Assignment

Subject: P&DC

Instructor: Sir Asif Laghari

Student: Mehmood Hassan

Id: CSC-22F-144

Topic: Energy Efficiency Considerations

Introduction

Parallel and distributed computing (P&DC) finds significant application in large-scale tasks in the field of artificial intelligence, simulations, pall computing, and financial modeling, and has become a necessary component of present-day computing. The similar system although offering unambiguous computational ability which has not been developed forward, the power consumption is high, a factor that raises concerns over the cost of operation, the sustainability of the tackle and environmental

concerns. In this respect, energy efficiency in P&DC is the determinant of the reduction in electric consumption and the increase in popularity of green artificial intelligence and mitigation of carbon footprint.

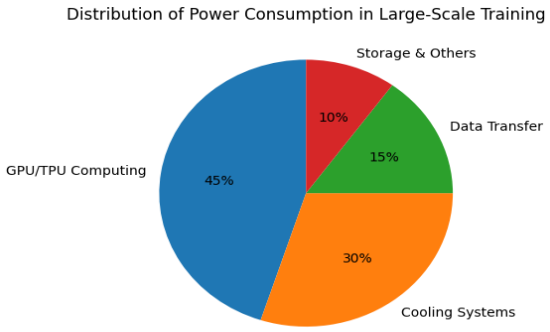
This paper discusses three key aspects of energy efficiency in P and DC, namely: (1) training power consumption on a large scale; (2) parallel algorithms that consume less energy; and (3) Green AI and reduced carbon footprint.

1. Power Consumption of Large-Scale Training.

Modern machine learning, in particular deep learning, requires a lot of processing power. Transformer based architectures such as GPT, BERT, and others, are large models that need huge datasets and billions of parameters to train. This means that large quantities of energy are required.

1.1 Conditions that lead to High Power Consumptions.

- **GPU and TPU Usage**: training at the scale of large models is dependent on the usage of high performance GPUs and TPUs. Each GPU has a power consumption of 300-400 watts and training often needs hundreds of these work units running simultaneously over many long periods.
- **Long Training Time**: Large models require days or even weeks of continued calculation which is highly energy consuming.
- **Memory and Data Transfer**: Distributed training has a considerable amount of data transfer as far as gradient synchronization and parameter updates are concerned. There are other processes of energy consumption in the communications..
- **Cooling Requirements**: To keep the operating temperatures at the optimal levels, the data centers require cooling systems.30-50 percent of the power consumed could be attributed to cooling.



The distribution of power among computation, cooling, data transfer, and storage in case of training a large model is depicted in Figure 1.

1.2 Real World Examples

- GPT-3 training consumed 1000 of kWh of electricity.
- The AlphaGo Zero training on the clusters of GPUs exhibited considerable consumptions of energy.
- Cloud services providers such as AWS, Google Cloud and Azure have invested in energy efficient infrastructure to reduce their costs and their carbon footprints.

1.3 Implications for P&DC

High power use affects:

- Expenses of operating data centers.
- Sustainability of the environment because of carbon emissions.
- Hardware life, excess heat may reduce the life of the components.
- Large scales of large computational loads.

2. Parallel Algorithms that use less energy.

Conventional algorithm design is concerned with time complexity. The energy efficiency is also important in the modern parallel computing. Very energy-inefficient algorithms used in these large-scale systems apply quickly. Efficient algorithms are created to minimize or cut down the computation time as well as energy consumption.

2.1 Definition

A parallel energy efficient algorithm is one which:

- Reduces the amount of the active processor utilization.
- Reduces the unnecessary computations.
- Restrains cross processor communication.
- Eliminates waiting time or synchronization time.
- Optimises use of resources.

2.2 Principal Methods

- **Load Balancing**: In ensuring the fair distribution of workloads, the idle CPUs are avoided wasting power that would otherwise go to waste.
- **Communication Overhead Reduction**: Compression data transfer between nodes, as well as communication-avoidance strategies are minimized.
- **Asynchronous Algorithms**: Use asynchronous processors so that they do not waste energy due to hurdles in synchronization.
- **Specialized Hardware**: FPGAs, TPUs, and general purpose GPUs are more efficient at some tasks than the general purpose ones.
- **Memory-Efficient Designs**: Memory operation often consumes more energy than consumption so optimization of memory access and locality result in less energy consumption.

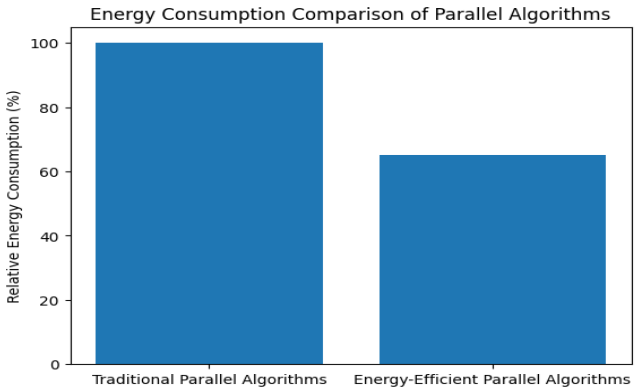


Figure 2 compares the relative energy consumption of traditional parallel algorithms with energy-efficient parallel algorithms.

2.3 Some Examples

- Parallel matrix Multiplication algorithm reduces the amount of communication among nodes.
- Asynchronous gradient descent methods are applied in distributed deep learning.
- Hardware-considered supercomputing clusters.

3. Green Artificial Intelligence and Carbon footprint.

Green AI is a subject of creating intelligent systems that would be highly efficient, and they would not harm the environment. The volume of the carbon generated by the big artificial intelligence models as they are being trained is now a problem. What we require in ensuring that artificial intelligence is good, to the earth, is green AI systems.

3.1 Importance of Green AI

Green AI is concerned with creating intelligent systems that can be extremely effective, and still have no damaging effects on the environment. It is an issue now that big artificial intelligence models produce such amount of carbon during the process of training them. It is the green AI systems that we require to ensure that artificial intelligence is good, to the earth.

3.2 Principles of Green AI

- **Efficiency Over Size**: I believe that it is preferable to concentrate on architectures that are efficient, and make use of smaller models with the same job and less energy. In this manner the models are more efficient such as the Efficiency Over Size idea. They also use less electricity which is not a bad idea, in Efficiency Over Size.
- **Model Compression**: helpful indeed since it contains techniques. Model Compression makes use of pruning, quantization and knowledge distillation among others to ensure that it works better. These properties in the Model Compression serve to minimise the requirements that Model Compression requires. This is well, as Model Compression since it simplifies things.
- **Renewable Energy Data Centre**: are important indeed. Google and Microsoft among others would like to reduce emissions. They exploit power and wind energy and hydropower to get

their data centers running. In this manner, Google and Microsoft and other firms can be able to supply energy to their Renewable Energy Data Centers without damaging the environment.

- **Carbon-Conscious Scheduling**: Schedule your work to the periods when the renewable energy is easily accessible.
- **ReUsing Pretrained Models an Idea**. We can take ready-built models of making new ones initially. This will be the tuning of pretrained models. Doing this can actually be of great help to the environment since it can cut to 90 percent of the energy that is consumed. It is an offer, to Reusing Pretrained Models.
- **Universal Frameworks**: TensorFlow, PyTorch, and JAX are typified by efficient Frameworks.

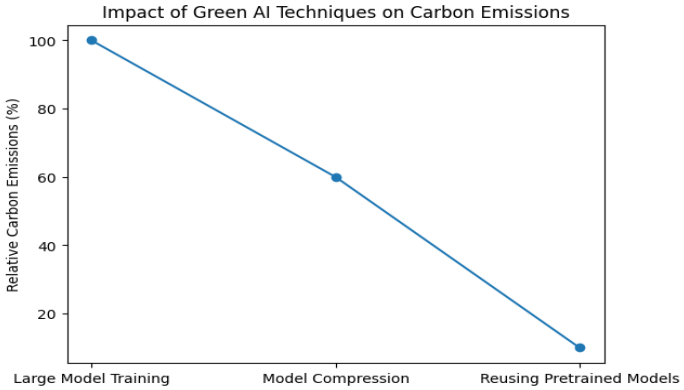


Figure 3 shows that how Green AI techniques help in reducing the carbon emissions in AI model development.

3.3 Benefits

- Considerable cut in carbon emissions.
- Lower operational costs.
- Quickened research processes and eco-friendly hardware.
- Promotes business sustainability and international environmental agenda.

4. Challenges with the Face and Future Projections.

The problems to solve are numerous, in terms of making P&DC energy efficient based on the advances made. There are still a number of obstacles on the way of achieving energy efficiency in P&DC:

- **Scalability**: Scalability is significant as we should ensure that algorithms are efficient and do not consume lots of energy even when they are implemented in a very large scale. We must make sure that scaling of algorithms is efficient to scale at energy. This is essential to the success of scalability of algorithms.
- **Heterogeneous Systems**: such systems are quite interesting. The usage of energy consumed by Central Processing Units, Graphics Processing Units, Tensor Processing Units and Field Programmable Gate Arrays is fairly balanced in Heterogeneous Systems. This is the energy

consumption balance of Central Processing Unit, Graphics Processing Unit, Tensor Processing Unit and Field Programmable Gate Arrays, which makes Heterogeneous Systems so handy.

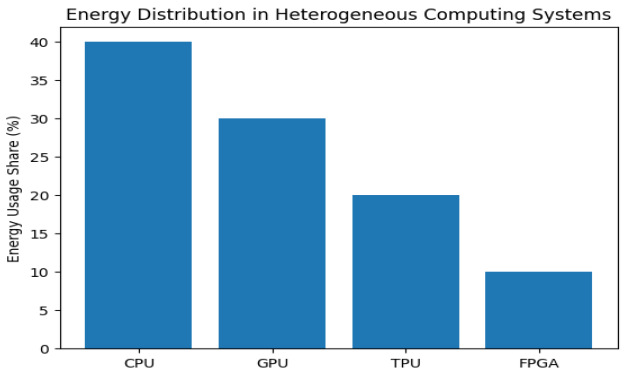


Figure 4 illustrates the energy consumption profile of heterogeneous components of computing.

- **Measurement and Benchmarking**: The standardisation of energy usage measures.
- **Policy and Incentives**: Publicizing the use of Green AI in the industry.

Future directions include:

- Creation of energy-sensitive scheduling systems of heterogeneous clusters.
- The renewably powered edge computing of distributed workloads.
- Adding AI energy monitoring into the clouds.
- Hardware accelerator innovation of AI and HPC with low energy.

Conclusion

Parallel, Distributed Computing must be energy. The reason is that we have to make more calculations and we have to consider the environment. It consumes much energy when we conduct large scale training. The reason behind this is that the computers are employed in a period of time when we have to transfer a lot of data and the computers become extremely hot.

We can also save energy by using the non-wasting computer power algorithms. Such are the algorithms that are quicker. Do not have the computers in the idle state. So there is the thing of Green AI. This aids in cutting down the things we introduce to the air by making the models to be more efficient with the use of natural energy and also by scheduling things such that the computers do not strain too much.

When we put these concepts together Parallel and Distributed Computing systems may be effective and even benevolent, to the environment. This implies that we are able to have technology and simultaneously take care of the earth. PDCS systems are capable of sustainable performance.

It is quite significant to use energy- methods. It is what we must do to grow Artificial Intelligence, cloud computing and High Performance Computing ecosystems. We must ensure that energy-practices are applied to Artificial Intelligence and High Performance Computing ecosystems to grow themselves.

References Recommended.

I used to find out about energy efficiency and Green AI and the amount of power computers consume when collaborating with one another by consulting some books and research articles. I have also browsed some channels. These items made me know the fundamentals of computer use of energy, as well as demonstrated to me how this is applied in the real world. They informed me of computing practices and Green AI and energy efficiency.

1 Strubell, E., Ganesh, A., and McCallum, A. (2019).

Deep Learning in Natural Language Processing Energy and Policy Consideration.

These models are quite expensive in terms of energy cost. It further discusses the issues that this has on the environment and the concerns that it would bring to those who formulate the policy. NLP models are a part of this. It is highly significant to be aware of the influence of NLP models and AI research on the world in life. This paper helps us do that.

2 Schwartz, R., Dodge, J., Smith, N. A., and Etzioni, O. (2020).

This paper concerns Green AI. The concept of green AI is compulsory. We must ensure that the AI models do not consume a lot of energy. People are currently putting an emphasis on creating AI models that are extremely precise and capable of doing a great number of things.. Another thing to consider is to develop Green AI models which consume less energy. In this article, we get to know the reasons why we should come up with energy- AI models. It provides a basic point, to make AI practices more sustainable and oriented towards Green AI.

3 Big Neural Network Training and Carbon Emissions.

In this paper, we will examine the carbon footprint associated with training neural networks. It discusses how the equipment, the location where all the computers, the energy they consume can help a lot in the quantity of bad stuff we inject to the air. The article is actually, on the carbon footprint of networks and what makes it larger or smaller.

4 Google Scholar

I searched Google Scholar to get research papers that have been reviewed by individuals. Such papers were required to be on how parallel algorithms can be made to consume energy, Green AI and big systems capable of training a lot of data. On Google Scholar, I was searching on energy- parallel algorithms, Green AI and large-scale training systems.

Website: https://scholar.google.com

5 Code - Green AI Section Papers.

It is a very useful tool since it bridges the gap between the research that academicians are engaged in and life practice. It can be used by individuals who are interested in information on the progress, in energy-efficient machine learning and sustainable AI models. The platform helps one see easily how these new ideas can be applied in the world. The focus of this platform is on energy-efficient ml and sustainable AI models.

Website: https://paperswithcode.com

6 Distributed Systems MIT OpenCourseWare.

To gain knowledge related to distributed systems and resources management, I used lecture notes and course materials of the MIT OpenCourseWare. This also enabled me to know how to make systems work properly. My interest was to understand the fundamentals of distributed systems, resource management and system level efficiency. I have read the materials of MIT OpenCourseWare.

Website: https://ocw.mit.edu

7 IEEE Xplore Digital Library

The IEEE Xplore provides individuals with an opportunity to read excellent research papers and journals related to parallel computing and energy-conscious system design and high-performance computing architectures. IEEE Xplore contains much information, concerning computing. IEEE Xplore allows people to know about energy- system design and high-performance computing architectures as well.

Website: https://ieeexplore.ieee.org