Purva Mimaansa | Refereed | Peer Reviewed
ISSN: 0976-0237 | Volume 12, Issue September 2021 | Impact Factor 4.765

ARTIFICIAL INTELLIGENCE CHIPS: BRIEF INTRODUCTION

NEHA
Assistant Professor in Electronics & IT, Sanatan Dharma College, Ambala Cantt

ABSTRACT

Artificial Intelligence is rapidly evolving. The quality of AI-based solutions has vastly improved in recent years. In terms of speed and performance, the software is increasing by leaps and bounds. Hardware upgrades are required to match the significant gain in software. So that a better AI system that can accomplish the work rapidly can be designed, particularly in applications that rely on real-time data. Artificial Intelligence Chips are explored in this paper. The architecture of the chips is discussed and an attempt is made to explain why AI chips outperform general-purpose CPUs.

KEYWORDS: AI, ANN, GPU, FPGA, ASIC

INTRODUCTION

Artificial intelligence is playing a significant role in present. In future also demand of AI based services and products will build numerous folds. So there is a consistent need of further developing AI innovation to satisfy this future interest. There is a need of more productive systems that can deal with volumes of information and work and work on it as quick as could really be expected. For instance self- driven cars need to work on real time data like pictures of individuals strolling in the city, vehicles passing by and output from various sensors quickly to settle on a best choice.This requires aproficient AI run software as well as AI specific hardware. A great deal of work has effectively been done on programming, now more work needs to be done on hardware.Accordingly many IT industries have begun creating AI chips (Kumar, Rahul, Supradip Baul: 2019). These AI chips not only increase the surface density to accommodate more transistors per surface area as compared to general purpose chips but are also very cost effective and can be designed according to specific application.

WHAT IS AI CHIP?

According to definition AI chips are specially designed Integrated Chips for Artificial Intelligence applications. For example artificial neural network (ANN) and Deep Learning based applications. AI chips are also called AI hardware or AI Accelerator (CemDilmegani: 2021). They contain billions of MOSFET transistors on them. They are used in robotics, internet of things and other data-intensive or sensor-driven tasks. They have multiple core design which performs low-precision arithmetic tasks, has novel dataflow architectures and in – memory computing capabilities (Merritt: 2016).

AI CHIP ARCHITECTURE

"AI chips" contains graphics processing units (GPUs), field-programmable gate arrays

(FPGAs) and certain types of application-specific integrated circuits (ASICs) especially designed for AI calculations.

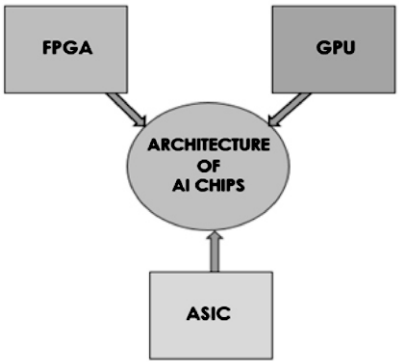

Figure: Architecture of AI Chip

GPUs: Graphics Processing Units (GPUs) is basically used for building and refining AI algorithms. It speeds up graphical processing through parallel computing instead of sequential computing used in General purpose chips. This is most widely used by deep learning software developers.

FPGAs: Field Programmable Gate Array (FPGA) are primarily used for inference i.e. to apply trained AI algorithms to real-world data inputs, a process. It gives freedom to design application specific chip by allowing hardware customization. It is easy to reprogram and reconfigure FPGA based chip which goes well with rapidly evolving AI. This feature allows designer to test algorithms quickly and deliver the product on time (VB Staff: 2020).

ASICs: (Application Specific Integrated Chip) can be built for both training and inference.It includes hardware Circuits which can be customized for specific algorithm. It is normally more efficient than FPGAs, but FPGAs are more programmable and allow for design optimization as AI algorithms evolve.

BENEFITS OF AI CHIPS OVER GENERAL PURPOSE CHIPS

General-purpose chips, such as central processing units (CPUs), can run AI systems, but AI chips run cutting-edge AI algorithms faster and more efficiently.They run the program parallelly with low precision so less number of transistors are required for the task this speeds up the memory access (Khan, Saif M., Alexander Mann: 2020).

For basic calculations General purpose chips use arithmetic blocks. These blocks works on serial processing which does not give god performance especially for deep learning techniques. These chips cannot perform a high number of simultaneous tasks. Whereas AI chips enable parallel processing thus enabling following benefits:

- **Higher Speed:** Many AI applications need to run sophisticated training models and algorithms. This is achieved in AI chips through parallel processing which speeds up the data processing ten times more than general purpose chips.

- **High Bandwidth Memory:**As AI chips use parallel processing it uses 4-5 times more bandwidth than general purpose chips (CemDilmegani: 2021).

CONCLUSION

AI chips are the need of the time for designing new generation AI systems which are faster, less costly and consumes less power. With the need of such systems demand of AI chips is going to increase in future. Many Tech Giants has already started manufacturing AI Chips. Although these AI Chips or AI hardware can be rented in case of small projects but still own hardware can bring down the overall project cost provided production is on large scale. AI chips are far better than general purpose chips making AI system more efficient and faster.

WORKS CITED

Kumar, Rahul, Supradip Baul. Artificial Intelligence Chip Market, May, 2019. Available: https://www.alliedmarketresearch.com/artificial-intelligence-chip-market

Khan, Saif M., Alexander Mann. "AI Chips: What They Are and Why They Matter. April 2020. Available: https://cset.georgetown.edu/wp-content/uploads/AI-Chips%E2%80%94What-They-Are-and-Why-They-Matter.pdf

CemDilmegani. "AI chips: In-depth guide to cost-efficient AI training & inference"July 1, 2021.

Merritt, Rick. Google Developing AI Processor. May 18, 2016. Available: https://www.eetimes.com/google-designing-ai-processors/

VB Staff. FPGA chips are coming on fast in the race to accelerate AI. Dec. 2020. Available: https://venturebeat.com/2020/12/10/fpga-chips-are-coming-on-fast-in-the-race-to-accelerate-ai/