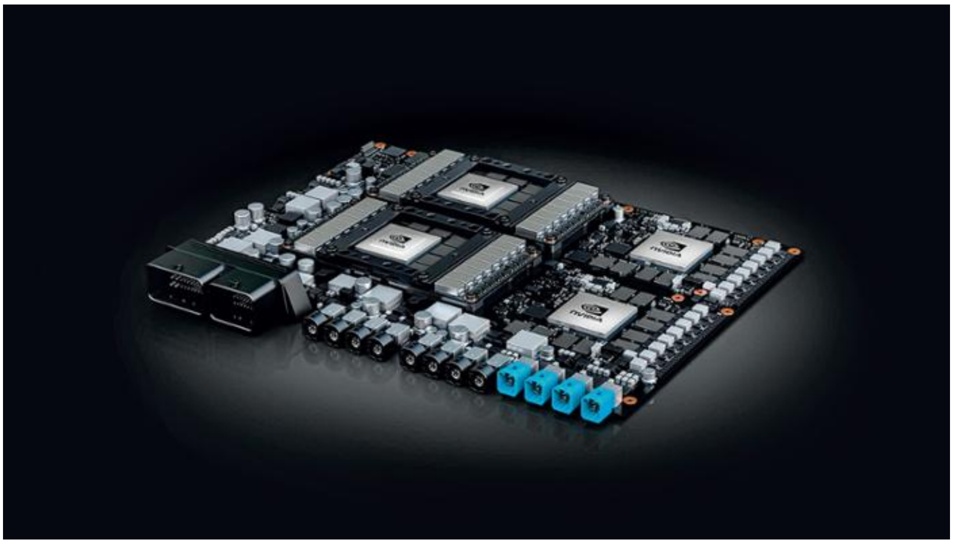
Technology 13/02/2020

Booster for AI calculations

Automated and autonomous driving functions are impossible to implement without AI. The required computing capacity is provided by special chips specialized in parallel computing. But researchers are also working on new, biologically inspired solutions as well as on quantum computers that promise even more computing capacity.

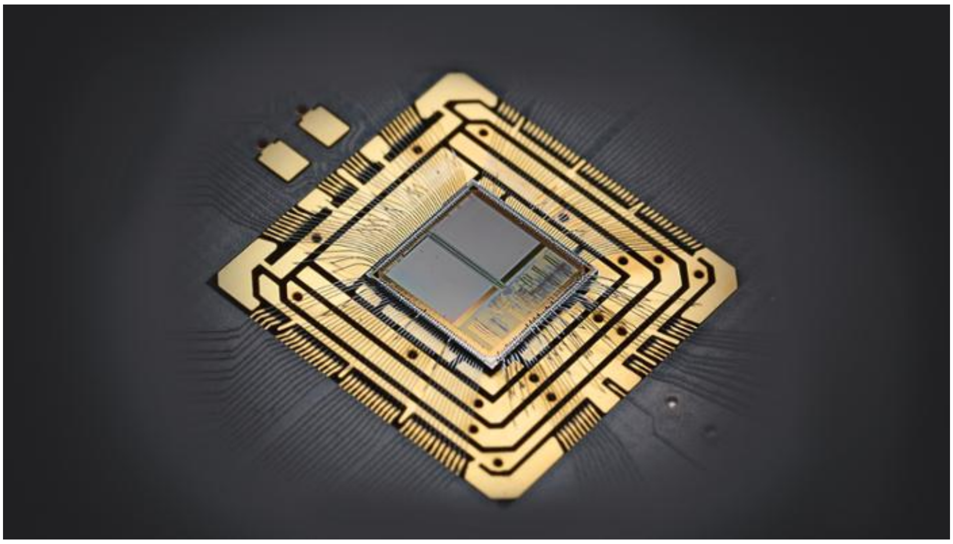


For decades, electronics have become increasingly prevalent in vehicles. Today, dozens of networked control devices control the engine, transmission, infotainment system and many other functions. Cars have long since become rolling computing centers—but now a new leap in computer power awaits them, because automated driving functions and autonomous driving require ever more powerful computers. And because the required performance cannot be achieved with conventional chips, the hour has come for graphics processors, tensor processing units (TPUs), and other hardware specially designed for the calculations of neural networks.
While conventional CPUs (central processing units) can be used universally, they lack the optimal architecture for AI. That is due to the typical calculations that occur during the training of and inference process with neural networks. “The matrix multiplications in neural networks are very elaborate,” explains Dr. Markus Götz of the Steinbuch Centre for Computing at the Karlsruhe Institute of Technology (KIT). “But these calculations are very amenable to parallelization—particularly with graphics cards. Whereas a high-end CPU with 24 cores and vector commands can perform 24 times 4 calculations per cycle; with a modern graphics card it’s over 5,000.”

Graphics processors (GPUs, graphics processing units) are specialized for parallel work from the outset and have an internal architecture tailored for that purpose: GPUs contain hundreds or thousands of simple computation modules for integer and floating-point operations, which can simultaneously apply the same operation to different data (single instruction multiple data). They are therefore able to execute thousands of computing operations per clock cycle—for instance to compute the pixels of a virtual landscape or the matrix multiplications for neural networks. So it’s no wonder that chips from the GPU manufacturer NVIDIA are currently ideally positioned as the workhorses for artificial intelligence in general and autonomous driving in particular. Volkswagen uses the US company’s hardware, among others. “You need special hardware for autonomous driving,” says Ralf Bauer, Senior Manager Software Development at Porsche Engineering. “GPUs are the starting point; later, application-specific chips will presumably follow.”
NVIDIA currently offers the Xavier processes for autonomous driving specifically. A silicon chip is outfitted with eight conventional CPUs and one GPU specifically optimized for machine learning. For automated driving on level 2+ (limited longitudinal and lateral control with enhanced functionality based on standard sensors compared to level 2), the Drive AGX Xavier platform is available, which can execute a

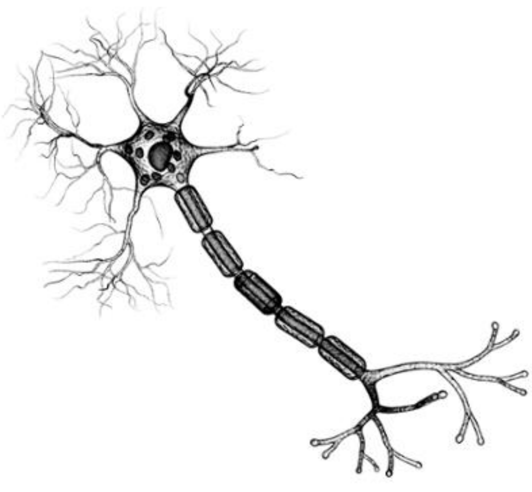
maximum of 30 trillion computing operations per second (30 TOPS, Tera Operations Per Second). For highly automated and autonomous driving, NVIDIA has the Drive AGX Pegasus (320 TOPS), under the control of which a test vehicle has driven as far as 80 kilometers without human intervention through Silicon Valley. As the successor to Xavier, NVIDIA is now developing the Orin GPU, though little is currently known about its performance data.

Not all automobile manufacturers utilize GPUs. In 2016, Tesla began developing its own processors for neural networks. Instead of graphics processors from NVIDIA, the US-based company has been installing its FSD (Full Self-Driving) chip in its vehicles since early 2019. In addition to two neural processing units (NPUs) with 72 TOPS apiece, it also contains twelve conventional CPU cores for general calculations and a GPU for post-processing of image and video data. The NPUs, like GPUs, are specialized in parallel and thereby fast execution of addition and multiplication operations.



Neuromorphic hardware from Heidelberg: this chip contains 384 artificial neurons and 100,000 synapses.

Nerve cells and artificial neurons



Nerve cells receive their signals from other neurons via synapses that are located either on the dendrites or directly on the cell body. Synapses can have either an excitatory or inhibitory effect. All inputs are totaled at the axon hillock and if a threshold is exceeded in the process, the nerve cell fires off a roughly millisecond-long signal that propagates along the axon and reaches other neurons.
Artificial neurons mimic this behavior more or less exactly. In conventional neural networks with multiple layers, each "nerve cell" receives a weighted sum as an input. It consists of the outputs of the neurons of the preceding layer and the weighting factor wi, in which the learning experience of the neural network is stored. These weighting factors correspond to the synapses and can also be excitatory or inhibitory. A configurable threshold value determines, like in a nerve cell, when the artificial neuron fires.

Info
Text: Christian Buck
Contributors: Ralf Bauer, Dr. Christian Koelen
Text first published in the Porsche Engineering Magazine, issue 02/2019

Link Collection

Link to this article
https://newsroom.porsche.com/en_AU/2020/technology/porsche-engineering-booster-ai-calculations-autonomous-driving-19924.html

Media Package
https://newsroom.porsche.com/media-package/porsche-engineering-booster-ai-calculations-autonomous-driving-

External Links
https://www.porscheengineering.com/peg/en/