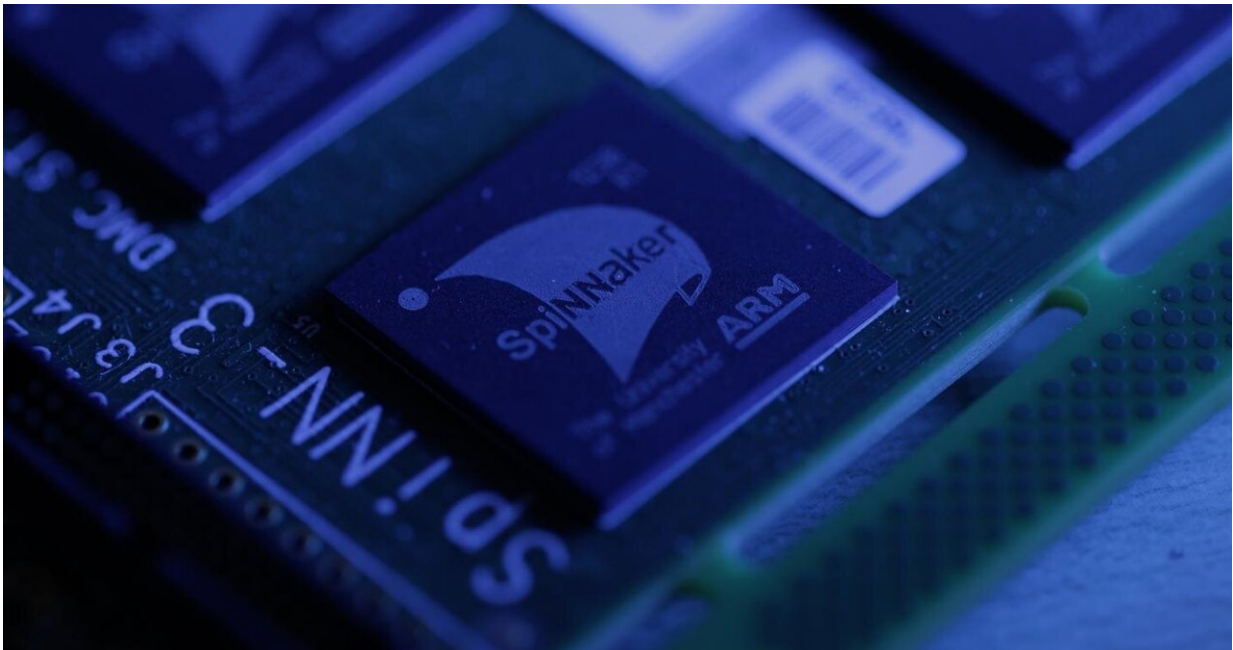# New approach found for energy-efficient AI applications

March 11 2021, by Christoph Pelzl

The algorithm will be implemented on brain-inspired computing systems, like the spike-based SpiNNaker (pictured here). SpiNNaker is part of the Human Brain Project's EBRAINS research infrastructure. Credit: Forschungszentrum Jülich

Most new achievements in artificial intelligence (AI) require very large neural networks. They consist of hundreds of millions of neurons arranged in several hundred layers, i.e. they have very 'deep' network structures. These large, deep neural networks consume a lot of energy in

the computer. Those neural networks that are used in image classification (e.g. face and object recognition) are particularly energy-intensive, since they have to send very many numerical values from one neuron layer to the next with great accuracy in each time cycle.

Computer scientist Wolfgang Maass, together with his Ph.D. student Christoph Stöckl, has now found a design method for artificial neural networks that paves the way for energy-efficient high-performance AI hardware (e.g. chips for driver assistance systems, smartphones and other mobile devices). The two researchers from the Institute of Theoretical Computer Science at Graz University of Technology (TU Graz) have optimized artificial neuronal networks in computer simulations for image classification in such a way that the neurons —similar to neurons in the brain—only need to send out signals relatively rarely and those that they do are very simple. The proven classification accuracy of images with this design is nevertheless very close to the current state of the art of current image classification tools.

Information processing in the human brain as a paradigm

Maass and Stöckl were inspired by the way the human brain works. It processes several trillion computing operations per second, but only requires about 20 watts. This low energy consumption is made possible by inter-neuronal communication by means of very simple electrical impulses, so-called spikes. The information is thereby encoded not only by the number of spikes, but also by their time-varying patterns. "You can think of it like Morse code. The pauses between the signals also transmit information," Maass explains.
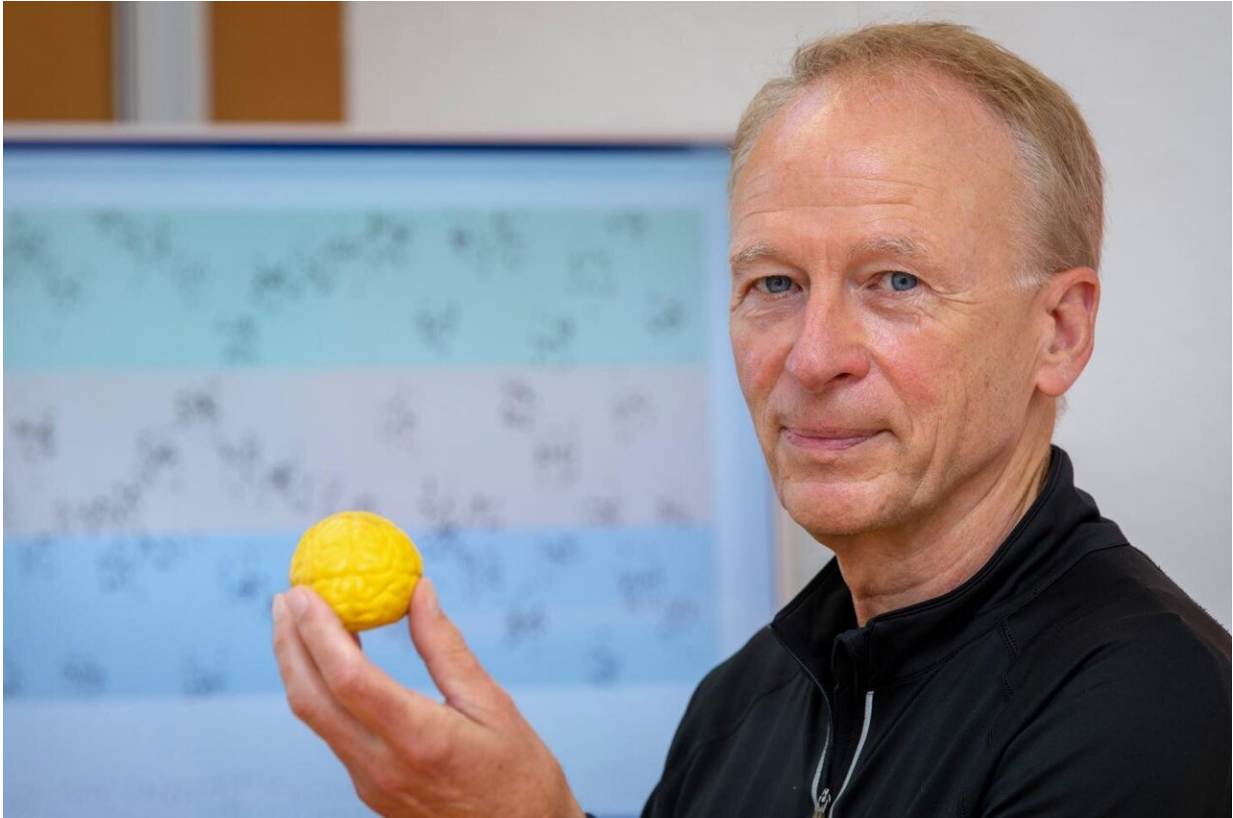
TU Graz computer scientist Wolfgang Maass is working on energy-efficient AI systems and is inspired by the functioning of the human brain. Credit: Lunghammer - TU Graz

Conversion method for trained artificial neural networks

That spike-based hardware can reduce the energy consumption of neural network applications is not new. However, so far this could not be realized for the very deep and large neural networks that are needed for really good image classification.

In the design method of Maass and Stöckl, the transmission of information now depends not only on how many spikes a neuron sends out, but also on when the neuron sends out these spikes. The time or the

temporal intervals between the spikes practically encode themselves and can therefore transmit a great deal of additional information. "We show that with just a few spikes—an average of two in our simulations—as much information can be conveyed between processors as in more energy-intensive hardware," Maass said.

With their results, the two computer scientists from TU Graz provide a new approach for hardware that combines few spikes and thus low energy consumption with state-of-the-art performances of AI applications. The findings could dramatically accelerate the development of energy-efficient AI applications and are described in the journal *Nature Machine Intelligence*.

More information: C. Stoeckl and W. Maass. Optimized spiking neurons can classify images with high accuracy through temporal coding with two spikes. *Nature Machine Intelligence*. (2021) DOI: 10.1038/s42256-021-00311-4

Provided by Graz University of Technology