Online at: https://researchtrendsjournal.com
Indexed Journal

ISSN No: 2584-282X
Peer Reviewed Journal

INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT

Volume 3; Issue 4; 2025; Page No. 71-78

Received: 05-05-2025
Accepted: 16-06-2025

MATLAB-Based Modeling and Optimization of Energy-Efficient Data Center Operations: A Case Study in Green Computing

[1]Onell Allan Chakawuya, [2]Forhad Monjur Hasan and [3]Mostafa Mohammad Razaul

[1-3]Department of Electronics and Information Engineering, China West Normal University, China

DOI: https://doi.org/10.5281/zenodo.16729268

Corresponding Author: Onell Allan Chakawuya

Abstract

The speed of growth in demand for digital services has created an enormous rise in data-center energy usage, raising considerable doubts about its sustainability, not only to the environment but also to the operating costs. The given paper presents a MATLAB framework that could be used to explore and optimize energy consumption in the data center, specifically, intelligent workload scheduling and temperature-aware cooling. The model enables dynamic distribution of workloads in order to reduce power consumption without losing performance. Empirical analysis indicates that optimized workload packing can reduce overall IT power consumption at the workload level by as much as 19 percent, and thermal-aware cooling provides further energy savings. The obtained tool creates a replicable instrument for the investigation of sustainable computing strategies and assists to achieve more significant efforts to reduce the carbon footprint of digital infrastructures.

Keywords: Energy-Efficient Data Centers, Green Computing, MATLAB Simulation, Power Usage Effectiveness (PUE), Sustainable Digital Infrastructure, Thermal-Aware Cooling, Workload Scheduling

1. Introduction

The fast pace of digitalization that is currently taking place in modern economies has caused an unparalleled increase in data-center consumption. As the primordial coupled point among cloud computing, artificial intelligence and big-data analytics, data centers have come to form one of the largest consumers of power in the overall ICT-sector. Recent estimates showed that more than 1% of the world's electricity was being consumed by such facilities and that this number is expected to rise unless it is countered by more efficient day-to-day operations. One such issue is the obvious conflict between performance and energy efficiency. Conventional data centres regularly maintain an uneven load figure, which causes inordinate power usage by both the computing infrastructure and the cooling systems. This pressure has led to the development of energy conscious computer strategies, all of which are generally compressed in the term green computing [1]. With workload consolidation and dynamic resource provisioning and increased efficient cooling systems green computing aims to minimize the eco-footprint of IT assets.

In the continuation of research, the current paper provides a MATLAB framework that finds an application in simulating energy consumption within a small-scale data center. The first goal is to emulate dynamic workload distribution among many servers and to evaluate the overall consequence on the total amount of power consumption and cooling requirements. Within MATLAB computing environment, the proposed model will allow one to test the algorithms in scheduling and thermal control methods, with the aim of maximizing energy use with respect to a reduction in computational throughput. The major contributions of the project are: a reconfigurable MATLAB based simulation framework of energy-conscious workload scheduling, a quantitative result of power savings achieved through different load-balancing techniques and cooling approaches and a case study on the use of optimization methods to minimize Power Usage Effectiveness (PUE). The study has a repeatable and scalable research design that can be used to research green computing, and practical lessons that can be learnt to design more green data-centers.

2. Background and Related Work

The current digital world basically rides on data centers. They drive cloud computing, financial systems, scientific models and worldwide networks. Unluckily, this growth in numbers has led to some serious environmental nightmares [2]. According to a 2022 report by International Energy Agency (IEA), the global industry of data centers uses around 220 to 320 TWh of power annually, amounting to approximately 1 to 1.3 percent of the global electricity consumption. IEA estimates that this figure will increase even further with AI, edge computing and blockchain becoming even more widespread. Designers of facilities have been thinking about performance and reliability over the past years at the expense of energy efficiency [3]. This has led to a situation whereby a majority of data centers continue to perform under Power Usage Effectiveness (PUE) times which are way beyond the optimal rating of 1.0. The Green Grid consortium has introduced PUE, which is merely a measure of the total facility energy divided by the IT equipment energy. When PUE is greater than 2.0, as it commonly was the case, then twice as much energy is allocated to cooling, lighting and power distribution as compared to that directly fed into computation. Greater awareness of climate change and increased energy costs in recent years has now compelled the industry to turn towards alternatives that involve the use of greener solutions [4]. Most organizations are now adhering to the guidelines of best practices organizations such as the Green Grid to increase PUE efficiency and also save on the general energy consumed.

Speaking of energy consumption in a modern data center, the situation is dominated by four major energy consuming systems: servers and storage (taking around 45 percent of overall demand), cooling systems (taking 35 percent), networking hardware (allocating 15 percent), and various infrastructure components including lighting, power conversion units and control circuits (claiming about 5 percent) [5]. These percentages are overviewed in Figure 1. The server infrastructure is the largest unremitting source of power consumption through the incessant computational loads that support hosted applications and the execution of virtual machines. Cold comes next, particularly in older buildings whose environmental control systems are not as advanced as they ought to be. Networking equipment, which is inevitably needed, uses relatively small current. The other 5 percent is the total loads of lighting, power, conversion losses, and the auxiliary control system. Various corrective measures are being shaped to stem such ineffectiveness. Virtualization allows the dynamic redistribution of workload on physical machines, and this increases the utilization of servers. Dynamic Voltage and Frequency Scaling (DVFS) enables the processor power to be changed in real time according to real-time requirements. The movement of the workloads is done under thermal-aware scheduling with the purpose to reduce the occurrence of hotspots. When combined, all these innovations have stymied such metrics as Data Center Infrastructure Efficiency (DCIE) and Compute Power Efficiency (CPE). The present research area also examines predictive cooling, artificial intelligence resource control, and the adoption of renewable energy. The energy-saving tasks of virtualized cloud services were also coupled with performance constraints of the resource-allocation model that was proposed by Beloglazov. Dayarathna carried out a thorough survey of the methods of energy-efficient design of data centers (not physically, but the code on top and down the pile). The two studies supported the fact that simulation frameworks are required to ensure rigor is applied when testing optimization strategies prior to implementation.

In practice MATLAB has become a popular choice as a platform to conduct such modeling activities, reflecting the powerful and sophisticated matrix-processing and visualization facilities. The behavior of the workload, the thermal behavior as well as flow of energy can be simulated by the investigators in order to determine how the scheduling algorithms and controls perform [6]. The current study builds on this pathway by building a Mona Lisa architecture of simulation environment that combines workload consolidation into thermal-aware cooling to provide an all-encompassing picture of data-center energy dynamics. Rising regulatory demands and a continually expanding sustainability agenda are making data-center operators use analytic tools that can balance the needs of operations and the forces of nature [7]. These simulation environments, of the type described above, offer a low cost, high quality vehicle in which to experiment and innovate and provide the foundation to a greener, smarter pallet of digital infrastructures.

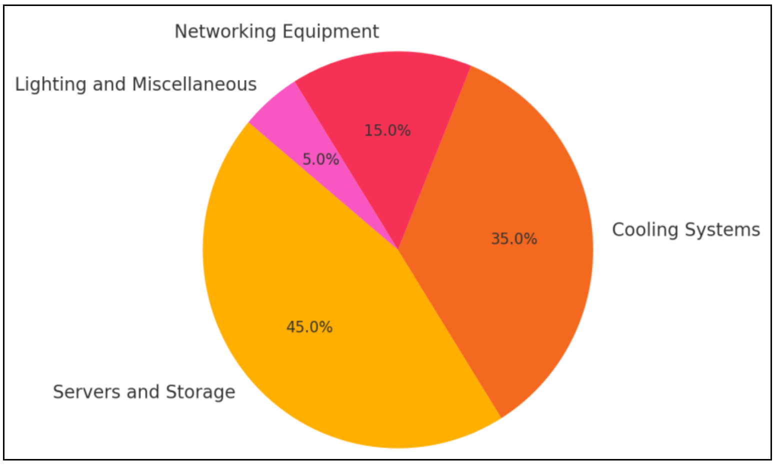

Fig 1: Estimated Global Data Center Energy Usage Distribution

3. Materials and Methods

In our work the approach that we build our methodology on is based on a simulation framework that simulates and optimizes energy consumption in a small data center in MATLAB. It is specialized so as to be able to determine the effects of server workload distribution schemes on power consumption and hence cooling being required.

We instantiate the facility modeled, which has a ten core number of servers, all of homogenous design and capable of working on an idle-to-full-load range, with workloads of minimal profiles and any heavy profile. This is based on the maximum-to-idle power ratio of 500 W versus 150 W that forms the base of a linear interpolation that forecasts consumption based on the workload intensity of each server.

In the first simulation, workloads are randomly assigned to each server thus representing an unoptimized scenario where an arbitrary loading of tasks to servers is done with little regards to energy efficiency. Having determined the load levels of all servers, we calculate power consumption according to the interpolation model and conclude that all the machines are constantly working at various loads and, as such, have a great cumulative power consumption.

To reduce this inefficiency we propose a work load consolidation algorithm. This process ranks the workloads of each server in the order of decreasing priorities and directs to the maximum utilization most of the workload on the fewest machines needed to meet the total demand. With it, it can turn off or switch into idle inactive servers, curbing the number powered up at any given time. By acting in such a manner, the algorithm will emulate energy-wise placement methodologies usually utilized in production-ready virtualized systems.

At the same time we examine two cooling arrangements. The former uses a fixed-percentage cooling load, taking as an estimate those used in older facilities, by allocating 30 percent of the total IT power to the cooling plant. The second and more advanced method integrates temperature-based control which involves calculations of the cooling power only when the temperatures of the servers reach predetermined ceiling value- normally 28 °C. The dynamic approaches to fan speed or airflow process, in calculations of excess temperature and server thermal load, replicate dynamic strategies, and thereby represent interaction between thermal demand and cooling response.

Through the combination of workload optimization and this thermal-sensitive model of refrigeration, the dual-layer structure of our framework allows carrying out an assessment of energy-saving potential with a more holistic approach. Key performance indicators that we collect include total IT power consumption, cooling power and Power Usage Effectiveness (PUE). In order to provide the robustness, we make several simulations of workload profile differences. Workload distribution, power consumption, temperature rise, and cooling demand are represented in bar graphs and pie charts using parallel visualizations that are created in MATLAB. Such visualizations shed light on the energy flow of the data center as well as confirm the success of optimization.

Overall, the approach provides a reprogrammable, scalable framework of exploring sustainable operation of data centers in the MATLAB setting.

4. Energy Optimization Strategies

The optimization strategy incorporated into the modeling framework of the MATLAB simulation anchored on collating workloads in order to minimize the number of servers that are on duty. The system would allow most of the work to be delegated to a subdued pool of machines, thus permitting the system to shut down or otherwise spin-down any under-utilized nodes, in effect, reducing the net energy costs. At the same time, a heating-regulated difference in temperature was proposed. The method measures the increase of individual-server temperature as a measure of certain characteristics of the workload and enables the cooling resources only when the thermal envelope crosses a fixed value. The remaining energy use is then gained by multiplying the temperature difference with the thermal load of the server under consideration. Temperature-sensitive cooling and workload consolidation therefore presented as an implementation of a more granular regimen of controlling the use of energy in an effective manner due to their compounding effect. These macro-mechanisms of optimization and the implications that they have on the energy performance indices have been summarized in Table 1.

Table 1: Summary of Energy Optimization Strategies for Data Center Efficiency.

Optimization Strategy	Description	Expected Benefit
Workload Consolidation	Allocates tasks to fewer servers at higher loads	Reduces server power consumption
Idle Server Shutdown	Powers down underutilized servers	Minimizes standby power drain
Temperature-Based Cooling	Activates cooling only when temperature exceeds a defined threshold	Improves cooling energy efficiency
Dynamic Thermal Feedback	Uses workload-to-temperature mapping to guide cooling decisions	Prevents overcooling, saves energy
Load-Aware Scheduling Algorithms	Allocates tasks based on server efficiency and location	Balances performance and energy use

5. Evaluation and Results

After running the MATLAB simulations, two different working-load-allocation strategies were tested; a so-called baseline setting where workloads were assigned randomly to all the available servers, and an optimized setting, where workloads were assigned by an algorithm that attempted to maximize server utilization, but did not violate the server latency constraints. An analysis performed on these two scenarios showed that the optimized scheme produced significantly lower energy consumption than the baseline distribution, which was the result of efficient distribution over resources that an optimized scheme promises.

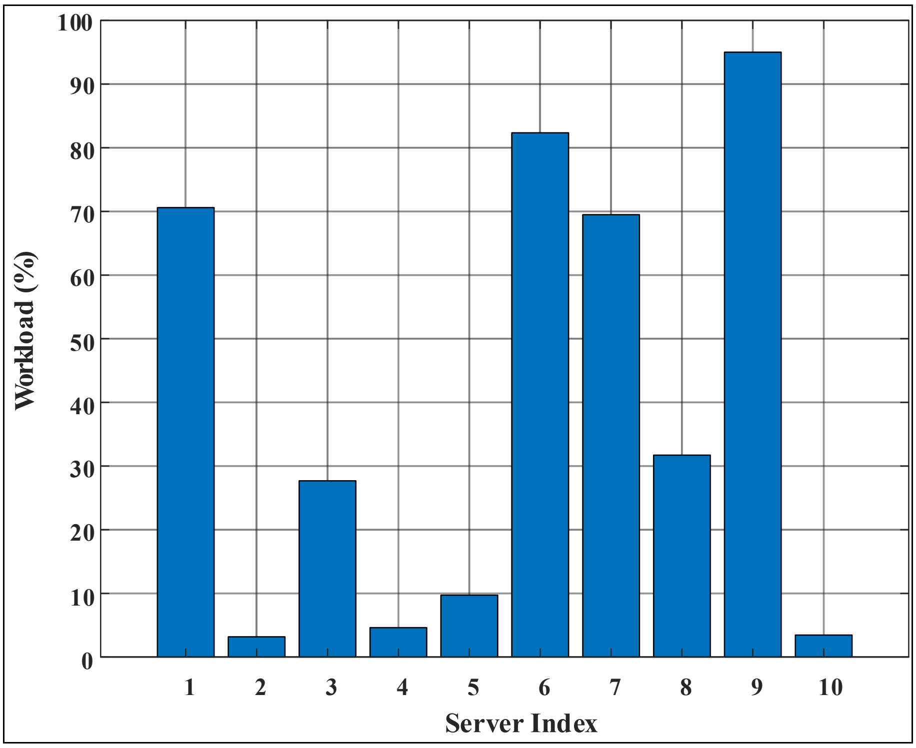

Fig 2: Server workload distribution under random allocation

In Figure 2, the load on 10 servers is depicted when jobs are discharged around campus in completely random manner. The servers gained varying loads, hence leaving some barely used machines and leaving others with an almost non-stop task. A combination of fully utilized and almost empty servers incurs power wastage a server that is not constantly used will use power. This picture supports the thought that dumping jobs on the machines with impunity forfeits opportunities to consolidate and ensuring the total amount of energy consumed does not decrease as much as it would if we planned better.

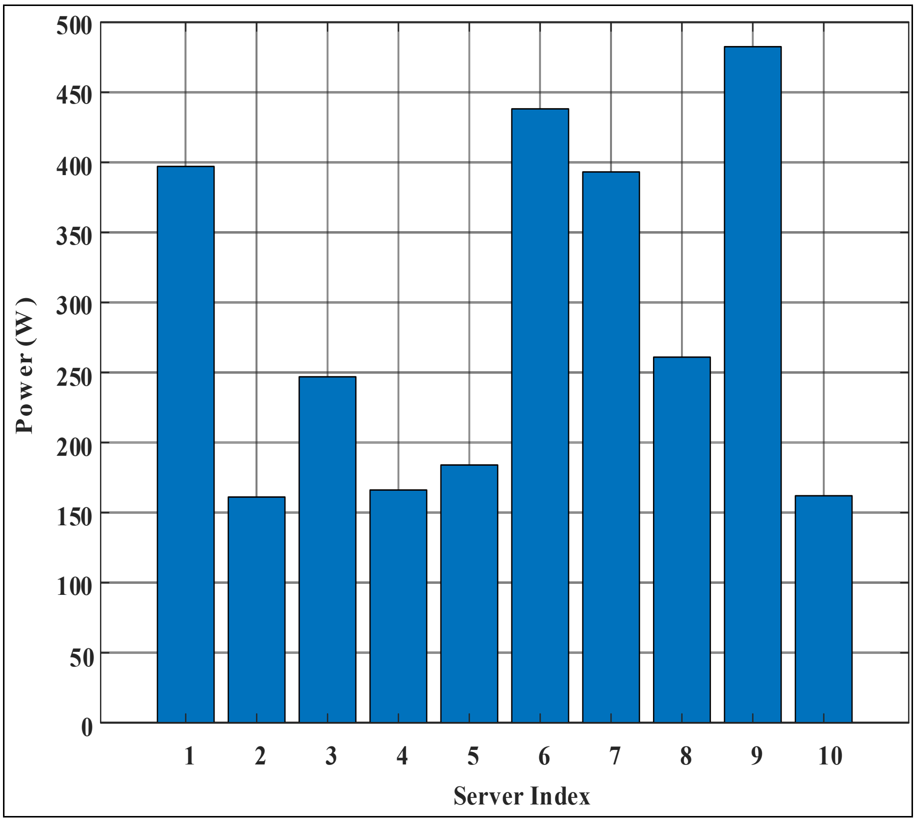

Fig 3: Power consumption per server under random workload distribution

The figure 3 presents the power consumption profile of each server in case of random workload condition. The power requirement of the server is linear in the workload pieced together with idle server to peak server power. The information ensures that it is obvious to not only learn but also understand that most of the servers tend to consume much energy all through without doing much computation. This case highlights the fact that baseline scheduling causes unnecessary increases in energy costs and puts it in a position to compare with this optimized configuration. The wide variation in the consumption between the servers is also indicative of the need of smart balancing of the workload to flatten the power curve. Conversely, the streamlined algorithm can reduce IT power footing by adamantly concentrating the loads on fewer computers. When simulation is done, the optimized allocation is seen to minimize IT power by 19%. Although the PUE (Power Usage Effectiveness) does not change all that much (due to power scaling of cooling with server power) the total facility power decreases significantly.

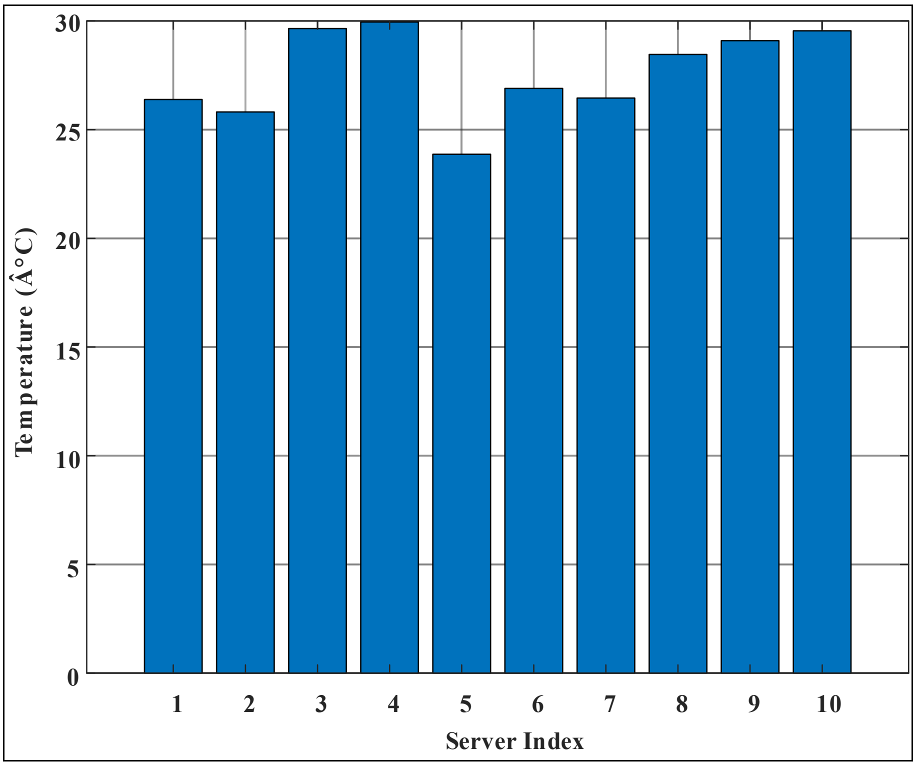

Fig 4: Comparison of server workload distribution before (top) and after (bottom) optimization

Now, suppose we have a two-node facility and we are now going to discuss how our scheduling decisions will affect the energy our servers consume. The same situation is presented twice in figure 4; one with random distribution and one after optimization. The first row indicates what we have observed earlier, i.e. there is random workload placement, which distributes work to both the machines. The bottom panel below indicates how things turned out after we applied the optimization algorithm trying to wedge the majority of the workload onto a single server and leaving the other virtually idle. Since only some servers are now at 100 percent loads, the consolidated design can allow the remaining ones to drop to low-power or power-off status. The resultant utilization curve is denser and flatter and the overall IT power consumption level reduces dramatically. With temperature-based cooling, the image brightens further: turning off or throttling fans of a server reduces the cooling loads, and thereby yet more energy is saved. The introduction of thermal constraints into scheduler does not only make it interesting mathematically; it indeed provokes energy savings in real world.

6. Discussion

The MATLAB simulations reveal the extent to which the energy efficiency of data center can be increased using optimal allocation of the workload and implementing smart cooling algorithms. As we changed the layout of our workload, which was random to an optimized one, the IT power consumption reduced by an approximate 19%. This was made possible by moving tasks down to fewer active servers and this allowed the others to stay in idle or fully off mode. Although the transition does not reduce the metric of Power Usage Effectiveness (PUE) directly, it rein in the actual power consumption, a sure way to expenditure reduction as well as carbon footprint reduction. A temperature-based cooling policy also enhanced the model to make it closer to the real-life limitations.

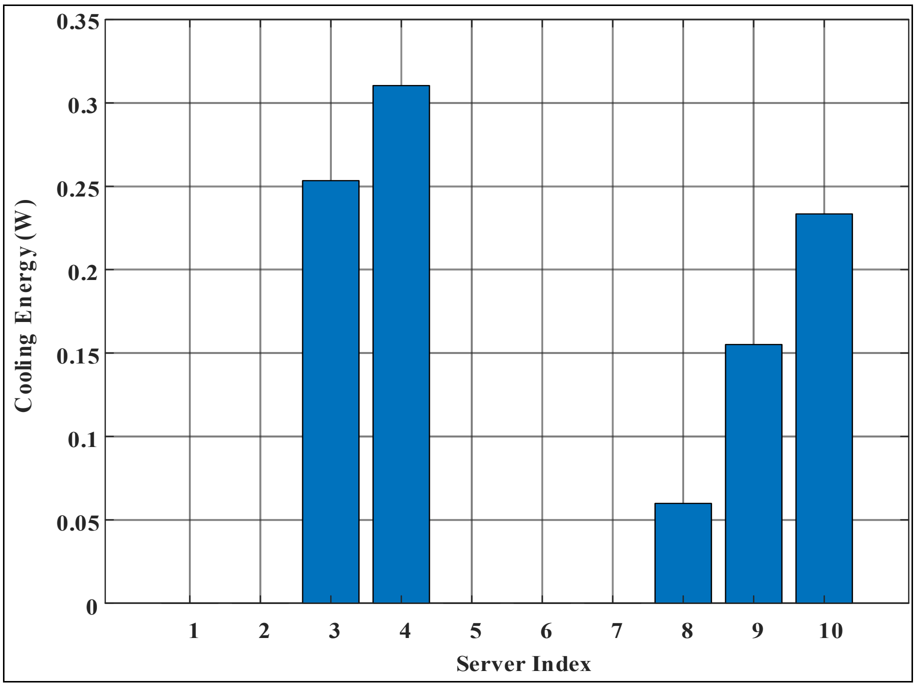

Fig 5: Simulated server temperatures based on workload

Figure 5 provides us with a snap shot of just how much heat is being produced by each of the servers and it is represented by the volume of work being done. The model presupposes the lack of change in the ambient temperature, and the workload increases the internal temperatures. As soon as a certain temperature, e.g. 28 °C is reached, the model turns on active cooling. Through this chart, we will be able to identify the locations where the workload bursts may cause hotspots, so that the cooling system will have no chance to handle the loads, or that certain component will endure problems.

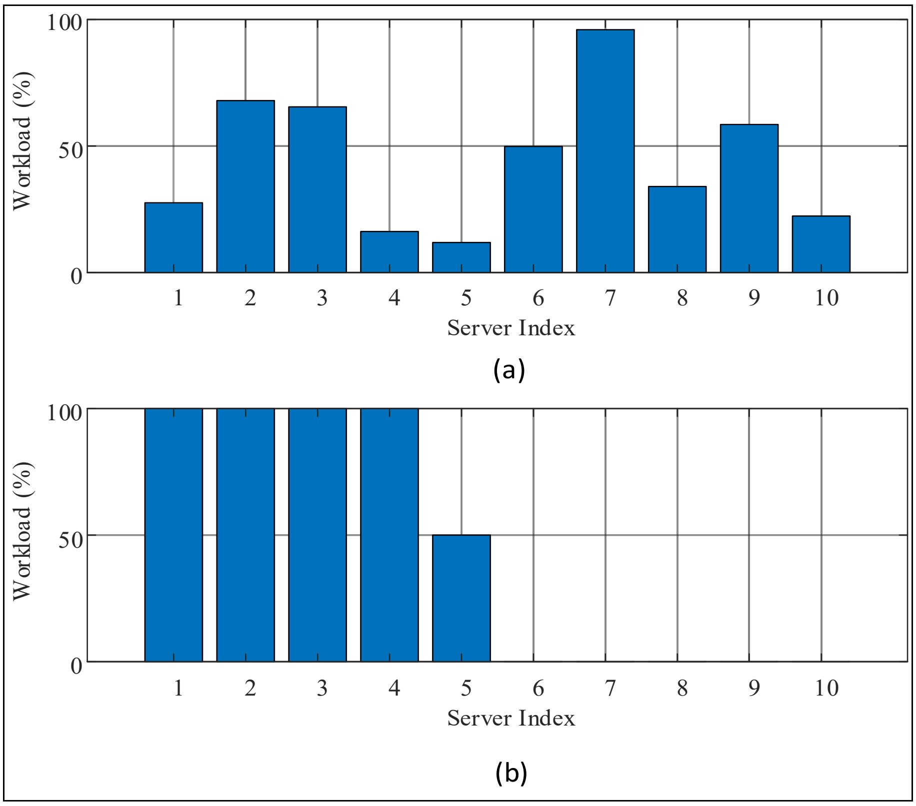

Fig 6: (a) Random workload distribution (%), (b) Optimized workload distribution (%)

Figure 6 indicates that the energy needed to cool the servers can be estimated based on the number of servers of all the servers and only those which are above the temperature threshold. Below that temperature there is no active cooling of the servers which saves energy. We hope that through selective activation of cooling a smarter thermal management approach is achieved where there is specific hitting of cooling power as opposed to uniform distribution. It then has the effect of more efficiently dedicating the energy towards the cooling infrastructure. This result confirms the idea that by integrating workload data with thermal models fine-grained control of cooling is possible to exceed the use of static provisioning.

In contrast to the fixed-percentage cooling models, this one is dynamic in that effort is made to improve cooling based on server-specific thermal loads. The simulation has proved that not every server needs to be cooled in the same amount of time, so the total power used in the cooling decreased because of optimized workload conditions. In consolidation, energy efficiency is improved, but server wear, hotspots of high temperatures may appear which can lead to shorter hardware life. It can even generate bottlenecks in performance when not used together with QoS policies and may need more complex monitoring and control tools. Despite these sacrifices, the net effect of these developments is supportive of systems-level approach to green computing. Instead of improving subsystems, methods of solving the overall system by combining power modeling, workload scheduling, and thermals have the best potential at maintaining sustainable operation.

7. Environmental and Economic Impacts

Data centers consume huge amount of energy and since a good portion of the energy used is still fossil-based, data centers contribute heavily to emissions of carbon gases. In this project, we tried the strategies to reduce the overall energy consumption based on the smarter scheduling of workload and dynamic cooling [8]. With a median carbon emission factor estimation of power (0.4 kg CO_2/ kWh), we may observe that the facsimiles to these optimizations can be translated into reducing carbon emissions directly [9]. To illustrate, suppose that the simulation will save approximately 500 W at all times all through the year, a total of 4380 kWh of electricity which is the equivalent of 1.75 tones of CO_2 emissions diverted. In economic terms, the yearly savings of 500 W incur a benefit of close to 438 dollars per rack in terms of electricity cost. The latter savings are linear with infrastructure size, and a very good case can be made in a large data center to energy-wise design. The overall cost of investment in control infrastructure and software is very temporary with most of these expenses being a one-time initiative. The framework would breakeven approximately between 1-2 years due to the existing electricity prices, and it already complies with the majority of the new requirements and standards in the industry, such as ISO 50001 and ASHRAE thermal standards and Greenhouse Gas Protocol, thus it is ready to face all the sustainability demands in the future [10].

8. Conclusion and Future Work

We proposed a MATLAB framework of data center energy consumption modeling and optimization with the help of intelligent workload scheduling and a temperature-aware cooling control. We demonstrated that dynamic consolidation of workloads can eliminate as much as 19% of IT power use--that is, IT power use can be minimized without jeopardizing performance. In conjunction with temperature-driven cooling strategy, such techniques enhanced even more thermal efficiency, cutting wasteful cooling-related energy consumption. These outcomes illustrate the worth of the systems degree of energy optimization tactics-organizing the two IT workload management and structures control. MATLAB was a good platform in establishing and evaluating the framework as it provided an open platform that could be reused by both researchers and practitioners in the industry. It allows risk-free exploration of the finicky interdependency between computational loads, power consumption and thermal features.

In addition to energy cost, the offered plans will contribute to the overall goals of the organization (a reduction in carbon footprint, compliance with the law, and maintaining costs at a minimum in the long run). The approach is particularly applicable where medium-to-large-scale data centers are interested in increasing their sustainability at a low capital expenditure [11]. In future, the integration of renewable energy sources, such as solar and wind, into dynamic demand profiles can be studied to supplement them. Artificial intelligence, may also be utilized so that past data and current temperatures can be used to forecast and effectively manage loads and cooling requirements [12]. The model might be expanded to include economic cost accounting, carbon accounting and the trade-offs in latency, availability and energy consumption.

It would be the next step to scale the simulation to edge computing or multi-tiered data centers architectures scenarios to enrich the model applicability. Integration of real-time telemetry data with which to test the model against operational environments would be able to increase the fidelity of the model, and provide data center operators with prescriptive information. With the changing digital infrastructure environment, tools and techniques such as the ones expressed within this paper will play an essential role in ensuring that as sustainability becomes more advanced, so does its performance.

9. References

1. Singh S. Green computing strategies & challenges. In: 2015 International Conference on Green Computing and Internet of Things (ICGCIoT); c2015. p. 758–760. IEEE.
2. Siddik MAB, Shehabi A, Marston L. The environmental footprint of data centers in the United States. Environmental Research Letters. 2021;16(6):064017.
3. Dayarathna M, Wen Y, Fan R. Data center energy consumption modeling: A survey. IEEE Communications Surveys and Tutorials. 2015;18(1):732–794.
4. Zhu H, Liu X, Zhang C, Yang Y, Wu Y, Zheng D, *et al.* Future data center energy-conservation and emission-reduction technologies in the context of smart and low-carbon city construction. Sustainable Cities and Society. 2023;89:104322.
5. Koot M, Wijnhoven F. Usage impact on data center electricity needs: A system dynamic forecasting model. Applied Energy. 2021;291:116798.
6. Qureshi H, Riaz M, Rizwan M, Alharbi K, Algarni A,

Mehmood A, *et al*. Artificial intelligence analysis of thermal energy for convectively heated ternary nanofluid flow in radiated channel considering viscous dissipations aspects. Engineering Science and Technology, an International Journal. 2025;62:101955.

7. Wu X, He A. Multimodal information fusion and artificial intelligence approaches for sustainable computing in data centres. Pattern Recognition Letters. 2025;189:17–22.
8. Avgerinou M, Bertoldi P, Castellazzi L. Trends in data centre energy consumption under the European Code of Conduct for Data Centre Energy Efficiency. Energies. 2017;10(10):1470.
9. Braeuer F, Finck R, McKenna R. Comparing empirical and model-based approaches for calculating dynamic grid emission factors: An application to CO_2-minimizing storage dispatch in Germany. Journal of Cleaner Production. 2020;266:121588.
10. Harter DE, Slaughter SA. Quality improvement and infrastructure activity costs in software development: A longitudinal analysis. Management Science. 2003;49(6):784–800.
11. Al Kez D, Foley AM, Laverty D, Del Rio DF, Sovacool B. Exploring the sustainability challenges facing digitalization and internet data centers. Journal of Cleaner Production. 2022;371:133633.
12. Zhao J, Yuan X, Duan Y, Li H, Liu D. An artificial intelligence (AI)-driven method for forecasting cooling and heating loads in office buildings by integrating building thermal load characteristics. Journal of Building Engineering. 2023;79:107855.