asynchronous and GALS (i.e. globally-asynchronous locally-synchronous) networks-on-chip offer the potential of significant benefits in performance, energy, reliability and scalability, since they eliminate the rigidity and overhead of the fixed-rate clock, and allow fine-grain assembly and communication of components. As such, they provide a flexible integrative medium for both high-performance computing (HPC) and complex system-on-chip (SoC) domains, allowing scalable and modular assembly and communication of Lego-like IP components. They also gracefully support dynamic voltage and frequency scaling (DVFS). This talk is in two parts. First, I will give a brief overview of asynchronous and GALS design, including recent industrial advances: Intel (Ethernet switch chips), IBM (TrueNorth neuromorphic computer), STMicroelectronics' STHORM processor, and others. In the second part, I will introduce our recent asynchronous switch designs for low-overhead GALS interconnection networks. Asynchronous networks-on-chip (NoCs) are an appealing solution to tackle the synchronization challenge in multicore systems. However, they have found only limited applicability so far due to two main reasons: the lack of proper design tool flows, as well as significant area footprint and performance overheads over synchronous counterparts. This talk proposes a largely unexplored design point for asynchronous NoCs, relying on transition-signaling (i.e. 2-phase handshaking) and single-rail bundled data encoding, which contribute to break the above barriers. In a post-layout comparison to a leading lightweight synchronous switch architecture, "xpipesLite," in identical low-power 40nm technology, the asynchronous switch achieved a 71% area reduction, up to 85% reduction in overall power consumption, and a 44% average reduction in energy-per-flit, while operating at over 900 MHz. Significant system latency benefits are also demonstrated. A semi-automated CAD tool flow was developed, using Synopsys Design Compiler and IC Compiler, which allows the creation of partially-reconfigurable macros. Recent acceleration techniques further improve latency by over 30%.
[This work appeared in DATE-13 and DAC-15. The former represents joint work with Prof. Davide Bertozzi's group, University of Ferrara, Italy.]

BIO: STEVEN M. NOWICK is a Professor of Computer Science and Electrical Engineering at Columbia University, and co-founder and former chair of the Computer Engineering Program. He received a Ph.D. in Computer Science from Stanford University in 1993, and a B.A. from Yale University. His main research area is on design methodologies and CAD tools for synthesis and optimization of asynchronous and mixed-timing (i.e. GALS) digital systems. His current projects include: scalable networks-on-chip (NoC's) for shared-memory parallel processors and embedded systems, ultra-low energy digital systems, fault tolerance, and low-power and robust global communication. Dr. Nowick is an IEEE Fellow, a recipient of an Alfred P. Sloan Research Fellowship, and NSF CAREER and RIA Awards. He received Best Paper Awards at the IEEE International Conference on Computer Design (1991, 2012) and the

IEEE Async Symposium (2000). He co-founded the IEEE "Async" Symposia series in 1994, and was its Program Committee Co-Chair and General Co-Chair. He was Program Chair of the IEEE/ACM International Workshop on Logic and Synthesis (IWLS), and program track/subcommittee chair at DAC, DATE and ICCD conferences. He is currently an associate editor of IEEE Design & Test magazine, ACM Journal on Emerging Technologies in Computer Systems, and IEEE Transactions on VLSI Systems, and former associate editor of IEEE Transactions on CAD. He was the selection committee chair of the ACM/SIGDA Outstanding Dissertation in EDA (OPDA) Award, and a member of the Best Paper award selection committees of ACM/IEEE DAC and ICCAD conferences. He also a recipient of the Columbia Engineering School Alumni Distinguished Faculty Teaching Award. He holds 12 issued US patents.

Session I: Interconnects (9:30am-10:45am)

09:30am-09:50am
"SI for Free: Machine Learning of Interconnect Coupling Delay and Transition Effects"
Andrew Kahng, Mulong Luo, Siddhartha Nath
UCSD

In advanced technology nodes, incremental delay due to coupling is a serious concern. Design companies spend significant resources on static timing analysis (STA) tool licenses with signal integrity (SI) enabled. The runtime of the STA tools in SI mode is typically large due to complex algorithms and iterative calculation of timing windows to accurately determine aggressor and victim alignments, as well as delay and slew estimations. In this work, we develop machine learning-based predictors of timing in SI mode based on timing reports from non-SI mode. Timing analysis in non-SI mode is faster and the license costs can be several times less than those of SI mode. We determine electrical and logic structure parameters that affect the incremental arc delay/slew and path delay (i.e., the difference in arrival times at the clock pin of the launch flip-flop and the D pin of the capture flip-flop) in SI mode, and develop models that can predict these SI-aware delays. We report worst-case error of 7.0ps and average error of 0.7ps for our models to predict incremental transition time, worst-case error of 5.2ps and average error of 1.2ps for our models to predict incremental delay, and worst-case error of 8.2ps and average error of 1.7ps for our models to predict path delay, in 28nm FDSOI technology. We also demonstrate that our models are robust across designs and signoff constraints at a particular technology node.

09:50am-10:10am
"Smart I/Os: A Data-pattern Aware 2.5D Interconnect with Space-Time Multiplexing"
Sai Dinakarrao, Kanwen Wang, Hantao Huang and Hao Yu
Nanyang Technological University

A data-pattern aware smart I/O is introduced in this paper for 2.5D through-silicon interposer (TSI) interconnect based memory-logic integration. To match huge many-core bandwidth demand with limited supply of 2.5D I/O channels when accessing one shared

memory, a space-time multiplexing based channel utilisation is developed inside the memory controller to reuse 2.5D I/O channels. Many cores are adaptively classified into clusters based on the bandwidth demand by space multiplexing to access the shared memory. Time multiplexing is then performed to schedule the cores in one cluster to occupy the supplied 2.5D I/O channels at different time-slots upon priority. The proposed smart 2.5D TSI I/O is verified by the system-level simulator with benchmarked workloads, which shows up to 58.85% bandwidth balancing and 11.90% QoS improvement.

10:10am-10:30am
"Compact Modeling and System Implications of Microring Modulators in Nanophotonic Interconnects"
Rui Wu[1], Chin-Hui Chen[2], Jean-Marc Fedeli[3], Maryse Fournier[3], Raymond G. Beausoleil[2], Kwang-Ting Cheng[1]
[1]UCSB [2]HP [3]CEA LETI

Silicon microring modulators are critical components in op- tical on-chip communications. In this paper, we develop the- oretical compact models for optical transmission, power con- sumption, bit-error-rate (BER), and electrical tuning of mi- croring modulators. The proposed theoretical models have been extensively validated by fabricated devices from a num- ber of designs and fabrication batches. Since the quality factor (Q) and the extinction ratio (ER) of the microring modulator are important to determine the BER and link power budget, we include accurate equations for the Q and the ER in our models. Based on the proposed models, we identify an extra power penalty for the electrical tuning, and an energy-efficient swing voltage for the microring modula- tor to achieve to minimum total energy consumption.

Session II: Timing and Clocking (11:00am-12:25pm)

11:00am-11:20am
"Clock Clustering and IO Optimization for 3D Integration"
Samyoung Bang[1], Kwangsoo Han[2], Andrew B. Kahng[2], Vaishnav Srinivas[2]
[1]Samsung [2]UCSD

3D interconnect between two dies can span a wide range of bandwidths and region areas, depending on the application, partitioning of the dies, die size, and floorplan. We explore the concept of dividing such an interconnect into local *clusters*, each with a *cluster clock*. We combine such clustering with a choice of three clock synchronization schemes (synchronous, source-synchronous, asynchronous) and study impacts on power, area and timing of the clock tree, data path and 3DIO. We build a model for the power, area and timing as a function of key system requirements and constraints: total bandwidth, region area, number of clusters, clock synchronization scheme, and 3DIO frequency. Such a model enables architects to perform pathfinding exploration of clocking and IO power, area and bandwidth optimization for 3D integration.

11:20am-11:40am
"On Fast Timing Closure: Speeding up Incremental Path-Based Timing Analysis with MapReduce"
Tsung-Wei Huang and Martin D. F. Wong
UIUC

Incremental path-based timing analysis (PBA) is a pivotal step in the timing optimization flow. A core building block analyzes the timing path-by-path subject to a critical amount of incremental changes on the design. However, this process in nature demands an extremely high computational complexity and has been a major bottleneck in accelerating timing closure. Therefore, we introduce in this paper a fast and scalable algorithm of incremental PBA with *MapReduce* – a recently popular programming paradigm in big-data era. Inspired by the spirit of MapReduce, we formulate our problem into tasks that are associated with keys and values and perform massively-parallel map and reduce operations on a distributed system. Experimental results demonstrated that our approach can not only easily analyze huge deisgns in a few minutes, but also quickly revalidate the timing after the incremental changes. Our results are beneficial for speeding up the lengthy design cycle of timing closure.

11:40am-12:10pm
(Invited) "An Analytic Evaluation of the Effects of Power Supply Noise on the Clock Distribution Network"
Peter Feldmann, J. Adam Butts, Timothy J. Correia, David E. Shaw
D.E. Shaw Research

Power supply noise is a significant source of jitter in clock distribution networks. This presentation focuses on one important source of power supply noise–induced jitter: the cumulative effect of noise over the long distribution path of the clock signal from its on-chip source (the PLL) to its destination (flip-flops, latches, memories). We formulate clock jitter as the solution of a differential equation that is solved analytically for relevant supply noise patterns and validated by independent Monte-Carlo simulations. Our analysis shows that the induced jitter amplitude depends strongly on the frequency spectrum of the supply noise. The results support a clock and power distribution co-design process, which must ensure that power supply resonances do not overlap with clock network sensitivities.

Invited Talk 1 (1:30pm-2:00pm)
"Signal-Integrity-Aware Power Delivery Modeling for Future-Generation I/O's"
Wei Kai Shih
Intel

No Abstract is available

Invited Talk 2 (2:00pm-2:30pm)
Modeling the Performance of Advanced Architectures

Kevin J. Barker
Pacific Northwest National Lab (PNNL)

As large-scale systems increase in complexity, sophisticated tools are required to navigate daunting application and system design and optimization spaces. Such tools and methods are becoming increasingly important as extreme-scale systems incorporate novel technologies, and as performance is no longer the sole metric of interest. We present an analytical modeling methodology that addresses many of the challenges posed in trying to gain an understanding of application behavior on current and future large-scale systems. In this talk, we provide a background of our modeling methodology that has been developed and successfully utilized at Pacific Northwest National Laboratory, as well as two case studies utilizing our approach to study the capabilities of future parallel systems. Additionally, we describe how an approach that has been developed to study application-specific performance has been adapted to explore the interaction between performance and energy consumption.

Session III: System Design (4:00pm-5:15pm)

4:00pm-4:20pm
"Power Line Communication for Hybrid Power/Signal Pin SoC Design"
Xiang Zhang[1], Yang Liu[2], Ryan Coutts[1] and Chung-Kuan Cheng[1]
[1]UCSD, [2]Xidian University

The number of available pins in ball grid array (BGA) of modern system-on-chips (SOCs) has been discussed as one of the major bottlenecks to the performance of the processors, for example many-core enabled portable devices, where the pack- age size and PCB floorplan are tightly constrained. A typical SOC package allocates more than half of the pins for power delivery, resulting in the number of IO pins for off-chip communications is greatly reduced. We observe that the requirement for the number of power and ground (P/G) pins is driven by the highest performance state and the worst design corners, while SOCs are in lower performance state for most of the time for longer battery life. Under this observation, we propose to reuse some of the power pins as dynamic power/signal pins for off-chip data transmissions to increase the off-chip bandwidth during SOC low performance state. Our proposed method provides 20Gbps bandwidth per hybrid pin pair, while providing minimum impact to the original power delivery network (PDN) design.

4:20pm-4:40pm
"Lynx: A Self-Organizing Wireless Sensor Network with Commodity Palmtop Computers"
Haifeng Xu, Melissa Bilec, William Collinge, Laura Schaefer, Amy Landis, Alex Jones
University of Pittsburgh

While the embedded class processors found in commodity palmtop computers continue to become increasingly capable, various wireless connectivity functions on them provide new opportunities in designing more flexible yet smarter wireless sensor networks

(WSNs), and utilizing the computation power in a way we could never imagine before. Designing Lynx, a self- organizing wireless sensor network (SOWSN), is our further step taken in exploiting the potential of palmtop computers. Fundamental functionalities such as automatic neighbor relation detection, link state maintenance, sensor integration, and multi- hop routing, together make a real world distributively managed WSN system implementation work quite well. And by combining with Ocelot, our mobile distributed computing engine, sensor nodes are now capable of collecting, recording, processing and sending data without any central server support. Significant energy saving is achieved by the Lynx and Ocelot combined system, compare to traditional power-hungry computer platforms such as BOINC when doing same tasks.

4:40pm-5:00pm
"Multi-Product Floorplan and Uncore Design Framework for Chip Multiprocessors"
Marco Escalante[1], Andrew Kahng[2], Michael Kishinevsky[1], Umit Ogras[3], Kambiz Samadi[4]
[1]Intel [2]UCSD [3]ASU [4]Qualcomm

Chip multiprocessors (CMPs) for server and high-performance computing markets are offered in multiple classes to satisfy various power, performance and cost requirements. As the number of processor cores on a single die grows, resources outside the "core", such as the distributed last-level cache, on-chip memory controllers and network-on-chip (NoC) interconnecting these resources, which constitute the "uncore", play an increasingly important role. While it is crucial to optimize the floorplan and uncore of each product class to achieve the best power-performance tradeoff, independent optimization may greatly increase the design effort, and undermine the savings ultimately achieved with a given total amount of optimization effort. This paper presents a novel multi-product optimization framework for next generation CMPs. Unlike traditional chip optimization techniques, we optimize the floorplan of multiple product classes at once, and ensure that the smaller floorplans can be obtained from larger ones by optimally removing, i.e., *chopping*, the unused parts.

ORGANIZING AND TECHNICAL COMMITTEES

General Chair
Rasit O. Topaloglu
IBM, USA

Technical Program Chair
Baris Taskin
Drexel University, USA

Finance Chair
Umit Ogras
Arizona State University, USA

Publicity Chair
Tsung-Yi Ho
National Chiao Tung University, Taiwan

Panel Chair
Cheng Zhuo
Intel, USA

Publications Chair
Joseph Zambreno
Iowa State University, USA

Steering Committee Members
Chuck Alpert
Cadence, USA
Deming Chen
University of Illinois at Urbana-Champaign, USA
Chung-Kuan Cheng
University of California, San Diego, USA
Andrew B. Kahng
University of California, San Diego, USA
Michael Kishinevsky
Intel, USA

Program Committee Members

Paul Bogdan
University of Southern California, USA

Tsung-Yi Ho
National Chiao Tung University, Taiwan

Hui-Ri Jiang
National Chiao Tung University, Taiwan

Taewhan Kim
Seoul National University, Korea

Selcuk Kose
University of South Florida, USA

Umit Ogras
Arizona State University

Payman Parkesh-Ha
University of New Mexico, USA

Emre Salman
Stony Brook University, USA

William Swartz, Jr
TimberWolf Systems, USA

Baris Taskin
Drexel University, USA

Rasit Topaloglu
IBM, USA

Evangeline F.Y. Young
The Chinese University of Hong Kong

Joseph Zambreno
Iowa State University

Cheng Zhuo
Intel, USA