Evaluating the position of neuromorphic computing in a von Neumann dominated industry

David Song, *Member, DFEC,*

***Abstract*—The advent of conventional computing is considered to be the most influential technological revolution in the past century, and although it has streamlined computation and data processing more than the early analog machines, the need for energy-efficient and faster computer architecture to process large amounts of data and solve specific problems with a massive dataset. The revival of artificial intelligence and machine learning demands a new architecture that an ideal Moore's Law timeline will never satisfy. Neuromorphic computing is a new idea with old roots in analog computing that could realize these advancements in computer science/data processing, but as a young area of study, there are no large scale data on neuromorphic computing's impact on the world. To do this, the current von Neumann architecture must be compared to the promises of neuromorphic computing from its circuit elements to its best applications.**

***Index Terms*—Artificial Intelligence, Computer Architecture, Neuromorphic Computing, von Neumann architecture**

I. Introduction

NO technology has more impact on our daily lives than the computer. After decades of unprecedented technical growth, the computing industry finds itself at a crossroads. Down one path, conventional digital computing based on the von Neumann architecture that boasts versatility and reliability; a developed multi-trillion dollar industry with a history of successes and improvements. Conventional computing has improved so rapidly since its explosive beginning that it is now wrestling with the issue of diminishing returns and atomic size limitations. On the other path lies neuromorphic computing; a computational framework riddled with uncertainty but promises accelerated change and the potential to make its mark on the world as a revolution. To determine if the disruptive innovation of neuromorphic computing is worth the challenges and roadblocks that lie ahead, this paper dives into the details of neuromorphic computing. How developed is neuromorphic computing today, and what is the status of its current development point to in the future? What prevents further development, and how will future development make a difference in the world? Many top chip manufacturing companies are also beginning to research and deploy neuromorphic computing because of its potential, but the most exciting developments seem to be from small yet innovative chip developers. At the moment, its applications have been hybrid processors, such as Apple's Neural Engine chip. It is based on conventional architecture but has neuromorphic components that excel in image processing and facial recognition. Neuromorphic computing is already proving its reach and possibility to influence the way people and organizations interact with their data and information but has yet to definitively prove itself in terms of widespread commercial viability.

Neuromorphic computing does not have a deep track record of the positive impact that conventional computing has provided over the past 50 years. There is substantial theorizing, design, and hardware implementation that pioneering engineers and scientists must develop before the general public begins to see the benefit of this technology. It is only natural for this unconventional computing paradigm to have more inertia than conventional computing because neuromorphic computing still requires inventing new technology. On the contrary, conventional computing can focus on improving upon existing products. This results in neuromorphic computing being at a competitive disadvantage to the von Neumann architecture in many tasks that the general public believes a computer should be able to accomplish right now. The von Neumann architecture excels in solving linear logic and mathematical algorithms compared to the current neuromorphic computer developments. Out of context, pursuing this technology is too niche for the resources the development process demands. Now, imagine asking the computer to solve those same problems with voice commands. Currently, conventional computing is better at calculating numbers quickly, but neuromorphic computing can create a completely different user experience that could be preferable to what conventional computing can produce. In addition to this, von Neumann architecture is approaching a plateau in growth, signified by bottlenecks such as the near-stagnation of Moore's Law, quantum effects, heat dissipation limitations, and the throughput bottleneck. Some entities in the industry decided to veer off the conventional path to seek greater rewards coupled with elevated risks.

II. Background

A. Analog vs. Digital computing

To explain neuromorphic computing, one must start with the framework of analog computers compared to digital computers. We live in the analog world, meaning there are an infinite amount of colors, tones, tastes, and smells even though our bodies cannot discern all of them. There is a continuous range; therefore, an infinite number of values for analog signals. Electronic analog computers use voltage and current modified by circuit elements such as resistors and capacitors as inputs and outputs. For example, to create an analog adder, one can connect two wires with electrical currents together into one resulting wire of a summed current, abiding by the principle of the conservation of electrical charge. On the other hand, digital signals have a finite set of values, made up of combinations of low voltages (0's) and high voltages (1's). Figure 1 shows a full adder to show the complexity of adding two 1-bit numbers together. Arithmetic is not the extent of analog computers. Analog computers can solve partial differential equations and have even helped the lunar lander get to the moon. It is faster and more energy-efficient than digital computers. Analog systems' ability to handle continuous data, which would take an infinite amount of processing power on a digital computer, is another benefit that analog computing holds over its digital counterpart. However, the infinite possibilities can become an issue for users because two outputs of the same inputs can never be the same for many reasons: human error, temperature, and wear. On top of this, no two analog computers of the same design will behave the same because of the variability of parts such as resistor values. Although the variability will not change the outputs drastically, analog computing is not precise. An analog computer cannot run the Microsoft Suite because it is fundamentally single-purpose. It is merely designed for a certain set of tasks and cannot go outside of its designed workspace. In the early history of computing architecture, these factors made analog computers fall to the wayside while digital computers' consistency and general-purpose gained the favor of the common consumer. The applications of computing systems were focused on precise outputs of ambiguous tasks (consider an accounting spreadsheet). Now, analog computers and the applications they are well suited for may be making a comeback.

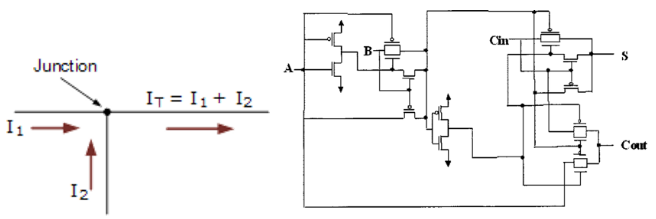

Fig. 1: Addition circuit in analog vs. digital [1][13]

Artificial Intelligence and Machine Learning (AI & ML) are emerging technologies that caught the world's attention not only because of science fiction and humanity's dreams of creating a human-like entity but also their vast range of applications in reality. Figure 2 shows the explosion in interest in AI can be seen by the number of patents filed in the past decade. It is clear that AI research and development is developing at a breakneck pace, but it raises the question of why it is only booming now.

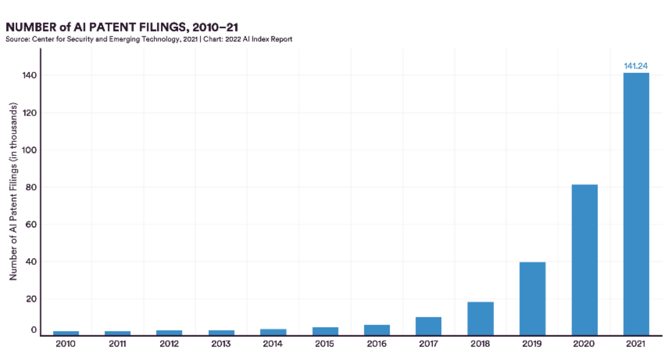

Fig. 2: Number of AI Patent Filings, 2010-2021 [24]

B. The First Artificial Neural Network: The Perceptron

The fundamental idea for a computer that mimics the operation of the brain first originated in the psychology discipline in 1958 but took over half a century before it resurged within the burgeoning computing industry. The perceptron was invented by Frank Rosenblatt at Cornell with the vision that computers would soon be able to see and understand language and become human-like. The perceptron's proof of concept is considered to be the foundation of modern neural networks. Simply stated, biological neurons are cells that take an electrical brain signal as an input and output another electrical brain signal [11]. Much like a digital signal, the neuron either spikes (turns on) or does nothing (stays off). Input signals of neurons come from multiple other neurons' outputs, passing through synapses that assign a "weight" to that particular input. All of those inputs are accumulated in the cell body, and if the amalgamation exceeds a threshold the neuron fires. Any signal below the threshold results in nothing. Having a basic understanding of neuron operation and structure in Figure 3 can aid in understanding why Rosenblatt designed the perceptron the way he did.

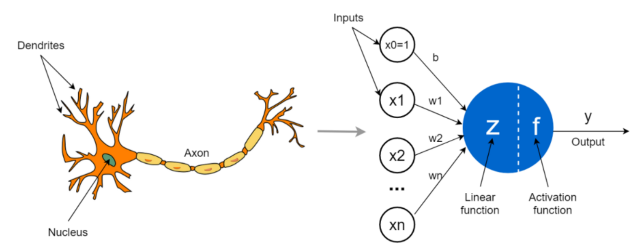

Fig. 3: Biological Neuron besides a single layer perceptron [22]

D. Rosenblatt's perceptron is more formally known as a photo-perceptron, which was intended to model an eye and its corresponding part of the brain to process images. The photo-perceptron's task was to categorize 2 different shapes. It receives stimuli from the sensory unit, called S-points where an all or nothing (1 or 0) response is assumed. Those signals are then transmitted to a set of association cells (A-units), where either excitatory (positive) or inhibitory (negative) weights augment the S-points. If the algebraic sum of the excitatory or inhibitory impulses/spikes is greater than a threshold, the response will fire, again in an all-or-nothing fashion. In the photo-perceptron, the system would become simplified to two responses for each of the two shapes. The benefit of this simplification is that responses of this organization are mutually exclusive, so if R1 occurs, it tends to inhibit R2 and visa-versa. The math is simple; the activations of each S-point are multiplied by their corresponding weights and then summed to complete the vector dot product and finally compared to a fixed value, resulting in a one or a zero. However, a single frame can become computationally intensive. As seen in (1), the calculations happening in this 2-response perceptron are basic addition and multiplications, but the number of operations can theoretically be scaled up to a greater number of S-points and A-units, increasing the resolution of the image captured and the precision of the association. Something to note is that the activations are not compared to an exact value, but check if it is greater than a set threshold value.

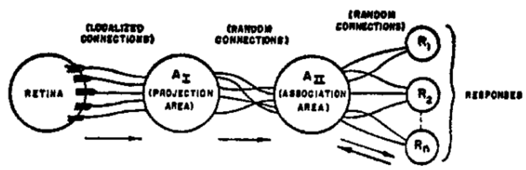

Fig. 4: General structure of the photo-perceptron [11]

$$x = \begin{cases} 1, & z \le \Theta \\ 0, & z > \Theta \end{cases}, \; z \equiv \begin{bmatrix} s_1 \\ s_2 \\ \vdots \end{bmatrix} \cdot \begin{bmatrix} w_1 & w_2 & \cdots \end{bmatrix}, \quad \begin{matrix} \theta \equiv set\ threshold\ value \\ s \equiv activation \\ w \equiv weig \end{matrix} \tag{1}$$

For example, if the perceptron's job is to discern between a circle and a square, the desired behavior of the output neuron is to fire when the S-points see a circle and not fire when it sees a square. First, Rosenblatt set all the weights to zero and started feeding the perceptron shapes to categorize. In our example case, if the output neuron correctly fires on a circle picture or doesn't fire when it sees a square, the weights remain unchanged. On the contrary, if the neuron does not fire when it should have, add the input activation to the weights for each corresponding S-point, and subtract the input activation when the neuron fires when it should not have. Then, repeat that process until the perceptron correctly identifies every image in the training dataset. The correction equation is shown in (2).

$$w_{new} = w_{old} + (t - x)s, \;\; t \equiv target\ output \tag{2}$$

Rosenblatt had high hopes for the perceptron and predicted that it would be capable of temporal pattern recognition and spatial recognition [11]. In the end, the perceptron did not live up to the media's expectations of creating a human-like machine, and the applications of the perceptron, such as computer vision demanded impossible computing capabilities in the 1960s [19]. Although Rosenblatt set the foundations of the artificial neural networks (ANN) we are using and researching today, Dr. Marvin Minsky and Dr. Seymour Papert criticized the lack of mathematical analysis of the perceptron and did not make the perceptron look like it held much promise within their analysis. This led to what was called the first AI winter. Many blame the 2 Massachusetts Institute of Technology professors for setting the AI/ANN research back for decades, but they indeed highlighted fatal flaws that needed to be addressed to make neural networks more powerful and useful. At the time, application developers could not work with the perceptron because of its limited capability and lack of general-purpose usage. Although Minsky and Papert's book stopped AI and ANNs in their track, the ANN's lacking utility ultimately led to the inhibition of adoption from the top consumers for computing systems. At this point, the Electronic Numerical Integrator and Computer (ENIAC), the first programmable, electronic, general-purpose computer has been completing more calculations in 15 years than in the entire span of humanity. During the first AI winter, Moore's Law was coined in 1965 and the home computer and video game consoles were being introduced to the market in the 1970s. It was clear that people valued the precision and reliability of digital computers over analog computers, much less the values of the perceptron.

C. ALVINN: the resurgence of AI

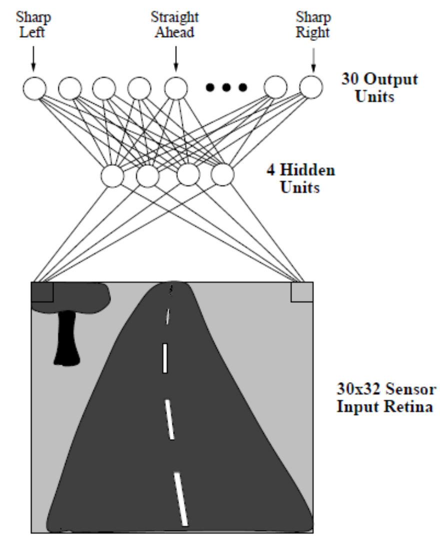

Fig. 5: ALVINN structure [6]

AI/ANN research's disappointing entrance into academia resulted in reduced funding and interest, slowing development to a near standstill in the field from the 1960s to the 1980s[6]. The underlying problem was still the lack of application in solving practical problems. In 1986, Ph.D. student Dean Pomerleau saw an opportunity to use the flexibility and uses of ANNs in domains characterized by high noise and variability. He developed ALVINN (Autonomous Land Vehicle In a Neural Network), an ANN designed to control the NAVLAB, a road-worthy truck designed by Carnegie Mellon University. Pomerleau stated "Autonomous navigation is a difficult problem for traditional vision and robotic techniques, primarily because of the noise and variability associated with real-world scenes. Autonomous navigation systems based on traditional image processing and pattern recognition techniques often perform well under certain conditions but have problems with others", marking one of the first times that ANNs were applied to make up for the shortcomings of conventional methods [6]. ALVINN's structure is similar to Rosenblatt's perceptron, but ALVINN has a hidden layer of neurons, another layer of neurons between the input and the output. ALVINN's input was a 30x32 image of the upcoming road, each of the input neurons was connected to a hidden layer of 4 neurons with variable weights. The output layer of 30 neurons dictated which way the NAVLAB turned. Since there is an extra layer, instead of a vector dot product, matrix multiplication is performed to go from one layer to the next as seen in (3). One can see how even just one more layer or neuron increases the number of computations because of neuromorphic computing's parallel nature.

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} * \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & w_{n3} & w_{n4} \end{bmatrix} * \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ w_{31} & w_{32} & \cdots & w_{3m} \\ w_{41} & w_{42} & \cdots & w_{4m} \end{bmatrix} \tag{3}$$

$n = 960, \quad m = 30$

D. ImageNet: Deepening ANNs

In Rosenblatt's perceptron training, Rosenblatt reset all the weights and manually adjusted the weights for each training example with one output unit. That may be feasible for a one output unit system but is practically impossible for a large NN. Backpropagation is the algorithm that determines how to alter weights and bias for a particular training example to cause the greatest decrease in cost. Cost functions are used to calculate the difference between the expected result and the output of the neural network, and they were a significant algorithmic discovery for AI/ANN systems. While ALVINN proved the power of backpropagation, AI still struggled to complete simple categorizing tasks. No one could prove whether the bottleneck lay in the hardware or the software, but in 2006 Dr. Fei-Fei Li hypothesized a different root cause to the limitations of modern AI/ANN systems and addressed it with ImageNet.

ImageNet is a database of 1.2 million high-resolution human-labeled images that Dr. Fei-Fei Li, professor of computer science at Stanford University, created because she asked if the shortcomings of ANN were not from the software or the hardware but the amount of training data received. From 2010 to 2017, ImageNet ran a contest called the ImageNet Large Scale Visual Recognition Challenge, where people created software to correctly detect and classify images from 1,000 different categories. The performance metric used to pick the winner was the top-5 error rate, which takes the top 5 output neuron activations and measures how frequently the correct category was included in those neurons.

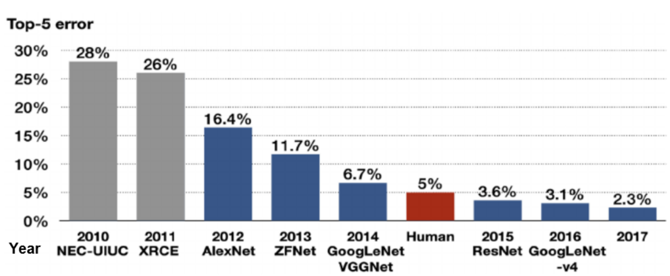

Fig. 6: Top 5 error rates of winning ILSVRC algorithms from the years 2010-2017
(VGGNet took 2nd place, but it is widely used in studies for its concise structure) [8]

In 2012, AlexNet took 1st place in a landslide victory, outperforming the 2nd place algorithm on 825 of the 1000 categories. We can also see the sharp drop in Top-5 error between 2011 and 2012, and what the University of Toronto team behind AlexNet did is increase the size/depth of the ANN. AlexNet was made of eight layers, 650,000 neurons, and 60 million parameters. This Convolutional Neural Network (CNN) is required in the realm of 700 million math operations for every single image, making the training extremely intensive. This was only possible because the team used graphical processing unit (GPU) hardware. GPUs are conventional von Neumann-based architectures designed for fast parallel computations, and their forte is solving repeatable problems with large amounts of data. Even then, 90 cycles through the 1.2 million image database took five to six days on two NVIDIA GTX 580 3GB GPUs. From 2012 onwards, the contestants followed AlexNet's lead and optimized their algorithms to 2.3% in 2017, better than human performance [14]. The trend is clear, to be practical, neural networks will demand increased size and depth for increased fidelity and performance.

TABLE 1
Tracking key indicators of the computing industry [2][4][10][20]

Year	Training cost ($)	# of parameters (M)	Feature size (nm)	Cap. Expenditure ($B)
2015	no data found	40	14	65.2
2016	no data found	83.6	10	67.8
2017	1112.64	86.1	10	95.6
2018	12.80	829	7	106.1
2019	19.00	829	7	102.5
2020	7.43	928	5	113.1
2021	4.59	7200	5	152

Table 1 shows several trends in the past 7 seven years of where the computing industry stands. The first column shows the lowest ImageNet training cost is representative of AI/ANN's increasing accessibility to anyone with the resources to train and utilize cutting-edge DNNs. These costs are from ImageNet algorithms that have at least 93% accuracy in ImageNet's ILSVRC. The minimum number of parameters were taken from a dataset of 577 ImageNet algorithms from 2014-2022, and the parameters listed had the highest Top 5 error rate of all the algorithms from that year. The number of parameters is indicative of the computational intensity of AI algorithms, but that is caveated by the fact that these algorithms are all image recognition and classification. The numbers are not as significant as the increasing trend. In the second half of the table are data pertaining to the status of conventional computing's development in past seven years. Feature size of any semiconductor technology is defined as the minimum length of the MOS transistor channel between the drain and the source or half the distance between cells in a dynamic RAM chip. This piece of information shows the current state of Moore's Law, and it is easy to see that the size is not halving every year anymore. Finally, the global semiconductor industry capital expenditure shows how much the industry is pouring into semiconductors to fabricate smaller and denser integrated chips. By comparing it to the feature size's trend and the highest transistor counts on a chip every year (not shown), one can only question if the capital can be better used in another technology.

III. Architecture

A. Architectural background of conventional computing with respect to solving AI Algorithms

Due to its historically desirable trait of effectively performing precise and repetitive calculations, it should be clear that the dominant computer architecture in today's age is the von Neumann architecture. John Louis von Neumann wrote the incomplete "First Draft of a Report on the EDVAC" in 1945, and it described a detailed design of a "very high speed automatic digital computing system" [18]. The key subsection that we are interested in is memory. Von Neumann recognized that "Any device which is to carry out long and complicated sequences of operations (specifically of calculations) must have a considerable memory" [18]. Staying true to his vision, von Neumann improved on the previous digital computer architecture by storing both the instruction set and data in the main memory, but he still failed to predict the issue of moving large amounts of data between the memory and the CPU (central processing unit).

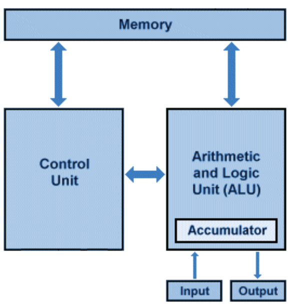

Fig. 7: von Neumann architecture [21]

Although the modularity, cost efficiency, and upgradability increased, the von Neumann architecture instructions are restricted to serial execution, making parallel computing essentially impossible at the machine code level. The architecture was not designed for a large amount of data, so every operation requires the CPU to fetch data from memory. For example, the 700 million operations to train AlexNet on one image could be three to four times as many data transfers on the bus between the main memory, CPU, and GPU. The volume of training data and using deep CNNs is very large; every weight, bias, and operation has to go through the bus, and even if the computation is fast, the transmission of the data is still restricting performance. This issue is called the von Neumann bottleneck, and it has been holding back conventional computing ever since the von Neumann architecture became the standard. In an attempt to push through the bottleneck, the computer industry focused efforts to increase the CPU's clock speed. However, by the early 2000s, the clock speed plateaued at 3GHz. In March 2022, Intel announced a new 5.5GHz processor called the Core i9-12900KS [23], but the rate of improvement still proves to have slowed tremendously from the 20th century.

Another cliff that conventional computing is approaching is the end of Moore's Law. Moore's Law states that the number of transistors on a chip doubles about every two years, which was first observed in 1965. Assuming a common form factor, Moore's Law implies that transistors were decreasing in size rapidly. Ever since Moore's Law was coined, the computing industry interpreted this as truth and did not think much about it until the progress started to slow down. More transistors on a single chip meant greater power consumption, and along with Dennard scaling, the electrical laws of the universe do not seem to be on humanity's side in creating more potent von Neumann-based computer chips. As more transistors were crammed on a chip, the heat from the circuit is more likely to destroy the silicon on the chip. Since the heat needs to escape the chip, cooling solutions are now required to keep up with the increased heat produced by the increasing density of transistors on chips. Dennard scaling is the principle that states that the dimensions of a device are proportional to its power consumption. This was great news to computer architects because that meant that as transistors became smaller, the power consumption and cost would follow the drop. By 2004, the transistor became so small that its operating voltages and currents were too small for the transistor to function reliably. This forces engineers to supply more power as transistors shrink. The industry shrank the transistor so much that Dennard's scaling did not apply anymore. Then, companies started putting multiple cores on a chip, but the transistor size and clock speed were lagging behind the number of cores on a chip. Developing solutions in the conventional computing world would be focused on power consumption because its computing power far exceeds human capability, but power consumption is what's restricting further growth in the conventional computing realm.

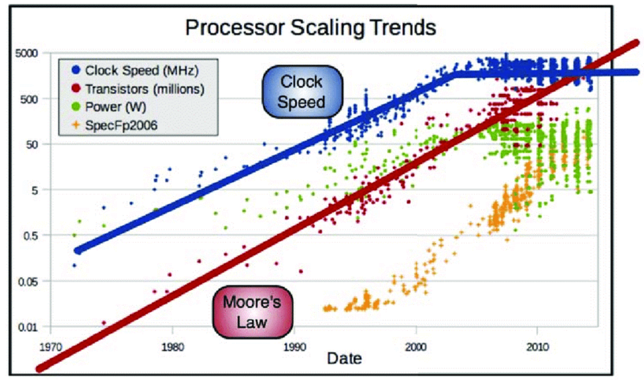

Fig. 8: CPU transistor densities, clock speeds power and performance from 1970-2015 [12]

B. Architectural background of neuromorphic computing with respect to solving AI Algorithms

Neuromorphic computing is a relatively new industry. It does not have the history and quantity of success stories or the amount of in-the-market examples that point to and provide a rapport for this new technology as does the von Neumann architecture. However, there have been efforts, primarily driven by new AI-specific applications in the past decade that show potential to increase the performance of solving AI.

Trends towards development and research in computer vision, language, speech, recommendation, reinforcement learning, and robotics push the AI industry further than Rosenblatt could have ever imagined.

Intel's flagship neuromorphic chip, the Loihi2 boasts 1 million neurons in a chip smaller than the size of a fingertip [15]. The Loihi was used in demonstrations such as Adaptive robot arm control, Visual-tactile sensory perception, Learning and recognizing new odors and gestures, and drone motor control with state-of-the-art latency in response to visual input. The power consumption on most of the demonstrations was said to be less than 1 watt, compared to the tens to hundreds that a conventional system of CPUs and GPUs consumes at extremely cold temperatures to mitigate heat problems [15]. The potential for mobile applications is apparent, and Apple is at the forefront of bringing neuromorphic computing to smartphones today. Apple's Neural Engine (ANE), is a hybrid modern neuromorphic architecture that was designed to make certain tasks like image processing and facial recognition more time and energy-efficient, preventing the iPhone from overheating on a conventional chip. The details are protected for proprietary reasons, but the biggest companies are not always the leaders in disruptive technology such as neuromorphic computing.

Mythic AI is a company that is credited with the first analog AI chip using in-memory analog computation. Using repurposed flash storage cells, Mythic does not use these cells as switches holding 1's and 0's, but as variable resistors [16]. Instead of putting a high voltage or no voltage on a cell's control gate, Mythic opted to carefully modulate the voltage applied to manipulate the number of electrons on each floating gate. The more electrons in the gate, the higher the resistance. Since flash memory cells are non-volatile, this allows those electrons to be trapped in the floating gate. Now, if one applies a small voltage to allow current to cross the channel, the current that flows is equal to or, where G is conductance, the reciprocal of resistance. The conductance is the scaling factor of that flash memory cell, allowing multiplication [26].

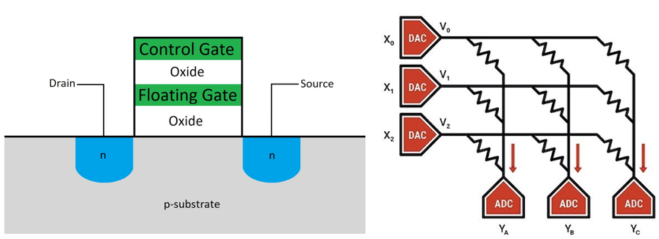

Fig. 9: Flash memory cell (Left) [9], Analog computing cell (Right) [5]

To create an analog matrix multiplying circuit in the application of running ANNs, the weights are written on the flash cells as their conductance. Then, input voltages are used as input activation values that are multiplied by the weights in the cells. The resulting currents are then accumulated at the end to complete the matrix multiplication. Mythic avoids the von Neumann bottleneck by embedding this process directly in the memory, and that results in the M1076 Analog Matrix Processor (Mythic AMP) topping out at 35 trillion operations per second with a 3W power profile [17]. Newer but similar technologies are being developed in 2022. For example, RRAM (Resistive Random Access Memory) is a non-charge-based device that can be used to store continuous data and carry out matrix multiplication and convolution more efficiently. Besides RRAM, spin-transfer-torque-magnetic random access memory (STT-MRAM), phase-change memory (PCM), conductive bridge random access memory (CBRAM), and optoelectronic RRAM (ORRAM) are emerging devices that are at the forefront of neuromorphic computer material studies [7].

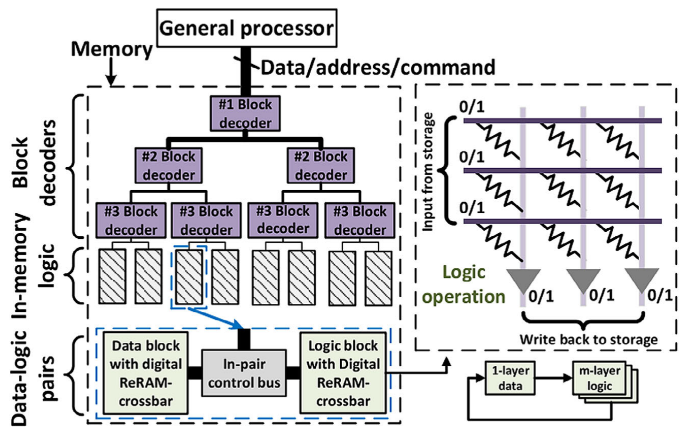

Fig. 10: Digital RRAM crossbar-based in-memory computing architecture [5]

IV. The Future of Neuromorphic Computing

The possibilities seem endless, but the future of neuromorphic computing holds many challenges. Some are similar to what conventional computing saw in its early stages, but others are unique to the neuromorphic architecture. Although, the energy efficiency of neuromorphic computers far exceeds that of conventional computing, its performance does not outperform deep learning approaches in terms of accuracy. This can incorrectly lead consumers to think that neuromorphic computing's low-energy expenditure is its only draw above conventional computing. Ever since backpropagation-based training became the standard, there has not been much diversification and leveraging of new architectures such as the one Mythic has developed. As more neuromorphic chips and computers become available, opportunities to explore on-chip training and learning will increase.

AI saw its first winter because of the lack of structure and analysis of AI systems. If the industry is not wary of this problem, neuromorphic computing may see another winter regarding the lack of benchmarking and performance metrics. Assessing the suitability of a system for a given algorithm or application will be crucial in selling products. If a company cannot convince consumers that their neuromorphic computer is better than another system, or if a potential buyer cannot discern the better neuromorphic system, chances are the potential buyer will remain as a potential buyer, leading the selling companies to a path to bankruptcy. The industry needs clearly established metrics and benchmarks to advertise its value and assess overall progress in this new field of study. Similar to conventional computing benchmarks, the difficulties lie in creating algorithms and programs that fully express the computer's full potential without narrowing the broader utility of the technology [Schuman]. Learning from the forefathers or conventional computing benchmarking, a possible solution to this problem is creating a suite of challenge problems that include both AI problems and non-AI use cases [3].

The complications do not end there. In Figure 10, a keen eye notices Analog to Digital Converters (ADC) and Digital ks Analog Converters. In the analog domain as mentioned previously, the output of an analog calculation is not repeatable and will differ from calculation to calculation. If one were to complete a sequence of many matrix multiplications using analog methods such as using Mythic's non-volatile memory cells, the answer will probably drift further and further away from the desired output at every stage of the inevitable distortion. To fix this, the analog output of the first processing block is converted to a digital signal and converted back to the analog domain pre-processing. The inherent lack of stability and repeatability haunts analog computers and for now, digital/analog hybrid computers seem to balance the two types of computing's fundamental deficiencies [26].

V. Conclusion

Neuromorphic computing has roots dating back to 1958 and Rosenblatt's perceptron, but the lack of application and capability left consumers disappointed. That led to the first AI winter where digital, general-purpose computers dominated the computing industry. However, AI saw a return with groundbreaking development such as the backpropagation training algorithm and a peek into the future of autonomous vehicles through ALVINN. Nonetheless, the established conventional computer's reliability and versatility were hard to give up for a rather unknown niche alternative. Usually, when a disruptive innovation presents itself to industries, the leading titans of those industries tend to ignore it. Kodak ignored the power of digital photography, Netflix essentially destroyed video rental companies, and Wikipedia rendered paper encyclopedias obsolete. In the case of neuromorphic computing, however, the leading companies in conventional computing are furiously competing with one another to bring about the new age of computing despite neuromorphic computing being a risky, unproven path. What changed was that the niche market of AI and ML algorithms transformed from a niche need for particular customers to a key emerging technology that these companies think has the potential to affect many aspects of life, from the high-level companies to the household. The future need for a better computer for AI algorithms came earlier than expected, and companies like Intel and Apple did not take any chances to miss this wave of public interest and investment. As with all emerging technologies, hindrances to progress are plentiful and must be addressed to sustain this innovation wave of neuromorphic computing. Neuromorphic computing's lack of established benchmarks and dependence on digital computing due to instability are a few of the many problems that the new computing method will face, but the plateau of Moore's Law and the accelerated growth of AI demand the computing industry's attention. The adoption of neuromorphic as a disruptive innovation has already begun, but only time will tell whether the investment and increased excitement will bear products that the general public will find value in.

Acknowledgment

Thanks to Dr. Heritsch for his continuous support throughout the production of this paper.

VI. References

[1] A. Shams, M. Bayoumi, J. Qian, and J. Wang, “Fig. 1. Commonly used 1-bit full-adder cells. (a) Conventional CMOS,” *ResearchGate*, 01-Jun-2000. [Online]. Available: https://www.researchgate.net/figure/Commonly-used-1-bit-full-adder-cells-a-Conventional-CMOS-full-adder-b-Transmission_fig1_3325506. [Accessed: 17-Apr-2022].

[2] “Capital spending in the semiconductor industry worldwide 2000-2021,” *Statista*. [Online]. Available: https://www.statista.com/statistics/864897/worldwide-capital-spending-in-the-semiconductor-industry/. [Accessed: 18-Apr-2022].

[3] C. D. Schuman, S. R. Kulkarni, M. Parsa, J. P. Mitchell, P. Date, and B. Kay, “Publisher Correction: Opportunities for neuromorphic computing algorithms and applications,” *Nat Comput Sci*, vol. 2, no. 3, pp. 205–205, 2022.

[4] “DAWNBench,” *Stanford.edu*. [Online]. Available: https://dawn.cs.stanford.edu/benchmark/ImageNet/train.html. [Accessed: 18-Apr-2022].

[5] D. Liu, H. Yu, and Y. Chai, “Low-power computing with neuromorphic engineering,” *Advanced Intelligent Systems*, vol. 3, no. 2, p. 2000150, 2021.

[6] D. Pomerleau, “Neural Network Vision for Robot Driving,” *CMU School of Computer Science*, 1995.

[7] D. V. Christensen *et al.*, “2022 roadmap on neuromorphic computing and engineering,” *Neuromorph. Comput. Eng.*, 2022.

[8] D.-Y. Kang, H. P. Duong, and J.-C. Park, “Application of deep learning in dentistry and implantology,” *Korean Acad. Oral Maxillofac. Implantol.*, vol. 24, no. 3, pp. 148–181, 2020.

[9] E. Europe, “Understanding Flash memory,” *EENewsEurope*, 18-Dec-2017. [Online]. Available: https://www.eenewseurope.com/en/understanding-flash-memory/. [Accessed: 18-Apr-2022].

[10] “Feature size,” *PCMAG*. [Online]. Available: https://www.pcmag.com/encyclopedia/term/feature-size. [Accessed: 18-Apr-2022].

[11] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain,” *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.

[12] G. A. Stewart, W. Lampl, and ATLAS Collaboration, “How to review 4 million lines of ATLAS code,” *J. Phys. Conf. Ser.*, vol. 898, p. 2, 2017.

[13] “https://www.electronics-tutorials.ws/dccircuits/kirchhoffs-current-law.html,” *Electronics-tutorials.ws*. [Online]. Available: https://www.electronics-tutorials.ws/dccircuits/kirchhoffs-current-law.html. [Accessed: 17-Apr-2022].

[14] “ImageNet Challenge 2012 Analysis,” *Image-net.org*. [Online]. Available: https://image-net.org/challenges/LSVRC/2012/analysis/. [Accessed: 18-Apr-2022].

[15] "Intel Labs' new Loihi 2 research chip outperforms its predecessor by up to 10x and comes with an open-source, community-driven neuromorphic computing framework Taking Neuromorphic Computing to the Next Level with Loihi 2," *Intel.com*. [Online]. Available: https://download.intel.com/newsroom/2021/new-technologies/neuromorphic-computing-loihi-2-brief.pdf. [Accessed: 18-Apr-2022].

[16] J. Hertz, "Mythic Redefines Edge AI by Combining Analog Processing and Flash Memory," *Allaboutcircuits.com*, 07-Jun-2021. [Online]. Available: https://www.allaboutcircuits.com/news/mythic-ai-redefines-edge-ai-by-combining-analog-processing-and-flash-memory/. [Accessed: 18-Apr-2022].

[17] K. Freund, "Mythic launches industry first analog AI chip," *Forbes*, 07-Jun-2021. [Online]. Available: https://www.forbes.com/sites/karlfreund/2021/06/07/mythic-launches-industry-first-analog-ai-chip/?sh=3776a0bc3e09. [Accessed: 17-Apr-2022].

[18] M. D. Godfrey, "First Draft Report on the EDVAC by John von Neumann," *IEEE Annals of the History of Computing*, vol. 15, no. 4, p. 27'43, Jan. 1993.

[19] M. Minsky and S. A. Papert, *Perceptrons: An introduction to computational geometry*. London, England: MIT Press, 2017.

[20] "Papers with code - ImageNet benchmark (image classification)," *Paperswithcode.com*. [Online]. Available: https://paperswithcode.com/sota/image-classification-on-imagenet?metric=Number%20of%20params. [Accessed: 18-Apr-2022].

[21] P. Lukas, "John von Neumann and Modern Computer Architecture," *NCLab*, 26-Feb-2017. [Online]. Available: https://nclab.com/john-von-neumann-and-modern-computer-architecture/. [Accessed: 18-Apr-2022].

[22] R. Pramoditha, "The concept of artificial neurons (perceptrons) in neural networks," *Towards Data Science*, 26-Dec-2021. [Online]. Available: https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc. [Accessed: 17-Apr-2022].

[23] S. Dent, "Intel says its new 5.5GHz i9-12900KS is the world's fastest desktop processor," *Engadget*, 29-Mar-2022. [Online]. Available: https://www.engadget.com/intel-says-its-new-55-ghz-i-9-12900-ks-is-the-worlds-fastest-desktop-processor-074143774.html. [Accessed: 18-Apr-2022].

[24] S. Huang, "Understanding AlphaGo: how AI thinks and learns (Advanced)," *Towards Data Science*, 21-Jan-2020. [Online]. Available: https://towardsdatascience.com/understanding-alphago-how-ai-thinks-and-learns-advanced-d70780744dae. [Accessed: 17-Apr-2022].

[25] "The AI index report – artificial intelligence index," *Stanford.edu*. [Online]. Available: https://aiindex.stanford.edu/report/. [Accessed: 18-Apr-2022].

[26] Veritasium, "Future computers will be radically different," 01-Mar-2022. [Online]. Available: https://www.youtube.com/watch?v=GVsUOuSjvcg. [Accessed: 18-Apr-2022].