



With us, you don't just talk the talk, you walk the walk.

Change starts with us.

Master Thesis / Intern (all genders) – Quantization- and Fault-Aware Training and Development for Neuromorphic Computing

The **Fraunhofer-Gesellschaft** currently operates 76 institutes and research institutions throughout Germany and is the world's leading applied research organization. Around 32 000 employees work with an annual research budget of 3.4 billion euros. The Fraunhofer Institute for Integrated Circuits IIS, located in Erlangen, is the largest institute of the Fraunhofer-Gesellschaft with more than 1 200 employees.

Fraunhofer IIS is very active in the fields of Embedded AI and **Neuromorphic Computing**. The **»Embedded AI**« group focuses on the development of concepts for analog and digital hardware architectures as well as the software frameworks required for the accelerated and energy-efficient computation of neural networks (DNNs and SNNs). In addition, the team creates and uses AI algorithms that run on resource-limited embedded processors.

The Neuromorphic DNN hardware accelerators developed at Fraunhofer IIS use analog compute cores for ultra-low energy consumption. However, weights and activations need to be severely quantized (down to signed 3-bit fixed-point), and – as for all analog hardware – the compute cores exhibit effects like deviation of components from their nominal values and noise. For these reasons, we have developed our own PyTorch-based library which extends Brevitas to tackle these effects already during the DNN training.

This work is focused on this internal Quantization- and Fault-Aware Training and Inference tool. The tasks involve developing our tool further to support new features like new NN models and layer types, e.g., LSTMs. Additionally, this work includes implementing and testing new methods to keep up with the state-of-the-art and expand on it in areas of training, neural architecture search and edge AI.

Are you interested in researching and developing the latest methods of Quantization- and Fault-Aware Training techniques in the exciting field of Neuromorphic Computing? Then have a look at our offer!

What you will do

- You extend the Fraunhofer training tool for new NN models and layer types.
- You train quantized yet robust NN models, e.g., for audio use cases, using the forementioned tool.
- Optional: You research new and novel methods to enhance the training and resulting models.

What you bring to the table

- You are currently studying electrical or computer engineering or a related field.
- You are experienced in Python programming.
- You have a solid background in Machine Learning and Neural Networks.
- You have a good understanding of Edge AI/TinyML (e.g. quantization, HW deployment).

What you can expect

- **Flexible** working hours
- Open and friendly team work
- Varied tasks with room for creativity
- Exciting seminars and events
- Networking with scientists
- Active contribution in applied research
- Interesting and innovative projects
- Mentoring program <u>»josephine®«</u> for talented female students

You can start from now on. Your full-time internship will last at least 3 months.

The thesis will be assigned and carried out in accordance with the rules of your university. For this reason, please discuss the thesis with a professor who can advise you over the course of the project.

We value and promote the diversity of our employees' skills and therefore welcome all applications - regardless of age, gender, nationality, ethnic and social origin, religion, ideology, disability, sexual orientation and identity.

Interested?

Apply <u>online</u> now (PDF: cover letter, CV, transcripts). We look forward to getting to know you!

Anne Weber Fraunhofer-Institute for Integrated Circuits IIS www.iis.fraunhofer.de/en

Location: Erlangen

