RESEARCH STATEMENT

Integrated circuits and hardware computing systems form the backbone of our information age. Integrating a growing number of digital and analog devices with shrinking footprints on the chip has been the essential avenue to higher performance, more functionality, and improved robustness. At the same time, integration has inevitably resulted in unprecedented design complexity and a myriad of electrical, thermal, and variability effects that must be accounted for in design and verification. In the meantime, the era of data-driven science and engineering has rendered a growing gap between the supply and demand of computing power. This calls for new computational models, circuits, and architectures to meet the escalating performance and energy efficiency needs.

Under the above context, my work has centered on:

1) *Analog, mixed-signal, RF, and memory circuits*, aiming at enabling modeling, verification, and optimization of the interfaces of our electronic systems with the "real world" and data storage, which are key bottlenecks in advancing computing capabilities, design robustness, and energy efficiency;

2) *Energy-efficient adaptive IC power delivery systems*, aiming at enabling design and analysis of large-scale systems that are responsible for distributing and regulating power, and hence playing a key role in robust low-power design;

3) *Computational neuroscience, brain-inspired computing*, and hardware machine learning systems, aiming at applying the learning and expertise from integrated circuit design to understanding the brain and applying this knowledge back to electronic systems to enable brain-inspired computing and hardware accelerated machine learning.

Modeling, verification & optimization of analog/mixed-signal/RF and memory circuits

My early work in this space focused on automated macromodeling of a broad family of time-invariant and time-varying nonlinear analog /RF circuits, developing theory and algorithms of model order reduction of weakly nonlinear systems with significantly improved accuracy and efficiency [1] (**2003 DAC Best Paper Award**). Read, write and standby operations of SRAMs are intrinsically dynamic processes, which the conventional static noise margins fail to capture. For the very first time, my research group proposed the new concepts of *dynamic* noise margins to characterize read-stability, writeability, and data retention of SRAMs with precise consideration of temporal effects, supported by the development of efficient transistor-level analysis techniques [2] (**2008 ICCAD Best Paper Award Nomination**). In a related direction, we developed the first analytical model revealing the dynamic properties of nanoscale memristive crossbar memories, providing design guidance for read/write circuit design [3]. On the transistor-level simulation front, my research group developed a number of parallel methods exploring rich parallelisms ranging from fine-grained matrix-level techniques to coarse-grained multi-algorithm parallelism on modern multi-core processor architectures and distributed platforms [4] (**2008 DAC Best Paper Award**). Based on my expertise in parallel CAD, I consulted for major semiconductor and EDA start-up companies such as Intel and helped establish parallel EDA a mainstream practice.

I have done significant work on verification and optimization of analog/mixed-signal (AMS) ICs, addressing tough design challenges and exploiting machine intelligence. One of these efforts was on automated detection of marginally-stable feedback loops formed by layout parasitics in large extracted industrial analog circuits, allowing for practical identification of design failures caused by noise susceptibility induced by these loops [5] (**2011 DAC Best Paper Award**). This technology has been adopted by Texas Instruments as a mandatory design sign-off step for a large portfolio of TI's high-precision analog ICs, and by Cadence Design Systems.

As demonstrated in my work [6, 7, 22, 23], statistical learning of complex parametric dependencies and failure mechanisms paves a powerful data-driven avenue for yield-aware performance modeling & optimization and verification of AMS circuits. I have worked towards bridging the widening gap between design complexity and verification coverage by developing the first hybrid framework integrating formal verification and statistical machine learning [8]. This new direction aims to detect extremely rare design failures while providing a degree of formal coverage. Along a similar line, recently I have explored Bayesian learning under the context of Bayesian optimization for robust rare failure detection [27] (**2018 ICCAD Best Paper Award Nomination**). These approaches have shown great promises and may be extended for rare failure detection of safety-critical systems such as automotive electronics.

Design and analysis of energy-efficient adaptive IC power-delivery systems

IC power delivery design presents key challenges and opportunities for low-power electronics. Towards analysis and verification of today's massive billion-node power delivery networks, we have developed a family of parallel algorithms and tools. Among these, my research group demonstrated the first GPU-accelerated power grid analysis tool, speeding up the runtime by more than two orders of magnitude [9] (**2008 ICCAD Best Paper Award Nomination**). More broadly, we developed a heterogonous GPU-CPU computing paradigm wherein intelligent algorithm design is capitalized to expose hidden data-parallel compute work in a non-data parallel application, producing large speedups via processing of the exposed data-parallel workload on massively parallel GPUs. This line of work stimulated a great deal of follow-up work from the EDA and GPU computing communities. One of my consulted start-up companies, Gear Design Solutions which developed parallel power grid analysis technologies, was acquired by ANSYS in 2015.

A significant focus of my work is placed on the full processing chain of voltage conversion, regulation, distribution, and their co-design, aiming at the ultimate power quality and efficiency. Part of these efforts has been towards enabling distributed on-chip voltage regulation, a recent industrial design trend. Distributed on-chip regulation provides fast responses to local current load changes and is well-positioned to address many challenges in power integrity and power efficiency. However, spatially distributing multiple voltage regulators within the same voltage domain strongly couples active regulators with the large passive power grids, making stability-ensuring design a daunting task. My research group showed that the phase/gain margins widely adopted in analog circuit design can produce completely misleading prediction of network stability. We developed the very first rigorous and feasible solution to overcome this roadblock to distributed voltage regulation under the framework of hybrid stability design, leading to results of both theoretical and practical significance [10] (**2012 ICCAD Best Paper Award**). Based on our newly introduced hybrid

stability margin metric, our approach provides additional design freedoms and embraces localized design, decoupling the design of active voltage regulators from that of large passive power grids while guaranteeing the stability of the entire network. This makes design of large power delivery networks trackable. Our more recent work minimizes the pessimism in hybrid stability analysis, thereby further improving regulation performance and design efficiency by reducing the amount of over-design [11] (**2016 DAC Best Paper Award**). Under a broader context, our on-going work explores the design of adaptive heterogenous voltage regulation consisting of multi-stages of on-chip linear voltage regulators and on-/off-chip switching converters.

Computational neuroscience, brain-inspired computing, hardware machine learning systems

At the intersection between engineering and biology, taking an interdisciplinary approach to increase our understanding of both electronic and biophysical systems can create large values. My focus in this area is on biophysically realistic modeling and computational techniques for understanding the bioelectrical processes and dynamics in biological brains. My group has developed brain models including biophysically-plausible cortical-thalamic models consisting of 22 neuronal cell types with detailed synaptic and intrinsic ionic characteristics and models at higher levels of abstraction [12, 13]. Apart from modeling microcircuitry or local neural activities, these models capture brain-wide behavior by encompassing 70 cortical regions based on brain imaging connectivity data. Advanced parallel numerical methods were developed to enable simulation of large-scale brain models with millions of neurons and hundreds of millions of states variables [12]. These capabilities provide an in-silico basis for understanding brain dynamic behavior, e.g. the underlying intrinsic and synaptic characteristics and local/global connectivity responsible for synchronized firing activities in the brain. I have collaborated with TAMU neuroscience faculty to find the causes of epilepsies and develop pharmaceutical treatments.

In the other direction, i.e. from biology back to engineering, I have strived to develop new architectures and circuits inspired by the computational principles in biological brains. This is motivated by the fact that with the slowing down of CMOS technology the mainstream von Neumann architecture is severely limited in meeting the escalating performance and energy efficiency needs in the era of data-driven science and engineering. This calls for disruptive paradigm shifts to close the widening gap between the supply and demand of computing power. I have endeavored to work on brain-inspired computing and hardware machine learning systems through the following lines of research: 1) embrace biologically plausible models of computation based on spiking neural networks (SNN), 2) explore computational powerful SSN learning architectures, and 3) address major learning and training challenges with SNNs, and 4) develop self-adaptive hardware architectures and circuits with integrated SSN or statistical learning capabilities.

The vision behind my SSN work is two-fold: SNNs closely mimic the spatiotemporal information processing of biological brains and their biological plausibility facilitates natural exploitation of brain-inspired computing; SNNs immediately lends themselves to ultra-low energy event-driven processing on neuromorphic hardware as initially demonstrated on the IBM TrueNorth and Intel Loihi neural processors. Nevertheless, to date SNNs are far from being completely understood, and their potentials are yet to be fully exploited to make SNNs a competitive choice for challenging real-world tasks. To this end, I have strived to raise the computing performance of SNNs by

exploring computational powerful network architectures and addressing major SNN learning/training challenges brought by discontinuous spiking activities. My group is exploring a range of architectures with varying complexities and computing capabilities: spiking feedforward fully-connected and convolutional networks (CNNs), recurrent liquid state machine (LSM) networks, hybrid feedforward and LSM networks, and networks with local recurrences and global feedbacks. Furthermore, we extended the standard single-stage LSM model into a deep LSM architecture (D-LSM) [14]. To address the training challenge, we developed an SSN error backpropagation algorithm that delivers the best recognition performances among all reported SNNs evaluated using the NMIST, neuromorphic MNIST (N-MNIST), and Extended MNIST (EMNIST) datasets, and the challenging 16-speaker spatio-temporal English letters of the TI46 Speech Corpus (**NIPS'18**) [15]. In parallel, brain-inspired Hebbian and on-Hebbian learning mechanisms immediately lead to local, parallel, distributed processing, and amenability to hardware realization, and are the key focus of our work. We have developed a family of spike-dependent supervised and unsupervised synaptic plasticity rules for training recurrent reservoir networks, demonstrating highly competitive results for speech/image classifications [16, 17]. Our latest work on intrinsic plasticity (IP) self-tunes the membrane time constant, capacitance, and leaky resistance of individual spiking neurons while maximizing the entropy of the output firing activities, i.e. information transfer capability, leading to significant performance improvements for challenging multi-speaker speech and image recognition tasks.

My neuromorphic hardware design work exploits the unique properties of SNN models of computation to develop energy-efficient circuits and architectures with integrated learning and runtime self-adaptation [18, 19, 20]. Large energy saving is achieved by embracing a number of opportunities: network architecture/learning algorithm/hardware co-design, event-driven processing exploiting spatial/temporal sparsity in firing activities, runtime allocation of computing resources based on data/task dependent firing activities and built-in learning performance prediction [21] (**ISCAS Honorary Mention Best Paper Award**).

Future work

Moving forward, I will build upon my past work to focus on important problems in hardware system design, computing, and neuroscience.

I will work towards developing machine learning (ML) enabled VLSI design methodologies and verifiable ML systems. As one example, I will expand my design verification research from the domain of integrated circuits to safety-critical systems in medical, transportation, vehicle, and aerospace applications, which are gaining importance due to their fast-paced adoption in the society. Ensuring reliability and safety is of paramount importance for these systems, however, becomes increasingly challenging to achieve as more automation and machine intelligence is integrated. I will tackle this verification challenge by exploring deeper synergies between formal verification and machine learning. In addition, I will work on robust machine learning for VLSI design and general deep learning applications [24].

In term of power delivery, the research focus will be to enable low-power systems powered by a variety of energy sources and regulated by multi-stages of on/off-chip voltage regulators. One

direction here is to achieve the ultimate power efficiency by adapting voltage regulation and power delivery networks with respect to real-time power supply and workload conditions. Autonomously-adaptive heterogeneous power networks show great promises. Such systems will be targeted by deploying low-cost on-chip voltage/current sensors and built-in machine learning to predict workload variations and by proactively adapting on-/off-chip voltage regulators and power distribution networks.

At the crossroad of computational neuroscience and brain inspired computing, I will investigate models of varying degrees of biological plausibility to help understand functions and information processing of the nervous system. The resulting biological understandings will be also used for architecting novel brain-inspired computational models. Integration of biological insights and theoretical learning principles in conjunction with co-optimization of learning architectures/mechanisms, hardware architectures, and circuit design while leveraging a target technology, e.g. CMOS, 3D integration [25, 26], or post-CMOS, can go a long way to engineer new generations of hardware computing systems.

While striking a balance between theoretical rigor and real-world relevance, in my future work I will bridge between theory, algorithms, and engineering design at the exciting intersections between hardware system design & verification, computational neuroscience, brain-inspired computing, and machine learning. I will support my research agendas by building multidisciplinary collaborations.

References

[1] Peng Li and Lawrence Pileggi, "NORM: compact model order reduction of weakly nonlinear systems," IEEE/ACM Design Automation Conference, pp. 472-477, June 2003 (Best Paper Award).

[2] Wei Dong, Peng Li, and Garng M. Huang, "SRAM dynamic stability: theory, variability and analysis," IEEE/ACM Int. Conf. on Computer-Aided Design, pp. 378-385, November 2008 (Best Paper Award Nomination).

[3] Yenpo Ho, Garng M. Huang, and Peng Li, "Nonvolatile memristor memory: device characteristics and design implications," IEEE/ACM Int. Conf. on Computer-Aided Design, pp. 485-490, November 2009.

[4] Wei Dong, Peng Li, and Xiaoji Ye, "WavePipe: parallel transient simulation of analog and digital circuits on multi-core shared-memory machines," IEEE/ACM Design Automation Conf., pp. 238-243, June 2008 (Best Paper Award).

[5] Parijat Mukherjee, Peter Fang, Rod Burt, and Peng Li, "Automatic stability checking for large linear analog integrated circuits," IEEE/ACM Design Automation Conference, pp. 304-309, June 2011 (Best Paper Award).

[6] Guo Yu and Peng Li, "Yield-aware hierarchical optimization of large analog integrated circuits," IEEE/ACM Int. Conf. on Computer-Aided Design, pp. 79-84, November 2008.

[7] Guo Yu and Peng Li, "Yield-aware analog integrated circuit optimization using Geostatistics motivated performance modeling," in Proc. of IEEE/ACM Int. Conf. on Computer-Aided Design, pp. 464-469, November 2007.

[8] Hanbin Hu, Qingran Zheng, Ya Wang, and Peng Li, "HFMV: Hybridizing formal methods and machine Learning for verification of analog and mixed-signal circuits," in Proc. of IEEE/ACM Design Automation Conference, June 2018.

[9] Zhuo Feng and Peng Li, "Multigrid on GPU: tackling power grid analysis on parallel SIMT platforms," in Proc. of IEEE/ACM Int. Conf. on Computer-Aided Design, pp. 647-654, November 2008 (Best Paper Award Nomination).

[10] Suming Lai, Boyuan Yan, and Peng Li, "Stability assurance and design optimization of large power delivery networks with multiple on-chip voltage regulators," IEEE/ACM Intl. Conf. on Computer-Aided Design, pp. 247-254, November 2012 (Best Paper Award).

[11] Xin Zhan, Peng Li, and Edgar Sanchez-Sinencio, "Distributed on-chip voltage regulation: theoretical stability foundation, over-design reduction and performance optimization," in Proc. of IEEE/ACM Design Automation Conference, pp. 54:1-54:6, June 2016 (Best Paper Award).

[12] Yong Zhang, Boyuan Yan, Mingchao Wang, Jingzhen Hu, Haokai Lu, and Peng Li, "Linking brain behavior to underlying cellular mechanisms via large-scale brain modeling and simulation," Neurocomputing, vo. 97, pp. 317-331, Nov. 2012.

[13] Boyuan Yan and Peng Li, "The emergence of abnormal hypersynchronization in the anatomical structural network of human brain," NeuroImage, volume 65, pp. 34-51, January 2013.

[14] Qian Wang, and Peng Li, "D-LSM: Deep liquid state machine with unsupervised recurrent reservoir tuning," The International Conference on Pattern Recognition, pp. 1158-1165, December 2016.

[15] Yingyezhe Jin, Wenrui Zhang, and Peng Li, "Hybrid macro/micro level backpropagation for training deep spiking neural networks," Conference on Neural Information Processing Systems (NIPS), 2018.

[16] Yong Zhang, Peng Li, Yingyezhe Jin, and Yoonsuck Choe, "A digital liquid state machine with biologically inspired learning and its application to speech recognition," IEEE Trans. on Neural Networks and Learning Systems, vol. 26, no. 11, pp. 2635-2649, Nov. 2015.

[17] Yingyezhe Jin, and Peng Li, "Calcium-modulated supervised spike-timing-dependent plasticity for readout training and sparsification of the liquid state machine," International Joint Conference on Neural Networks, pp. 2007-2014, May 2017.

[18] Yu Liu, Yingyezhe Jin, and Peng Li, "Exploring sparsity of firing activities and clock gating for energy-efficient recurrent spiking neural processors," IEEE/ACM Intl. Symp. on Low Power Electronics and Design (ISLPED), pp.1-6, July 2017.

[19] Yingyezhe Jin, Yu Liu, and Peng Li, "SSO-LSM: A sparse and self-organizing architecture for liquid state machine based neural processors," In Proc. of IEEE/ACM International Symposium on Nanoscale Architectures, pp. 55-60, July 2016.

[20] Yu Liu, Yingyezhe Jin, and Peng Li, "Online adaptation and energy minimization for hardware recurrent spiking neural networks," in ACM Journal on Emerging Technologies in Computing Systems, vo. 14, no. 1, pp. 11:1 – 11:21, Jan. 2018.

[21] Qian Wang, Youjie Li, and Peng Li, "Liquid state machine based pattern recognition on FPGA with firing-activity dependent power gating and approximate computing," in Proc. of IEEE Intl. Symposium of Circuits and Systems, pp. 361-364, May 2016 (ISCAS Honorary Mention Best Paper Award).

[22] Parijat Mukherjee, Chirayu Amin, and Peng Li, "Approximate property checking of mixed-signal circuits," IEEE/ACM Design Automation Conference, pp. 1-6, June 2014.

[23] Honghuang Lin and Peng Li, "Relevance vector and feature machine for statistical analog circuit characterization and built-in self-test optimization," IEEE/ACM Design Automation Conference, pp. 11:1-11:6, June 2016.

[24] T.J. Chang, Yukun He, and Peng Li, "Efficient two-step adversarial defense for deep neural networks," International Conference on Learning Representations (ICLR), 2019 (submitted).

[25] Bon Woong Ku, Yu Liu, Yingyezhe Jin, Peng Li, and Sung Kyu Lim, "Area-efficient and Low-power Face-to-Face-bonded 3D liquid state machine Design," IEEE/ACM Conf. on Computer-Aided Design, November 2018.

[26] Bon Woong Ku, Yu Liu, Yingyezhe Jin, Sandeep Samal, Peng Li, and Sung Kyu Lim, "Design and architectural co-optimization of monolithic 3D liquid state machine-based neuromorphic processor," IEEE/ACM Design Automation Conference (DAC), pp. 165:1–165:6, June 2018.

[27] [ICCAD'18] Hanbin Hu, Peng Li, and Jianhua Z. Huang, "Parallelizable Bayesian optimization for analog and mixed-signal rare failure detection with high coverage," in Proc. of IEEE/ACM Conf. on Computer-Aided Design, November 2018 (Best Paper Award Nomination).