e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:11/November-2021 Impact Factor- 6.752 www.irjmets.com

REVIEW RESEARCH PAPER NEUROMORPHIC COMPUTING AND APPLICATIONS

Hitesh Dureja*1, Yash Garg*2, Dr. Rishu Chaujar*3, Bhavya Kumar*4

*1UG Student (2K20/EP/055, Dept. Of Applied Physics, Delhi Technological University)

*2UG Student (2K20/EP/121, Dept. Of Applied Physics, Delhi Technological University)

*3Professor (Delhi Technological University)

*4Ph.D. Student (Delhi Technological University)

ABSTRACT

A neuro-inspired computing chip emulates the structure and operation of the biological brain and represents an innovative and assured approach to the development of intelligent computing[1]. When processing Artificial Intelligence workloads, these neuro-inspired computing chips are expected to provide advantages in power efficiency and computing power over traditional systems[3]. Over the past few years, a spread of neuro-inspired computing chips have been developed[11]. Several neuro-inspired innovations have been incorporated into these chips at various levels, from the hardware to the circuitry to the architecture[7]. This is still at a beginning stage in the development of neuro-inspired computing chips, so exploring the hurdles and opportunities for the field is crucial[10].

The origins of neuro-inspired computing chips and recent progress in the domain is studied by us[5]. We classify four critical metrics for deciding the performance of the fragments: computing density, energy efficiency, computing accuracy, and learning capability[8]. We then explore the challenges and co-design principles of developing large-scale chips based on non-volatile memory (NVM)[2]. We also address the future electronic design automation (EDA) toolchain and propose a technological roadmap to develop large-scale neuro-inspired computing chips[1].

I. INTRODUCTION

Human brains are capable of processing extensive amounts of information while consuming little energy. In response to a need, the brain turns up computation, but it immediately returns to a baseline. As far as silicon-based computers are concerned, such efficiencies have never been achieved. Massive amounts of electricity are required to process huge amounts of data[4].

Deep neural networks (DNNs) and chips have achieved considerable improvements in accuracy on a variety of large-scale classification tasks, some even surpassing human performance[6]. Nevertheless, in order to achieve better training efficiency, the DNNs model's parameters rise exponentially, resulting in hundreds of millions of parameters and large training datasets stored in the memory. In the traditional, von Neumann-based computer architecture, the data needs to be moved back and forth between memory, and the processor, resulting in limited hardware energy efficiency for these machine learning workloads[14,10,8].

The concept of adaptive parallel processing in biological neural networks (BioNNs) using neuro-inspired computing is proposed to eliminate the energy-intensive and inefficient transmission of von Neumann-based platforms.

The two main paradigms of neuro-inspired computing are digital-bit-encoded artificial neural networks (ANNs) and spike-timing-encoded spiking neural networks (SNNs). Deep neural networks (DNN) and convolutional neural networks (CNNs), as well as recurrent neural networks (RNNs), have been successful in logical computing, machine vision, intelligent search, and automatic driving.[2,3]

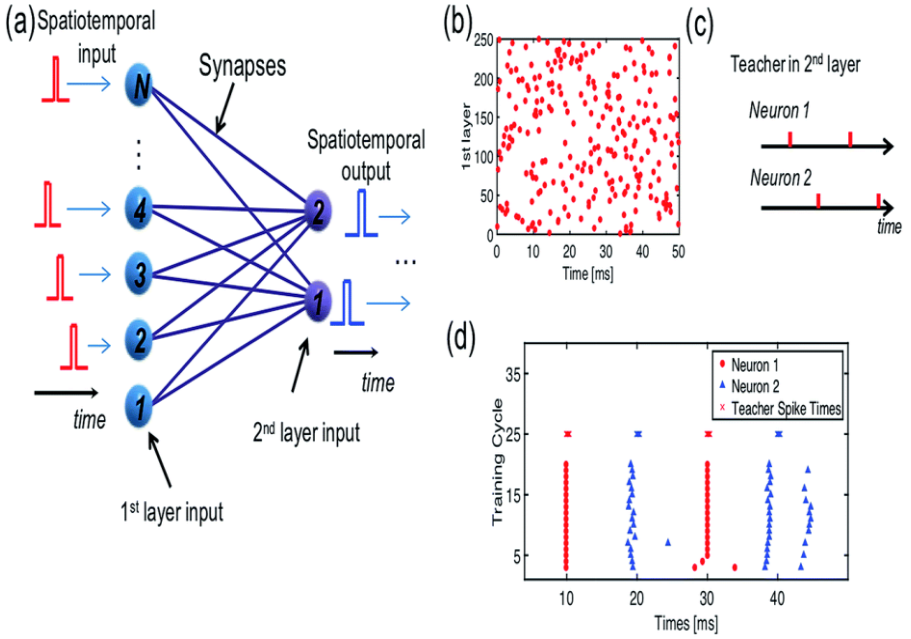
Algorithms of the brain. Human brains are dynamic, reconfigurable systems composed of neurons that are interconnected through synaptic connections[12]. For physicists, it exhibits many fascinating phenomena such as energy minimization or entropy minimization, phase transitions, criticality, self-oscillation chaos, synchronization, stochastic resonance, and many more[1]. Since the beginning, physicists have participated in theoretical efforts to explain the brain's algorithms, making contributions to both the fields of computer science and computational neuroscience. Statistical physics, nonlinear dynamics, and complex systems theory helped

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:11/November-2021 Impact Factor- 6.752 www.irjmets.com

cast light on neural mechanisms that allow learning. Hopfield networks and Boltzmann machines, which inherit from Ising spin systems, are the most famous ones, but many others have been proposed, exploiting, in particular, nonlinear dynamics for computing[3,8,10].

II. PHYSICS FOR NEUROMORPHIC COMPUTING

Neuromorphic computing influences the brain to create energy-efficient hardware for information processing, capable of highly sophisticated tasks. Systems built with regular electronics obtain gains in pace and energy by mimicking the dispersed topology of the brain. Scaling up such arrangements and improving their energy usage, performance, and speed by several orders of magnitude requires a revolution of hardware. Neuromorphic computing could be greatly enhanced by incorporating more physics in algorithms and nanoscale materials. We review impressive results that leverage physics to improve the computing capabilities of artificial neural networks, using resistive switching elements, photonics, spintronics, and other technologies. We study the pathways that could affect these approaches to maturity towards low-power, miniaturized chips that could infer and acquire in real-time.



The performance conditions of NVM devices for neuro-inspired computing chips are largely dependent on particular systems and applications. The number of analog states determines the weight tuning precision. It has been reported that precision of at least eight equivalent bits is required for training a relatively big neural network45, such as ResNet46.

Current electronics are not enough. In the brain, neurons which can roughly be viewed as performing processing -- have straight access to memory, supported by synapses. Current electronics, on the contrary, intrinsically separate memory and computing into discrete physical units, between which data must be transferred back and forth. This "von Neumann bottleneck" is an issue for artificial intelligence algorithms, which require reading substantial amounts of data at every step, performing complex operations on this data, and then writing the results back to memory. It slows down computing and considerably enhances the energy loss for learning and inference.

The standard model in neuromorphic computing is, therefore, to take inspiration from the topology of the brain to build circuits formed of physical neurons interconnected by physical synapses that perform memory in-situ, in a non-volatile way, thus drastically cutting the need to move data around the circuit and providing huge gains in speed and energy efficiency. This is unfortunately complicated by using Complementary Metal Oxide Semiconductor (CMOS) technology alone. Dozens of transistors are needed to imitate each neuron, and additional outside memories are required to execute synapses. CMOS-based artificial neurons and synapses are typically several micrometers wide. The number of physical neurons and synapses that can be blended into a CMOS chip is inherently limited by the chip area. This is problematic because the performance of neural

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:11/November-2021 Impact Factor- 6.752 www.irjmets.com

networks increases with the number of neurons and synapses: ideal image recognition algorithms today involve millions of neurons and synapses on an average. Large numbers of neurons and synapses can be achieved by gathering chips. In addition, the whole system becomes heavy, and much energy is wasted in the interconnects. Nanodevices that can imitate important features of neurons and synapses at the nanoscale, such as non-linearity, memory, and learning, are required to build low-power chips comprising several millions of neurons and synapses.

Finally, it is difficult to achieve a high degree of interconnection between neurons using CMOS technology only. The brain highlights an average of 10,000 synapses per neuron. Such connectivity is impossible to reproduce with current electronics. CMOS technology is mostly confined to two dimensions (2D), fanout is limited, and it is difficult to efficiently and fairly supply energy to components in the circuit. On the contrary, the brain is tri-dimensional (3D), neuron axons and dendrites provide high fan-in/fan-out, and blood efficiently distributes energy to the entire system.

III. SPIKING NEURAL UNIT (SNN)

In spiking neurons, the input is filtered in some way, usually a low pass filter, and they fire when a state variable exceeds a threshold. The spike train is calculated using Dirac delta functions where tk is the spike time. We wish to express learning in an SNN as minimization of a loss function across an extensive number of training samples, similar to old machine learning. A learning process includes finding sets of synaptic weights that allow scattered representation to be performed as well as decreasing the sum of all scattered coding losses in the sparse coding case. Learning in an SNN naturally proceeds online, where training samples are sent to the network sequentially.

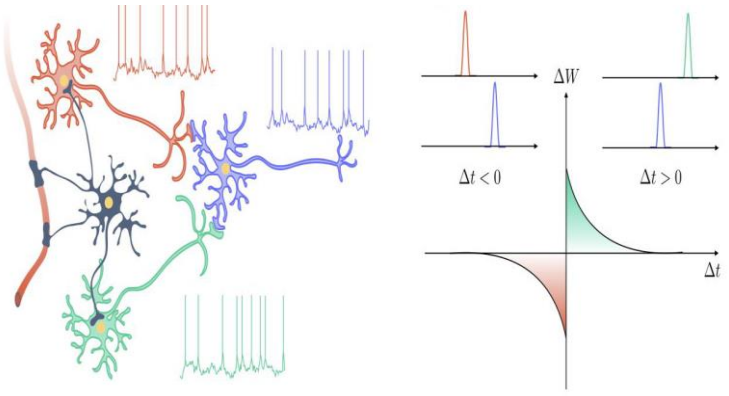
$$\dot{v}_i(t) = -\frac{1}{\tau_v} v_i(t) + u_i(t) - \theta_i \sigma_i(t)$$

$$u_{i(t)} = \sum_{j \neq i} w_{i,j} (\alpha_u * \sigma_j)(t) + b_i$$

IV. NEUROSCIENCE CONCEPTS AND HARDWARE

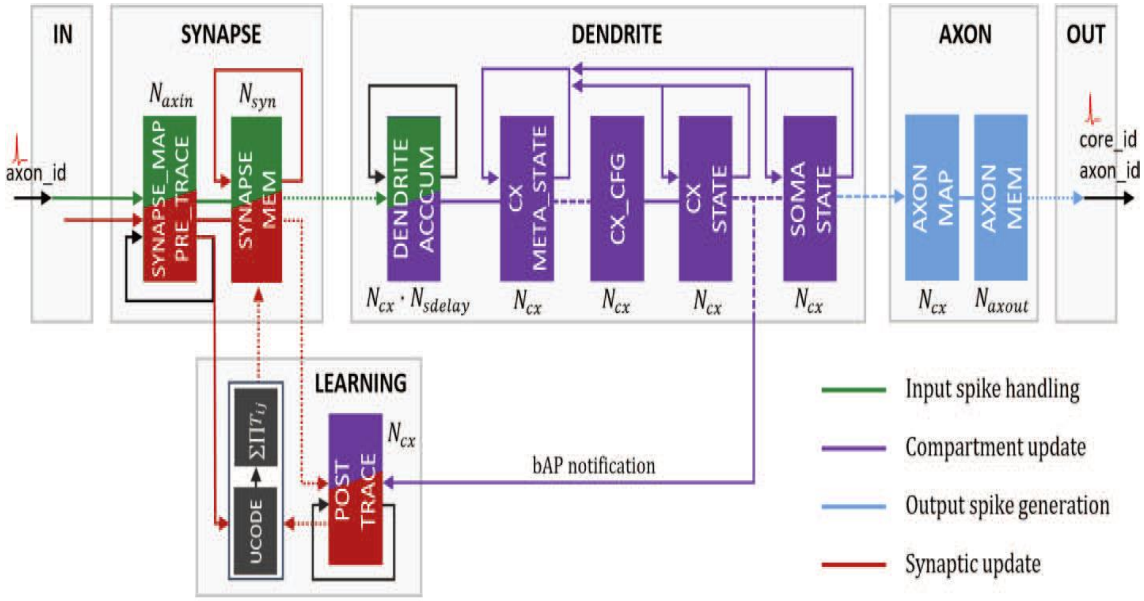
Artificial Intelligence algorithms cannot compete with the complexity of the brain. Biological neurons are more than nonlinear functions. A spike is leaky, feature memory is stochastic, and it can oscillate and synchronize. Several functional slots integrate signals coming from different areas over a space that is spatially extended.

Biological synapses are more than analog weights. Some synapses transmit only a fraction of the spikes they receive, so they can be extremely stochastic. Often overlooked brain components play a critical role and are important to imitate: dendrites appear to be able to execute very complex computations, whereas astrocytes are involved in neuronal regulation. Artificial neural networks can benefit from all these attributes and thus are more appealing to implement.



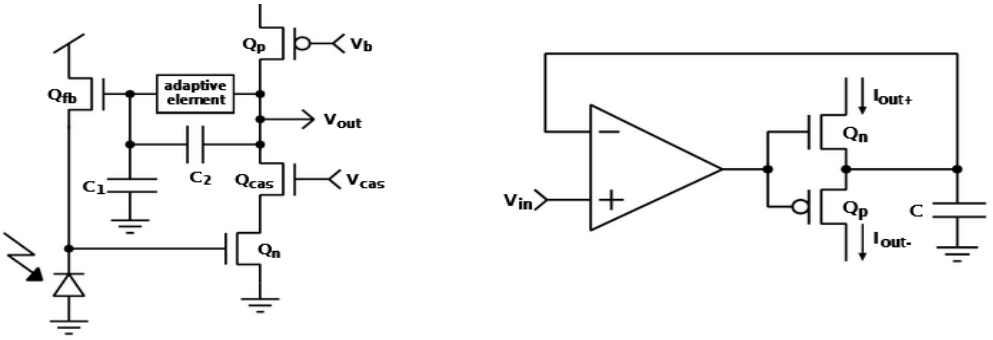
Through the related physical effects and intrinsic qualities of materials, all of these ideas have great potential if they can be realized at very low energy. The purpose of this research field was to exploit the exponential dependence of transistor leakage current on voltage. In the last decade, a huge variety of other physical phenomena, depicted, have been used to imitate interesting qualities of synapses and neurons.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:11/November-2021 Impact Factor- 6.752 www.irjmets.com



V. LOGIC GATES AND APPLICATIONS

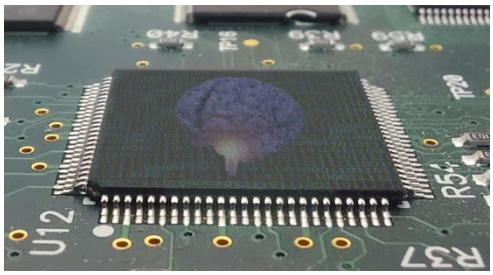
A human brain performs synaptic operations every second. Therefore, it has an efficiency of about 3x1014 operations per joule. Despite the slow and noisy parts in our heads and bodies, the human brain can perform complex computations in real-time. In comparison with the computing efficiency of a digital microprocessor, it can be observed that the brain is no less than seven orders of magnitude more systematic . Additionally, neural networks, based on the real brain, have a great deal of connectivity, which is expected to contribute to the high efficiency of biological systems due to their fault tolerance. The neuron achieves the equivalent of a logical OR operation on the excitatory inputs. By interpreting pulse behavior as a logical value of 1, we can realize how the OR gate functions by using neurons with two excitatory inputs and the output feedback as a piece of inhibitory information. If the excitation ceases, the neuron returns to its comfortable state, which corresponds to a logical value of '0'.



VI. NEURO-INSPIRED CHIP

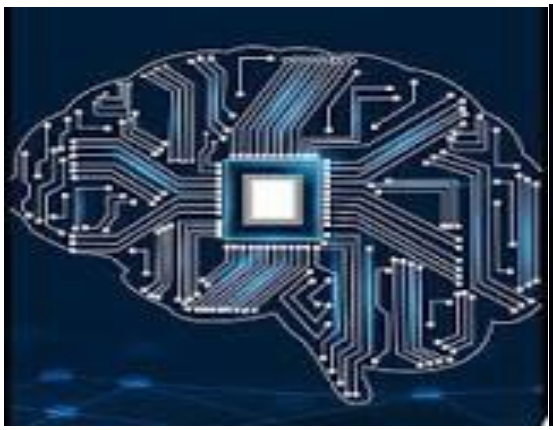
Artificial synapses and neurons are bound to neuromorphic chips, which simulate the action spikes within the human brain. These chips handle all of this processing on their own. This appears in smarter, far more energy-efficient computing systems. An energy-efficient approach to AI computing workloads could be provided by neuro-inspired computing chips combine features derived from neural systems. We analyze four key metrics for benchmarking neuro-inspired computing chips, including computing density, energy efficiency, computing accuracy, and on-chip learning capability, and discuss co-design principles, from the device to the algorithm level, for neuro-based computing chips using non-volatile memory.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:11/November-2021 Impact Factor- 6.752 www.irjmets.com



VII. NEUROMORPHIC SENSORS

The concept of neuromorphic systems can be extended to sensors. A neuromorphic camera is one example. Researchers are rapidly creating different variations of neuromorphic designs, as they did in the early days of neural networks. As a result, the boundary between research tools for neuromorphic computing and tools for industrial adoption is still blurry from a practical perspective. Generally speaking, neuromorphic hardware can be divided into digital, analog, and mixed-signal circuits. Several architectures have been proposed for performing neurons and synapses in hardware. Digital circuitry is most commonly used to implement neuromorphic architectures. Among the main advantages of this type of circuitry is the ease of development, low power dissipation, and reusability. The main reason for the popularity of digital neuromorphic architectures over their analog counterparts is the low development cost. An example is the Digital Neural Array (DNA), which is an array of digital neurons on a large scale. DNAs target both FPGAs and ASICs. FPGAs, for example, allow reprogramming, while ASICs offer higher density and more reliable performance despite a lower level of flexibility. Analog circuitry is less expressed than digital architectures, in spite of its greater suitability for designing neuromorphic systems. The physical properties of analog implementations are similar to those of neuromorphic architectures and, as with SNNs, these architectures are robust to noise, making them an ideal hardware implementation. Such physical characteristics include reliability and asynchronous operation. The Field-Programmable Analog Array (FPAA) is the best device for analog circuits. The Field-Programmable Neural Array, a custom design aimed at neuromorphic applications, uses programmable components to mimic neuron and synapse functions.



VIII. NEURO-INSPIRED COMPUTING USING PCRAM TECHNOLOGY

A dissection of BioNN reveals that the basic tasks of neuro-inspired computing are to replicate fundamental synapses, neurons, and their synaptic behaviors using hardware technology. Throughout the decades, there have been many implementations of neuro-inspired computing using PCRAM technology. Here, we present comprehensive discussions about how PCRAM is utilized for neural-inspired computing, including the fundamental electric-induced conductance mechanism and advanced techniques to simulate biological components and behavior, as well as current state-of-the-art in intelligent applications built on PCRAM.

ELECTRIC-INDUCED CONDUCTANCE CONTROLLABILITY

Based on biological synaptic behavior, a key precondition for implementing BioNN using hardware is the continuous regulation of conductance. The feasibility of PCRAM originates from the electric-induced controllable phase change that occurred at the active region. By applying appropriate electrical stimulation to

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:11/November-2021 Impact Factor- 6.752 www.irjmets.com

the active area, crystallinity and amorphous states can be controlled bilaterally and continuously. The process of crystalline-to-amorphous change accompanied by prompt quenching is triggered by high and narrow pulses, which is akin to synaptic depression. Additionally, the reverse process of mimicking synaptic potentiation involves long and moderate heating, which typically involves applying a lower and wider pulse. PCRAM can therefore be reliably controlled with special programming pulses in a bidirectional manner, i.e., multilevel reduction in conductance and cumulative enhancement in conductance.

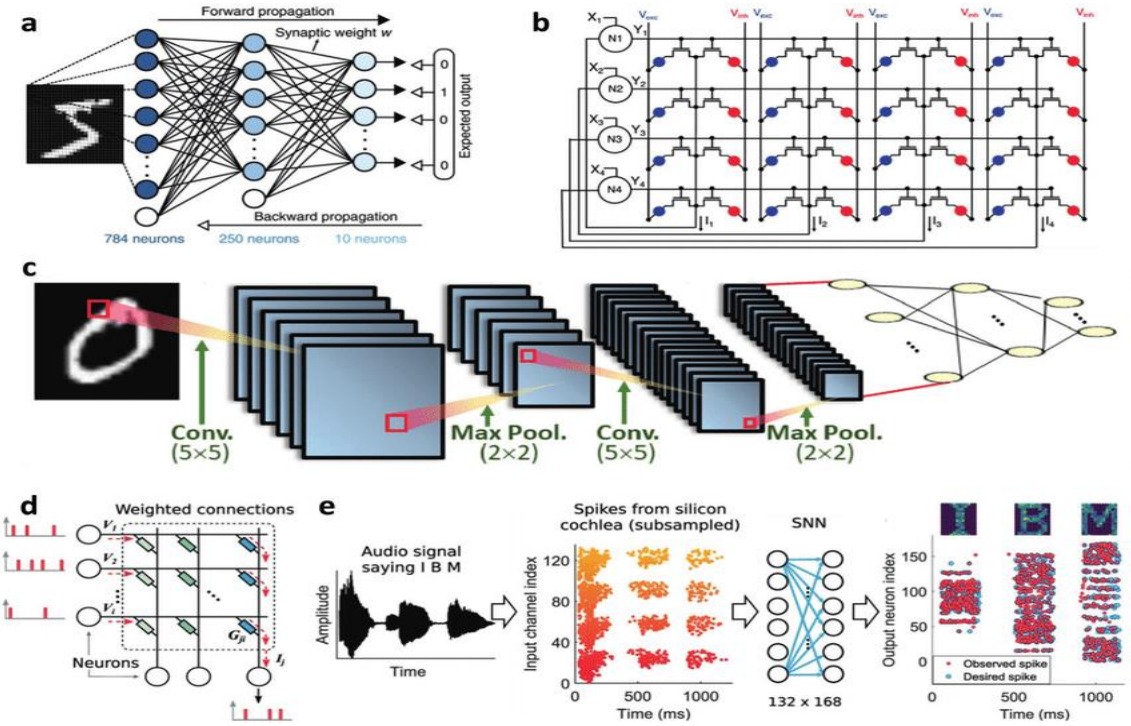
In order to holp up a wide range of workloads, SNN connectivity needs to be flexible and well-provisioned. Some networks may call for dense, all-to-all connectivity while others may call for sparse connectivity; some may have uniform graph degree patterns, others power-law distributions, some may require precision synaptic weights, for example, to support learning, while others can be done with binary connections. As a network grows, algorithm performance scales along with it, as revealed by neuron counts as well as neuron-to-neuron fanout degrees. Biologically, this rule holds true. Based on the O(N2) scaling connectivity state in many fan outs, it becomes an enormous challenge to support networks with high levels of connectivity today using integrated circuit technology.

To address the challenge, it supports a range of features to relax the sometimes severe constraints that other neuromorphic designs had inflicted on the programmer:

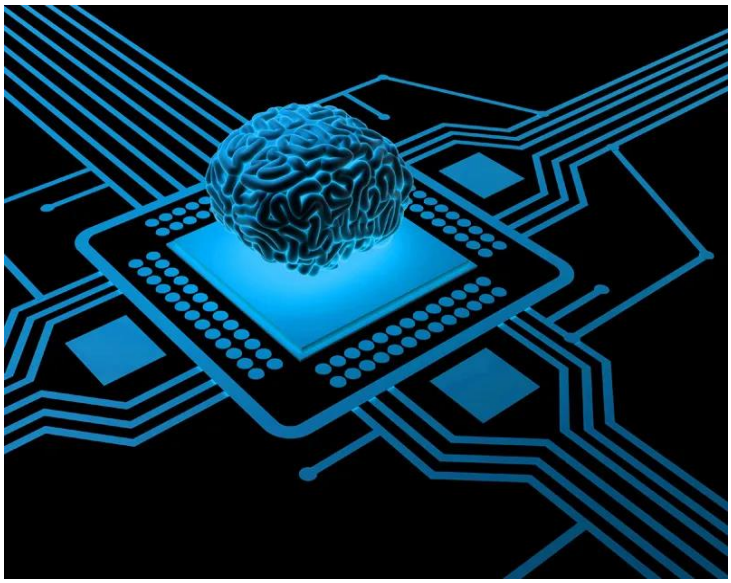
a. Sparse network concentration. In addition to thick matrix connectivity, it also supports three sparse matrix compression models in which neuron indices are calculated using index data stored with synapse state variables.
b. Core to core multicast. As the network connectivity requires, any neuron may send a unique spike to any number of objective cores.
c. Variable synaptic formats. Loihi supports any weight precision between one and nine bits, signed and unsigned, and there is no restriction on how the weight accuracy is set.

a. PCRAM Neuron

The LIF mechanism in BioNN dominates information transmission through the neuron. According to the Hodgkin–Huxley model and various threshold-based neuronal models, the LIF mechanism is driven by complex electrochemical mechanisms combined with the bilayer structure of the membrane. Postsynaptic potentials from dendrites modulate membrane potential stochastically. The sum of potentials above the firing threshold causes neurons to export new potential via axons.



e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:11/November-2021 Impact Factor- 6.752 www.irjmets.com



IX. CONCLUSION

Artificial intelligence is advancing rapidly throughout the world. We will have exponentially more devices connected to the cloud by the end of 2030 than we do now. In addition to more devices on the cloud, there are more data traffic and more hardware requirements. In spite of the current devices, data traffic remains high, and a considerable amount of power and resources are being consumed. Further, it is concerning to know that support structures of information technology are not growing as fast as the technology itself, and we may find ourselves in a situation sooner rather than later when no amount of hardware or software can handle the amounts of data that smart devices will generate. These problems can be solved by neuromorphic computing, which handles much larger volumes of data with drastically reduced energy consumption for the same. Artificial synapse is the groundbreaking technology that makes neuromorphic computing a reality. The artificial synapse is still being researched to fully understand and utilize its benefits. Memristors are the best examples and proof of the level of advancement that we have achieved in the recent past to take us further into the world of artificial intelligence. Although neuromorphic computing has physical and technical limitations and challenges, it is growing at a significant rate and is expected to change computing for good within the next few years.

X. REFERENCES

[1] Zhang, W., Gao, B., Tang, J. et al. Neuro-inspired computing chips. Nat Electron 3, 371–382 (2020). https://doi.org/10.1038/s41928-020-0435-7

[2] Wang, Q.; Niu, G.; Ren, W.; Wang, R.; Chen, X.; Li, X.; Ye, Z.; Xie, Y.; Song, S.; Song, Z. Phase change random access memory forneuro-inspired computing.Adv. Electron. Mater.2021, 2001241.

[3] Wang, P., Yu, S. Ferroelectric devices and circuits for neuro-inspired computing. MRS Communications 10, 538–548 (2020). https://doi.org/10.1557/mrc.2020.71

[4] ellamraju, S., Kumari, S., Girolkar, S., Chourasia, S., Tete, A.D.:Design of various logic gates in neural networks. In: Annual IEEEIndia Conference (INDICON), pp. 1–5 (2013)

[5] Maas, W.: Networks of spiking neurons: the third generation ofneural network models. Neural Netw.10, 1659–1671 (1997)

[6] Danijela Markovic,Alice Mizrahi,Damien Querlioz,Julie Grollier:

[7] https://www.researchgate.net/publication/339840879_Physics_for_Neuromorphic_Computing

[8] Buesing et al.(2011) Neural Dynamics_as_Sampling_A_Model_for_Stochastic Computation_in_Recurrent Networks_of Spiking Neurons

[9] Yellamraju, S., Kumari, S., Girolkar, S., Chourasia, S., Tete, A.D.: Design of various logic gates in neural networks. In: Annual IEEE India Conference (INDICON), pp. 1–5 (2013)

[10] Zhang, Y.; Qu, P.; Ji, Y.; Zhang, W.; Gao, G.; Wang, G.; Song, S.; Li, G.; Chen, W.; Zheng, W.; et al. A system hierarchy forbrain-inspired computing.Nature2020,586, 378–384

[11] S. Yu, "Neuro-inspired computing with emerging nonvolatile memorys," in Proceedings of the IEEE, vol. 106, no. 2, pp. 260-285, Feb. 2018, doi: 10.1109/JPROC.2018.2790840.

[12] Z. Yu, A. M. Abdulghani, A. Zahid, H. Heidari, M. A. Imran and Q. H. Abbasi, "An Overview of Neuromorphic Computing for Artificial Intelligence Enabled Hardware-Based Hopfield Neural Network," in IEEE Access, vol. 8, pp. 67085-67099, 2020, doi: 10.1109/ACCESS.2020.2985839.

[13] X. Si, J. Chen, Y. Tu, W. Huang, J. Wang, Y. Chiu, W. Wei, S. Wu, X. Sun, R. Liu, S. Yu, R. Liu, C. Hsieh, K. Tang, Q. Li, and M. Chang: A twin-8T SRAM computation-in-memory macro for multiple-bit CNN-based machine learning. In 2019 IEEE International Solid- State Circuits Conference — (ISSCC), San Francisco, CA, USA, 2019, pp. 396–398.

[14] Ao P, Wu H, Gao B, Tang J, Zhang Q, Zhang W, Yang J J and Qian H 2020 Fully hardware-implemented memristor convolutional neural network Nature

[15] S. Schmitt, J. Klaehn, G. Bellec, A. Gruebl, M. Guettler, A. Hartel, S. Hartmann, D. Husmann, K. Husmann, V. Karasenko, in 2017 Int. Joint Conf. on Neural Networks (IJCNN), IEEE, Piscataway, NJ 2017.