CS/ECE 570
Survey on GPU Accelerators

Motivation

Due to the massively parallel processing capabilities of these specialized processors, GPU accelerators have grown in popularity in recent years. Numerous computer applications, such as machine learning, deep learning, scientific simulations, data analytics, and general-purpose processing, can benefit significantly from GPU acceleration.The capacity to accelerate the processing of massive amounts of data is one of the primary reasons for utilizing GPU accelerators. This is crucial for machine learning and deep learning applications because these techniques train massive neural networks by analyzing lots of data. GPU accelerators can dramatically accelerate the learning process, enabling the creation and deployment of machine learning models more quickly.The capacity of GPU accelerators to carry out complex simulations with a high degree of accuracy and realism is another reason to use them. In order to describe complicated physical systems, scientific simulations like molecular dynamics, fluid dynamics, and computational fluid dynamics need to process a lot of data. These simulations can be considerably accelerated by GPU accelerators, allowing for quicker scientific research innovation and discovery.

Additionally, general-purpose processing and data analysis can both be accelerated by GPU accelerators. The ability to analyze data fast and effectively is vital because of the rising volume of data being produced nowadays. GPU accelerators may significantly speed up general-purpose computation and data analytics, enabling quicker insights and decision-making.Another motive is the ability to bring GPU power to the network edge, enabling low latency and high bandwidth processing. This is accomplished by using GPU accelerators in edge computing. This creates new opportunities for a variety of applications, including automated factories, smart cities, and self-driving cars.Overall, the use of GPU accelerators has the potential to significantly speed up a variety of computing tasks, allowing for the quicker and more effective processing of large amounts of data and opening up new perspectives and scientific discoveries in data analytics, general-purpose computation, and scientific research.

Background

The background section of report on GPU accelerators should provide an overview of the history, development, and current state of GPU accelerators. Here are some key points you may want to include in this section:

The origins of GPU accelerators: The use of graphics processing units (GPUs) for general-purpose computation dates back to the early 2000s, when researchers first began to explore the idea of using the massive parallel processing capabilities of GPUs for tasks beyond computer graphics. The development of GPU accelerators: Over the years, the capabilities of GPUs have continued to evolve, with hardware manufacturers developing specialized GPU accelerators that are optimized for a wide range of computing applications.The current state of GPU accelerators: Today, GPU accelerators are widely used in a variety of fields, including machine learning, deep learning, scientific simulations, data analytics, and general-purpose computation. The use of GPU accelerators is becoming increasingly popular due to the massive parallel processing capabilities of these specialized processors. Advancement in GPU accelerators: Advancements in deep learning and AI have also led to the development of more sophisticated GPU accelerators that are specifically designed for deep learning tasks, such as NVIDIA's Tesla V100 and A100.

The role of GPU accelerators in the current HPC market: GPU accelerators have become an integral part of High Performance Computing market, and are being used to accelerate a wide range of applications, including simulations, data analytics, and machine learning.The role of GPU accelerators in Edge computing: With the increasing amount of data being generated, the ability to process this data quickly and efficiently is becoming increasingly important. GPU accelerators can provide significant speed-ups for data analytics and general-purpose computation, allowing for faster insights and decision making. With the advent of edge computing, the use of GPU accelerators in edge devices is becoming increasingly popular.Conclusion of the background: This section should give an overview of the development of GPU accelerators and their current state, and should give a clear picture of the role of GPU accelerators in the current high performance computing market.

Timeline:

Week-3:

Motivation, Background work on GPU accelerators

Week-4 to 5:

Literature survey on the use of GPU accelerators in different fields such as machine learning, scientific simulations, data analytics, and general-purpose computation, as well as the latest developments and trends in the field.

Week-6 to 7:

Collect data on the costs and performance of different GPU accelerator products, including hardware and software, and compare them to alternatives such as CPU-based computing.

Week-8 to 9:

Analyze the data collected from the literature review and survey, and draw conclusions about the current state of GPU accelerators and their applications. Write up the report, including an introduction, background, methodology, results, and conclusion sections.

Week-10:

Prepare presentation

Roles:

Manikanta Ranganath: Responsible for collecting information on GPU accelerators in the field of machine learning and data analytics and collecting information about GPU accelerators products which are launched recently.

Bharath Kumar Reddy Gangavaram: Responsible for collecting information on GPU accelerator architectures, and the latest trends in the architectural changes in GPU-based accelerator products.

Mahendra Kodidala: Responsibility includes collecting the data and comparing the different accelerator products, listing the alternative CPU- based computing models.

References:

"A Survey of GPU-Accelerated High-Performance Computing for Deep Learning" by Sainbayar Sukhbaatar, Jason R. Mitchell, and Rob Fergus.

"Accelerating Deep Neural Networks on GPUs" by Hao Li, and Roger Grosse.

"Hardware Accelerator Design for Machine Learning" by Li Du, and Yuan Du.

"NVIDIA cuDNN" by Nvidia Corp.