IBM researchers are developing a new computer architecture, better equipped to handle increased data loads from artificial intelligence. Their designs draw on concepts from the human brain and significantly outperform conventional computers in comparative studies. They report on their recent findings in the *Journal of Applied Physics*, from AIP Publishing.

Today's computers are built on the von Neumann architecture, developed in the 1940s. Von Neumann computing systems feature a central processer that executes logic and arithmetic, a memory unit, storage, and input and output devices. Unlike the stovepipe components in conventional computers, the authors propose that brain-inspired computers could have coexisting processing and memory units.

Abu Sebastian, an author on the paper, explained that executing certain computational tasks in the computer's memory would increase the system's efficiency and save energy.

"If you look at human beings, we compute with 20 to 30 watts of power, whereas AI today is based on supercomputers which run on kilowatts or megawatts of power," Sebastian said. "In the brain, synapses are both computing and storing information. In a new architecture, going beyond von Neumann, memory has to play a more active role in computing."

The IBM team drew on three different levels of inspiration from the brain. The first level exploits a memory device's state dynamics to perform computational tasks in the memory itself, similar to how the brain's memory and processing are co-located. The second level draws on the brain's synaptic network structures as inspiration for arrays of phase change memory (PCM) devices to accelerate training for deep neural networks. Lastly, the dynamic and stochastic nature of neurons and synapses inspired the team to create a powerful computational substrate for spiking neural networks.

Phase change memory is a nanoscale memory device built from compounds of Ge, Te and Sb sandwiched between electrodes. These compounds exhibit different electrical properties depending on their atomic arrangement. For example, in a disordered phase, these materials exhibit high resistivity, whereas in a crystalline phase they show low resistivity.

By applying electrical pulses, the researchers modulated the ratio of material in the crystalline and the amorphous phases so the phase change memory devices could support a continuum of electrical resistance or conductance. This analog storage better resembles nonbinary, biological synapses and enables more information to be stored in a single nanoscale device.

Sebastian and his IBM colleagues have encountered surprising results in their comparative studies on the efficiency of these proposed systems. "We always expected these systems to be much better than conventional computing systems in some tasks, but we were surprised how much more efficient some of these approaches were."

Last year, they ran an unsupervised machine learning algorithm on a conventional computer and a prototype computational memory platform based on phase change memory devices. "We could achieve 200 times faster performance in the phase change memory computing systems as opposed to conventional computing systems." Sebastian said. "We always knew they would be efficient, but we didn't expect them to outperform by this much." The team continues to build prototype chips and systems based on brain-inspired concepts.