Gourav Datta

https://www.linkedin.com/in/gourav-datta-959571a3/

Email : gdatta@usc.edu
Mobile : +1-213-519-2003

EDUCATION

- **University of Southern California** — Los Angeles, CA
  PhD in Electrical Engineering with focus on ML Hardware; Major GPA: 4/4; advisor: Dr. Peter Beerel — *August 2018 – August 2023*
  Major courses taken: Design for Testability (A), Computer Systems Organization (A), Advanced Computer Architecture (A), Analysis of Algorithms (B+), VLSI System Design (A), Probability and Statistics (A), Deep Learning (A), Database Systems (A), Beyond-CMOS Materials and Devices (A-), Mathematical Foundations for System Design (A)

- **Indian Institute of Technology (IIT) Kharagpur** — Kharagpur, India
  B.Tech in Electrical Engineering (minor in Electronics); GPA: 9.00 /10.0 (Class Rank: 1) — *August 2014 – May 2018*
  Graduate courses taken: Architectural Design of IC, VLSI CAD, Computer Architecture, Semiconductor Devices

INTERNSHIP EXPERIENCE

- **Amazon Alexa AI, Applied Science Internship** — Los Angeles, CA
  ***Dyadic Facial Motion Generation**; Mentors: Dr. Vivek Yadav, Dr. Yue (Rex) Wu* — *May 2022 - August 2022*
  - Proposed a novel style-aware framework for modeling facial motion in dyadic conversations that surpasses existing works in qualitative and quantitative motion realism. (Patent submitted; paper under review in ICASSP'24 and accepted in IJCAI'23 workshop.)
  - Enabled non-deterministic prediction by learning a discrete latent representation of realistic listener motion with a novel style-aware motion-encoding VQ-VAE, thereby capturing the multimodal nature of dyadic interactions.

- **Amazon Alexa AI, Applied Science Internship** — Remote, CA
  ***Multimodal Active Speaker Detection**; Mentors: Dr. Vivek Yadav, Dr. Pradeep Natarajan* — *June 2021 - August 2021*
  - Developed a novel, accurate, and multimodal (audio and video) training framework for active speaker detection (ASD) to help smart home assistants like Alexa to better interact with humans. (Work **accepted in ICASSP'22**)
  - We significantly improve the trade-off between accuracy and model size compared to existing ASD benchmarks.

- **Stealth Startup, Senior Research Engineering Internship** — Portland, OR
  ***Logic Design using Ferroelectrics**; Manager: Dr. Sasikanth Manipatruni* — *May 2020 - August 2020*
  - Developed a new class of ferroelectric logic based on low voltages, high charge density and non-volatility to power the next generation of computing with a 100x improvement in energy efficiency compared to CMOS Boolean logic.

- **Apple Inc., CPU Design Implementation** — Cupertino, CA
  ***Logic Synthesis for Timing**; Manager: Shashank Shastry* — *May 2019 - August 2019*
  - **Synthesis**: Explored *synthesis techniques* to improve timing of a design given its' RTL and used design metrics like WNS, TNS, etc. to evaluate their benefit.
    - *Functionally efficient cell usage*, which can improve placement and routing, thereby improving TNS.
    - *Fanout optimization using flop and logic duplication* to improve WNS without impacting density and power.
  - **Timing closure**: Worked closely with the methodology and the CAD team to implement the above techniques and come up with new recipes to close setup and hold timing violations.

- **INRIA Research Centre (Undergraduate Research Internship)** — Rennes, France
  ***C-based Synthesis Modelling for RISC-V processor**; advisor: Dr. Olivier Sentieys* — *May 2017 - July 2017*
  - Designed an in-order core micro-architecture supporting 32-bit RISC-V instruction set from C-based specifications using High Level Synthesis(HLS) tools.
  - Validated the design with real time RISC-V benchmarks and synthesized the core down to gate level.
  - Compared simulation results (clock frequency, area, power) with some existing RISC-V designs, such as Rocket core.

GRADUATE RESEARCH EXPERIENCE

- **University of Southern California** — Los Angeles, CA
  Rethinking the whole Computing stack (Circuit-Architecture-Algorithm Co-Design) — *Aug. 2018 - Present*
  Synchronization and Computational units for Single Flux Quantum (SFQ) technologies
  - Designed an efficient gate level pipelined 32-bit ALU for SFQ processors. Parallel 4-bit Sklanskey adder blocks are used as building blocks in a block-skewed manner to improve throughput for data dependent operations.
  - Designed a metastability-resilient clock domain crossing FIFO synchronizer (inspired by CMOS 2-flop synchronizers) that simulations show delivers 1000 reduction in logical error rate at 30 GHz compared to state-of-the-art designs.

Design of a Superconducting L_1 Cache Memory

- Developed a novel direct-mapped SFQ Cache that promises higher throughput than the state-of-the-art based on circular FIFOs. Our design enables direct access to memory cells and split and merge elements to reduce overhead.
- Completed JSIM models to compare the JJ complexity and feasible performance to the baseline. Initial analysis shows the potential for significant increase in performance and power with a reasonable increases in JJ complexity.

Modeling and Characterization of Metastability in Single Flux Quantum (SFQ) Logic

- Analyzed the impact of setup time violations and metastability in SFQ circuits, comparing the derived analytical models to their CMOS counterparts, and curve fitting the model to simulations using the existing SFQ5ee process.
- Developed new techniques to reduce the average latency in metastability-tolerant SFQ synchronizers, and evaluated their effects on the layout and critical margin of the design.

Algorithm-hardware co-design for In-sensor Analog Computing for Computer Vision applications

- Proposed a novel in-sensor computing paradigm based on algorithm-hardware co-design that embeds all the computations required in modern CNN layers including convolution, batch norm and ReLU inside image sensors.
- Benchmarked power and performance for the system level acceleration of CNNs with the proposed paradigm.
- Enabled an order of magnitude reduction in bandwidth and energy delay product for tinyML, hyperspectral image recognition, and multi-object tracking tasks with negligible accuracy degradation.

Hardware Acceleration of Bayesian Neural Networks (BNN) using Stochastic Memristors

- Developed learning algorithms for BNNs to be implemented in memristive crossbar array based in-memory computing hardware that exploits the experimentally demonstrated cycle-to-cycle variability of the devices.

Algorithm-hardware Co-design for Energy-efficient Inference with Spiking Neural Networks (SNN)

- Developed a novel DNN-to-SNN conversion and SNN fine-tuning algorithm that reduces the conversion error for ultra-low latencies by accurately capturing the DNN and SNN activation distributions with CNNs and LSTMs.
- Proposed a novel application of our SNN training framework in compute-heavy 3D CNNs for hyperspectral imaging.
- Proposed a Hoyer regularization based novel training strategy to jointly optimize the SNN membrane potential distribution and the relative placement of the threshold with ultra low number of time steps.
- Compared to non-spiking implementations, these SNN works reduce the total inference energy (by up to 100x) and latency (by up to 10x) on both digital and neuromorphic accelerators for a wide range of CV/NLP/speech tasks.
- Proposed a novel architecture that processes SNNs using an output-stationary dataflow model along with a hybrid on-chip memory. Implemented a hardware model using custom RTL specifications on Kintex7 FPGA platform.

Hardware Acceleration of Vision Transformer Inference at the Edge

- Developed a configurable hardware accelerator for Vision Transformer (ViT) based edge inference using novel head-level pipelines and inter-layer multi layer perceptron (MLP) optimizations.
- Our design supports several commonly used vision transformer models with minimal overhead in control logic. It achieves 90% utilization efficiency while consuming only 0.88W of power at 150 MHz to process Swin Transformer.

Enabling ISPless Low Power Computer Vision

- Proposed to invert the image signal processing (ISP) pipeline, that converts RGB images (abundantly available to train CNNs) to its raw counterparts, and enable CNN training on raw images to support in-sensor computing.
- Developed an energy-efficient form of analog in-sensor demosaicing that may be coupled with in-sensor CNN computations to reduce bandwidth/energy by 4x on average compared to traditional ISP-processed vision systems.
- Developed a novel few-shot learning technique to improve the performance of downstream tasks (e.g. 20% mAP increase in objecy detection from PASCALRAW dataset) with raw images captured by real sensors.

Aggressive Pooling for bandwidth reduction in in-sensor computing

- Developed a novel pooling layer that efficiently aggregates dependencies between non-local activation patches during down-sampling, and consists of consists of patch embedding, multi-head self-attention, spatial-channel restoration.
- Proposed method aggressively down-samples the activation maps in the initial layers (providing up to 22x reduction in memory consumption) and surpasses the test accuracy of existing pooling methods with diverse CNN backbones.
- Proposed method enables the deployment of CNNs in memory-constrained devices (e.g. micro-controllers) because the initial activation maps require a large on-chip memory for high-resolution images required for complex CV tasks.

RESEARCH GRANTS

1. Co-authored a successful proposal funded by DARPA with grant number HR00112190120 (PI: Dr. Ajey Jacob, Co-PIs: Dr. Peter A. Beerel and Prof. Akhilesh R. Jaiswal) on In-Pixel Intelligent Processing for Multi-Object Tracking.
2. Co-authored a successful proposal funded by SAMSUNG (Sole PI: Dr. Peter A. Beerel) on In-Sensor Computing based Hardware-Software Co-Design for both low- and high-level vision tasks.
3. Co-authored a successful proposal funded by Intel (Sole PI: Dr. Peter A. Beerel) on Energy-efficient Spiking Neural Networks for complex vision tasks.

SELECTED PUBLICATIONS [Google Scholar]

1. **G. Datta**, S. Kundu, H. Cong, P. A. Beerel, "qCDC: Metastability-Resilient Synchronization FIFO for SFQ Logic," *International Superconductive Electronics Conference (ISEC) 2019.*
2. S. Kundu, **G. Datta**, P.A. Beerel, M. Pedram, "qBSA: Logic Design of a 32-bit Block-Skewed RSFQ Arithmetic Logic Unit," *International Superconductive Electronics Conference (ISEC), 2019.*
3. A. Abdelhadi, D. Chen, H. Cheng, **G. Datta**, Y. Zhang, P. A. Beerel, M. Greenstreet, "Two-Phase Asynchronous to Synchronous Interfaces for an Open-Source Bundled-Data Flow," *International Symposium on Asynchronous Circuits and Systems (ASYNC) 2019.*
4. **G. Datta**, P.A. Beerel, "Modeling and Characterization of Metastability in Single Flux Quantum (SFQ) Synchronizers," *in International Symposium on Circuits and Systems (ISCAS) 2020.*
5. **G. Datta**, P.A. Beerel, "Single Flux Quantum (SFQ) First-in-first-out (FIFO) Synchronizers: New Designs and Paradigms," *in IEEE Transactions on Applied Superconductivity.*
6. **G. Datta**, Y. Lin, B. Zhang, P.A. Beerel, "Metastability in Superconducting Single Flux Quantum(SFQ) Logic," *accepted in IEEE Transactions on Circuits and Systems-I.*
7. S. Kundu, **G. Datta**, M. Pedram, P.A. Beerel, "Spike-Thrift: Towards Energy-Efficient Deep Spiking Neural Networks by Limiting Spiking Activity via Attention-Guided Compression," *accepted in Winter Conference on Applications of Computer Vision (WACV) 2020.*
8. **G. Datta**, S. Kundu, P.A. Beerel, "Training Energy-efficient Deep Spiking Neural Networks with Single-Spike Hybrid Input Encoding" *accepted in International Joint Conference on Neural Networks (IJCNN), 2021.*
9. **G. Datta**, P.A. Beerel, "Can Deep Neural Networks be Converted to Ultra Low-Latency Spiking Neural Networks?" *accepted in Design, Automation, and Test in Europe (DATE) Conference, 2021.*
10. S. Kundu, **G. Datta**, M. Pedram, P.A. Beerel, "Towards Energy-Efficient Deep Spiking Neural Networks via Attention-Guided Compression," *to be submitted to IEEE Transactions on Computers, 2022.*
11. **G. Datta**, S. Kundu, A.K. Jaiswal, P.A. Beerel, "ACE-SNN: Algorithm-Hardware Co-Design of Energy-Efficient and Low-Latency Deep Spiking Neural Networks for 3D Image Recognition" *Frontiers in Neuroscience, 2021.*
12. **G. Datta**, S. Lin, P.A. Beerel, "A High Performance and Robust FIFO Synchronizer-Interface for Crossing Clock Domains in SFQ Logic" *under review in IEEE Transactions on Circuits and Systems-II, 2021.*
13. **G. Datta**, S. Kundu, Z. Yin, R. T. Lakkireddy, J. Mathai, A. Jacob, P.A. Beerel, A. Jaiswal, "P^{2}M: A Processing-in-Pixel-in-Memory Paradigm for Resource Constrained TinyML Applications" *Scientific Reports, 2021.*
14. **G. Datta**, Z. Yin, A. Jacob, A. Jaiswal, P.A. Beerel, "Toward Efficient Hyperspectral Image Processing inside Camera Pixels" *accepted in ECCV Workshop on Distributed Smart Cameras 2022.*
15. **G. Datta**, S. Kundu, Z. Yin, J. Mathai, Z. Liu, Z. Wang, M. Tian, S. Liu, R. T. Lakkireddy, A. Schmidt, W. Abd-Almageed, A. P. Jacob, A. R. Jaiswal, P. A. Beerel, "P^{2}MDeTrack: Processing-in-Pixel-in-Memory for Real Time, Energy-Efficient, and Large-Scale Object Detection and Tracking" *VLSI-SoC 2022 (**Best paper award nomination**).*
16. **G. Datta**, H. Deng, R. Aviles, Z. Liu, and P. A. Beerel, "Bridging the Gap between Spiking Neural Networks and LSTMs for Latency and Energy Efficiency" *accepted in ISLPED 2023.*
17. **G. Datta**, F. Chen, S. Kundu, and P. A. Beerel, "Efficient Deep Learning via Self-Attention-based Aggressive Downsampling" WACV 2023.
18. **G. Datta**, Z. Liu, Z. Yin, L. Sun, A. Jaiswal and P. Beerel, "Enabling ISPless Computer Vision" *WACV 2023.*
19. F. Chen, **G. Datta**, P. A. Beerel, "PoolDyn: Aggressive Pooling with Dynamic Convolutional Neural Networks for On-Device Multi-Object Detection" *in submission; title altered for anonymity.*
20. **G. Datta**, R. Aviles, S. Chen, Y. Lin, M. Li, P. A. Beerel, "Proposal for an Ultrafast Energy-Efficient Superconducting Cache Memory for General Purpose Workloads" *to be submitted to IEEE Transactions on Applied Superconductivity 2022.*
21. **G. Datta**, T. Etchartt, V. Yadav, V. Hedau, P. Natarajan, "Efficient Active Speaker Detection using Self and Multimodal Transformers" *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.*
22. M. Kaiser*, **G. Datta***, Z. Wang, A. P. Jacob, P. A. Beerel, A. R. Jaiswal, "Neuromorphic P^{2}M: Processing-in-Pixel-in-Memory Paradigm for Neuromorphic Image Sensors" *accepted in Frontiers in Neuroinformatics.*
23. **G. Datta**, Z. Liu, M. Kaiser, S. Kundu, J. Mathai, Z. Yin, A. P. Jacob, A. R. Jaiswal, P. A. Beerel, "In-Sensor & Neuromorphic Computing are all you need for efficient computer vision" *in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.*

* denotes equal contribution

24. S. Nag, **G. Datta**, S. Kundu, N. Chandrachoodan, P. A. Beerel, "ViTA: A Vision Transformer Inference Accelerator at the Edge" *accepted in ISCAS 2023.*
25. **G. Datta***, Z. Liu*, P. A. Beerel, "Hoyer Regularizer is all you need for Ultra Low Latency Spiking Neural Networks" *in submission; title changed for anonymity.*
26. Y. Hu, Y. Liu, X. Ye, **G. Datta**, S. Mutnuri, N. Asavisanu, N. Ayanian, K. Psounis, P. A. Beerel, "FireFly: Synthetic Dataset for Ember Detection and Tracking in Wildfire" *in ICCV Workshop on Artificial Intelligence for Humanitarian Assistance and Disaster Response (AIHADR) 2023.*
27. M. Kaiser*, **G. Datta***, S. Sarkar, S. Kundu, Z. Yin, M. Garg, A. P. Jacob, P. A. Beerel, A. R. Jaiswal, "Technology-Circuit-Algorithm Tri-Design for Processing-in-Pixel-in-Memory" *accepted in GLSVLSI 2023.*
28. **G. Datta**, B. Huang, N. Sekhar, V. Yadav, S. Lin, A. Jaiswal, Y. Wu, and P. Singhal, "Let's Listen with Style: Modeling Non-Deterministic Dyadic Facial Motion Generation" *IJCAI Workshop on Spatio-Temporal Reasoning and Learning 2023, and under review in ICASSP 2023.*
29. **G. Datta***, Z. Liu*, P. A. Beerel, "Training Ultra-Low-Latency Spiking Neural Networks from Scratch" *under review in ICASSP 2024.*
30. S. Sarkar, X. Ye, **G. Datta**, P. A. Beerel, "FixPix: Fixing Bad Pixels with Deep Learning" *under review in ICASSP'24.*
31. Y. Hu, **G. Datta**, K. Beerel, P. A. Beerel, "Let's Roll: Synthetic Dataset Analysis for Pedestrian Detection Across Different Shutter Types" *under review in ICASSP 2024.*
32. F. Chen*, **G. Datta***, P. A. Beerel, "Resiudal Encoded Distillation for Efficient Computer Vision" *in submission; title changed for anonymity*

UNDERGRADUATE PROJECTS

- **Embedded System Design Mini Project**: Feb 2017 - Apr 2017
  - Acquied, amplified, filtered and displayed ECG signals using STM32F4 MCU and Altera Cyclone IV FPGA.
  - The FIR filter and UART transmitter/receiver is implemented on the FPGA. The FPGA applies FIR filtering to the signals and sends them back to the MCU.
- **VLSI Design Mini Project**: May 2016 - Jul 2016
  - Performed RTL design for Hamming distance based pattern matching (with and without parallelism) using IC compiler.
  - Designing an efficient general purpose Multi-Cycle CPU that supports arithmetic, logical and memory read/write operations. A 1024-bit SRAM is utiized as Data memory. Used python to read assembly code and generate vector file for control signals.
- **Design for Testability Mini Project**: Implemented an ATPG algorithm(PODEM) and a Deductive Fault Simulator for some ISCAS'85 bencmark circuits. Achieved high fault coverage(>99%) for all circuits. Sept 2018 - Present
- **Design of a low power analog correlator for neural spike sorting (B.Tech Project)**: Aug 2017 - May 2018
  - Designed an analog correlator for neural spike sorting, involving the use of switched capacitor circuits along with digital control. 6-T SRAM cells were designed to store the spike templates.
  - A low noise front end amplifier was integrated with the correlator to reduce the effect of background noise in spike sorting.

ACADEMIC ACHIEVEMENTS

- Adjudged one of the winners of the prestigious 2022-2023 **Ming Hsieh PhD scholar program** at USC.
- Awarded the **best Graduate Research Assistant Award** from the ECE department at USC.
- Recipient of the prestigious **Annenberg Fellowship** that is awarded to top incoming graduate students at USC.
- Adjudged the **academically best outgoing student** in my batch of Instrumentation Engineering at IIT Kharagpur. Also, received **2 additional academic excellence awards** during the academic session 2017-2018.
- Selected as one of the **finalists in the Qualcomm Innovation Fellowship** (2022) North America competition.
- Recipient of the prestigious 2022 IEEE CSC **Graduate Study Fellowship in Applied Superconductivity** (only 6 students are selected worldwide every year).
- Recipient of a **best paper nomination** at the VLSI-SoC conference 2022.
- Awarded the **ACM travel grant** to participate in Student Research Competitions at ICCAD'22 & DAC'23.
- Awarded the **most novel research project award** in the Deep Learning course (EE-599) at USC.
- Awarded the **Annenberg Top Off Travel/Research Award** from USC in 2022.

TECHNICAL SKILLS

- **Programming Languages**: C, C++, Python. **Simulation Environments**: Cadence Virtuoso, Synopsis (PnR Tool), Catapult HLS, MATLAB, Modelsim, Xilinx ISE. **HDLs**: Verilog, System-verilog. **ML Tools**: Pytorch, Tensorflow

TEACHING & MENTORING EXPERIENCE

- **Teaching Assistant** at IIT Kharagpur for two undergraduate courses, Signals and Networks and Analog Electronics during the Academic Session 2016 - 2017.
- **Student mentor** of Student Welfare Group of IIT Kharagpur, having mentored 5 undergraduate students in their academic courses related to Electrical Engineering.
- **Graduate Mentor** in the USC Viterbi Graduate Student Association, having served as guest speakers in seminars related to cracking research internships and jobs.
- **Guest TA** in the Asynchronous VLSI Design course at USC, where I helped the students get acquainted with the basics of spiking neural networks for the course project.
- **Mentored** 3 PhD, 15 MS and 2 undergraduate students in directed research on superonductive electronics and deep learning, and co-authored top-tier publications with 8 of them.

PROFESSIONAL SERVICE

- **Journal Reviewing:** IEEE Transactions on Circuits and Systems-I & II, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, IEEE Internet of Things, IEEE Systems, IEEE Transactions on Applied Superconductivity, Science Advances, Frontiers in Neuroscience, Frontiers in Neurorobotics, Patterns - Cell Press, Neurocomputing, Neural Networks, MDPI
- **Conference Reviewing:** IEEE International Symposium on Circuits and Systems (ISCAS), IEEE Midwest Symposium on Circuits and Systems (MWSCAS), Winter Conference on Applications of Computer Vision (WACV), British Machine Vision Conference (BMVC), International Conference on Learning Representations (ICLR), Design Automation Conference (DAC), IEEE International Conference on Accoustics, Speech and Signal Processing (ICASSP)
- **Technical Program Committee:** International Conference on VLSI Design (VLSID) 2024, International Symposium on Quality Electronic Design (ISQED) 2024

References

- Dr. Peter A. Beerel
  Research Director, Information Sciences Institute
  Professor & Associate Chair, Computer Engineering Division
  University of Southern California
  +1-213-740-4481, **pabeerel@usc.edu**
- Dr. J. Joshua Yang (IEEE Fellow)
  Professor of Electrical & Computer Engineering
  University of Southern California
  +1-213-740-4709, **jjoshuay@usc.edu**
- Dr. Massoud Pedram (IEEE Fellow)
  Professor of Electrical & Computer Engineering
  University of Southern California
  +1-213-740-4218, **pedram@usc.edu**
- Dr. Ajey P. Jacob
  Director of Advanced Electronics
  Information Sciences Institute, University of Southern California
  +1-703-248-6171, **ajey@isi.edu**