Development of a Neuromorphic Computing System

Luping Shi[1*], Jing Pei[1], Ning Deng[2], Dong Wang[1], Lei Deng[1], Yu Wang[3], Youhui Zhang [4], Feng Chen [5], Mingguo Zhao[5], Sen Song[6], Fei Zeng[7], Guoqi Li[1], Huanglong Li[1], Cheng Ma[1]

[1]Center for Brain-Inspired Computing Research (CBICR), Optical Memory National Engineering Research Center, Department of Precision Instrument,[2]Institute of Microelectronics, [3]Department of electronic engineering , [4]Department of computer science and technology, [5]Department of automation , [6]Department of biomedical engineering , [7]Department of materials science and engineering, Tsinghua University, Beijing 100084, China

Tel: (86) 10 62771685, Fax: (86) 1062771685, Email: lpshi@tsinghua.edu.cn

Abstract

Although a variety of solutions for neuromorphic systems based on different hardware technology and software programming schemes, there has yet to be a common accepted one. Based on some recent findings in brain science, we propose a new design rule for developing a brain inspired computing system. We design and fabricate a neuromorphic chip, named 'Tianji' chip. A multi-chip architecture-based PCB board has been demonstrated. The detailed hardware implementation and software programming scheme are presented in this paper.

Introduction

Neuromorphic computing is very attractive for developing future high performance and intelligent computer. In the history, computer is no doubt one of the greatest inventions. Based on computers we have built a digital universe in which we can connect and communicate with each other anywhere and anytime. During last half century, scaling both CPU and memory, two main components in von Neumann architecture-based computers, has been being the critical driving force of computer development. However, the separated computation and storage leads to low efficiency and the scaling will approach its physical limitation in 15 to 20 years. It is the time to find a new paradigm to further develop computer technology.

State-of-the-art neuromorphic systems

Modha et al. [1], have demonstrated a building block of compact modular core for large-scale neuromorphic system architecture. The neurosynaptic core combines digital neurons with large synaptic array. Based on it, a scalable, efficient, and flexible non–von Neumann architecture using silicon technology was demonstrated. The aim is to build neurosynaptic supercomputers by creating systems with hundreds of thousands of cores, hundreds of millions of neurons, and hundreds of billions of synapses.

Furber et al. [2], have demonstrated a parallel thousand-core computer which is suitable for modeling large-scale spiking neural networks with bio-plausible real-time performance. A system with up to 2500 processors, to present the real-time event-driven programming model was demonstrated. The aim is to simulate the behaviors of one billion neurons that is equal to about 1% as many as that in a human brain.

In addition, a lot of other efforts have been put to develop various solutions for neuromorphic computing systems, such as analog circuits-based type [3], mixed-analog-digital type [4, 5], non-volatile memory based or memristor-based type [6-11], etc. Each solution has its special advantages and constrains, as a result till now there has not been a common accepted one. It is also necessary to look for some new technological solutions.

Theoretical consideration for the development of a brain inspired computing system

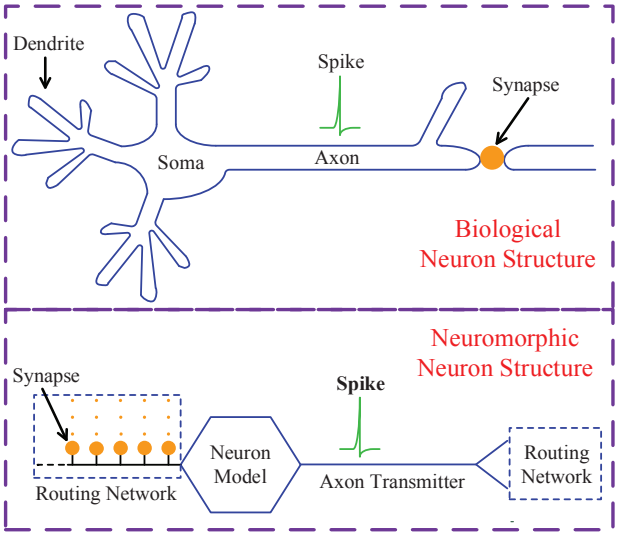

Figure 1. Schematic of the biological neuron structure and the neuromorphic neuron structure.

In order to design a neuromorphic computing system, also termed as a kind of brain inspired computing system, we set forward a new system design rule. Typically as illustrated in Figure 1, the neuromorphic design rule, including element or

system model, hardware architecture, routing network, computing algorithms, was built based on the current findings in brain science [12-17], such as the neuron and synapse structure and connection configuration, the connectivity characteristics of the mammalian brain, neuron cluster, hierarchical information flow, whole brain model, and so on.

Hardware implementation of 'Tianji' chip

Based on the above concerning, we have designed and fabricated a neuromorphic chip called 'Tianji' chip. As shown in Figure 2, the proposed neuromorphic computing chip architecture comprises building block of multicores, a two-dimensional array of cores, with a digital hardware foundation. All six cores are connected by a mesh routing network, and execute a parallel operation. Event packets, generated by spiking or analog neurons in cores, are transmitted to the target cores through the routing network. A functional neuromorphic system is comprised of dozens or hundreds of interconnecting chips. Inter-chip communication is implemented in the same manner as on-chip communication. The chip supplies researchers with a set of primitives for building massively parallel neuromorphic models.

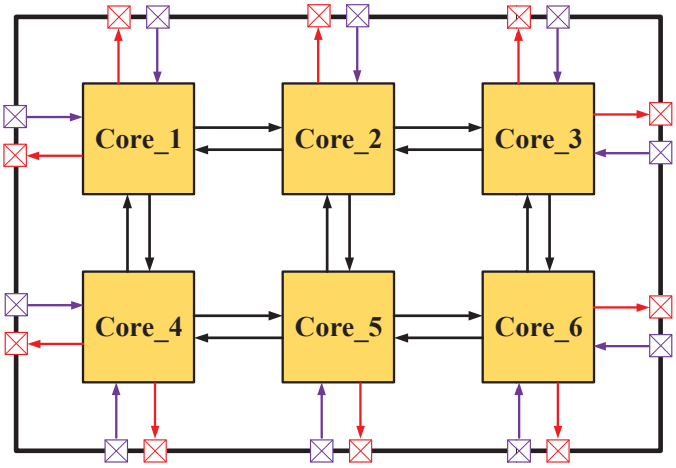

Figure 2. The schematic of parallel computing paradigm of the 'Tianji' chip.

Figure 3 illustrates the element structure of the core of 'Tianji' chip which consists of the neuron block, synaptic network, router, synchronizer and parameter manager, as well as input and output interfaces. The Neuron block performs the integration of dendritic inputs, leak subtraction, and the threshold checking in a TDM (Time Division Multiplexing) manner. When a neuron's potential exceeds the predefined threshold, a spike signal will be sent to the router. Synaptic network performs multiplication of its stored synaptic weights with the status vector of pre-inputs, and sends the results to the neuron circuits as dendritic inputs. Router is in charge of intra-core and inter-core data transfer, i.e. it receives event packets from the neuron circuits or the adjacent cores and delivers them to the target cores or local core's synchronizer. Synchronizer parses the event packets received from router and saves the corresponding synapse status in the specified buffer memory. Then it delivers synapse status vector to the synaptic network, but only at the rising edge of a global trigger signal. Parameter manager receives core parameters from the host, including synaptic weights, neuron parameters and routing addresses, and then transmits them to the target modules.

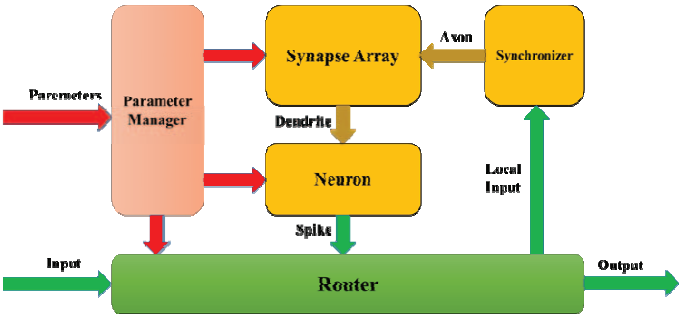

Figure 3. The logic block diagram of the core, which is the basic unit of the 'Tianji' chip.

Figure 4 shows the structure of the synchronizer within which the synchronizer controller between input and output interface communicates with the buffer memory of synapse status. When a local input packet is received, the controller parses it to get the target axon row address. The parsing results of all the axon addresses in a predefined time window, synapse status vector, is then stored in the synapse status buffer memory. At the next rising edge of the global trigger signal, the controller retrieves the synapse status vector from the synapse status memory and then sends it out to synaptic network.

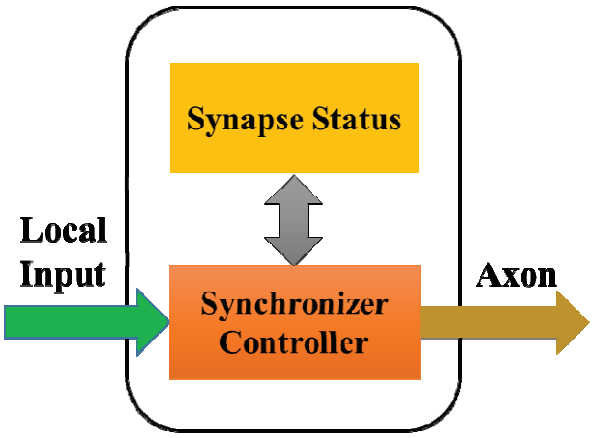

Figure 4. The synchronizer consists a synchronizer controller and a synapse status buffer memory.

Figure 5 illustrates the timing configuration of the synchronizer. The local input and synchronizer data are the input and output of the synchronizer, respectively. A fixed-frequency clock signal is used as the global trigger signal to synchronize the time steps of multiple cores.

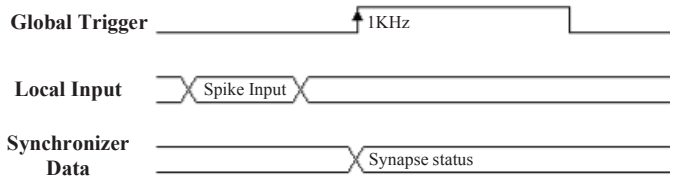

Figure 5. Illustration of the specific timing configuration of the synchronizer.

Figure 6 illustrates an adaptive parameter management module. It is responsible for receiving the core parameters, including synaptic weights, neuron parameters and target address (or termed as routing table), from a host computer or a host FPGA. The parameters are then downloaded into the synaptic network, neuron block and router to initialize the core.

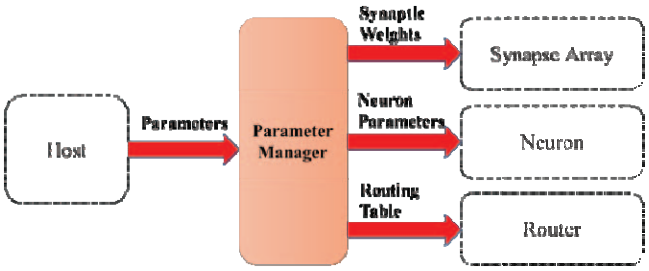

Figure 6. An adaptive parameter management module.

Figure 7 illustrates the contour of IC design. The neuromorphic chip has been implemented in 120 nm CMOS technology. Six different colors represents six cores in one chip, while the shaded areas denotes memory in which synaptic weights, neuron parameters, target addresses (routing table) are stored. Figure 8 shows a photo of the 'Tianji' chip.

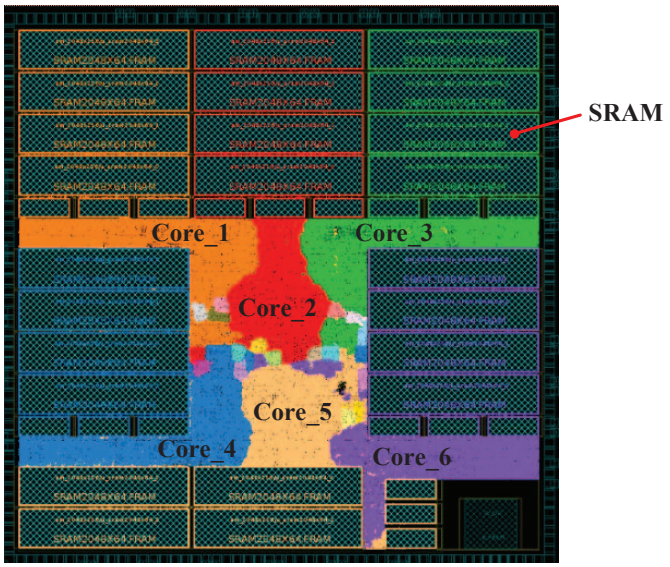

Figure 7. The contour of IC design of 'Tianji' chip.

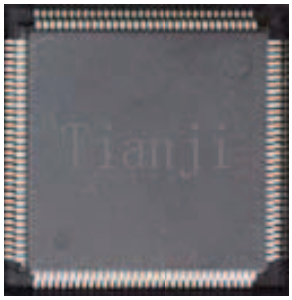

Figure 8. A photo of the 'Tangji' Chip.

Figure 9 depicts the multi-chip PCB board system for developing various potential applications, such as fast computing, hearing and vision processing, robot control, and so on. The FPGA chip serves as the main controller and the communication medium with the host computer through the USB port or RJ-45 Ethernet port on board. Chips, forming a 4×4 two-dimensional array, receive input data or configuration data (core parameters) from the FPGA controller through the Config Data Bus; while the FPGA obtains these data from the 1Gbit FLASH or the host computer. Output data flow of chips is also transferred to the FPGA through the data bus, and then stored in the FLASH or sent to the host computer.

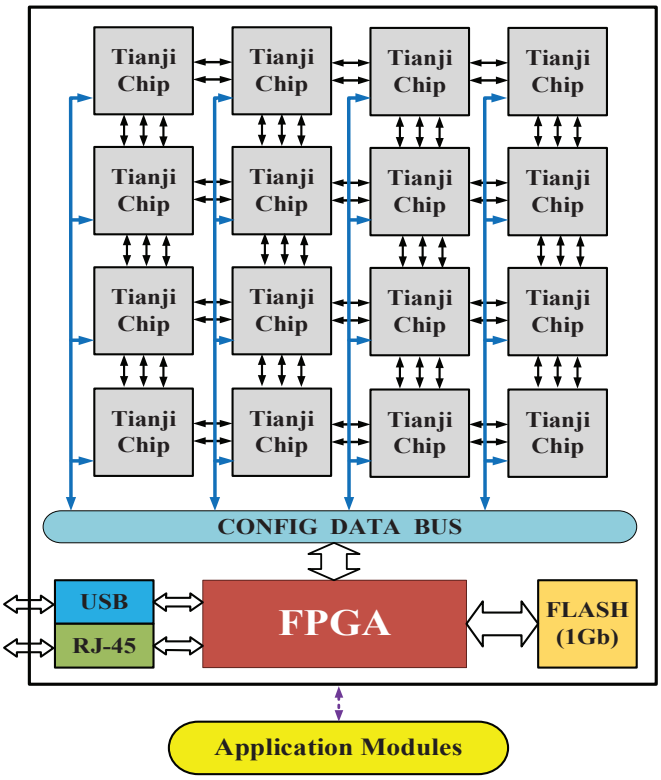

Figure 9. A multi-chip architecture-based PCB board.

Software Programming

In order to operate 'Tianji' chip, we developed a software system that can use various neural network models as a reconfigurable directed graph and map the latter onto the hardware system for high use efficiency. We also designed a simulation platform based on computer cluster and FPGA-based systems so that it can conduct simulations based on different models. The nodes hold computational state, as well as exchange information. All nodes execute one computational step in a parallel manner. The specific clock of graph and coding scheme is set so that we can apply it for some real-time applications.

Conclusion

A new design rule of neuromorphic computing system based on the current findings in brain science was proposed and the 'Tangji' chip was designed and fabricated based on it. A multi-chip architecture-based PCB board was implemented. The corresponding software system and simulation platform were developed for further applications on the chip.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61475080), the Study of Brain-Inspired Computing System of Tsinghua University (No. 20141080934) and Beijing program on study of functional chip and related core technologies of ten million class of brain inspired computing system.

References

[1] P. A. Merolla et al., "A Million Spiking-Neuron Integrated Circuit with A Scalable Communication Network and Interface", Science 345, 668-673 (2014).

[2] S. B. Furber et al., "The SpiNNaker Project," Proc. IEEE 102, 652–665 (2014).

[3] E. Chicca et al., "Neuromorphic Electronic Circuits for Building Autonomous Cognitive Systems," Proc. IEEE 102, 1367-1388 (2014).

[4] J. Schemmel et al., "A Wafer-Scale Neuromorphic Hardware System for Large-Scale Neural Modeling,"Proc. IEEE Int. Symposium on Circuits and Systems, 1947-1950 (2010).

[5] B. Benjamin et al., "Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations," Proc. IEEE 102, 699-716 (2014).

[6] G. Snider et al., "From Synapses to Circuitry: Using Memristive Memory to Explore the Electronic Brain," Computer 2, 21-28 (2011).

[7] S. Park et al., "Electronic System with Memristive Synapses for Pattern Recognition, " Sci. Rep. 5, 10123 (2015).

[8] M. Prezioso et al., "Training Andoperation of an Integrated Neuromorphic Network Based On Metal-Oxide Memristors," Nature 521, 61-64 (2015).

[9] B. L. Jackson et al., "Nanoscale Electronic Synapses Using Phase Change Devices," ACM J. Emerg. Technol. Comput. Syst. 9, 12 (2013).

[10] D. Kuzum et al., "Synaptic Electronics: Materials, Devices and Applications," Nanotechnology 24, 382001 (2013).

[11] O. Bichler et al., "Visual Pattern Extraction Using Energy-Efficient '2-PCM Synapse' Neuromorphic Architecture," IEEE Trans. Elec. Dev. 59, 2206–2214 (2012).

[12] C. Eliasmith et al., "A Large-Scale Model of the Functioning Brain," Science 338, 1202-1205 (2012).

[13] W. Maass et al., "Noise as a Resource for Computation and Learning in Networks of Spiking Neurons," Proc. IEEE 102, 860-880 (2014).

[14] M. Okun et al., "Diverse Coupling of Neurons to Populations in Sensory Cortex," Nature 521, 511-515 (2015).

[15] F. Anselmi et al., "Deep Convolutional Networks are Hierarchical Kernel Machines," arXiv preprint arXiv:1508.01084, 2015.

[16] N. Hiratani and T. Fukai. "Mixed Signal Learning by Spike Correlation Propagation in Feedback Inhibitory Circuits," PloS Comput. Biol. 11, e1004227 (2015).

[17] S. R. Kheradpisheh et al., "Bio-inspired Unsupervised Learning of Visual Features Leads to Robust Invariant Object Recognition," arXiv preprint arXiv: 1504.03871, 2015.