LONDON JOURNAL
OF
INTERDISCIPLINARY SCIENCES

eISSN 3049-7809

A Novel Method for Optimising Energy Efficiency in High-Performance Computing Systems

Aarah Sardesai*
Saif Al Dulaimi**
Merve Gokgol***

Abstract

This paper introduces a new method to promote energy efficiency in computing systems. The rapidly growing demand for computational power in high-performance computing (HPC) systems is accompanied by a significant increase in energy consumption. This research investigates the many ways to minimise energy consumption in HPC systems without sacrificing computational performance. Building upon previous research we combined and refined ideas to develop an optimised approach. Our method utilises a tri-modular framework incorporating an Energy-Efficient Hardware Design, Resource Management, and Optimisation, and Server Virtualisation. The first module employs a method of designing smart energy-efficient hardware. This method incorporates leakage reduction techniques such as clock gating and sleep states. The second module uses a technique focused on optimising server software. This method is based on Dynamic Voltage and frequency scaling (DVFS) for power management in data centres. The third module is based on server virtualisation and employs software to create multiple virtual machines on a single physical server allowing for a significant reduction in energy and hardware costs. The methods used in these three modules are systematically integrated to produce a more efficient consumption of electricity. This will allow for the computing system to minimise energy consumption without compromising computational power.

Keywords: DVFS (Dynamic Voltage and Frequency Scaling), Server Virtualizations, High-Performance Computing (HPC) systems, Energy Efficiency

https://doi.org/10.31039/ljis.2025.4.307

*Researcher, USA, sardesai.aarah@gmail.com
**Researcher, USA
***HS of Endeavor, Austin, USA

1. Introduction

Perhaps one of the greatest marvels of engineering is the computer. From phones, laptops, consoles, TVs, etc. we rely on computer systems daily. However, many people who use these devices are unaware of the many different pieces that function behind them and the power needed to operate them. Under the sleek exteriors of these technologically advanced devices lies a complex network of components that form High-Performance Computing Systems(HPCs) that connect all the devices. While HPCs are great at connecting devices, they are still raw within the field of efficiency and consume an incredible amount of global electricity. This paper aims to resolve this issue by manifesting a novel method to increase the efficiency of HPCs towards a greener future.

1.1 What are Computer Systems?

Computers are used daily ranging from phones, laptops, consoles, and even TVs. However, most people who use these devices do not know how many different parts work behind them and the intensity of energy required to run the devices. A computer system is a complex array of many units which is capable of computing a large number of things ranging from mere arithmetic problems to complicated models. Fundamentally, a computer system is a combination of hardware and software that performs data processing functions, executes instructions, and provides usable output. Hardware is the physical components of the computer such as CPU, memory, storage devices, and I/O devices; on the other hand, Software is the programs that control and use the hardware through operating systems and applications. All together, these elements allow for the excellent functionality of the devices we use every day.

1.2 What are Servers?

A server is a computer or system that acts as a central hub in a network, processing requests from other computers, known as clients. Clients send requests over a network, and the server responds with the requested information or service. This fundamental interaction is the cornerstone of the client-server model. Servers can provide a wide range of services, including file sharing, database management, web hosting, and application hosting. A single server can handle multiple clients simultaneously, allowing for efficient resource sharing and distribution of services. Conversely, a client can interact with multiple servers, accessing different resources or services as needed. This flexibility and scalability make the client-server model a cornerstone of modern computing infrastructure.

When observing a physical server, it is essentially a powerful computer designed to handle specific tasks within a network. It's the foundation on which many digital services and applications are built. These are usually housed within data centres along with storage devices, networking equipment, and security systems. To stay within a stable, secure, and efficient environment for the operation of IT services.

1.3 How Much Energy is Used in Computer Systems?

Our reliance on computers comes with a hidden cost: energy consumption. A personal computer that is commonly used by users consumes electricity in the range of 30 to 70 watts per day while data centres, which are facilities housing servers that support the internet and digital services, are significantly larger consumers of power that can use up to 500 MW or more (Masanet et al. , 2020; IEA, 2023). The global scale of computer usage translates to a substantial energy demand.

For instance, a computer used for six hours daily consumes approximately 2.1 kilowatt-hours (kWh) weekly, 9.1 kWh monthly, and 109.5 kWh annually (IEA, 2023). This consumption differs based on various factors, for instance, the power supply system and the display. In addition, the type of computer and its settings: a complex workstation with a video card, for example, a desktop, will consume more energy than a portable computer, such as a laptop. Older computers may use more energy because the components they use may be less energy efficient (IEA, 2023).

This energy consumption is further magnified by data centres which are known to be extremely power-hungry even compared to other IT systems because of the density of servers and the never-ending need for cooling power alongside backup power. Worldwide, data centres utilised between 240 to 340 Terawatt hours (TWh) of energy in 2022 alone, which is one million times more than a personal computer (IEA, 2023).

The leading issue concerning computer systems is the high amount of energy that is used, both in individual and centralised ways. The energy consumption from using one computer in a day is significant and one can imagine the numerous computers used globally in homes, offices, and schools. This underscores the importance of developing energy-efficient computing technologies. Knowledge about such factors and their mitigation at both individual and organisational levels can lead to energy conservation and therefore the protection of the environment (ServerWatch, 2023).

1.4 How Can We Make Them More Efficient?

To mitigate these issues, we can focus on increasing the efficiency of HPCs. Research can guide strategies to optimise server performance and reduce energy consumption. One strategy that was used by Barham et al. (2004) involves Server Virtualization. This approach uses software to create multiple virtual machines on a single physical server, allowing each Virtual Machine to run its operating system and applications. Barham et al.(2004) demonstrated that Virtualization can allocate workloads into fewer physical servers and can significantly cut energy and hardware costs.

Another technique focuses on optimising server software. Beloglazov et al. (2012) investigated dynamic voltage and frequency scaling (DVFS) for power management in data centres. DVFS adjusts the operating voltage and frequency of a server's CPU based on the real-time workload demands. This allows the computer to reduce its energy consumption during periods of low consumption without sacrificing its performance at peak times.

2. The Future of Computing Systems

Computers will progressively need higher power demands for their enhanced performances and functions as the years go by due to technological developments that aid our lives. But given energy demands for these ever-so-growing computer systems, it's critical to reduce the consumption effectively. This issue has become challenging and rather well-known as data centres and computer infrastructures shape global electricity use. With the growth of the digital age, we are seeing an exponential increase in size and complexity.

Over time, these technologies become determining factors that propel economic development and incorporation of new trends in fields such as healthcare, finance, and entertainment. However, this growth has brought with it a new challenge of handling their impact on the environment and this is especially so given that data centres, which are significant in processing and storing information around the world, are infamous for their energy consumption (ServerWatch, 2023). This has led to the rising concern that future computing has to be sustainable and the incorporation of innovative designs and best practices in energy consumption technology. Due to this, our team has decided to formulate a tri-pronged approach targeting the different aspects of the problem.

2.1 Module 1: Energy-efficient Design

Efficiency is a vital factor in the design of high-power-consuming devices in the contemporary computer system. Attempting to address the continually developing problem of data centres as well as other global systems requiring the use of computers as our central focus, a solution, in this case, lies in improving the actual hardware. This approach includes elements such as sleep states and clock gating that are aimed at preventing leakage of power when various components are not operated. Sleep states provide ways for power-hungry components of the system to shut down or be managed in a low-power mode at some point in time without presenting a negative impact once the system is called upon to wake up. Clock gating on the other hand occurs when the clock signal is disabled in certain parts of the circuit so that only the required part is powered, thus minimising power consumption. Integrating such energy-saving measures into hardware platforms enables one to have substantial control over the total energy consumption in computing. This also improves the sustainability of these systems and guarantees that they can indeed accommodate the growing future needs in the future world of computing (Ratković, Bežanić, Ünsal, Cristal, & Milutinović, 2015).

2.2 Module 2: Optimising Server Technology (DVFS)

To add to the energy-efficient techniques, optimising server technology can provide a significant boost to the existing software, causing a shift in the amount of electricity. As seen by the experiment done by Beloglazov et al. (2012), dynamically shifting the operating frequency and voltage of a processor or other electronic component based on the current workload can be an effective power management technique. By reducing the voltage and frequency during periods of low activity, DVFS can reduce power consumption during periods of low workload, leading to lower energy costs and reduced heat generation.

Conversely, when the system is under heavy load, DVFS can increase the voltage and frequency to improve performance. This flexibility enables systems to adapt to varying workloads, optimising both energy efficiency and responsiveness. This technique has become increasingly important in the context of server technology, where energy efficiency is a critical consideration due to the high density and continuous operation of servers in data centres (Shao et al., 2010).

2.3 Module 3: Server Virtualization

Server virtualisation is a strategy that enables the hosting of multiple operating systems on a single physical server. This is accomplished through the segmentation of virtual machines, each of which functions as if it has its physical resources of hardware. Consolidating several physical servers into a single virtual environment is especially important for organisations because it allows them to significantly reduce hardware costs, optimise resource utilisation, and improve overall system performance (Barham et al., 2004).

This is one of the primary benefits of server virtualisation because it allows for increased server utilisation. Loading a physical server with multiple virtual machines (VMs) allows for a more efficient allocation of resources to an organisation's various applications. This means that the critical application can receive more CPU, memory, or storage capacity while other applications share the remaining resources. Furthermore, it reduces hardware constraints that can be overcome by distributing workloads across multiple Virtual Machines (VMs) (Lin et al., 2016).

3. Conclusions

A brief review of the three generic approaches that are the pillars on which modern computing systems could be enhanced in terms of energy efficiency is as follows: Energy-efficient design, optimised server technology, and server virtualisation. There are two types of power management techniques: sleep states, which reduce the power consumption of devices while they are inactive, and clock gating, which reduces power consumption by removing the clock signal from a device when it is not in use. Extending the concept of power management, there is DVFS, which stands for Dynamic voltage and frequency scaling; it changes the processor's capacity as needed for a given task. However, it contributes positively to resource utilisation because many operating systems can be run on the physical server. As previously stated, if all of the described strategies are implemented concurrently, energy efficiency is likely to be realised, as will other operating costs of the existing systems.

By combining the above approaches, we will be able to formulate a solution that is comprehensive and addresses all of the pressing issues that data centres face in the modern day. This is especially important in the modern day, where high data density environments necessitate various types of energy for both environmental compliance and practical business applications. As data centres grow in size and complexity, the ability to reduce power consumption while maintaining impressive efficiency using this approach. This can serve not only to reduce the environmental impact of organisations' operations but also to gain benefits for the organisations themselves.

Acknowledgements

This article was presented at the 13th London International Conference, held in Istanbul on 24-26 July 2024, Turkiye. It has been published in the proceedings of the London International Conferences under No. 11(2024).

References

Ratković, I., Bežanić, N., Ünsal, O. S., Cristal, A., & Milutinović, V. (2015). An Overview of Architecture-Level Power- and Energy-Efficient Design Techniques. In Advances in computers (pp. 1–57). https://doi.org/10.1016/bs.adcom.2015.04.001

Beloglazov, A., Buyya, R., Lee, Y. C., & Zomaya, A. (2011). A Taxonomy and Survey of Energy-Efficient Data Centers and Cloud Computing Systems. In Cloud Computing and Distributed Systems (CLOUDS) Laboratory, Department of Computer Science and Software Engineering, the University of Melbourne, Melbourne, VIC 3010, Australia (Vol. 82, pp. 47–108). https://doi.org/10.1016/B978-0-12-385512-1.00003-7

Shao, Z., Iyer, V. S., & Xie, Y. (2010, April). DVFS-based dynamic thermal management for multi-core processors. In 2010 Design, Automation and Test in Europe Conference & Exhibition (DATE) (pp. 1617-1622). IEEE

Data centres & networks - IEA. (n.d.). IEA. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networs

Lin, C.-T., Huang, Y.-M., & Wen, Y.-C. (2016). Server consolidation techniques in virtualized data centers: A survey. The Journal of Supercomputing, 72(1), 105-12

Barham, P., Dragunov, B., Harris, E., Kay, R., & Newell, A. (2004). Xen: A virtual machine management system for x86-based PCs. ACM SIGOPS Operating Systems Review, 38(5), 186-199.

Monteclaro, A. J. (2023, November 20). Data Center Power Consumption: What You Need to Know.ServerWatch.https://www.serverwatch.com/virtualization/data-center-power-consumption/