Subhankar Pal | Curriculum Vitae

2260 Hayward Street, Ann Arbor, MI 48109

subh@umich.edu • subhankarpal.com

Biography

I am a fifth-year PhD student in Computer Science and Engineering at the University of Michigan, Ann Arbor. I am seeking full-time positions in industry research starting in Fall 2021. My primary experience and research interests lie in computer architecture, hardware-algorithm co-design, low-level system software, and hardware-software interfacing.

My thesis aims to bridge the gap between general-purpose processors and fixed-function accelerators for irregular and mixed dense/sparse workloads. Various emerging, large-scale applications in cloud computing, datacenters and HPC have experienced a rising demand for high energy-efficiency, in addition to high performance. Fixed-function hardware is the *de facto* choice for energy-efficient execution for a given application. However, unlike general-purpose processors, fixed-function accelerators are constrained by limited programmability and high non-recurring costs. The problem is exacerbated by the gap in the pace of algorithmic development, and the turnaround time to design, fabricate and test new accelerator hardware. My thesis proposes a vertically-integrated solution that uses hardware reconfiguration to adapt to the nature of the application and deliver near-accelerator-level efficiencies, while remaining as programmable as a GPU through a custom software stack that abstracts the details of the underlying hardware. Through studies conducted at various stages — architectural simulation, FPGA emulation, and chip prototyping — my research has demonstrated the efficacy of the proposed system, particularly for applications involving sparsity and those composed of multiple phases, where the diversity in each phase is captured by a specific configuration of the hardware.

My primary-author work has been recognized at reputed venues in computer architecture, system modeling, and hardware design, such as HPCA, PACT, IISWC and VLSI. Beyond my primary area, I have also delved into the domains of real-time scheduling, exploiting sparsity in deep neural network (DNN) accelerators, compiler optimizations for DNN accelerators, system-level application characterization, reliability for 3D processors, and analysis of emerging transistor technologies.

Education

- **University of Michigan** — **Ann Arbor, MI, USA**
  Doctor of Philosophy in Computer Science and Engineering — *2016–2021 (expected)*
  CGPA: 4.0/4.0
  Thesis topic: *"Towards Closing the Programmability-Efficiency Gap using Software-Defined Hardware"*
  Advisor: *Prof. Ronald G. Dreslinski*
  Relevant coursework: Applied Parallel Programming with GPUs (EECS 598)

- **University of Michigan** — **Ann Arbor, MI, USA**
  Master of Science in Computer Science and Engineering — *2016–2018*
  CGPA: 4.0/4.0
  Relevant coursework: Computer Architecture (EECS 470), Microarchitecture (EECS 573), Data-Centric Systems (EECS 598), Parallel Computer Architecture (EECS 570), Special Topics in Computer Architecture (EECS 670), Operating Systems (EECS 482), Advanced Compilers (EECS 583), Machine Learning (EECS 545)

- **Birla Institute of Technology and Science – Pilani** — **Hyderabad, TS, India**
  Bachelor of Engineering in Electrical and Electronics Engineering, graduated with Distinction — *2010–2014*
  CGPA: 9.6/10.0
  Relevant coursework: Digital Electronics and Computer Organization, Microprocessor Programming and Interfacing, Machine Learning, Image Processing, Analog and Digital VLSI Design

Research Experience

- **University of Michigan** — **Ann Arbor, MI, USA**
  Graduate Student Research Assistant, CADRe Group — *September 2016–Present*

– Leading the architecture and hardware-software interfacing efforts for the DARPA Software-Defined Hardware (SDH) project comprising the University of Michigan, University of Edinburgh, and Arizona State University, with the goal to build a reconfigurable, programmable, manycore accelerator for HPC/datacenter applications called Transmuter
– Presently developing a software runtime consisting of a machine-learning based model for dynamic reconfiguration on Transmuter, that enables the hardware to adapt to the application and data based on a set of performance counters
– Developed multiple simulation and emulation infrastructures for the SDH system, including a functional simulator, a performance simulator and a trace-driven simulator using gem5
– Designed a low-level software application programming interface (API) in C++ for programming and implementation of workloads on the Transmuter architecture, in addition to the host-side API for orchestrating Transmuter
– Mapped multiple fundamental kernels in linear algebra, signal processing and machine learning on Transmuter

– Developed an open-source tool called HetSim, a trace-driven, synchronization and dependency-aware architectural simulator that accelerates the simulation speed of workloads on large-scale systems with 1000s of cores
– Assisted students from Arizona State University and the University of Arizona on the use of HetSim for design-space exploration of their solution for DARPA's Domain-Specific System-on-Chip (DSSoC) program

– Led the architecture and chip front-end development of the OuterSPACE sparse matrix multiplication accelerator, a single-program, multiple-data driven accelerator that uses heterogeneous cores and a reconfigurable memory hierarchy

– Developed energy optimization techniques for a heterogeneous task scheduler used in autonomous vehicles on top of STOMP, an open-source discrete event simulator for evaluating scheduling techniques, in collaboration with IBM Research

– Implemented and characterized the processing element hardware design for the sparse tensor processing unit (STPU) architecture that adapts the well-known TPU architecture for sparse matrix operations

– Characterized traditional and emerging desktop applications, such as cryptocurrency mining and virtual reality games, on a high-end system to evaluate how software has evolved to harness the parallelism offered by contemporary hardware

- **University of Bremen** — **Bremen, HB, Germany**
  Research Assistant, Institut für Theoretische Elektrotechnik und Mikroelektronik (ITEM) — *May 2013–July 2013*
  – Performed coding and characterization of low-power encoders and decoders for very deep sub-micron buses
  – Implemented and characterized various prior bus encoding and decoding techniques in hardware, revealing a trade-off between these parameters to aid chip designers

- **Bhabha Atomic Research Centre** — **Mumbai, MH, India**
  Research Intern, Division of Remote Handling and Robotics — *May 2012–July 2012*
  – Developed a drive mechanism for a piezoelectric transducer based dispenser used for the manufacture of bio-chips
  – Implemented microcontroller code to generate pulses meeting the specs of the dispenser, a GUI to generate and control the pulse over UART, and an analog amplifier circuit to feed the pulse to the transducer

Industry Experience

- **IBM Corporation** — **Yorktown Heights, NY, USA**
  Research Intern, Software Architecture — *May 2019–August 2019*
  – Developed a compiler extension for smart scratchpad management in deep learning accelerators that support both static graph-based and eager-execution based neural network frameworks
  – Explored different algorithms for selective quantization of weights and activations in mixed-precision neural networks
  – Enhanced the compiler runtime for IBM's RAPID DNN accelerator by adding feedback loops between each tool in the DeepTools framework for runtime error detection and handling

- **Advanced Micro Devices, Inc.** — **Boxborough, MA, USA**
  Co-Op Engineer, Architecture Research — *May 2017–September 2017*
  – Innovated techniques for improving the performance and energy efficiency of the instruction fetch pipeline front-end for next-generation AMD CPUs as part of the exascale computing effort
  – The tasks involved understanding the CPU simulation infrastructure and embedding fine-grained profiling hooks for deeper insights into the branch target buffer (BTB) access patterns, resulting in a two-point solution for performance/power enhancement; the first targeted at reducing BTB misses and enhanced the overall micro-instructions per cycle, and the next solution reduced the power consumed by the BTB structure by reducing number of lookups

- **NVIDIA Corporation** — **Bangalore, KA, India**
  ASIC Design Verification Engineer — *July 2014–July 2016*

– Worked on pre-silicon verification of GPUs that involved running full-chip tests and debugging issues in the domain of PCI-Express, clocks and resets, and enhancement and maintenance of full-chip testbench covering these features
– Actively performed initial bring-up of Maxwell and Pascal GPUs, focusing on testing and validation of new features
– Collaborated with the post-silicon validation team to redesign a utility widely used internally for debug of issues during the post-silicon validation and bring-up phases
– Responsible for automating the GPU power measurement/analysis flow by implementing a GUI and socket-driven backend

- **NVIDIA Corporation** — **Bangalore, KA, India**
  ASIC Engineering Intern — *January 2014–June 2014*
  – Developed two GUI-based tools from scratch to assist post-silicon debug of GPUs
  – Debugged PCIe issues encountered with the chips using a PCI Express logic analyzer, and enhanced existing post-silicon debug tools that assisted in quicker and a more convenient debug experience

Teaching Experience

- **University of Michigan** — **Ann Arbor, MI, USA**
  Graduate Student Instructor — *January 2018–April 2018*
  I was the Graduate Student Instructor (GSI) with Prof. Satish Narayanasamy for EECS 570, which is an advanced graduate-level course in Computer Architecture. I was responsible for teaching a discussion section, answering student questions online and in-person, tutoring students on parallel programming paradigms, designing exam questions and grading exams/assignments/projects.

Conference Publications

- S. Kim, M. Fayazi, A. Daftardar, K.-Y. Chen, J. Tan, **S. Pal**, T. Ajayi, Y. Xiong, T. Mudge, C. Chakrabarti, D. Blaauw, R. Dreslinski and H.-S. Kim, *"Versa: A Dataflow-Centric Multiprocessor with 36 Systolic ARM Cortex-M4F Cores and a Reconfigurable Crossbar-Memory Hierarchy in 28nm"*, *under review*.
- **S. Pal**, S. Venkataramani, V. Srinivasan and K. Gopalakrishnan, *"OnSRAM: Efficient On-Chip Scratchpad Management in Deep Learning Accelerators"*, 2021 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), *to appear*.
- A. Amarnath, **S. Pal**, H. Kassa, A. Vega, A. Buyuktosunoglu, H. Franke, J.-D. Wellman, R. Dreslinski and P. Bose, *"AVSched: Mission-Aware Scheduling in Heterogeneous SoCs for Autonomous Vehicles"*, *under review*.
- S. Feng, J. Sun, **S. Pal**, K. Kaszyk, X. He, D.-h. Park, J. Morton, D. Blaauw, H.-S. Kim, T. Mudge, M. Cole, M. O'Boyle, C. Chakrabarti and R. Dreslinski, *"CoSPARSE: A Software and Hardware Reconfigurable SpMV Framework for Graph Analytics"*, 2021 IEEE/ACM Design Automation Conference (DAC), *to appear*.
- **S. Pal**, K. Kaszyk, S. Feng, B. Franke, M. Cole, M. O'Boyle, T. Mudge and R. Dreslinski, *"HetSim: Simulating Large-Scale Heterogeneous Systems using a Trace-Driven, Synchronization and Dependency-Aware Framework"*, 2020 IEEE International Symposium on Workload Characterization (IISWC), Virtual, 2020, pp. 13-24. [Link]
- **S. Pal**, S. Feng, D.-h. Park, S. Kim, A. Amarnath, C.-S. Yang, X. He, J. Beaumont, K. May, Y. Xiong, K. Kaszyk, J. Morton, J. Sun, M. O'Boyle, M. Cole, C. Chakrabarti, D. Blaauw, H.-S. Kim, T. Mudge and R. Dreslinski, *"Transmuter: Bridging the Efficiency Gap using Memory and Dataflow Reconfiguration"*, 2020 ACM International Conference on Parallel Architectures and Compilation Techniques (PACT), Virtual, 2020, pp. 175-190. [Link]
- Y. Xiong, J. Zhou, **S. Pal**, D. Blaauw, H.-S. Kim, T. Mudge, R. Dreslinski and C. Chakrabarti, *"Accelerating Deep Neural Network Computation on a Low Power Reconfigurable Architecture"*, 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Sevilla, 2020, pp. 1-5. [Link]
- J. Bagherzadeh, A. Amarnath, J. Tan, **S. Pal** and R. Dreslinski, *"R2D3: A Reliability Engine for 3D Parallel Systems"*, 2020 IEEE/ACM Design Automation Conference (DAC), San Francisco, 2020, pp. 1-6. [Link]
- X. He, **S. Pal**, A. Amarnath, S. Feng, D.-H. Park, A. Rovinski, H. Ye, Y. Chen, R. Dreslinski and T. Mudge, *"Sparse-TPU: Adapting Systolic Arrays for Sparse Matrices"*, 2020 International Conference on Supercomputing (ICS), Barcelona, 2020, pp. 1-12. [Link]
- A. Soorishetty, J. Zhou, **S. Pal**, D. Blaauw, H. Kim, T. Mudge, R. Dreslinski and C. Chakrabarti, *"Accelerating Linear Algebra Kernels on a Massively Parallel Reconfigurable Architecture"*, 2020 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Barcelona, 2020, pp. 1558-1562. [Link]
- **S. Pal**, D.-h. Park, S. Feng, P. Gao, J. Tan, A. Rovinski, S. Xie, C. Zhao, A. Amarnath, T. Wesley, J. Beaumont, K.-Y. Chen, C. Chakrabarti, M. Taylor, T. Mudge, D. Blaauw, H.-S. Kim and R. Dreslinski, "A 7.3 M Output Non-Zeros/J Sparse Matrix-Matrix Multiplication Accelerator Using Memory Reconfiguration in 40 nm", 2019 Symposia on VLSI Technology and Circuits (VLSI), Kyoto, 2019, pp. C150-C151. [Link]

- S. Feng, **S. Pal**, Y. Yang and R. Dreslinski, "Parallelism Analysis of Prominent Desktop Applications: An 18-year Perspective", 2019 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), Madison, 2019, pp. 202-211. [Link]
- **S. Pal**, J. Beaumont, D.-h. Park, A. Amarnath, S. Feng, C. Chakrabarti, H. S. Kim, D. Blaauw, T. Mudge, R. Dreslinski, "OuterSPACE: An Outer Product based Sparse Matrix Multiplication Accelerator", 2018 IEEE International Symposium on High-Performance Computer Architecture (HPCA), Vienna, 2018, pp. 724-736. [Link]
- A. Amarnath, S. Feng, **S. Pal**, T. Ajayi, A. Rovinski and R. Dreslinski, "A Carbon Nanotube Transistor based RISC-V Processor using Pass Transistor Logic", 2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED), Taipei, 2017, pp. 1-6. [Link]
- **S. Pal**, C. K. Vudadha, P. S. Phaneendra, S. Veeramachaneni and S. Mandalika, "A New Design of an N-Bit Reversible Arithmetic Logic Unit," 2014 IEEE International Symposium on Electronic System Design (ISED), Surathkal, 2014, pp. 224-225. [Link]

Journal Publications

- J. Bagherzadeh, A. Amarnath, J. Tan, **S. Pal** and R. Dreslinski, "A Holistic Solution for Reliability of 3D Parallel Systems", *under review*.
- D.-h. Park, **S. Pal**, S. Feng, P. Gao, J. Tan, A. Rovinski, S. Xie, C. Zhao, A. Amarnath, T. Wesley, J. Beaumont, K.-Y. Chen, C. Chakrabarti, M. Taylor, T. Mudge, D. Blaauw, H.-S. Kim and R. Dreslinski, "A 7.3 M Output Non-Zeros/J, 11.7 M Output Non-Zeros/GB Reconfigurable Sparse Matrix-Matrix Multiplication Accelerator", IEEE Journal of Solid-State Circuits (JSSC), 2020, 55(4), pp. 933-944. [Link]

Archived Articles

- A. Vega, A. Amarnath, J.D. Wellman, H. Kassa, **S. Pal**, H. Franke, A. Buyuktosunoglu, R. Dreslinski and P. Bose, "STOMP: A Tool for Evaluation of Scheduling Policies in Heterogeneous Multi-Processors", 2020, arXiv preprint arXiv:2007.14371. [Link]

Workshop Presentations

- **S. Pal**, K. Kaszyk, M. Cole, M. O'Boyle, R. Dreslinski, "HetSim: Simulating Large-Scale Heterogeneous Systems using a Trace-driven, Synchronization and Dependency-Aware Framework", 2020 Workshop on Modeling and Simulation of Systems and Applications (ModSim), 2020.
- **S. Pal**, "Towards Closing the Programmability-Efficiency Gap using Software-Defined Hardware", 2020 IEEE/ACM Design Automation Conference (DAC) PhD Forum, 2020.

Patents

- J. Kalamatianos, A. Yalavarti, V. Agrawal, **S. Pal**, V. Srinivasan, "Filtered Branch Prediction Structures of a Processor", U.S. Patent Application 16/109,195. [Link]

Open-Source Contributions

- **HetSim: Simulating Large-Scale Heterogeneous Systems using a Trace-driven, Synchronization and Dependency-Aware Framework**
  A trace-driven, synchronization and dependency-aware framework for fast and accurate pre-silicon performance and power estimations for heterogeneous systems with up to thousands of cores. (IISWC 2020)
  Role: Primary author and developer *Repository*: https://github.com/umich-cadre/HetSim-gem5
- **STOMP: A Tool for Evaluation of Scheduling Policies in Heterogeneous Multi-Processors**
  A simulator for fast implementation and evaluation of task scheduling policies in multi-core/multi-processor systems with a convenient interface for "plugging" in new scheduling policies in a simple manner. (arXiv 2020)
  Role: Contributor *Repository*: https://github.com/IBM/stomp

Talks

- Presentation on the Transmuter work at NVIDIA Research, held virtually (Jan 2021)
- Guest lecture on accelerators for sparse computation for EECS 598: Parallel Programming with GPUs (Nov 2020)
- Multiple tutorials on writing efficient parallel code for the Transmuter architecture at the University of Michigan and

Arizona State University (2018–2020)

- Presentation on the HetSim work at IISWC, held virtually (Oct 2020)
- Presentation on the Transmuter work at PACT, held virtually (Oct 2020)
- Presentation on the ongoing work on HetSim at the ModSim workshop, held virtually (Aug 2020)
- Presentation on thesis topic at the DAC PhD forum, held virtually (Jul 2020)
- Presentation on the sparse matrix accelerator chip at VLSI in Kyoto, Japan (Jun 2019)
- Talk on the SDH project at the University of Edinburgh, UK (Feb 2019)
- Guest lecture on accelerators for sparse computation for EECS 598: Accelerated Systems for AI and Health (Dec 2019)
- Guest lecture on the OuterSPACE work for EECS 570: Parallel Computer Architecture (Dec 2019)
- Presentation on the OuterSPACE work at HPCA in Vienna, Austria (Feb 2018)
- Invited talk to share experiences as a DAAD-WISE scholar held at the University of Hyderabad, India (Apr 2014)

Notable Projects

- *'Performance-Optimized Implementation of CNN inference on an NVIDIA GPU'*

  – Implemented tiled 2D convolution to accelerate two layers of a convolutional neural network using CUDA,
  – Incorporated features such as reduced thread divergence through smarter work distribution, and design space exploration with different block and grid dimensions; optimized the code using the nvprof visual profiler

- *'Using ACCEPT to Supplement Stale Value Approximation on CMPs'*

  – Extended prior work on approximating loads using stale values to perform approximation on selective loads; this used the ACCEPT compiler, which annotates loads that are deemed safe by analyzing side-effects of code blocks as APPROX
  – Developed LLVM infrastructure to propagate the APPROX-annotated loads to ARM machine code, and ran the generated binary through a modified gem5 simulator to perform stale-value approximation on the annotated loads

- *'Development of Components of a Basic Operating System'*

  – Worked in a small group to develop a multi-threaded disk scheduler, a thread management library, a virtual memory manager and a networked file system

- *'Compiler Support for a High-Bandwidth, Fixed-Throughput Regular Expression Accelerator'*

  – Added support for processing nested Kleene operators (*/+/x) in HARE, a hardware accelerator for matching regular expressions against large in-memory logs, while keeping the changes in throughput and area minimal

- *'Formal Verification of the MSI Cache Coherence Protocol'*

  – Implemented and formally verified the MSI cache coherence protocol using the Murphi tool, with additional optimizations, namely spontaneous self-downgrading and cruise-missile invalidation

- *'Design and Implementation of a 2-Way Out-of-Order Superscalar Processor'*

  – As part of a group, implemented a MIPS R10K-style out-of-order processor based on the Alpha ISA using SystemVerilog; the highlights were: 2-way superscalar pipeline, store-to-load forwarding, multiple outstanding load misses, set-associative, multi-ported, write-back data cache, branch history table with bimodal predictors and next N-line I-Cache prefetching

- *'Bacterial Colony Detection for Testing Antibiotic Sensitivity'*

  – Implemented a MATLAB codebase for the detection of different types of bacteria in agar culture from their images

Technical and Personal Skills

- **Programming Languages/Tools:** C, C++, Verilog, Python, Tk, Perl, Shell, gem5, PAPI, LLVM, MATLAB
- **Operating Systems:** Linux, Windows, MacOS
- **CAD/EDA Tools:** Verdi, Design Compiler, Virtuoso, MultiSim, LabVIEW, PSpice, Keil μVision
- **Version Control Tools:** Git, Perforce, Subversion

Awards and Honors

- **Rackham Travel Grant** (2019): Financial award for presenting at VLSI Symposium in Kyoto, Japan
- **U-M Tuition Waiver** (2016–present): Full tuition waiver with benefits working as Graduate Student Research Assistant with Prof. Ronald Dreslinski
- **Top Contributor** (2015): Honored by NVIDIA for excellent performance as an employee for 2014-2015
- **Merit-Cum-Need Scholarship** (2014): Issued by BITS-Pilani for merit and financial need
- **DAAD-WISE Scholarship** (2013): Selected among 163 students from India by the German Academic Exchange Service

(DAAD) for a funded internship in Germany during the summer of 2013

- **"Best Student" Scholarship** (2012): Awarded by the Angiras Foundation, TX for merit and financial need
- **Certificate of Merit** (2010): Issued by CBSE for outstanding performance in Computer Science in All India Senior School Certificate Examination

Mentorship and Service

- External review committee member, International Symposium on High Performance Computer Architecture, 2021
- Member of the steering committee for the Computer Architecture Student Association (CASA), which is a community run by and for graduate and undergraduate students interested in computer architecture
- CSEG Wellness Buddy Program that pairs seasoned graduate students with incoming graduate students through the semester (*Fall 2020–Present*)
  - Wentao Zhang, zwtao@umich.edu; Master's
- Michigan's *Lunch and Lab* Mentorship Program that involves lunch meetings and future follow-ups between a graduate student and incoming graduate students or junior/senior undergraduate students (*Fall 2018–Present*)
  - Matthew Shannon, mattshan@umich.edu; Bachelor's
  - Sahas Dendukuri, sahasd@umich.edu; Bachelor's
  - Yang Du, duyung@umich.edu; PhD
  - Jonathan Brasch, jbrasch@umich.edu; Bachelor's
  - Tianyuan Wu, wutian@umich.edu; Master's
  - Nishant Patel, mnishant@umich.edu; Bachelor's
  - Enhao Zhang, ehzhang@umich.edu; Master's
- Worked closely with and mentored a senior undergraduate student, Kyle May (kylemay@umich.edu), mainly for the SDH program, over a period of three academic terms
- Mentored an intern, Anand Thati (thati.anand9@gmail.com), in the ASIC Verification team at NVIDIA for four months