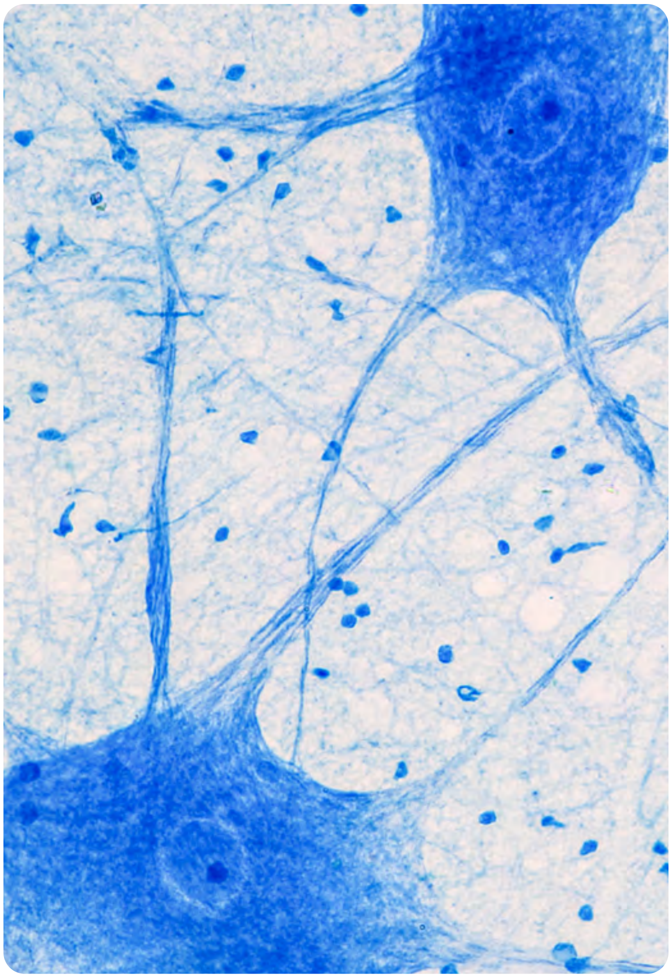
Neuromorphic Engineering Enabled by High Performance Hardware and Custom FPGAs

EXXACT

Contents



Background

New machine learning models powered by artificial intelligence (AI) and trained on large neural networks have had a significant impact worldwide. While AI can perform highly complex tasks and display human-like intelligence, the training of these complex models and their neural network structure has departed from traditional biological intelligence that they were once modeled after.

An artificial neural network (ANN) consists of layers of interconnected nodes called neurons, organized into multiple layers of numerically represented parameters called weights. These weights adjust during training by making predictions against real-world representative training data to enable the model to continuously learn and make predictions with input data alone. ANNs train by adjusting the weights of neuron nodes iteratively, necessitating constant mathematical computation powered by high-performance GPUs and CPUs. As AI models continue to grow larger, the energy and hardware requirements also increase.

The **Integrated Systems Neuroengineering (ISN) Lab** at the **University of California, San Diego** recognized the need for a transformative approach. The ISN Lab is part of the **Shu Chien-Gene Lay Department of Bioengineering** in the **Jacobs School of Engineering** and the **Institute for Neural Computation** at UC San Diego, and their research involves new methods to computing biologically inspired systems that not only reduces energy consumption, but also enhances efficiency & performance of neural networks. The brain serves as the template for the ultimate computer, propelling both voluntary and involuntary actions, igniting creativity, and reacting to unexpected stimuli with a mere 12 watts of power per day. The lab's mission is to bridge the gap between biological intelligence and computer-generated intelligence by developing a brain inspired neural network prioritizing energy efficiency over the power-hungry GPUs and AI accelerators.

©2023 Exxact Corporation

Challenges & Objectives

At the core of ISN's mission lies the formidable challenge of addressing the ever-increasing energy & computational demands of current complex AI models, such as generative AI and Large Language Models (LLMs), which seek to emulate human-like intelligence. While traditional neural networks have their place, they do not accurately emulate the true nature of the brain.

Developing a brain-inspired computing platform capable of constructing biologically inspired neural networks and challenging conventional machine learning paradigms is no easy feat. The brain utilizes a complex neural network where interconnected neurons respond to stimuli, firing when the action potential is reached and causing other neurons to also fire. The neural chain reaction is what defines the intelligence of neurons. The study of neuromorphic cognitive computing aims to research solutions for complex cognitive tasks by implementing a computer-emulated spiking neuron that captures the characteristics of a biological neural network. Despite 30 years of industry interest in neuromorphic cognitive computing, spiking neurons, bio-inspired neural networks, & proof-of-concept models have been confined to small projects with high entry barriers, such as prosthesis and neural control devices.

The ISN Team at UC San Diego employed research in Neuromorphic Engineering, Machine Learning, and Brain Interfacing to develop a new computing platform closely resembling the structure of a biological neural network. ISN’s project can help solve the increasing energy demand required to run AI models and enable researchers to further the study the brain via a testbed for a spiking neural network (SNN).

©2023 Exxact Corporation

The Solution

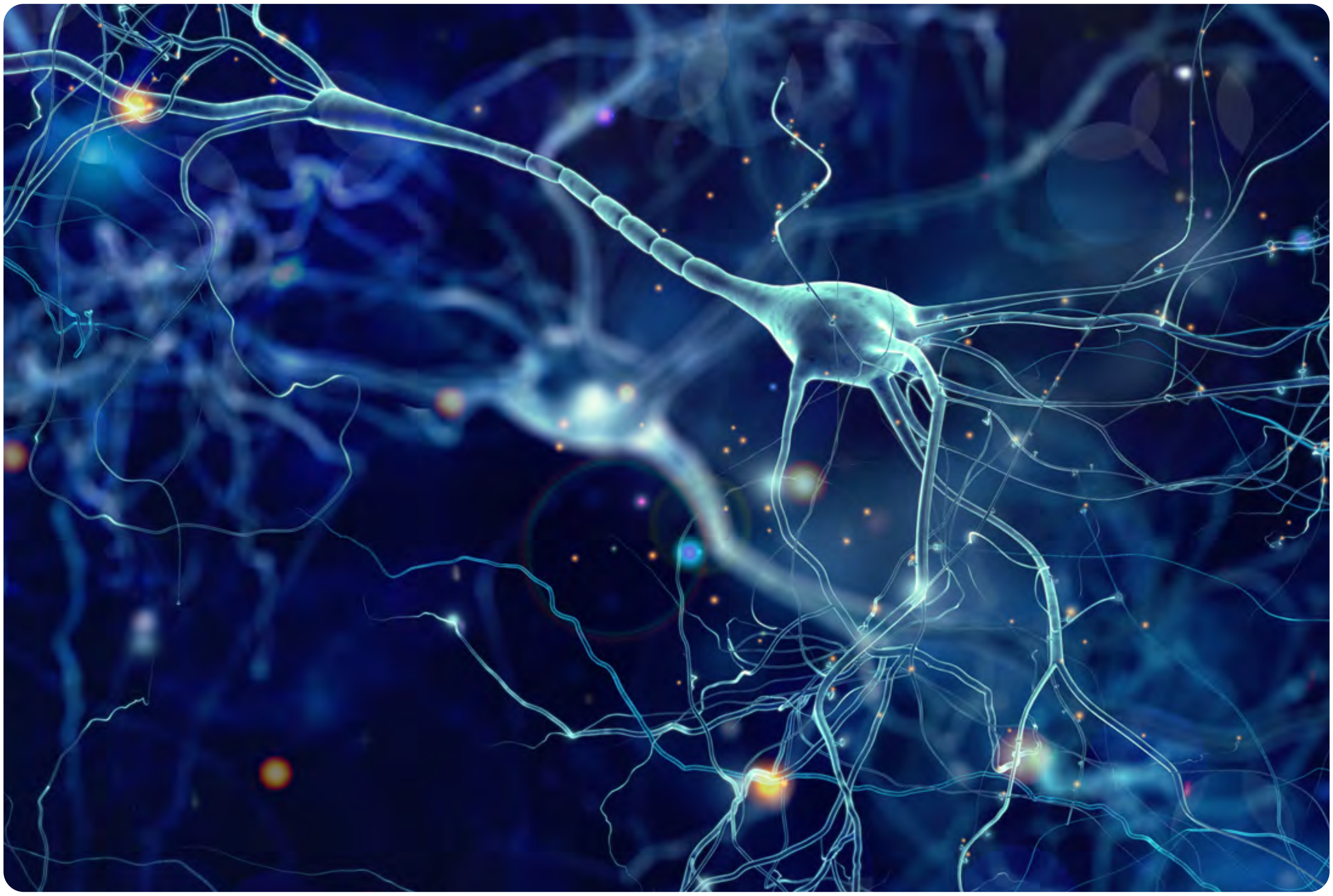
When it comes to hardware selection, ISN chose Exxact Corporation for its flexibility. Developing a neuromorphic cognitive computing platform detracts from traditional AI computing hardware like GPUs. Instead, ISN utilizes specially programmed FPGAs (field programmable arrays) designed in house, for fast hardware interconnect to create a spike-based neuromorphic system. With help of Exxact's engineers, ISN configured multiple server platforms powered by dual AMD EPYC processors, ample storage, sufficient memory, and AMD Infinity Fabric, while also capable of housing up to 10 dual-wide PCIe devices. Opting for a no-GPUs option, ISN can slot in their custom spiking-neuron FPGAs in the unpopulated PCIe slots.

The Solution

The computing platform is comprised of a rack with 5 Exxact compute servers equipped with 8 custom FPGAs each for a total of 40 FPGAs all interconnected to deliver an impressive 160 million neurons and 40 billion synapses. Exxact served as a pivotal hardware provider in enabling ISN’s computing goals, developing very fast, very responsive, and network accessible compute platform for emulation of a biological brain.

> “Exxact was very patient with a lot of questions from one of us with very little server configuration experience allowing us to select the platform that best suits our needs, allowing us to add our own ‘one-off’ specialized FPGAs rather than having a preconfigured system.”
>
> - Steve Deiss, Development Engineer at ISN

To mitigate the increased energy demand, ISN's spiking neural network takes inspiration from the brain's neural structure, incorporating temporal dynamics and neuron-like behaviors. ISN's neuromorphic computing platform distinguishes itself from traditional neural networks in two fundamental aspects. With a parallelized approach, multiple spiking neurons can fire and trigger other spiking neurons to also fire, a manner reminiscent of biological neurons. Secondly, due to the event-driven nature of spiking neurons, computations are initiated only when spikes exceed an adjustable weighted threshold. Similar to ANN, each spiking neuron also carries a weight, which is updated and stored in memory. The ISN neuromorphic computing platform delivers in a more energy efficient platform compared to the traditional artificial neural network computing platforms that perform constant matrix calculations during the training and inference stages of the AI model.



How ISN is Impacting Neuroscientific Computing

ISN places a strong emphasis on collaboration with neuroscientists and their drive to construct the first computing testbed for emulating biological neurons. To facilitate this groundbreaking endeavor, they have made their neuromorphic cognitive computing platform publicly accessible, offering users a user-friendly GUI to submit jobs via the San Diego Supercomputer Center Neuroscience Gateway (NSG).

Experiments performed on ISN's spiking neural network range from developing new methods for powering artificial intelligence models to research on the biological brain. Researchers submit jobs via NSG to run workloads using ISN's bio-inspired neural network, enabling researchers worldwide to experiment with computer-emulated spiking neurons and temporal dynamics research.

The computing platform has profound impact in the AI compute industry for continuous research towards a more energy-efficient AI training and deployment. An approach adopting spiking neurons allows for a more dynamic and event-based processing, veering away from the traditional artificial neural networks that are better suited for machine learning type tasks. ISN's computing platform provides a means to research the use of SNNs for sensory type AI tasks like robotics control and brain control interfacing as well as the potential to power current AI such as image recognition & language models. By offering a neuromorphic computing platform to the community, research and development of fascinating scientific results are limitless, boundless, and world changing.

ISN Lab's work possesses the potential to advance AI research, enriching and unlocking new horizons in neural psychiatry and neurocognitive computing. While research and the capabilities of spiking neurons, computational neuroscience, and neuromorphic systems are still in their infancy, democratizing the specialized FPGAs and HPC resources, as well as a new computing paradigm, brings the world closer to new technological advancements.

©2023 Exxact Corporation

References

G. Hota, G. Frank, K. Wang, A.Uppal, O. Olajide, J. Liu, S. Bansal, K. Yoshimoto, Q. Wang, S. Deiss, and G. Cauwenberghs, "Reconfigurable Event-Driven Spiking Neuromorphic Computing Near High-Bandwidth Memory," IEEE Biomedical Circuits and Systems Conf. (BioCAS'2023), Toronto Canada, Oct. 19-21, 2023.

G. Frank, G. Hota, K. Wang, A. Uppal, O. Olajide, J. Liu, S. Bansal, K. Yoshimoto, Q. Wang, S. Deiss, and G. Cauwenberghs, "Hardware-Software Co-design for Large-Scale Reconfigurable Event-Driven Neuromorphic Computing," ACM Int. Conf. on Neuromorphic Systems (ICONS'2023), Aug. 1-3, 2023.

N. Mysore, G. Hota, S.R. Deiss, B.U. Pedroni, and G. Cauwenberghs, "Hierarchical Network Connectivity and Partitioning for Reconfigurable Large-Scale Neuromorphic Systems," Frontiers in Neuroscience, vol. 15, pp. 797654:1-13, 10.3389/fnins.2021.797654, 2022.

G. Hota, N. Mysore, S. Deiss, B. Pedroni, and G. Cauwenberghs, "Hierarchical Multicast Network-On-Chip for Scalable Reconfigurable Neuromorphic Systems," 2022 IEEE Int. Symp. Circuits and Systems (ISCAS'2022), pp. 481-485 , Austin TX, May 28 - June 1, 2022.

B.U. Pedroni, S.R. Deiss, N. Mysore, and G. Cauwenberghs, "Design Principles of Large-Scale Neuromorphic Systems Centered on High Bandwidth Memory," 2020 International Conference on Rebooting Computing (ICRC'2020), pp. 90-94, 2020.

Links

https://isn.ucsd.edu
https://be.ucsd.edu
https://jacobsschool.ucsd.edu
https://inc.ucsd.edu/
https://ucsd.edu

INTEGRATED SYSTEMS NEUROENGINEERING

©2023 Exxact Corporation