Maxence Bouvier, PhD

Email : maxence.bouvier.pro@gmail.com
LinkedIn | GoogleScholar
Mobile : +41-78-210-71-94

AI & ML Research Scientist with 7+ years of experience at the hardware/software interface, leading innovations in ML-powered chip design at Huawei. Expertise spans EDA automation, transformer-based AI models, computer vision, and energy-efficient HW accelerators.

EXPERIENCE

Huawei — Zurich, Switzerland
AI & HW Research Scientist — *May 2024 - Present*

- **Team Leader**: Built and led a multidisciplinary team of experts to develop innovative solutions for reducing switching activity and power consumption in Huawei's GPUs.
- **ML for Chip Design**
  - Automated the synthesis and simulation of millions of designs using containerized, open source EDA tools.
  - Designed a transformer-based neural network for discovering low switching activity multiplier units. The model is trained to predict the power consumption of a multiplier unit. The model is then inverted to generate low-power multipliers. - (*1 paper (wip).*)
- **ML for Advanced Synthesis**: Developed a predictor-driven synthesis framework achieving up to 21% QoR improvement and 14x faster execution, significantly advancing chip design optimization methods. - (*2 papers.*)
- **Characterization of ML Workloads Acceleration**
  - Developed a simulation platform to accurately map (tiling and multi-core scheduling) tensor operations onto Huawei's Ascend "Cube" tensor accelerator.
  - Utilized the simulator to benchmark tensor reshaping and vector reordering strategies, proposing novel software-level optimizations that effectively reduce power consumption.

SONY — Zurich, Switzerland
Senior AI Research Engineer — *Aug 2023 - Apr 2024*

- **Sparsity Exploitation in Transformers**
  - Engineered an asynchronous PointNet-based embedding, enabling continuous spatio-temporal data conversion into dense tensors for seamless, continuous feeding of Transformer models. - (*1 paper, 1 patent.*)
  - Designed an NPU-compatible, block-wise sparse scaled dot-product attention module for highly efficient flash attention in Transformers, achieving more than 50% FLOPs reduction during inference and higher accuracy.
- **SLAM Enhanced AI Training**: Enhanced performance by incorporating a cutting-edge SLAM pipeline for multi-modal training process, achieving 6x faster model convergence and a 15% accuracy improvement.

AI Research Engineer — *June 2022 - Aug 2023*

- **SW/HW Co-Design Automation with Neural Architecture Search**
  - Built an AI-driven, Hardware-Aware Neural Architecture Search framework. Reduced model FLOP cost to 8% of the original, with only a 4% accuracy loss.
  - Integrated a Design Space Exploration software in the NAS loop to estimate energy and latency of model execution.
- **Transformer Hardware Acceleration Survey**: Conducted a literature study, featured in CTO's strategic report.
- **Vision Transformer for Image Generation**
  - Implemented an AI model leveraging CNN and Transformer architectures to realize advanced frame generation. - (*1 patent.*)
  - Built a live demo of the model, from image sensor to application. This led to 2 major collaborations with other teams.
  - Packaged the model as an API to simplify sharing across teams and projects.
- **Software Maintainer**: Responsible for the CI of a few Python libraries shared among teams.

STMicroelectronics — Grenoble, France
Digital IC Design Engineer — *Apr 2021 - May 2022*

- **CPU Design and Automation**
  - Created a toolbox to automate component assembly of the Trace and Debug subsystem with ARM's Armv9-A SoC modules.
  - Developed an RTL generator for STM32 MPU SoC, streamlining the design of a multi-clock-domain reset and clock-control system for over 300 peripherals.
- **CPU Benchmarking**: Conducted CoreMark benchmarking on a multi-core MPU SoC, highlighting significant performance gains (up to 6x) through compiler updates.

CEA LETI — Grenoble, France
Doctoral Researcher on AI and Digital IC Design — *Apr 2018 - Apr 2021*

- **Neuromorphic Hardware Survey**: Conducted a comprehensive literature review on scalable, distributed, multi-chip neuromorphic hardware, leading to a widely cited publication in ACM JETC. - (*1 paper*.)
- **ULP NPU Design**: Built (RTL design, synthesis and layout) an ultra-low-power sparse AI accelerator, setting energy efficiency records (2.86pJ/OP in 28nm) and enabling seamless integration for 3D-stacked imagers. - (*1 paper, 2 patents.*)
- **EB VIO/SLAM Pipeline and Object Detection Innovation**: Developed an Event-Based VIO/SLAM pipeline with ego-motion compensation, leading to a solution for detecting moving objects. - (*1 patent.*)

IBM Research Yorktown Heights, NY, USA
Intern IC Design Engineer *Feb 2017 - Aug 2017*

- Automated wafer-scale memory device characterization, reducing execution time from days to hours.
- Contributed to the optimization of PCM technologies for Compute-in-Memory-based AI acceleration. - (*1 paper, 1 patent.*)

Skills

- **Languages**: Python, C/C++, SystemVerilog, MATLAB, VHDL
- **Libraries**: PyTorch, MLFlow, ONNX, CUDA, OpenCV, ROS, Dash, Flask, concurrent
- **Hardware & EDA Tools**: Yosys, OpenRoad, Verilator, Synopsys DC, Cadence Innovus
- **Software & DevOps**: Docker, Git, Continuous Integration (CI)

Education

- **Grenoble Alpes University** | *Ph.D. in Computer Science* *Apr 2018 – Apr 2021* | Grenoble, France
- **EPFL** | *"M.Eng. in Electronics (Highest Honors)"* *Sep 2015 – Sep 2017* | Lausanne, Switzerland
- **Grenoble Institute of Technology** | *B.Eng. in Electronics* *Sep 2012 – Sep 2015* | Grenoble, France

Patents

- M. Bouvier, et al., "Apparatus, method, and computer program for processing visual event data," *WO2024200170A1*, 2024.
- M. Bouvier, A. Valentian, "Observation system and associated observation method," *US2023196779A1*, 2023.
- F. Carta, et al., "Pulsing synaptic devices based on phase-change memory to increase the linearity in weight update," *US11557343B2*, 2023.
- M. Bouvier, A. Bige, "Device for compensating for movement of an event sensor, and associated systems and methods," *WO2022117535A1*, 2022.
- M. Bouvier, A. Valentian, "Device for compensating movement of an event-driven sensor and associated observation system and method," *US2022101006A1*, 2022.

Publications

- F. Arnold, et al., "Late Breaking Results: The Art of Beating the Odds with Predictor-Guided Random Design Space Exploration," *2025 62nd ACM/IEEE Design Automation Conference (DAC)*, 2025.
- C. M. Turrero, et al., "ALERT-Transformer: Bridging Asynchronous and Synchronous Machine Learning for Real-Time Event-based Spatio-Temporal Data," *Proceedings of the 41st International Conference on Machine Learning (ICML)*, 2024.
- M. Bouvier, et al., "Scalable pitch-constrained neural processing unit for 3D integration with event-based imagers," *2021 58th ACM/IEEE Design Automation Conference (DAC)*, 2021.
- M. Bouvier, "Study and design of an energy-efficient perception module combining event-based image sensors and spiking neural network with 3D integration technologies," *Ph.D. Dissertation, Université Grenoble Alpes*, 2021.
- M. Bouvier, et al., "Spiking neural networks hardware implementations and challenges: A survey," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 2019.
- P. Vivet, et al., "Advanced 3D technologies and architectures for 3D smart image sensors," *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2019.