TrueNorth: IBM’s newest chip

Jaleesa Dmytrow
200300170
Task: E. The Future
dmytrowj@uregina.ca
February 22, 2016

Introduction

Building chips using von Neumann’s architecture has hit a roadblock; if we want to make faster processors we need to build chips with more advanced techniques. This is where IBM’s TrueNorth CMOS chip fits in; it is not built like traditional chips. The TrueNorth chip is built to mimic the workings of a brain, complete with neurons and synapses. Traditional computers focus on language and analytical thinking such as solving problems quickly and correctly [3]. On the other side, neuromorphic chips can handle senses and pattern recognition similar to the right side of a human brain [3]. Neuromorphic chips are event-driven and only run when a task needs to be completed. This results in a cooler operating environment and lower power usage [3].

History of Computer chips

The chips used in today’s computers are based on von Neumann’s architecture which is seven decades old [1]. In these conventional chips the memory, control units and processing are all separate. A chip today is made to perform logical operations in sequence and to manipulate numbers, nothing too complex that requires large amounts of data [1]. For a simple task such as adding two variables together conventional chips have to grab the two variables from memory, add them together, then put the result back into memory. These separate tasks take a lot of time and use a lot of power. To speed up and reduce power consumption within a chip, engineers have been shrinking transistors, communication lines and other pieces of a chip [1]. Although this has provided some benefit there is no way to continue shrinking these chips without losing important parts of a chip. This is the reason that chips based on von Neumann’s architecture cannot handle vision or language recognition without constraints.

Constraints of conventional chips

A conventional chip follows the von Neumann architecture for creating chips. These chips have a lot of constraints which include power consumption, scalability and the time a chip takes to compute a task [2]. These constraints are important when talking about processing image and voice recognition. As stated in the section above called “History of computer chips”, conventional chips can process voice and image but the algorithms needed take a large amount of time and power. There is a demonstration by Google that used 16,000 processing chips to recognize cats and human faces in video clips [1]. These 16,000 processing chips used about 100 kilowatts of power [1]. This amount of power is quite extreme as a human brain does the same task as the 16,000 chips however it only uses about one tenth of a watt [1]. There is a significant difference between the 16,000 chips and a human brain.

Another constraint is the scalability of conventional chips. The memory, processing and control unit parts of a chip are all separate in the von Neumann architecture. With this design, which is still used today, the scalability of these chips is possible but does not make the actual chips faster. For example if a person builds a computer with 4 chips instead of 2 chips the computer would be faster at handling tasks in parallel but the individual chips would not be faster. This constraint ties in with the power constraint as the more chips you add the more power you require.

Lastly, the constraint of time is important for conventional chips. Each chip has separated the three main parts into separate sections. This causes a lot of time to be wasted by going back and forth. As stated above in the section “History of computer chips”, engineers have being trying to shrink the distance that the data has to travel. The current computer chip cannot be manipulated using this method anymore as important pieces of a chip could be lost if anymore ‘shrinking’ occurs. Even with this method of shrinking the distance between the three sections there is still extra time needed to access the individual areas one at a time.

IBM’s TrueNorth Chip

TrueNorth has “5.4 billion transistors wired together to form an array of 1 million digital ‘neurons’ that talk to one another via 256 million ‘synapses’”[1]. These neurons and synapses, or transistors and cores, are built to act like a brain. The TrueNorth chip is made to be able to handle voice and image recognition better than other chips that are more conventional [1]. Conventional chips are made to compute answers from algorithms and provide answers to questions. It would not be able to easily distinguish between two different images. These conventional chips could eventually notice the difference but to find the difference the chip would need to use complex algorithms which take a large amount of power to run [1].

How does TrueNorth work?

TrueNorth is made of 4096 neurosynaptic cores in a 2-d array that has 1 million neurons and 256 million synapses [2]. This chip has a computational performance peak of 58 giga-synaptic operations per second (GSOPS) [2]. When running at the computational performance peak the chip has the efficiency of 400 GSOPS per Watt of power [2]. When comparing the above facts to the demonstration by Google (mentioned in the section called "Constraints of conventional chips") it is clear that the amount of power used by the TrueNorth chip is quite small. TrueNorth is an event-driven architecture which means that there are no global clock networks or collocated memory (which minimizes the distance any data has to travel) [2]. By removing these parts IBM's chip is not constantly using power if no tasks are being performed. Removing power hungry parts of a chip allows TrueNorth to achieve the low power status which makes the chip desirable for many different devices.

The TrueNorth chip does not work with known programming languages such as C++ or Java but needs to be programmed in its native language, Corlet [2]. By building a specific programming language for the chip IBM and their partners had to create classes to teach this new programming language. These classes are held at SyNAPSE University [3]. Learning a new programming language can cause a large learning curve however when TrueNorth is actually released to the public and used in various types of computers we will have more functional devices.

Constraints of IBM's Chip

The constraints for IBM's TrueNorth chip are the same as conventional chips. The difference is how TrueNorth handles the constraints. The constraint of power is different for TrueNorth as it is a low-power chip. This means that to perform a task the chip will use the least amount of power necessary. The constraint regarding time is changed for TrueNorth, also. This chip has control units, memory and processing in one spot [2]. This eliminates the time needed to go from memory to processing back to memory, which is needed in conventional chips. The last constraint worth mentioning is scalability. IBM has created TrueNorth to have processing, memory and control units in one spot which makes scalability very easy. Each TrueNorth chip is highly parallel which means that multiple tasks can be done at the same time using multiple cores. As the chip is highly parallel it would be easy to assume that by adding more cores there would be no disruption to the chip. This idea makes TrueNorth highly scalable. With all of these constraints being addressed it is possible to say that TrueNorth does not have to worry about the main constraints that interfere with conventional chips.

What does the future look like?

With IBM's TrueNorth chip the way our devices and computers work could be drastically changed. For example if TrueNorth was placed within a smart phone there would be no need for complex algorithms to run applications that recognize voice and images, the chip could do that. If TrueNorth was placed in a watch the owner of the watch could set the time by voice command. This kind of functionality could help someone with arthritic hands to set up a watch without asking for help. Another unique way the TrueNorth chip could help change the way technology helps people is to create a device for a person with limited vision. If the chip within a device can recognise cars, people, bikes or anything blocking the path a person with poor vision could walk without fear of walking into something. When new chips are built they are usually placed in technology that is already out there. These TrueNorth chips could be used in new ways to make things better for people who need some assistance in their daily lives.

Conclusion

IBM's TrueNorth chip is one of the first neuromorphic chips created that has the potential to change the way computers work. TrueNorth is a chip that has multiple cores that handles the memory, control units and processing all in one spot. This fact alone minimizes many of the constraints from conventional chips. The power consumption is lower for TrueNorth as there is no unneeded global clock or distance that data has to travel. Both of these features that are in conventional chips require power to perform their tasks. With TrueNorth being an event-driven chip power will only be used when the chip is completing a task. Another constraint that is minimized from having the three sections in one spot is the time constraint. If data does not need to travel to different sections each time something is needed the time to complete a task is drastically reduced. Lastly the constraint of scalability is lessened. TrueNorth is highly scalable as each core is functional without other pieces of the chip which means that if other cores are added there would be no disruption within the chip. TrueNorth can handle tasks such as image and voice recognition much better than conventional chips which could change how computers could be built. As technology grows to handle neuromorphic chips like IBM's TrueNorth chip we could see the end of von Neumann architecture being used.

References:

[1]R. Service, "The brain chip", *Science*, vol. 345, no. 6197, pp. 614-616, 2014.

[2]F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G. Nam, B. Taba, M. Beakes, B. Brezzo, J. Kuang, R. Manohar, W. Risk, B. Jackson and D. Modha, "TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 10, pp. 1537-1557, 2015.

[3]D. Modha, "IBM Research: Brain-inspired Chip", *Research.ibm.com*, 2016. [Online]. Available: http://www.research.ibm.com/articles/brain-chip.shtml. [Accessed: 22- Feb-2016].