Sujan Kumar Gonugondla

sujan.gonugondla@gmail.com, (217) 848-4808

EDUCATION

- **University of Illinois at Urbana-Champaign** Illinois, USA
  Ph.D in Electrical and Computer Engineering *Aug. 2014 - June 2020*
  Thesis: Cross-layer methods for energy-efficient inference using in-memory architectures
  Adviser: Prof. Naresh Shanbhag

- **Indian Institute of Technology Madras** Chennai, India
  Masters in Technologies (M.Tech) in Electrical Engineering *Aug. 2009 - July 2014*
  Thesis: Optical Phase Lock Loops
  Bachelors in Technologies (B.Tech) in Electrical Engineering
  Minor in Physics

EXPERIENCE

- **Research Scientist II, Amazon.com Services LLC** *June 2020 - ongoing*
  Manager: Dr. Kamin Whitehouse
  - Developing machine learning algorithms for on-device implementation in a brand new Amazon Alexa product.

- **Research Assistantship, Adviser: Prof. Naresh Shanbhag** Urbana-Champaign, IL
  - **Deep In-Memory Computing Architectures for Machine Learning** *July 2014 - Dec 2018*
    - This project explores deep in-memory architecture (DIMA), where computations are embedded into the memory arrays. DIMA drastically reduces memory read and data access costs by performing mixed signal computations in the periphery of the memory array.
    - Designed a DIMA based architecture to accelerate Convolutional Neural Networks (CNN) that shows over 4.9× and 2.4× improvements in energy-efficiency and throughput. [ICASSP 2015, JETCAS 2018]
    - Designed a multi-functional SRAM based DIMA IC on 65 nm CMOS to implement machine learning algorithms such as SVM, k-NN, template matching and matched filtering while showing 10× energy savings and 5×. throughput enhancement. [JSSC 2018]
    - Proposed an optimization framework for a programmable instruction set architecture (ISA) with DIMA for machine learning acceleration. [ISCA 2018]
  - **Application Specific ML accelerators using In-Memory Architectures** *Feb. 2015 - Dec 2019*
    - Designed the world's first dedicated IC for random forest algorithm using DIMA on 65 nm CMOS showing 6.8× improvement in energy-delay product.[ESSCIRC 2017, JSSC 2018]
    - Designed a dedicated key-word spotting IC that achieves state-of-the-art decision latency of 39.9 μs with a decision energy $< 0.5\,\mu$J/dec which translates to more than 24× savings in the energy-delay product (EDP) of decisions over existing KWS ICs. [CICC 2019]
  - **Variation-tolerant machine learning classifier via on-chip training** *May 2016 - Aug 2018*
    - Studied the use of on-chip learning to achieve variation-tolerance in mixed-signal machine learning architectures. We designed a DIMA based classifier IC with on-chip SGD based trainer on 65 nm CMOS which adapts to changing environment and process parameters. It demonstrates 2.4× energy reduction over previous ICs. [ISSCC 2018, JSSC 2018]
  - **Machine learning acceleration using Beyond-CMOS devices** *July 2016 - May 2020*
    - Analysed GDOT a dot-product nano-function using graphene transistors. GDOT implementing a Gaussian blur shows up to 10^4 greater signal-to-noise ratio (SNR) over CMOS based implementations. [VLSI-T Symposia 2016]
    - Designed a NAND flash based DIMA for large scale machine learning acceleration showing 8× reduction in energy consumption. [ISCAS 2018]

- Designed an ML accelerator using MRAM devices that achieves 4.5× and 70× lower energy and delay than their CMOS counterparts [ISCAS 2019]
- Proposed SWIPE algorithm that achieves high-accuracy writes for RRAM crossbar-based machine learning accelerators at 5×-to-10× lower cost than standard program-verify methods. [ICCAD 2020]

- **Fundamental Limits of In-Memory Architectures** *July 2018 - Present*
  - Developed methodologies to analytically relate circuit, architecture and process parameters to energy, SNR and throughput of in-memory architectures.
  - Analyzed the fundamental limits of energy and precision of in-memory architectures. [ICCAD 2020]
- **Low Bit-width Quantization of Deep Nets** *July 2018 - Present*
  - Designed an iterative precision reduction algorithm to train deep neural networks with mixed/low bit-width quantization of weights and activation and demonstrated improvements over State-of-the-art techniques via experiments on ImageNet, CIFAR-100, SVHN, and CIFAR-10 datasets.
- **Energy Efficiency via Statistical Error Compensation** *Jan. 2015 - Oct. 2015*
  - Developed a novel statistical error compensation (SEC) technique called algorithmic error cancellation (AEC) for designing robust and energy-efficient computational kernels on scaled process technologies. AEC compensates for timing error rates of up to 73% while achieving energy savings of up to 27.7%. [ICASSP 2016]

- **Research Intern - Kilby Labs at Texas Instruments** Dallas, Texas
  Manager: Dr. Mahesh Mehendale *May 2018 - Aug 2018*
  - Worked on training and compression techniques for deep neural network implementation on edge devices.
  - Developed techniques to train quantized neural network and to map it on a resource constrained platform.

- **Masters thesis, Advisers: Prof. Anil Prabhakar and Prof. Sankaran Aniruddhan** Chennai, India
  Optical Phase Locked Loops (OPLLs) *May 2013 - June 2014*
  - Designed an optical phased lock loop (OPLL) to synchronize phase and frequency of a semiconductor laser and evaluated different feedback circuit topologies.
  - Improved stability and acquisition range of optical phased locked loops using Type II feedback circuits.

- **Assistive Data Gloves for Speech Impaired** Chennai, India
  Adviser: Prof. Nagendra Krishnapura *July 2013 - Sept. 2013*
  - This project was part of Texas Instruments Analog Design Contest (TI-ADC).
  - Designed affordable data gloves for the speech impaired that uses pattern recognition algorithms and converts sign language to speech. [TIEEC 2013]

- **Design Engineer Intern, Texas Instruments India** Bangalore, India
  Manager: Amit Ashara *May 2012 - July 2012*
  - Worked with the validation team of TI's Stellaris microcontroller.
  - Identified and isolated bugs on analog comparator and ADC components of the microcontroller.

HONORS AND AWARDS

- SSCS Predoctoral Achievement Award 2020
- M. E. Van Valkenburg Graduate Research Award 2019.
- Honorable Mention in MICRO Top-Pics 2018.
- Best Paper Award - Neural Systems and Applications, ISCAS 2018.
- Finalist in Qualcomm Innovation Fellowship 2018.
- Dr. Ok Kyun Kim Fellowship 2018.
- Analog Devices Outstanding Student Designer Award 2018.
- Shannon Award for Excellence in Systems Research 2016.
- Best Student Paper Award- Design and Implementation of Signal Processing Systems, ICASSP 2016.

- MCM scholarship at IIT Madras, 2009-2013.

PROFESSIONAL SERVICES

- **Reviewer for :** IEEE Transactions on Circuits and Systems-I (TCAS-I) • IEEE Transactions on Circuits and Systems-II (TCAS-II) • IEEE International Symposium on Circuits & Systems (ISCAS) • IEEE International Conference on Artificial Intelligence Circuits & Systems (AICAS) • IEEE Journal of Solid-State Circuits (JSSC) • IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS) • IEEE Access • IEEE Transaction in Very Large Scale Integrated Circuits (T-VLSI) • IEEE Electron Device Letters

RELEVANT COURSEWORK

- **UIUC:** Random Processes, Advanced Digital Signal Processing, VLSI for Communication & Signal Processing, Machine Learning in Silicon, Topics in Image Processing, Machine Learning, Learning: Algorithms & Models • MDP & Reinforcement Learning.
- **IIT Madras:** Analog IC Design, RF IC Design, Digital IC Design, Device Modeling, VLSI Technology, Data Structures and Algorithms.

PROGRAMMING SKILLS

- MATLAB, LaTeX, Python, Verilog, Cadence and C++.

AUTHORED BOOKS

1. Mingu Kang, **Sujan K. Gonugondla**, and Naresh R. Shanbhag, "Deep In-memory Architectures for Machine Learning." Springer, 2020.

PUBLICATIONS

1. **Sujan K. Gonugondla**, Ameya Patil, and Naresh Shanbhag, "SWIPE: Enhancing Robustness of ReRAM Crossbars for In-memory Computing," *International Conference on Computer Aided Design (ICCAD)*, 2020.

2. **Sujan K. Gonugondla**, Charbel Sakr, Hassan Dbouk, and Naresh Shanbhag, "Fundamental Limits on the Precision of In-memory Architectures," *International Conference on Computer Aided Design (ICCAD)*, 2020.

3. Mingu Kang, **Sujan K. Gonugondla**, and Naresh Shanbhag, "Deep In-Memory Architectures in SRAM: An Analog Approach to Approximate Computing," *Proceedings of IEEE*, 2020.

4. Hassan Dbouk, **Sujan K. Gonugondla**, Charbel Sakr, and Naresh Shanbhag, "A 0.44 uJ/decision 39.9 us/dec Recurrent Attention In-memory Processor for Keyword Spotting," *IEEE Journal of Solid-State Circuits (JSSC)*, 2020.

5. Hassan Dbouk, **Sujan K. Gonugondla**, Charbel Sakr, and Naresh Shanbhag, "KeyRAM: A 0.34 uJ/decision 18 k decisions/s Recurrent Attention In-memory Processor for Keyword Spotting," *Custom Integrated Circuits Conference (CICC)*, 2020.

6. Chandrasekhar Radhakrishnan, **Sujan K. Gonugondla**, "Adaptive Filtering in In-Memory-Based Architectures," *Asilomar Conference on Signals, Systems and Computers*, 2019.

7. Ameya Patil, Haocheng Hua, **Sujan K. Gonugondla**, Mingu Kang, and Naresh Shanbhag. "An MRAM-based deep in-memory architecture for deep neural networks," *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2019.

8. **Sujan K. Gonugondla**, Mingu Kang, and Naresh R. Shanbhag, "A variation tolerant in-memory machine learning classifier with an on chip trainer," *IEEE Journal of Solid-State Circuits (JSSC)*, 2018.

9. Mingu Kang, Sungmin Lim, **Sujan K. Gonugondla** and Naresh R. Shanbhag, "An In-Memory VLSI Architecture for Convolutional Neural Networks," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS)*, 2018.

10. Mingu Kang, **Sujan K. Gonugondla**, Sungmin Lim, and Naresh R. Shanbhag, "A 19.4 nJ/decision, 364K decisions/s, in-memory random forest multi-class inference processor," *IEEE Journal of Solid-State Circuits (JSSC)*, 2018.

11. Prakalp Srivastava, Mingu Kang, **Sujan K. Gonugondla**, Sungmin Lim, Jungwook Choi, Nam Sung Kim, Vikram Adve, and Naresh R. Shanbhag, "PROMISE: An end-to-end design of a programmable mixed-signal accelerator for machine learning algorithms," *ACM/IEEE Annual International Symposium on Computer Architecture (ISCA)*, 2018. (**MIRCO Top-Pics Honorable Mention**)

12. **Sujan K. Gonugondla**, Mingu Kang, Yongjune Kim, Mark Helm, Sean Eilert, and Naresh R. Shanbhag, "Energy-efficient deep in-memory architectures for NAND flash memory," *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018 (**Best Paper Award - Honorable Mention**).

13. Mingu Kang, **Sujan K. Gonugondla**, Ameya D Patil, and Naresh R. Shanbhag, "A multi-functional in-memory inference processor using a standard 6T SRAM array," *IEEE Journal of Solid-State Circuits (JSSC)*, 2018.

14. **Sujan K. Gonugondla**, Mingu Kang, and Naresh R. Shanbhag, "A 42pJ/decision, 3.12 TOPS/W robust in-memory machine learning classifier with on-chip training," *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2018.

15. Mingu Kang, **Sujan K. Gonugondla**, and Naresh R. Shanbhag, "A 19.4 nJ/decision 364K decisions/s in-memory random forest classifier in 6T SRAM array," *IEEE European Solid-State Circuits Conference (ESSCIRC)*, 2017.

16. Ning Wang, **Sujan K. Gonugondla**, Ihab Nahlus, Naresh R. Shanbhag, and Eric Pop, "GDOT: A graphene based nano-function for dot-product computation," *IEEE Symposium on VLSI Technology (VLSIT)*, 2016.

17. **Sujan K. Gonugondla**, Byonghyo Shim, and Naresh R. Shanbhag, "Perfect error compensation via algorithmic error cancellation," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016 (**Best Student Paper Award**).

18. Mingu Kang, **Sujan K. Gonugondla**, Naresh R. Shanbhag, and Min-Sun Keel, "An energy-efficient memory-based high-throughput VLSI architecture for convolutional networks," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

19. Celestine Preetham, Girish Ramakrishnan, **Sujan Kumar**, Anish Tamse, and Nagendra Krishnapura, "Hand talk: implementation of a gesture recognizing glove," *Texas Instruments India Educators' Conference (TIIEC)*, 2013

PRESENTATION AND TALKS

- International Conference of Computer Aided Design, Virtual Event, United States 2020
- International Solid-State Circuits Conference, San-Francisco, California 2018
- Coordinated Science Laboratory Student Conference, Urbana, Illinois 2016, 2018
- European Solid-State Circuit Conference, Leuven, Belgium 2017
- TechCon, Semiconductor Research Cooperation (SRC), Austin, Texas 2017
- Systems On Nanoscale Information fabriCs, Champaign, Illinois 2015, 2016, 2017, 2018