SURIM OH

soh31@ucsc.edu

RESEARCH INTEREST

My research interests lie broadly in computer architecture and hardware-software interface. I have recently been working on the topic of fetch-directed instruction prefetching in modern processors to improve the CPU performance for datacenter workloads with large instruction footprints.

EDUCATION

University of California, Santa Cruz — *Santa Cruz, CA, USA*
PhD Student in Computer Science and Engineering — *Sep 2020 - Present*

· Advisor: Professor Heiner Litz, Total GPA: 4.0/4.0

Seoul National University — *Seoul, South Korea*
Master of Science in Computer Science and Engineering — *Feb 2015 - Feb 2017*

· Advisor: Professor Bernhard Egger, Total GPA: 3.52/4.0

· Thesis: *Hierarchical Manycore Resource Management Framework using Control Processors* [pdf]

Sogang University — *Seoul, South Korea*
Bachelor of Science in Computer Science and Engineering — *Feb 2011 - Feb 2015*

· Total GPA: 3.70/4.0 (95.8/100), Summa Cum Laude

· Exchange student at **Northern Arizona University** — *AZ, USA, Spring 2014*

WORK EXPERIENCE

Meta — *Sunnyvale, California, USA*
ASIC Engineer Intern, Architecture — *June 2024 - Sep 2024*

Akeana — *San Jose, California, USA*
PhD Intern — *June 2023 - Sep 2023*

SAP Labs Korea — *Seoul, South Korea*
Developer — *Jan 2018 - Sep 2020*

Hyundai Motor Company R&D Division — *Hwaseong, South Korea*
Engineer — *Feb 2017 - Jan 2018*

SKILLS

C, C++, Python, Bash, Scarab simulator, Intel Perf, DynamoRIO, Verilog, Magic, OpenCL, Intel PCM, LLVM, HSIM-TQSIM (System C-based) manycore simulator

PUBLICATIONS

ISCA 2024: **Surim Oh**, Mingsheng Xu, Tanvir Ahmed Khan, Baris Kasikci, Heiner Litz. UDP: Utility-Driven Fetch Directed Instruction Prefetching. *In the 51th International Symposium on Computer Architecture (ISCA)*, 2024

M.S. Thesis: **Surim Oh**. Hierarchical Manycore Resource Management Framework using Control Processors. *Seoul National University*, Seoul, South Korea, February 2017. [pdf] [slides]

TPDS 2019: Younghyun Cho, **Surim Oh**, and Bernhard Egger. Performance Modeling of Parallel Loops on Multi-Socket Platforms using Queueing Systems. In *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, in press, available online, September 2019. [pdf]

MULTIPROG 2017: Younghyun Cho, **Surim Oh**, and Bernhard Egger. Cooperative Parallel Runtimes for Multicores. *Presented at the 10th International Workshop on Programmability and Architectures for Heterogeneous Multicores*, January 2017. [pdf]

CATC 2016: **Surim Oh**, Younghyun Cho, and Bernhard Egger. Efficient Resource Management for Many-cores with Centralized L2 Caches using Distributed Control Processors. *Presented at the 7th Compiler, Architectures and Tools Conference*, September 2016. [pdf] [slides]

PACT 2016: Younghyun Cho, **Surim Oh**, and Bernhard Egger. Online Scalability Characterization of Data-parallel Programs on Many Cores. In *Proceedings of the 25th International Conference on Parallel Architectures and Compilation Techniques (PACT)*, September 2016. [pdf]

JSSPP 2016: Younghyun Cho, **Surim Oh**, and Bernhard Egger. Adaptive Space-shared Scheduling for Shared-memory Parallel Programs. *Presented at the 20th Workshop on Job Scheduling Strategies for Parallel Processing, May 2016. In Lecture Notes in Computer Science (LNCS)*, Volume 10353, July 2016. [pdf]

PATENTS

US Patent 2018A: Bernhard Egger, **Surim Oh**, Younghyun Cho, Dong-hoon Yoo. Method of processing OpenCL Kernel and Computing Device Therefor. *US Patent 20180181443A1*, June 2018. Worldwide applications in KR, EP, CN, JP including US.

US Patent 2018B: Bernhard Egger, Younghyun Cho, **Surim Oh**. Dong-hoon Yoo. Computing devices and methods of allocating power to plurality of cores in each computing device. *US Patent 20180246554A1*, August 2018. Worldwide applications in KR, CN including US.

RESEARCH AND PROJECT EXPERIENCE

University of California, Santa Cruz *Santa Cruz, CA, USA*
PhD Student *Sep 2020 - Present*

· Working on CPU frontend in modern processors.

- Studied a state-of-the-art Fetch Directed instruction Prefetching (FDIP) on a microprocessor simulator, Scarab, and the performance impact of FDIP on frontend-bound applications with large instruction footprints.
- Introduce Utility-Driven FDIP (UDP) by learning the performance impact of running-ahead distance of FDIP and the usefulness of instruction cache lines for optimal running-ahead distance and filtering mechanisms.

SAP Labs Korea *Seoul, South Korea*
Developer *Jan 2018 - Sep 2020*

· SAP HANA Database Management System

- Table replication technology between HANA-to-otherDB or HANA-to-HANA for scaling out mixed OLTP/OLAP workloads in main-memory databases.
- Contributed to design and implementation of table replication framework to generate and collect replication logs in transactional order with minimum source side cost.

– Design and implementation of log formats and protocols for querying the logs from any other database systems to be compatible with other general database systems.

Hyundai Motor Company R&D Division *Hwaseong, South Korea*
Engineer *Feb 2017 - Jan 2018*

· Vehicle data monitoring system

– Data monitoring framework to collect in-vehicle data from a great number of distributed ECUs and to save the data in a remote centralized server.

– Contributed to design and development of the data collection framework that allows other research engineers to exploit specific data suitable for research in related fields.

– Design and development of a merge tool for video data from testing cars to develop Advanced Driver Assistance Systems technology.

Computer System And Platform Laboratory, Seoul National University *Seoul, South Korea*
Graduate Researcher *Feb 2015 - Feb 2017*

· Worked on resource management on manycore SoC architecture.

– Developed manycore resource management techniques in consideration of runtime scheduling overhead and power consumption exploiting architecture support of tiny control processors. - *CATC 2016, M.S. Thesis, US Patent 2018A, US Patent 2018B.*

– Distributed dynamic scheduling overheads through hierarchical core resource allocation for simultaneously running OpenCL applications.

– Worked on implementation on top of a TQSIM-HSIM (Timed QEMU-based & SystemC-based) manycore simulator.

– Collaborated with four other research groups from Seoul National University; architecture design, manycore simulator, programming model, and applications with 96-core Samsung research processor (research project with Samsung Advanced Institute of Technology).

– Presented at HumanTech Paper Award organized by Samsung (not awarded).

· Contributed to performance modeling on NUMA architectures.

– An analytical performance model based on queueing theory to estimate the resource utilization on multiprocessor systems for parallel programs - *PACT 2016, TPDS 2019.*

– Developed a tool for the performance model to query hardware performance counters related to the NUMA interconnection network in Intel/AMD systems.

AWARD

SnuMAP: 2nd prize in 10th Open Source Software World Challenge *Seoul, South Korea*
SNU Manycore Profiler for Big-Data with SNU team *December 2016*

TEACHING EXPERIENCE

Computer Architecture (CSE120), UC Santa Cruz *Fall 2023, Winter 2023, Fall 2021*
Teaching Assistant
Computer Architecture, Seoul National University *Spring 2016, Fall 2015*
Teaching Assistant

REFERENCES

Prof. Heiner Litz (University of California, Santa Cruz)
E-Mail: hlitz@ucsc.edu