

Andreas Falkenberg, PhD

Principal Compiler Architect — Hardware-Software Co-Design Consultant

Escondido, CA 92026 · 760.705.6234

andreas@falkenbergtech.com

linkedin.com/in/andreas-falkenberg-91b8571 · github.com/afalkenberg1

scholar.google.com/citations?user=LbZS3pAAAAAJ

SUMMARY

Principal compiler architect specializing in production compiler systems for AI accelerators and custom processors — LLVM/MLIR depth with hands-on ASIC/SoC, FPGA/SystemC and signal-processing experience, bridging research, hardware architecture and product delivery. Available for one 30–40 h/week contract engagement.

EXPERIENCE

05/2026 – 08/2026	PiePeople Consulting — Google / YouTube contract SystemC / FPGA Validation Consultant SystemC- and FPGA-based validation of a complex video-processing pipeline; isolating hardware/software issues across simulation and hardware environments.
05/2025 – 07/2026	BrainChip · Laguna Hills, CA Principal AI Compiler Architect Owned architecture and roadmap of the LLVM/MLIR AI compiler — PyTorch/ONNX models via Torch-MLIR across RISC-V, Arm and x86 hosts plus the BrainChip accelerator; hardware-aware graph partitioner; INT4–FP32 quantization taken from research into the production compiler.
08/2023 – 01/2025	AMD – Xilinx · San Jose, CA Compiler Software Engineer — Ryzen AI Group C++/Python compiler software for the Neural Network Engine across ONNX, Torch-MLIR, MLIR and Linalg; RMSNorm, ROPE and BF16 matmul kernels; patent filed on quantization using underutilized hardware features.
01/2025 – 05/2025	Ectron Corporation · San Diego, CA Part-Time Interim CTO / Strategic Advisor CEO-led business development, contract acquisition and product rollout.
08/2022 – 05/2023	d-Matrix · Santa Clara, CA Principal Compiler Software Architect LLVM/MLIR dialect and pass development for accelerator targets; prototype graph partitioner distributing computation across constrained execution resources.
2020 – 08/2022	Metawave · Carlsbad, CA Senior Principal Software and AI Lead Radar signal processing and object-classification AI — accuracy 69% → 95%+ (AWARE); SPECTRA verification framework; NXP S32R45 radar kernels; multiple patents filed.
2008 – 2020	Ostendo Technologies · Carlsbad, CA Digital Frontend ASIC Manager and Architect Custom processors, instruction sets and HW/SW partitioning for multiprocessor imaging SoCs; VHDL implementation and cycle-accurate workload analysis; built and led the ASIC team.
05/2023 – 08/2023 earlier 2014–2020	Center for Neuroacoustic Research · contract Audio Library / Pipeline in C++ Production C++ audio pipeline and microtuning algorithm — a cornerstone of CNR's daily operations.

SKILLS

LLVM/MLIR	TORCH-MLIR	ONNX	C++	PYTHON	VHDL	SYSTEMC	FPGA	ASIC/SOC	ISA DESIGN	QUANTIZATION
GRAPH PARTITIONING	DSP	HW-SW CO-DESIGN								

EDUCATION & ACADEMIC

Dr. rer. nat. (PhD), Computer Science — University of Hagen, Germany · 1998

PhD (EE) — University of South Wales, Pontypridd, UK · 2007

Diplom-Informatiker — Technical University of Dortmund, Germany

Adjunct professor & lecturer — RheinMain UAS, University of San Diego, UCSD Extension, Münster UAS (AI, databases, telecom, OOP).

EARLIER

Chief software/compiler architect roles at Nulife Technologies and Ellipsis Digital Systems · WCDMA/TDMA baseband software at Siemens organizations · DSP algorithms for Audi / Bang & Olufsen · forensic valuation software (Longstaff, Black-Scholes, Finnerty) for Vianello · patent-litigation review and technical consulting.

PATENTS & PUBLICATIONS

52 patents — **33** granted. Selected: US20260187184A1 (quantized ops on underutilized hardware features) · US2024323765A1 (balanced model distribution) · US20220034996A1 (pipelined FFT) · US10,528,004 / US10,310,450 (light-field display).

23 publications. Selected: IEEE ISSPIT 2018 · MIKON 2022 · CS&IT 2025 (kernel-fusion ×2). Full lists on request.