Changing the World with Research Results

A scientist sits in his quiet chamber and does research ... and then? In the interview, Dr. Jens Krüger talks about how research findings find their way into companies and from there into people’s everyday lives. He is Fraunhofer expert for the strategic research field of "Next Generation Computing". This stands on three pillars: the first pillar is based on classical architectures. The second pillar is neuromorphic computers, which function in much the same way as our brain, and the third pillar is quantum computers.

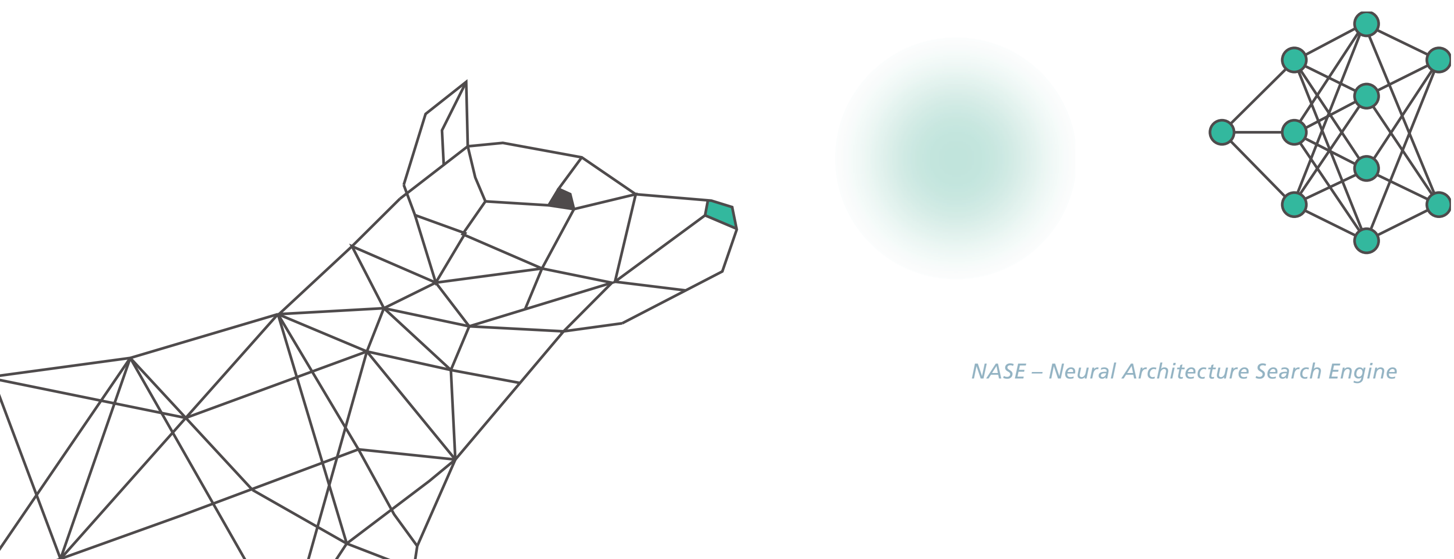
Please briefly summarize what drives you as a researcher?

I am a curious person. I want to try out new ideas and develop them further into products that will then benefit our society and the economy. This goes from the development of highly efficient processors to the optimization of artificial neural networks for mobile devices. One example are smart watches that record your ECG. In this way, the device can detect an approaching heart attack and trigger the alarm at an early stage. This technology has the potential to save people’s lives.

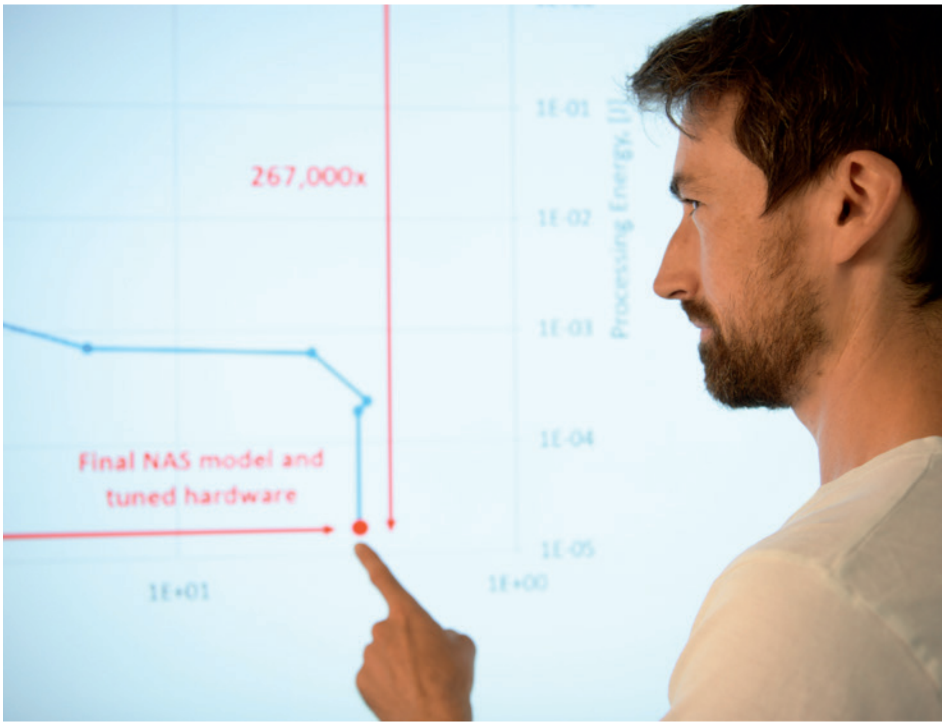
National science competitions often give research impulses. In March 2021, your team won a prize in the pilot innovation competition "Energy-efficient AI systems" of the Federal Ministry of Education and Research (BMBF). What was that about?

The task was to develop the most energy-efficient AI hardware possible that detects cardiac arrhythmias and atrial fibrillation in ECG data with at least 90 percent accuracy. We entered the competition with was called HALF, which stands for "Holistic AutoML for FPGAs”. Our approach was to have an holistic automated machine learning (AutoML) optimization of the neural network model and the FPGA implementation. We investigated the interdependence of the energy consumption of the hardware and the neural network topology.

The choice of network has a considerable influence on the hardware complexity – and thus on the required energy and vice versa. We have optimized these dependencies and developed a new methodology that finds not only more energy efficient models, but also reduces the development time for optimal neural network topologies.



NASE – Neural Architecture Search Engine



What happened after you won the competition?

We now make our expertise available to companies so that they can develop their products accordingly. The AI chip in the competition was merely a test platform. In the follow-up project, we are now working directly with a manufacturer to develop the next generation of devices, which will then be used in clinical trials.

However, we can support all industries, because almost everyone faces the challenge that data volumes are constantly increasing and AI can help to process them. This is economically interesting for almost everyone, for example, for the automotive industry or the telecommunications industry. This gave rise to the software product »NASE« (Neural Architecture Search Engine).

NASE makes scientific expertise available to companies. How does this work?

Every job is individual, but one thing is clear: for us, efficiency starts with the algorithm. We use state-of-the-art methods of automatic neural network search to develop networks that can be efficient with respect to many aspects at the same time. We consider peculiarities of the underlying platform and incorporate them into the network design. The algorithm then adapts the networks to the hardware. For example we can provide our experience, the technology, and computing capacity. Companies provide us with the data sets that are relevant to them and define the requirements, such as accuracy and speed. We then use our supercomputers and our framework to find the best model. The network is then ready for immediate use.

To meet the demand for ever more and faster computing power your team is part of the European Processor Initiative (EPI) which develops highly efficient processors for Europe. What contribution does the Fraunhofer ITWM here?

Our contribution is the so-called Stencil and Tensor Accelerator (STX), which we are developing together with Fraunhofer IIS based on an architecture from ETH Zürich. We focus on the efficient execution of highly parallelizable applications with specific access patterns, as they occur in many applications – from fluid dynamics, climate and weather prediction to imaging techniques. Real-world applications are expected to become more energy efficient, easier to program, and lower cost. Interested parties can already test their own codes on our simulator. Next year, the next generation of test chips will be available. By 2025, we want to have the first complete system up and running. A major challenge, but also an important step toward a new national and European industry for high-performance processors and accelerators.

www.itwm.fraunhofer.de/NASE_EN

Contact

Dr. Jens Krüger
Speaker Division Management
"High Performance Computing"
Phone +49 631 31600-4541
jens.krueger@itwm.fraunhofer.de