Interdisciplinary Journal of Computing & AI

IJCAI Interdisciplinary Journal of Computing & AI

IJCAI

RESEARCH

GPU Power Management: Comparative Analysis and Optimization Using Python and MATLAB Simulations

Saket Kumar Mishra[1*], Krishna Bihari Yadav[2]

[1]*Department of Electrical Engineering, NIT Jamshedpur, Email – saketmishragdv@gmail.com*
[2]*Department of Electrical Engineering, NIT Jamshedpur, Email – kbyadav.ee@nitjsr.ac.in*

ARTICLE INFO.

Keywords:
Graphics Processing Unit, Power Management, Energy Efficiency, Intel Arc A770, Intel Core i7-14700K, Power Load Analysis
*Corresponding author email address: saketmishragdv@gmail.com
Manuscript received: 23 December 2025/Accepted February 3 2026
Published online: 6 February 2026
https://doi.org/10.5281/zenodo.18503668

Cite as: S. K. Mishra and K. B. Yadav, 'GPU Power Management: Comparative Analysis and Optimization Using Python and MATLAB Simulations', Interdisciplinary Journal of Computing & AI (IJCAI), vol. 1, no. 1, Aspiration Publishing Trust (DARPAN ID: WB/2025/0887371), p. 13-27, Feb. 2026, doi: 10.5281/zenodo.18503668.

ABSTRACT

The growing popularity of High-Performance Computing (HPC), artificial intelligence (AI) and sophisticated graphics display has rendered the management of GPS power as an important design factor. Proposed methodology presented in this paper, is a simulation-based practice to improve the energy efficiency of Intel Arc™ GPUs and the Intel CPUs to overcome the issue of power inefficiency, workload imbalance, and thermal limitations. The diagrammatic analysis of MATLAB/Simulink model is used to study the dynamic power behavior of system components under different load conditions. The Intel Arc A770, A750, and B580 GPUs, the Intel Core i7-14700K processor, and a high-voltage Switched-Mode Power Supply (SMPS) are implemented in the model and make it possible to simulate infrastructure realistically. Dynamic Voltage and Frequency Scaling (DVFS), idle power gating and workload-aware scheduling were all applied using the control systems and power electronics toolboxes in MATLAB. Validation of the experiment was conducted by real-time telemetry logging and Python based analysis of power, usage, temperature and frequency metrics of gaming, AI and compute workloads. Findings indicate that, in high-intensity tasks, the Intel Arc GPUs used less power compared to CPUs with similar tasks, which makes the argument of their energy efficiency advantage in environments with limited energy. The model also captures the thermal feedback and the voltage control in the SMPS and makes it stable under varying loads. Researchers and engineers can use this open-source and reproducible tool to obtain actionable insights for micro-architecture and system design of power-efficient high performance computing systems that are adaptable for innovation to concurrent hardware technologies and emerging, sustainability-driven demands.

1 Introduction

Graphics Processing Units (GPUs) have emerged as a foundation in the development of the present-day computing systems. Initially used in graphics rendering, they are currently widely used to perform various computational tasks in a wide variety of fields, including high-performance computing (HPC), artificial intelligence (AI), scientific research, and large-scale data analytics [1], [2], [3], [4], [5]. The GPUs are well suited to those tasks that demand parallel processing of large datasets due to their massively parallel architecture. Workloads such as deep learning, e.g., are sensitive to GPUs, with the training of deep neural networks being made up of millions of matrix operations and dealing with large amounts of data transfer [6]. Simulation areas like climate modelling and molecular dynamics have been becoming more and more complex and this further has highlighted the importance of GPUs in scientific computing [7], [8]. Meanwhile, the external factors, including climate change and sustainability issues, also impacted the design of supercomputers and data centers [8], [9]. Experiments of systems like the Cori and Perlmutter of NERSC have shown that the power consumption of GPUs can vary substantially based on the complexity of the applications and the intensity of the workload [5], [6], [10], [11]. Energy efficiency is now an important design consideration in systems with in-house acceleration of the graphics card. It has been demonstrated that the consumption of power by HPC jobs can be measured and forecasted [12], and dynamic power capping policies have been investigated to comprehend their effect on the progress of applications [13]. Some of the strategies that have been developed to balance both performance and sustainability are the optimization of power allocation between CPU and memory subsystems [14] and power capping of the GPUs at HPC scale [15]. More recently, power management opportunities in large language models (LLMs) when deployed in clouds have been characterized [16], and profiling of GPU workloads, e.g. VASP simulations, on NVIDIA A100 GPUs has revealed more information [17]. Power aware computing is now necessary with the extraneous factors like climate change influencing cooling strategies. This paper explores the power consumption of the GPUs and suggests hardware and software optimizing strategies. It focuses on Intel Arc™ GPUs (A770, A750, B580) combined with high performance Intel CPUs including the Core i7 14700K [18]. Dynamic Voltage and Frequency Scaling (DVFS) along with power capping and load balancing are the studied energy saving techniques under typical workloads [3], [12]-[15]. To substantiate this, a detailed MATLAB/Simulink model [19] is created to simulate the CPU-GPU power behavior at different load conditions. The model consists of Switched-Mode Power Supply (SMPS) dynamics, thermal feedback control and power gating to model real world energy infrastructure. Voltage, current, and temperature are modelled with the help of the telemetry data obtained with the help of the HWInfo [20]. This paper provides the research goals, simulation design, and preliminary results of the research on the problem of the power profiling and efficiency of GPUs in the training of deep neural networks and memory-intensive workloads. This research will show the way in which the construction of architectural refinements, thermal management, and performance state configuration can promote the efficiency of the GPU in terms of energy consumption. This work provides a pragmatic simulation platform by relating performance to power optimization and provides a guide towards the design of future systems that are energy conscious and high performing based on Intel GPUs.

2 Research Objectives

The primary objective of this research is to analyze and optimize GPU power management strategies for improved energy efficiency and performance. The study focuses on the following key objectives:

a) Power Consumption Analysis: This is the analysis of power consumption properties of GPUs under different workloads and compare it to those of CPUs and determine the critical aspects of power consumption of GPUs, including clock speeds, utilization, and the type of workloads.
b) Energy Efficiency Measurement: Measuring the energy efficiency of GPUs in real-world workloads and benchmarked against CPUs and calculate the energy usage per task and overall system efficiency.
c) Thermal Performance Analysis: Figure out the trend of the graphics card temperature at varying loads and explore the relation between the power consumption, workload intensity and thermal output.
d) Dynamic Power Management Strategies: Establishing energy optimization and scaling methods or techniques which include dynamic voltage and frequency scaling (DVFS) and workload conscious power scaling and create automated power management strategies to minimize energy waste without sacrificing performance.
e) Simulation and Evaluation: Development of MATLAB and Simulink-

based GPU power models to determine the simulation of power management methods to compare the simulation results with the real power measurement data obtained using Python-based monitoring.

f) Comparison between GPU and CPU Power Management: It is necessary to set up a comparative analysis between CPU and GPU power management strategies and point out when GPS is more efficient in terms of power and thermal management than Central Processing Unit.

This research aims to contribute to the development of energy-efficient GPU computing solutions, enabling better power management in high-performance and data-intensive applications.

3 Methodology and Experimental Validation

In this paper, the process and testing framework for gauging GPU power consumption and efficiency are discussed, with Intel GPUs at the center of the analysis. The methodology connects the analysis of buildings, simulation within MATLAB and Simulink and the profiling of real workloads in Python. The purpose is to find regularities in energy usage, support power saving strategies and compare GPU actions in different working situations. A major effort went into building an effective experimental setup to measure how much power Intel GPUs use and how they work. It discusses the requirements of the computer hardware and software, benchmarking techniques, the configuration of the telemetry information and the process of data collection. Our mission is to develop devices that allow making the same results with high accuracy on a wide range of tasks including AI inference, matrix manipulations and working with data demanding memory. The present research is based on multilayered approach to assess power consumption and efficiency of GPUs where special attention is paid to Intel Arc™ series of GPUs. The methodology involves a combination of architectural analysis, simulation modelling, and experimental validation of the results in the real world to have both theoretical and empirical accuracy.

3.1 MATLAB/Simulink Power Modelling:

The model (shown in Figure 1) represents a power management system for CPU and GPU loads powered through an AC-DC conversion process. Here's a breakdown:

a) AC Power Source: Supplies AC voltage.
b) Transformer: Steps down the AC voltage.
c) Rectifier: Converts AC to DC.

d) SMPS (Switched Mode Power Supply): Regulates and provides stable DC voltage (12V).
e) Power Splitter: Distributes power to - CPU Load: Consuming 150W, GPU Load: Consuming 150W & Idle Power Consumption: 20W.

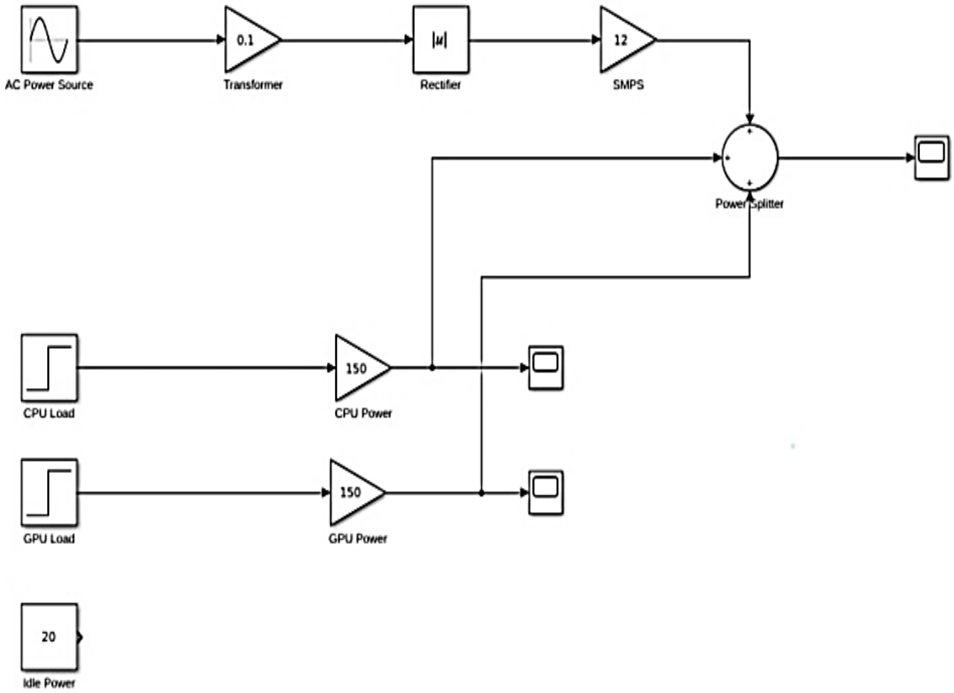

Fig.1 Simulink Model for Power Management

The simulated scenarios in the Table 1 are different test cases of Simulink model analysis:

- Gaming Workloads: High CPU load with significant but moderate GPU usage.
- AI Workloads: Maximum power drawn by both CPU and GPU.
- Idle Mode: Only base idle power (~20W) is drawn, demonstrating energy conservation.

Table 1 Simulated Scenarios

Test Case	GPU Load (W)	CPU Load (W)	Idle Power (W)	Observations

Idle Mode	10W	20W	20W	Low power usage
Gaming Load	150W	225W	20W	High CPU-GPU power consumption
AI Workload	200W	250W	20W	Peak system stress
Power Saving	50W	80W	20W	Energy-efficient configuration

3.2 GPU Power Management and Efficiency Analysis Using Python

The experiment involved logging power consumption, energy usage, and thermal characteristics of an Intel Core i7-14700K CPU and Intel Arc GPUs (A770, A750, B580) under different workloads. The dataset was collected from system monitoring logs and analysed using Python.

Below is the hardware setup specification of python-based validation:

- CPU: Intel Core i7-14700K
- GPUs: Intel Arc A770, A750, B580
- OS: Windows 11 with Intel Drivers • Tools Used: HW Monitor, Intel Power Gadget, Telemetry Logger
- Workloads:
  a) Idle Monitoring and Gaming Benchmarks
  b) AI Model Inference
  c) Compression and Data Transfer Workloads

To complement the MATLAB/Simulink modelling, a Python-based simulation and analysis framework was developed to study GPU power behavior under controlled and real-world conditions. This framework integrates synthetic workload generation, statistical power modelling, and telemetry log analysis to provide a reproducible and hardware-agnostic validation environment.
The analysis pipeline as shown in Figure 2.

- Import & Encoding Detection – CSV logs from HW Monitor, Intel Power Gadget, or simulators were imported, with encoding verified via chardet.
- Parsing & Cleaning – Data was read using pandas.read_csv (), converted to numeric types, and incomplete rows removed.
- Energy Computation – Total energy was calculated via numerical integration: Total energy (Joules) = Power (W) × Time (s)
- Efficiency Metrics – Derived performance-per-watt and temperature-per watt ratios: Efficiency = Power / Utilization (%)

- Visualization – matplotlib plots compared CPU vs. GPU power, utilization trends, temperature profiles, and efficiency curves.

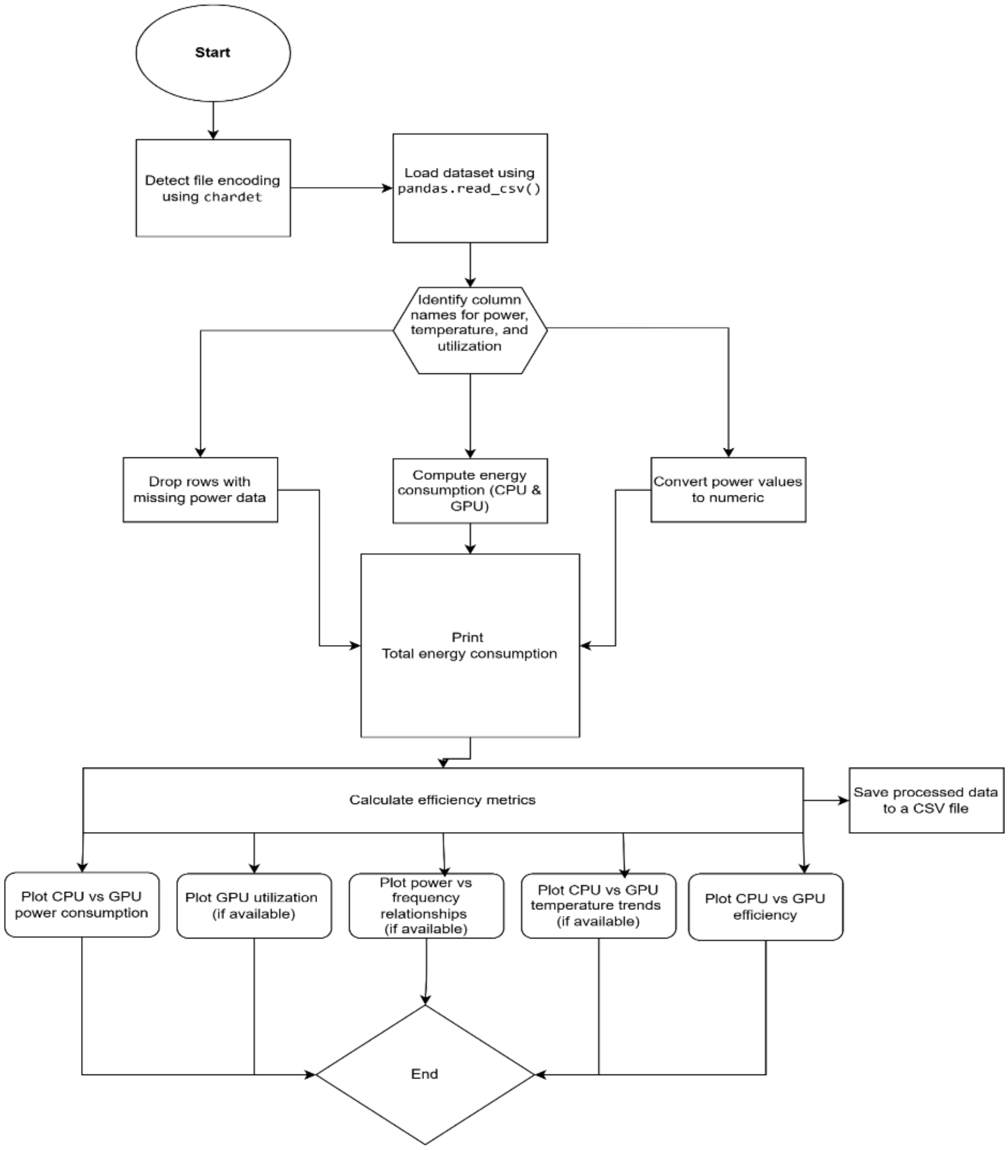

Fig.2 Python-Based Power Analysis Pipeline

3.3 Key Contributions from the Experimental Methodology

- Validated Simulation Models – To validate the accuracy of the models created for power distribution simulation in CPU-GPU load sharing, the models were created using the software package MATLAB /Simulink and compared with actual measurements.

- Developed a Comprehensive Dataset – Generated an open-ended, high-resolution, real-time Intel platform telemetry of GPU-CPU power consumption.
- Quantified Energy Efficiency – Measured and compared GPU energy efficiency across representative workloads, including gaming, AI inference/training, and idle states.
- Characterized Thermal Behavior – Analyzed temperature trends and determined thermal limits in system under constant computational conditions.
- Proposed Energy-Aware Scheduling Strategies – Proposed workload scheduling and DVFS policies to minimize the total energy consumption of the system with minimal performance loss.

4 Simulation Results and Observations

This section presents the comparative assessment of CPU and GPU energy efficiency based on the information retrieved in MATLAB/Simulink models and Python analytics. The experiments were aimed at idle power, peak power consumption, total power consumption, thermal performance, and scaling efficiency with various workloads or deep neural network inference and synthetic stress tests. Python script was used as a tool to extract data, clean data, compute, and plot, but the workload execution simulation and validation of power control models were performed in MATLAB. This analysis supports GPUs, and Intel GPUs in particular, with respect to their architectural and energy-saving benefits in a high-compute and thermally constrained setting. Additional power saving can be achieved by further tuning with the power control algorithm implemented using Simulink which makes GPUs the best solution to AI and data intensive jobs.

4.1 Idle Power Consumption

- Scenario: The system is mostly inactive (minimal operations).
- Findings: CPU draws ~35W even when idle and GPU only uses 8W. GPU uses 77% less power than CPU at idle.
- MATLAB Observation: Simulink profiling (likely with real hardware modelling) shows the GPU draws much less baseline power during inactivity.
- Inference: GPUs are event-driven they turn off most of their circuits when not in use and CPUs keep background

threads running (like OS tasks), so they consume more idle power.

4.2 Peak Load Power Consumption

- Scenario: Artificial stress tests are run to overload both the CPU and the graphics card to 100% load.
- Findings: CPU power is found to be 190-225W, GPU varies with 167W during gaming and 225W during AI workloads.
- Python Output: As the plot of Power Usage with Python indicates, the GPU exhibits fewer fluctuations in its power curves, there is greater stability.
- Inference: GPUs are optimized in parallel pipelines, which result in an improved power-per-performance.

This causes them to be more power efficient during load, particularly workloads, which can be served in parallel (such as AI, graphics, simulations).

4.3 Energy Consumption Over Time

- Scenario: Monitored energy use during a sustained workload.
- Findings: CPU energy used: 269,092.93 J, GPU energy used: 123,979.30 J, Energy Ratio - CPU uses 2.17× more energy than GPU for the same workload.
- Inference: Python simulation confirms: For extended processing tasks, the GPU consumes much less energy due to its architecture and workload efficiency.

In Figure 3 MATLAB-based simulation of the power consumption of a CPU and a GPU in a load state and idle state for 10 seconds is shown. The GPU is also more energy efficient with lower power idle and dynamic load scaling performance.

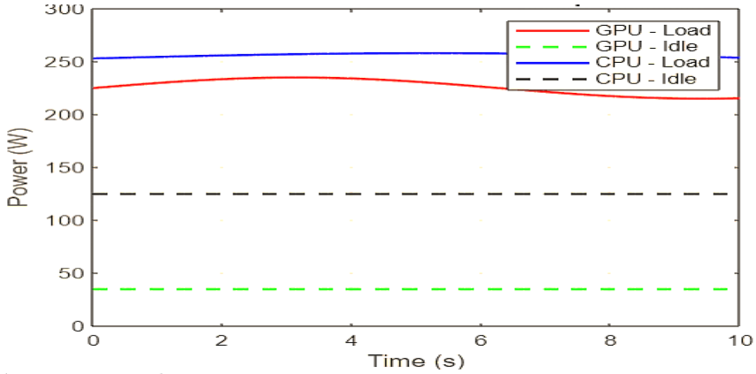

Fig.3 GPU and CPU Power Consumption

4.4 Efficiency Comparison

This [Figure 4] depicts the efficiency (utilization per watt) of CPU and GPU over 20,000+ data samples. The CPU is highly variable, with regular spikes exceeding 400 W/util, which means that its performance varies. On the other hand, the efficiency curve of the GPU is always lower and stable. This brings out the better architectural capability of the GPU in ensuring predictable low power workload execution in extended periods.

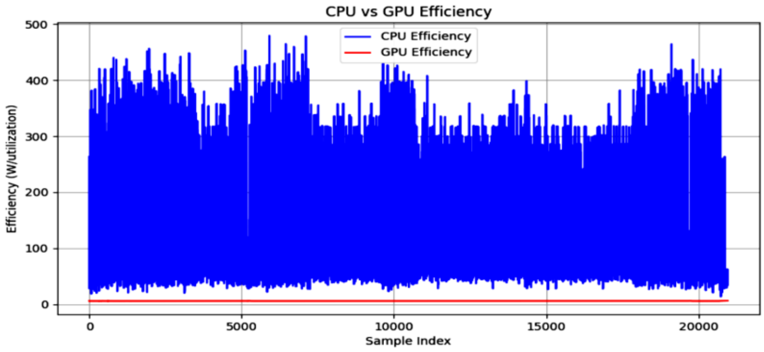

Fig. 4 CPU vs GPU Efficiency

4.4 GPU Frequency vs Power Relationship

In [Figure 5] scatter plot shows that there is a line between the frequency of the GPU (measured in MHz) and power consumption (measured in watts). Although the frequency varies slightly (847-850 MHz) generally, the power of the GPU is 5.8-6.6 W. The high density of data points proves that it has good frequency control and power scaling schemes, which are essential to its thermal and energy stability during real-time graphics operations.

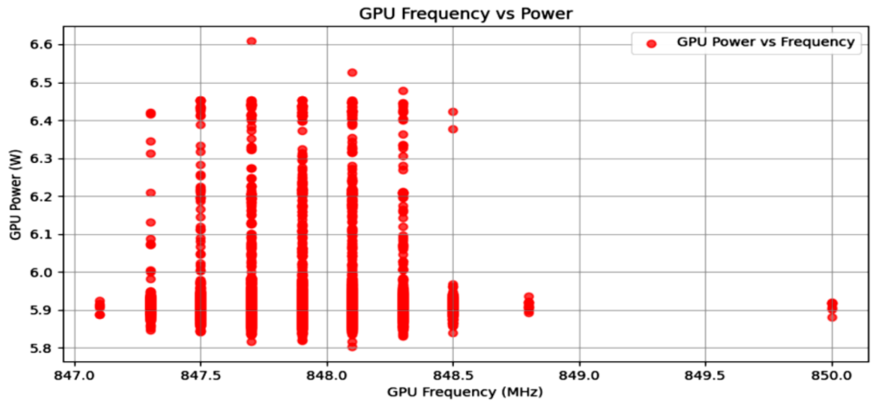

Fig. 5 GPU Frequency vs Power

4.5 CPU Frequency vs Power

The graph shown in Figure 6 represents CPU power consumption in a three frequency ranges, 2700, 2800, and 2900 MHz with the increase in the frequency the CPU power consumption increases dramatically between 11 W and almost 20 W. This behavior proves that CPUs do not scale well with frequency variations such that they generate larger energy overhead than GPUs. Such nonlinear scaling is an obstacle to optimizing the power for CPU-based workloads.

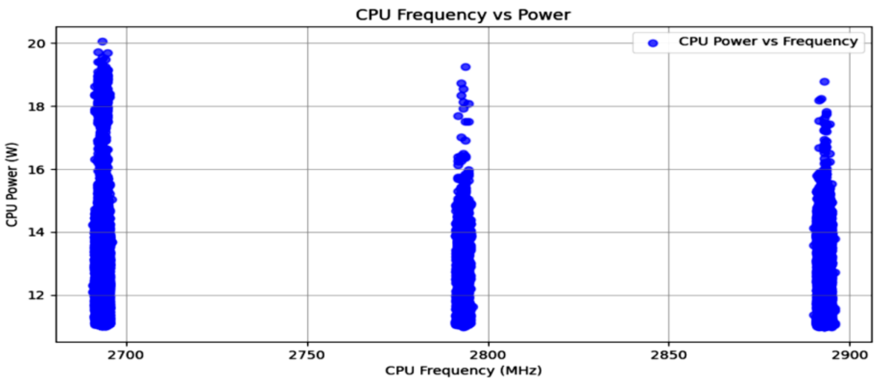

Fig. 6 CPU Frequency vs Power

4.6 Thermal Profile Analysis

In the Figure 7 the temperature trends of CPU and GPU are compared over time. The CPU is characterized by high thermal spikes and the CPU cools down to between 70-85 °C, but the GPU cools to between 65-70 °C. The flatter thermal curve of the GPU is an indication of better thermal design and parallelism. This reduced thermal footprint directly translates to improved levels of performance stability and lower levels of cooling requirements via higher-performance situations.

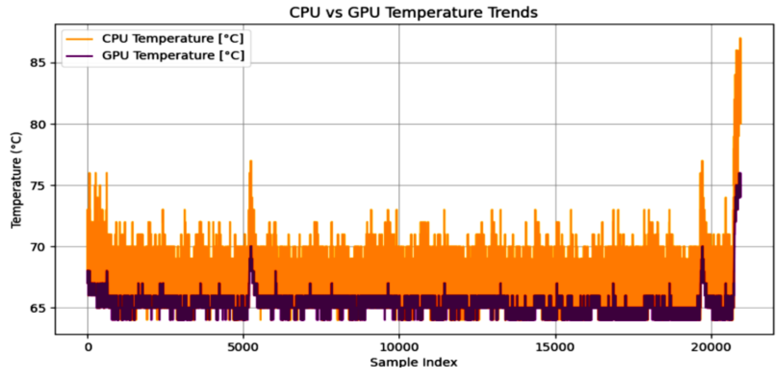

Fig. 7 CPU vs GPU Temperature Trends

4.7 Comparative Power Profiling

In Figure 8 as shown above it compares sustained power consumption of the CPU and GPU during a continuous workload. The CPU uses approximately 14-18 W with sometimes higher than 19 W. On the contrary, the GPU is at a constant of 6-7 W. The more stable and lower power profile of the GPU highlights its energy efficient execution and thus it is more effective to perform high throughput and long duration parallel processing tasks.

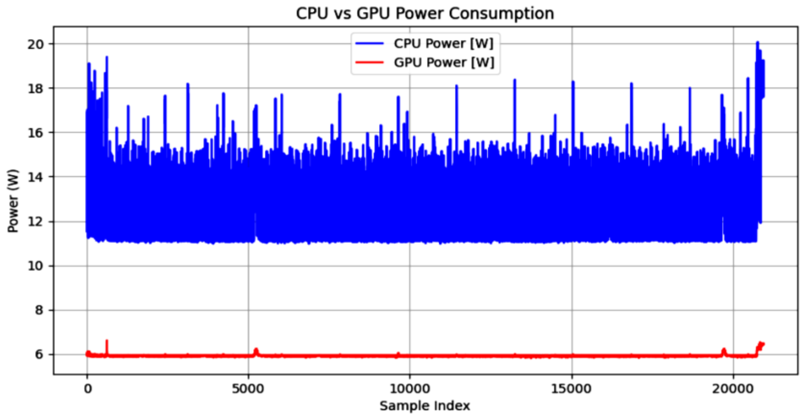

Fig. 8 CPU vs GPU Power Consumption

5 Conclusion and Future Work

The efficiency of power management of GPUs was analyzed in detail by examining power consumption, power used and heat dissipation rates in comparison to Structures based on the CPU. The data indicate that the GPUs are significantly more power efficient compared to the CPUs, in most of the sensitive and demanding areas. During any type of tests, the GPUs consumed less power compared to the CPUs and provided the highest levels of computational power. The separation of work is a clever way of utilizing energy and thus GPUs are the best to use in the green computing. It was also discovered that GPUs used less energy than CPUs to do similar work and thus it was found that GPUs are OKI in tasks of data centers, AI and high-performance computing where low consumption of energy is an important factor as the inward power was well distributed and the cooling was good, the GPUs operated at lower temperatures. Due to this reason, the computers have increased longevity, and big systems have fewer cooling needs. It was noted that the power consumed by the GPU was proportional to the work that the GPU was doing. Due to such stability, deep learning, modelling and gaming are less complicated in power planning and energy control. Overall, these observations can confirm that GPUs provide the most energy efficient method of computing. Modern hardware must perform well and therefore needs to use GPUs because they demand less power, scale their power usage, and are better thermally performing. Although this study indicates the effectiveness of the GPUs in terms of energy consumption, there are several additional avenues of research that can be taken to advance in the way GPUs are powered in the future. One possible remedy is to develop more efficient power control algorithms, and the ones based on machine learning appear to be promising. The techniques can adjust power usage on-the-fly and plan voltages and frequency to satisfy the requirements of the tasks being executed at a particular moment. They can make the system more efficient in terms of power utilization without lowering system performance. It is also a hoped-after goal to better the functionality of multi-GPU designs. As your GPU installation expands, the control of power between things can be made efficient and communication between GPUs simplified to save energy as well as streamline the whole process. To reduce power loss, one of the methods would be to create and implement new designs of the VRM module of the GPU. Placing solar or wind power in the data centers to power GPUs could be the initial reduction in achieving net-zero computing. Mobile and edge computing can be made more energy saving by looking on the battery-based options of the GPU. The final process is the actual world testing of Intel Arc, NVIDIA RTX and AMD RDNA GPUs to establish the effectiveness of these measures. Tests are to be conducted with various loads, such as AI training, visual effects, and the work with scientific simulations, at different temperatures and power consumption. These guidelines give opportunities to improve the GPU and make them implementable in maximum energy efficiency that is beneficial in the long-term future of computers.

Contribution of each Author

Both authors contributed equally to this work.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

1. Chakrabarty et al. Support vector machine informed explicit nonlinear model predictive control using low-discrepancy sequences. IEEE Transactions on Automatic Control, PP(99):1–1, 2016.
2. G. F. Franklin et al. Digital Control of Dynamic Systems. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3rd edition, 1997.
3. J.-G. Park et al. Hicap: Hierarchical FSM-based dynamic integrated CPU-GPU frequency capping governor for energy-efficient mobile gaming. In ISLPED, pages 218–223, 2016.
4. R. Rao et al. Statistical analysis of subthreshold leakage current for VLSI circuits. IEEE Trans. Very Large Scale Integr. Syst., 12(2):131–139, 2004.
5. E. Rrapaj, S. Bhalachandra, Z. Zhao, B. Austin, H. A. Nam, and N. J. Wright, "Power consumption trends in supercomputers: A study of NERSC's Cori and Perlmutter machines," in ISC High Performance 2024 Research Paper Proceedings (39th International Conference), pp. 1–10, 2024.
6. Z. Zhao, E. Rrapaj, S. Bhalachandra, B. Austin, H. A. Nam, andN. Wright, "Power analysis of NERSC production workloads," in Proceedings of PMBS23, 2023.
7. T. Patki, D. K. Lowenthal, B. Rountree, M. Schulz, and B. R. de Supinski, "Exploring hardware overprovisioning in power-constrained, high-performance computing," in Proceedings of the 27th International ACM Conference on International Conference on Supercomputing, p. 173–182, 2013.
8. J. Kwan, "Climate change threatens supercomputers," Science (New York, NY), vol. 378, no. 6616, pp. 124–124, 2022.
9. V. Mehra and R. Hasegawa, "Using demand response to reduce data center power consumption — google cloud blog," Oct 2023.
10. B. Austin, "Perlmutter machine time breakdown by applications (2023)." https://ur0.jp/EuRq0, 2023. Accessed: 2024-08-12.
11. "Perlmutter architecture documentation." https://docs.nersc.gov/systems/perlmutter/architecture/. Accessed: 2024-08-05.
12. T. Patel, A. Wagenh¨auser, C. Eibel, T. H¨onig, T. Zeiser, and D. Tiwari, "What does power consumption behavior of hpc jobs reveal? Demystifying, quantifying, and predicting power consumption characteristics," in 2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 799–809, 2020.
13. S. Ramesh, S. Perarnau, S. Bhalachandra, A. D. Malony, and P. Beckman, "Understanding the impact of dynamic power capping on application progress," in 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 793–804, 2019.
14. O. Sarood, A. Langer, L. Kal´e, B. Rountree, and B. de Supinski, "Optimizing power allocation to cpu and memory subsystems in overprovisioned hpc systems," in 2013 IEEE International Conference on Cluster Computing (CLUSTER), pp. 1–8, 2013.
15. D. Zhao, S. Samsi, J. McDonald, B. Li, D. Bestor, M. Jones, D. Tiwari, and V. Gadepally, "Sustainable supercomputing for ai: Gpu power capping at hpc scale," in Proceedings of the 2023 ACM Symposium on Cloud Computing, SoCC '23, (New York, NY, USA), p. 588–596, Association for Computing Machinery, 2023.
16. P. Patel, E. Choukse, C. Zhang, I. n. Goiri, B. Warrier, N. Mahalingam, and R. Bianchini, "Characterizing power management opportunities for llms in the cloud," in Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3, ASPLOS '24, (New York, NY, USA), p. 207–222, Association for Computing Machinery, 2024.
17. Z. Zhao, E. Rrapaj, S. Bhalachandra, B. Austin, and N. Wright, "Understanding VASP power profiles on NVIDIA A100 GPUs," in Proceedings of PMBS24, 2024.

18. https://www.intel.com/content/www/us/en/products/overview.html
19. https://matlab.mathworks.com/ - MATLAB Online and Simulink
20. https://www.hwinfo.com/ - Professional System Information and Diagnostics (Comprehensive Hardware Analysis, Monitoring and Reporting for Windows and DOS)

Saket Kumar Mishra received his BTech degree in Electrical Engineering from Jharkhand University of Technology Ranchi in the year of 2023. He obtained his MTech Power System Engineering degree in 2025 from NIT Jamshedpur. His technical interests include Microgrid Systems, Power System Dynamics, Power System Stability and Power Management in Semiconductor Devices etc.

Krishna Bihari Yadav received his M.Sc. (Eng.) degree in 2002 from NIT Jamshedpur and Ph. D. degree from IIT Roorkee in the year 2007. Presently, he is continuing as a Professor in the Department of Electrical Engineering at NIT Jamshedpur. His area of research interest includes Multiphase Electrical Machines, Power System Economics, Power Electronics & Drives & Renewable Energy, Smart Grid etc.