Behnam Khaleghi

bkhaleghi@ucsd.edu linkedin.com/in/behnam-khaleghi cseweb.ucsd.edu/~bkhalegh/

Research Interests

I have a broad research interest in developing domain-specific accelerators, spanning from FPGA prototyping and ASIC design to leveraging unorthodox computing approaches, with a slant toward trade-offs for energy efficiency.

Education

UC San Diego **Sep 2017 – Sep 2022**
PhD in Computer Science and Engineering *La Jolla, California*

Sharif University of Technology **Sep 2014 – Feb 2016**
MSc in Computer Engineering (Data Storage and Networks Lab by Prof. Hossein Asadi) *Tehran, Iran*

Sharif University of Technology **Sep 2008 – June 2013**
BSc in Computer Engineering (Data Storage and Networks Lab by Prof. Hossein Asadi) *Tehran, Iran*

Relevant Coursework

- Advanced VLSI
- Low Power Design
- SoC Design
- Computer Architecture
- Electronic System Design
- Testability
- Reconfigurable Compute
- Electronic Sys. Design
- HDL Languages
- Advanced Algorithms
- Robotics Mathematics
- Data Analytics

Technical Skills

Programming Languages: Python, C++, Java
HDL: Verilog, Xilinx HLS
EDA Tools: Design Compiler, Innovus, SiliconSmart, HSPICE, Modelsim, Xilinx Vivado/Vitis
ML Framework: PyTorch

Work Experience

Qualcomm **Oct 2022 – present**
Senior Engineer at Qualcomm Multimedia R&D San Diego, California

- Working on efficient methods and hardware for computer vision applications.

Qualcomm **Oct 2020 – Jan 2021**
Research & Development Intern San Diego, California

- Developing a new image classification method by combining light-weight CNN and hyperdimensional computing (HD).
- Implementing and evaluating the HD classifier in Verilog, and estimating the performance, area, and power.

Sharif Advanced ICT Research Center **Sep 2013 – Sep 2014**
Developer Tehran, Iran

- Developing custom placement and routing algorithms for Xilinx FPGAs enabling reliable design implementation (e.g. modular redundancy) and mitigating differential power analysis (DPA) attacks.
- Developing automated tool to protect Hardware Trojan insertion in FPGAs.

Select Research Projects

HDnn: Learning on the Edge with CNN-HD Combination (TSMC Tape-out) **UCSD**

- A new algorithm for CNN compression using patterned weight clustering that uses only a few so-called seed filters, which also facilitates operation reuse. Achieved higher parameter and operation saving compared to state of the art.
- Designing an efficient architecture with maximal data reuse to effectively implement the proposed patterned CNNs.

Brain-Inspired Hyperdimensional Computing for Edge Applications (DARPA, SRC, and NSF) **UCSD**

- Developing parameterized FPGA implementations for HD, including a tool to generate HLS codes based on user specs.
- Developing new algorithms with trainable, flexible, compact, and robust ASIC implementations for HD, with up to three orders of magnitude less energy consumption versus other ML alternatives, while being capable of continuous learning.

Opportunistic Energy & Performance Efficiency in ASIC and FPGA Designs **UCSD**

- Integrating the thermal/power estimate of ASIC/FPGA in EDA flow for performance/energy gain by reducing margins.
- Developing an emulation tool to examine the applications' accuracy under non-ideal condition (e.g. over-scaled voltage).
- Developing a workload-aware energy saving mechanism in multi-FPGA platforms by load distribution and DVFS.

Big Data Acceleration using FPGAs and Processing-in-Memory (SRC) **UCSD**

- Porting the compute-intensive stages of bioinformatics workload to FPGA. Proposed and implemented a new two-stage read alignment algorithm that improves over existing popular tools by 40×.
- Implemented novel HD learning using FPGA and PIM for TB scale dataset; achieved over 150× speedup vs CPU.
- Integrated ASIC and FPGA cores within memory and flash hierarchy for near-memory learning.

Reliability of Embedded Systems under Voltage Scaling, Aging, and Thermal Effect **KIT, Germany**

- 2014 and 2015 summer intern in Karlsruhe Institute of Technology, Chair for Embedded Systems (Prof. Jörg Henkel)
- Examined the impact of degradations on circuits performance and implemented design- and run-time techniques to combat (e.g., approximate computing). Two of the publications nominated for best paper.

Honors and Awards

- Qualcomm Innovation Fellowship finalist (anomaly detection in sensor networks using HD computing).
- Best paper nomination in DAC 2017 and DATE 2017 conferences.
- Graduate Research Fellowship (UCSD).

Select Publications and Patents

Google scholar link for full list.

- B. Khaleghi *et al.*, SALIENT: Ultra-Fast FPGA-based Short Read Alignment, ICFPT 2022 (Best Paper Nomination)
- B. Khaleghi *et al.*, PatterNet: Explore and Exploit Filter Patterns for Efficient Deep Neural Networks, DAC 2022
- B. Khaleghi *et al.*, GENERIC: Highly Efficient Learning Engine on Edge using Highdimensional Computing, DAC 2022
- U. Mallappa *et al.*, TermiNETor: Early Convolution Termination for Efficient Deep Neural Networks, ICCD 2022
- A. Dutta *et al.*, HDnn-PIM: Efficient in Memory Design of HD Computing with Feature Extraction, GLSVLSI 2022
- B. Khaleghi *et al.*, tiny-HD: Ultra-Efficient Hyperdimensional Computing Engine for IoT Applications, DATE 2021
- S. Salamat *et al.*, Multiplication-free Neural Network by Migration to Residue Number Systems, ASP-DAC 20210
- B. Khaleghi *et al.*, Prive-HD: Privacy-Preserved Hyperdimensional Computing, DAC 2020
- S. Salamat *et al.*, Workload-Aware Opportunistic Energy Efficiency in Multi-FPGA Platforms, ICCAD 2020
- B. Khaleghi *et al.*, FPGA Energy Efficiency by Leveraging Thermal Margin, ICCD 2019
- H. Amrouch, B. Khaleghi *et al.*, Towards Aging-induced Approximations, DAC 2017 (Best Paper Nomination)
- H. Amrouch, B. Khaleghi *et al.*, Optimizing Temperature Guardbands, DATE 2017 (Best Paper Nomination)
- H. Amrouch, B. Khaleghi *et al.*, Reliability-Aware Design to Suppress Aging Effects, DAC 2016
- Methods, Circuits, and Articles of Manufacture for Searching within a Genomic Reference Sequence for Queried Target Sequence using Hyper-Dimensional Computing Techniques, US Patent App, 2022
- Methods and Systems Configured to Specify Resources for Hyperdimensional Computing Implemented in Programmable Devices using a Parameterized Template for Hyperdimensional Computing, US Patent App, 2022

Work Authorization: Permanent resident