Artificial Intelligence: The Analog Future
Why is Analog Computing the Future of AI?

Orson W Hannath
COMP1002 Fundamentals of Computational Intelligence
Flinders University

Hann0265@flinders.edu.au

Abstract

Artificial Intelligence has an array of applications from image detection to unlock your phone through to smart technology for your home. As the uses of AI expand and consumers become complacent with the idea that technology will keep getting better whilst maintaining a compact size companies must adapt and find new ways to compute and apply their AI. One proposed way of which is going Analog.

1 Introduction

One step forward, two steps back. A phrase often used to express difficulty in progression. This phrase can be used to represent the way that Artificial Intelligence (AI) is advancing in today's day and age. AI is being used in a plethora of different applications from the self-driving cars of Tesla (Ajitha & Nagra 2021) to performant facial recognition programs with up to 94.7% detection rate (Aitkenhead & McDonald 2003). One downside to AI is that the more inputs required to determine an outcome the more computationally intensive it becomes (Shafiee et al. 2016). This is due to the fundamental nature of Convolutional Neural Networks (CNN's) which solve an array of dot-product math equations to determine a series of weighted outputs (Shafiee et al. 2016). The question is however how and why is analog computing going to be the future of the AI industry. The answer of which boils down to a simple observation made by Gordon Moore in 1965 in which he theorized that the number of transistors per silicon chip will double every year (Schaller 1997). This is a major problem for digital computing as we know it as there will be a point at which the progression of silicon chips plateaus and the need for more computing power will not be met by the supply available from a single silicon chip (Waldrop 2016). This means that with the current technology computing chips will either need to get larger, be built for specific tasks or take on an entirely new form of technology (Track, Forbes and Strawn, 2017). One other option that is currently being explored takes us back to Analog methods of computing which are currently being applied to the Neural Network (NN) project DaDianNao specializing in complex computations of matrices via an Analog approach (Shafiee et al. 2016).

2 Limitations of Progress at the Rate of Moore's Law

As the world progresses and mankind becomes complacent with the idea of technology evolving at an exponential rate technology companies need to keep up. Gordon Moore in 1965 came up with a theory known as Moore's Law which describes how fast technology companies will develop silicon chip technology for digital computers (Schaller 1997). Moore's Law states that tech companies will double the number of transistors per silicon chip each year essentially doubling the performance from one year to another (Schaller 1997). This theory however has physical limitations with the current technology available (Waldrop 2016).
Waldrop (2016) explains in their article 'The Chips are Down for Moore's Law' that there would be major implications as silicon chip transistors reach a "2–3-nanometre limit, where features are just 10 atoms across". One implication would be at a Quantum Level as electrons behavior would be susceptible to quantum uncertainties deeming them useless for computations (Waldrop 2016). Track, Forbes and Strawn (2017) agree with the idea that Moore's Law level progression must come to an end predicting that the untimely demise would

occur within the next 10-15 years a theory which is backed by Gordon Moore himself.

3 Analog vs Digital Computing

The computing world as we know it in the 21st century is magnitudes ahead of the computing world prior to the 21st century. The computers that we have now are digital completing computations via 1's and 0's computing at speeds that are getting exponentially better. However, this was not always the case the earliest form of computers were Analog harnessing natural phenomena to complete computations such as the Greek Astrolabe which was not uncommon in the 17th century during the Middle Ages (North 1974).

Digital Computers use a binary number system composed of 1's and 0's to complete computations. These computations are completed via the use of logic gates (Woods & Holdsworth, 2002). These gates are composed of resistors and transistors each of which have an extremely important role in the functionality of a logic gate (Woods & Holdsworth 2002). By combining logic gates into very specific sequences logic chips such as adders can be created which are able to compute basic addition arithmetic (Woods & Holdsworth 2002). These logic gates are essentially conditional switches which take inputs and compare them to their conditional statement and then either output a 0 or a 1 depending on whether the condition has been met or not (Woods & Holdsworth 2002).

Analog Computers on the other hand represent a wide array of techniques as it is more so a category of computers rather than a type. These computers can use naturel phenomena such as Electronics, Mechanics and Hydraulics. For example, back in the 1950's Electrical Engineer A. Phillips created a hydromechanical analog computer which was able to essentially solve differential equations with an accuracy of ±4% using hydraulic and mechanical principals to simulate the flow of currency in an economy (Bissell 2007).

4 Convolutional Neural Networks

AI comes in many forms one of which is image recognition neural networks. Image Recognition AI generally use a Convolutional Neural Network (CNN) as this type of Neural Network (NN) is best suited for grid layout input data similar to that of a pixel grid in an image (Yamashita et al. 2018). CNN's are generally composed of three layers convolution, pooling and fully connected layers. Convolution and Pooling are said to extract features from the input whereas the fully connected layer is said to convert the extracted features into a final classification (Yamashita et al. 2018). Each node within the convolutional layer is given a kernel which is a matrix of n-by-n size which is filled with random floating-point numbers prior to training. These kernels essentially scan across the input matrix iterating over each location taking a dot-product between the input at location (x,y) and the kernel, this value is then sent to the next layer in the NN (Yamashita et al. 2018). In the case of a CNN the pooling layer generally down-samples the input from the convolution layer via either finding the max value or the average value from a filter matrix of size n-by-n (Yamashita et al. 2018). The last step in the CNN is to convert the data from the pooling layer to something that can represent an output this is called the fully connected layer (Yamashita et al. 2018). Each node in the fully connected layer of the CNN takes inputs from each node in the pooling layer of which each input is adjusted by some weight which is learnt by the NN via a process such as backpropagation (Yamashita et al. 2018).

Once a NN has been trained and its weights and kernels have been determined it is ready to be deployed. In the deployment phase it is possible to pass in inputs and have the expected outputs come out after the NN completes a series of dot-product equations.

5 Analog Convolutional Neural Network

Due to the nature of CNN's and how they compute their outputs has led to the pioneering of the use of Analog computers in AI (Shafiee et al. 2016). The current project on the forefront of this technology is DaDianNao which as described by Shafiee et al. (2016) "adopts a near data processing approach, where a specialized neural functional unit performs all the digital arithmetic operations and receives input weights from adjacent eDRAM banks". But what does this really mean. Essentially at the root of the DaDianNao system are tiles/nodes formed mainly of memristor crossbar arrays. These arrays are not only able to store the synaptic weights of the NN but they are also able to compute the dot-product equations (Shafiee et al. 2016). One shortfall of the DaDianNao system is that the arrays are not easily re-programed meaning that the NN's must be preprogrammed to one set series of synaptic weights (Shafiee et al. 2016). Once the outputs from the tile/node have been computed they are then stored in eDRAM so they can be sent to the next layer in the network (Shafiee et al. 2016). The advantages to this form of NN is that the computations are completed in-situ meaning the overall system needs to make less calls to memory than a traditional digital NN making it faster (Shafiee et al. 2016). One thing to note about the DaDianNao system is that its key overheads are the Analog to digital converters and the digital to analog converters (Shafiee et al. 2016). At the time of publication Shafiee et al. (2016) claimed that the "64-chip DaDianNao has already been shown to have 450x speedup and 150x lower energy than an NVIDIA K20M GPU".

6 Applications of an Analog Neural Network in The Modern Era

To continue the bridging of NN's into our daily life's they must have hardware implementations. In industries such as aerospace and military analog hardware implementation is essential as their applications require them to be of small volume, reduced weight, shock and vibration resistant and have hive execution rates (Draghici 2000). One reason as to why analog AI is going to be incorporated into technology as stated by Draghici (2000) is that it is a cost-effective technique at a large scale. Some fields that image classification CNN's are currently being used in are "autonomous robots or vehicles, surveillance and security systems, medical applications, industrial processes and transport" (Gatet et al. 2008).

A team from Jet Propulsion Laboratory and Irvine Sensors Corp have displayed the modern era use of analog AI technology via a "3-dimensional neural network circuit aimed at target recognition for missile interception application" that they have developed (Draghici 2000). The technology was demonstrated to the US army after a contract had been won by Irvine Sensors with aims of demonstrating the feasibility of the NN and its applications (Draghici 2000).

Another example of an Analog CNN being used in the modern era is as surface detection sensors (Gatet et al. 2008). The current standard practice is to use physical sensors such as optical, infrared, ultrasonic or radar (Gatet et al. 2008). However, Gatet et al. (2008) suggests that NN's "have become an effective solution to improve the robustness and the data processing of integrated sensors". The benefit to applying NN's to the classification of surfaces is due to current techniques being limited by distance and/or incidence angle (Gatet et al. 2008). Using NN's along side well-known techniques it is possible to classify surfaces with a range of around 1m allowing also for varying incidence angles which is a baseline requirement for road surface detection in driverless vehicles (Gatet et al. 2008).

7 Discussion and Conclusions

Overall, pressures such as the implications of denser and denser silicon chips are forcing companies to come up with solutions allowing for advancements to be made in the tech sector. One of these solutions is the use of analog computing being applied in AI NN's. This is a sector that appears to have a bright future as it paves the way between advanced technologies and consumer markets.

Part of this journey will involve companies taking analog AI in their own directions using current practices and technology. These directions could be based on current trends globally or specific tasks of which the companies are trying to achieve. Some of the main areas that will likely be further developed for the technology are, transforming the analog NN circuits from a fixed architecture to a programable architecture via the use of programable weights and scalability and connectivity via the interfacing of NN's with existing circuits (Draghici 2000).

Finally, it is clear that Analog AI is the way of the future, but the question really is when will we be seeing this technology applied in our day to day life and how is it going to benefit us the consumer.

References

Aitkenhead, M & McDonald, A 2003, "A neural network face recognition system", *Engineering Applications Of Artificial Intelligence*, vol. 16, no. 3, pp. 167-176.

Ajitha, P & Nagra, A 2021, "An Overview of Artificial Intelligence in Automobile Industry – A Case Study on Tesla Cars", *Solid State Technology*, vol. 64, no. 2, p. 506.

Bissell, C 2007, "Historical perspectives - The Moniac A Hydromechanical Analog Computer of the 1950s", *IEEE Control Systems Magazine*, vol. 27, no. 1, pp. 69-74.

Draghici, S 2000, "NEURAL NETWORKS IN ANALOG HARDWARE — DESIGN AND IMPLEMENTATION ISSUES", *International Journal of Neural Systems*, vol. 10, no. 1, pp. 19-42.

Gatet, L, Tap-Beteille, H & Lescure, M 2008, "Analog Neural Network Implementation for a Real-Time Surface Classification Application", *IEEE sensors journal*, vol. 8, no. 8, pp. 1413-1421.

North, J 1974, "THE ASTROLABE", *Scientific American*, vol. 230, no. 1, pp. 96-107.

Schaller, R 1997, "Moore's law: past, present and future", *IEEE Spectrum*, vol. 34, no. 6, pp. 52-59.

Shafiee, A, Nag, A, Muralimanohar, N, Balasubramonian, R, Strachan, J & Hu, M et al. 2016, "ISAAC: a convolutional neural network accelerator with in-situ analog arithmetic in crossbars", *ACM SIGARCH Computer Architecture News*, vol. 44, no. 3, pp. 14-26.

Track, E, Forbes, N & Strawn, G 2017, "The End of Moore's Law*", Computing In Science & Engineering*, vol. 19, pp. 4-6.

Waldrop, M 2016, "The chips are down for Moore's law", *Nature*, vol. 530, no. 7589, pp. 144-147.

Woods, C & Holdsworth, B 2002, *Digital Logic Design*, 4th edn, Elsevier, pp. 1, 30-40.

Yamashita, R, Nishio, M & Do, R 2018, "Convolutional neural networks: an overview and application in radiology.", *Insights Imaging*, vol. 9, pp. 611-629.