Vignesh Suresh

Email: vv15@illinois.edu
Website: https://publish.illinois.edu/vigneshsuresh

Education

Ph.D., Electrical and Computer Engineering Jan 2021 - Present
University of Illinois, Urbana-Champaign
Advisor: Prof. Sarita Adve

B.Tech., Electronics and Communication Engineering Jul 2013 - Jun 2017
Vellore Institute of Technology, India

Publications

- **Mozart: Taming Taxes and Composing Accelerators with Shared-Memory.**
  *International Conference on Parallel Architectures and Compilation Techniques (**PACT 2024**).*
  Vignesh Suresh, Bakshree Mishra, Ying Jing, Zeran Zhu, Naiyin Jin, Charles Block, Paolo Mantovani, Davide Giri, Joseph Zuckerman, Luca Carloni and Sarita Adve.
- **A 12nm Linux-SMP-Capable RISC-V SoC with 14 Accelerator Types, Distributed Hardware Power Management and Flexible NoC-Based Data Orchestration.**
  *IEEE International Solid-State Circuits Conference (**ISSCC 2024**)*
  Maico Cassel dos Santos, Tianyu Jia, Joseph Zuckerman, Martin Cochet, Davide Giri, Erik Jens Loscalzo, Karthik Swaminathan, Thierry Tambe, Jeff Jun Zhang, Alper Buyuktosunoglu, Kuan-Lin Chiu, Giuseppe di Guglielmo, Paolo Mantovani, Luca Piccolboni, Gabriele Tombesi, David Trilla, John-David Wellman, En-Yu Yang, Aporva Amarnath, Ying Jing, Bakshree Mishra, Joshua Park, **Vignesh Suresh**, Sarita Adve, Pradip Bose, David Brooks, Luca P. Carloni, Kenneth L. Shepard, Gu-Yeon Wei.

Experience

- **Graduate Intern**, Qualcomm, San Diego Jun 2024 - Aug 2024
  Mentor: Bob Rychlik
  Summary: Formalizing, evaluating and extending the coherence implementation in Snapdragon SoCs.
- **Research Assistant**, University of Illinois, Urbana-Champaign Jan 2021 - Present
  Mentor: Prof. Sarita Adve
  Summary: Design of scalable and performant control and data interfaces for heterogeneous systems.
- **Silicon Validation Engineer**, Intel Corporation, India Jan 2017 - Jan 2021
  Mentors: Suresh Vasu, Joydeep Maitra
  Summary: Pre-silicon validation, post-silicon validation, and power and performance (PnP) validation for edge SoCs.

Talks

- **Mozart: Taming Taxes and Composing Accelerators with Shared-Memory.**
  International Conference on Parallel Architectures and Compilation Techniques (PACT), Oct 2024
- **Towards a Flexible and Low-overhead Memory Hierarchy for Heterogeneous Systems**
  Applications Driving Architectures (ADA), University of Michigan, Aug 2022
- **Cache Coherence Flexibility in Hardware with Spandex and ESP**
  Prof. Luca Carloni's Group, Columbia University, Apr 2021

Skills

- **Programming languages**: C, C++, SystemVerilog, Python, SystemC, CUDA, VHDL
- **Tools**: EDA tools (RTL simulation, High-Level Synthesis, FPGA synthesis)
- **Frameworks**: PyTorch

Experience Details

- **Graduate Intern**, Qualcomm, San Diego
  - Formalized the memory model and coherence protocol implemented in Snapdragon SoCs. The formalization was also useful in showing analytically that the coherence implementation can functionally support new uses like multi-GPU.
  - Proposed a lightweight extension to the coherence implementation to support the NPU – IDF submitted.
  - Conducted a competitive analysis of coherence and unified memory support in recent industry products. Proposed enhancements to Qualcomm products based on this analysis.
- **Research Assistant**, University of Illinois, Urbana-Champaign
  - Composable acceleration in shared-memory SoCs:
    * Demonstrated the performance impact of *acceleration taxes* on fine-grained acceleration.
    * Implemented *Mozart* – an SoC architecture for efficiently composing disaggregated accelerators through shared-memory. Mozart includes an *Accelerator Synchronization Interface* to minimize control taxes through shared-memory synchronization, and Spandex coherence to minimize data taxes through direct producer-consumer forwarding.
    * Evaluated Mozart with RTL prototyped on FPGA, using real-world applications like 3D spatial audio and mini-ERA.
  - EPOCHS:
    * Collaboration between IBM, Columbia, Harvard and UIUC to enable scalable development of multi-accelerator SoCs.
    * Contributed to (1) the RTL implementation of the Spandex coherence protocol using the ESP framework, capable of booting multi-core Linux on RISC-V CPUs, and (2) the development of 3D spatial audio accelerators.
- **Silicon Validation Engineer**, Intel Corporation, India
  - Power and Performance (PnP) Validation:
    * Worked alongside SoC performance architects to evaluate the PnP of the vision processing unit (VPU) for an edge vision SoC using use cases such as ResNet50 and MobileNet.
    * Identified and worked with architects to resolve a showstopper bug involving large voltage droops during compute intensive workloads, such as early layers of ResNet50.
  - Post-Silicon Validation:
    * Functional validation of the VPU and media (image and video) codecs using directed and end-to-end tests for an edge vision SoC.
  - Pre-Silicon Validation:
    * Functional validation (including end-to-end 5G NR use cases) for a 5G modem SoC.
    * Functional validation of an ARM subsystem, including boot, coherence and low power modes.

Relevant Coursework

- Computer System Organization (CS433)
- Parallel Computer Architecture (CS533)
  - **Project**: Optimizing a GeMM Accelerator with HLS optimizations for parallelism and pipelining.
- System-on-chip Design (ECE527)
  - **Project**: Designing a hardware accelerator for PageRank using the Xilinx Pynq platform.
- Architectures for Mobile and Edge Computing (CS598SG)
  - **Project**: Algorithmic and hardware optimizations for scene reconstruction in InfiniTAM.
- Advanced Topics in Computer Architecture: Immersive Computing Systems (CS534)
  - **Project**: Hardware acceleration for NeRF with an MLP accelerator.
- Compiler Construction (CS426)
- Real-World Algorithms for IoT and Data Science (ECE434)
- Deep Learning for Computer Vision (CS444)
  - **Project**: Designing a simpler approach for multi-view consistent instance segmentation with NeRF.
- AI Efficiency: Systems & Algorithms (CS598AIE)
  - **Project**: System Implications of Real-time Multi-task Foundation Models in Extended Reality