An Experiment in Using the HEPscore Benchmark to Examine and Improve Data Center Operational Energy and Carbon Performance

YICHEN GAO, Massachusetts Institute of Technology, USA
NOMAN BASHIR, Massachusetts Institute of Technology, USA
CHRIS HILL, Massachusetts Institute of Technology, USA
JEREMY GREGORY, Massachusetts Institute of Technology, USA

In this short paper we describe ongoing work to quantitatively evaluate opportunities for revenue neutral improvements to research computing data centers operational energy and carbon performance. We outline how we are using the HEP score high-energy physics benchmark suite as a tool for quantitatively mapping the landscape of heterogeneous fleets of servers of different ages. We illustrate the use of HEP score in a real-world data center experiment which we are currently undertaking. The experiment is still ongoing, but we expect it to show that for the sorts of older machines that we suspect can still be found in many research environments the cumulative energy and carbon costs over periods as shorts as 24 months can be greater than the combined capital and energy costs of new replacement machines that provide the same computational capacity. We speculate that for many research organizations with limited budgets, the practice of over extending the life of hardware could well be financially and environmentally counter productive. We suspect this scenario is particularly prevalent in research and academic organizations. In these organizations operational energy and emissions costs are often hidden for individual researchers. We argue that, if our ongoing experiment hypothesis is validated, then organizations like universities that typically externalize energy and emissions costs for physical hardware based research computing may want to impose proactive life-cycle management practices on computer equipment. These practices could be based on the approach we describe in this article. The practices would, we hypothesize, result in exclusively positive outcomes. Operational costs associated with energy purchases and environmental emissions would potentially be reduced. Research productibity would be increased.

CCS Concepts: • **Computer systems organization → Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: datasets, neural networks, gaze detection, text tagging

1 INTRODUCTION

The data center industry is booming alongside rapid AI growth. McKinsey & Company projected 10% annual industry growth through 2030, with $49 billion global spending on new facilities, and a 20% annual increase in the hyperscale market [Bangalore and Srivathsan 2023]. Data centers are vital to today's economy, powering search engines, e-commerce, and AI. Global expansion is driven by the U.S. (40% of the market) and significant growth in Latin America, Europe, and Asia-Pacific. Sustainability of data centers became a focus due to rising energy and water usage, with impending U.S. climate regulations pushing companies to disclose carbon emissions. Providers are innovating for higher density and sustainability [Schaap 2024].

Compute needs are continually growing in research organizations, just as in commercial and consumer computing. In many research organizations, however, funding streams for computational infrastructure are often not growing as fast as potential needs, which in many cases may be fundamentally insatiable. This can lead research groups to retain computer hardware for long periods of time even as new hardware is acquired. This seems particularly common in the recent

---

Authors' Contact Information: Yichen Gao, Massachusetts Institute of Technology, , USA, ygao7@mit.edu; Noman Bashir, Massachusetts Institute of Technology, , USA, nbashir@mit.edu; Chris Hill, Massachusetts Institute of Technology, , USA, cnh@mit.edu; Jeremy Gregory, Massachusetts Institute of Technology, Cambridge, USA, jgregory@mit.edu.

multi-core era where computing performance increases do not come predominantly from increases in CPU clock speed.

If a key per-core metric like clock speed remains relatively steady, keeping a working system running for as long as possible may appear prudent. However, computers do not readily display *energy usage per unit of compute*. This, gas mileage like number, has continued to fall, exponentially even as clock speeds have stayed relatively constant. As a result, over time, the practice of retaining hardware for long periods can lead to significant hidden energy and emissions costs relative to replacing hardware. Ultimately these costs integrated over a few years can potentially exceed the cost of hardware replacement.

This paper presents a project to look quantitatively at this topic. It leverages the high-energy physics HEP score benchmark code as a stable measure of energy and emissions performance of computer systems over time Group [2023]. In the paper we describe the benchmark and how we use its information. We present preliminary data gathered from a representative hardware fleet and describe an ongoing experiment to examine these issues in a real-world scenario.

Our work contributes a possible quantitative framework and for answering questions about when to replace old systems in research computing environments. In many cases, for cost reasons, these environments are based on physical hardware operated by university staff. In these environments there does not appear to us to be a satisfactory existing framework for answering this question. We believe the framework and implementation we describe can be applied in any research computing environment. While specific numbers on energy and emissions profiles and hardware mix may vary, the formula and implementation we describe is general. They can form the basis for focusing stake holder decision making in any research computing organization.

1.1 Decision Model

The model aims to address assumptions surrounding the financial and environmental impact of research computing and recognizes its significance within the broader scope of carbon emissions. By focusing on the replacement and optimization of computing resources, the project seeks to minimize the environmental footprint while maximizing computational efficiency. To achieve these goals, we've developed a model for regular equipment renewal that can be tailored to any data center's needs. Users can input variables specific to their data centers, such as electricity cost, carbon emissions per kWh of electricity, PUE, etc. This model is accessible to individuals without background knowledge of sustainable computing and is guided by factors such as workload optimization, financial constraints, and processor lifecycle considerations.

Central to our methodology is the collection and analysis of data on processors in data centers. The main requirements for data collection include the model of processors in the data center, and the model potential replacement processors, along with their TDP, purchase price, and HEP score benchmark for each model. Systematic organization of collected data enables users to assess current operational costs from both financial and environmental perspectives, establish a baseline for evaluating alternative options, and calculate the financial payback time if the data center gets renewed. The model allows for the discovery of alternative pathways to improve the performance of data centers by comparing the financial and environmental impact of replacing specific old processors with new ones.

1.2 Carbon operational Cost

Acknowledging sustainability's broader scope, our methodology incorporates an assessment of embodied carbon in processors. However, our primary focus lies on carbon operational costs due to their long-term impact. By prioritizing carbon operational costs, we gain insights into the enduring sustainability of MIT's research computing infrastructure.

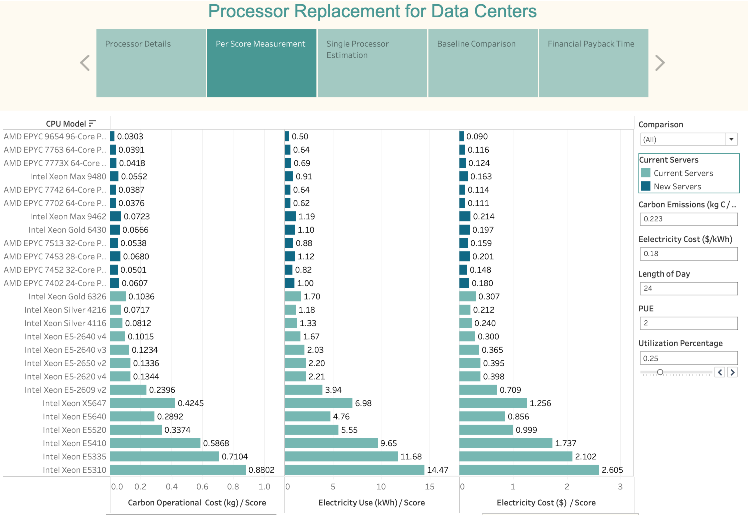
1.3 Benchmarking

To ensure consistency and reproducibility, we employ the HEPScore23 (HS23) benchmark provided by the Benchmarking Working Group [Group 2023]. This benchmark accurately measures performance under scientific research experiment requirements, aiding in the effective evaluation of computing resources and hardware procurement.

2 DASHBOARD INTERFACE

The dashboard provides a comprehensive toolkit for evaluating processor attributes, performance metrics, and environmental impact indicators within MIT's data center infrastructure. Users can compare processor models based on computational power, analyze their environmental footprint, and assess financial implications. Interactive features enable users to make informed decisions regarding processor selection, optimizing computational efficiency while considering sustainability and financial factors. Through its Performance Comparison, Sustainability Analysis, and Financial Analysis sections, the dashboard empowers users to navigate data center renewal initiatives with clarity and effectiveness.

The interface is designed to be clear and easy to use. Users can switch between different sections by clicking the topic names on the top bar. In the center, columns display various values and data for each corresponding CPU model. The "Current Server" and "New Servers" sections are color-coded for easy differentiation between the models currently in use in the data center and potential replacement models. You can also use the filter to compare several chosen models. Users can click on each row to view detailed information about that model. User input parameters are listed on the right side, allowing users to change variables within a set range to calculate specific values by customizing the environment.



Comparison
(All)
Carbon Emissions (kg C / ..
0.223
Eelectricity Cost ($/kWh)
0.18
Length of Day
24
PUE
2
Utilization Percentage
0.25

Fig. 1. Processor Replacement Tableau Dashboard Interface of the Per Score Measurement Section.

2.1 Performance Comparison

This section provides a comprehensive comparison of processor attributes and performance metrics, aiding users in selecting the most suitable processor model for their computational needs. Users can visualize and analyze the performance of different processor models based on desired computational power, measured by the HEPiX score. Interactive features allow users to select a baseline processor and observe performance differences among various models. By leveraging this section, users can make informed decisions regarding processor selection to optimize computational efficiency. The calculation of the electricity use each year of a processor is shown below:

$$E = \mathrm{TDP} * t * 365 * \mathrm{PUE} * (0.3 + 0.7 * r_u) * \frac{1}{1000}$$
$$E = \text{Electricity Use (kWh) / Year}$$
$$\mathrm{TDP} = \text{Thermal Design Power}$$
$$t = \text{Length of Day}$$
$$\mathrm{PUE} = \text{Power Usage Effectiveness}$$
$$r_u = \text{Utilization Rate}$$

Since even if the utlization of a Processer is 0%, it will still has power consumption. Here, we set it as having a default 30% consumption even if the processor is not computing. The rest of 70% are scaled up to a range of 1 so that the user can control the utlization percentage on the dashboard form 0 to 1. Even if the user set the utlization rate to 0, they will still see the electricity concumption of the processors.

2.2 Sustainability Analysis

In this section, users delve into the environmental impact of each processor model through a detailed sustainability analysis. Graphical representations illustrate carbon emissions and electricity consumption, enabling users to assess the environmental footprint of each model. Parameters such as carbon emission rate, electricity cost, and hours of operation per day can be fine-tuned to gain insights into environmental impact variations. Crucial sustainability metrics, including carbon operational cost, electricity cost, and pounds of coal burned per year for each model, are provided to facilitate environmentally conscious decision-making [Agency 2024] . Let C =Carbon Operational Cost (kg) / Year, E = Electricity Use (kWh) / Year, and r_b = Carbon burning rate (kg / kWh). The equation we use to calculate the carbon operational cost is:

$$C = E * r_b \tag{1}$$

2.3 Financial Analysis

The Financial Analysis section offers insights into the financial implications of selecting different processor models. Users input the purchase price of each processor and subsequently evaluate the financial payback time based on savings in electricity costs. Through interactive features, users can compare the number of replaced processors required for each model, aiding in assessing overall financial feasibility. By utilizing this section, users gain a comprehensive understanding of the financial considerations associated with each processor model, facilitating cost-effective decision-making in data center renewal initiatives. The calculation of the Financial Payback Time is shown below:

$$T_f = \frac{P_{new} * N}{E_{old} * N_{old} - E_{new} * N_{new}}$$

$$T_f = \text{Financial Payback Time}$$
$$P_{new} = \text{Purchase Price of the New Processor}$$
$$N_{old} = \text{\# Number of Old Processor}$$
$$N_{new} = \text{\# Number of New Processor}$$
$$E_{old} = \frac{\text{Electricity Cost (\$)}}{\text{Year}} \text{of the Old Processor}$$
$$E_{new} = \frac{\text{Electricity Cost(\$)}}{\text{Year}} \text{ of the New Processor}$$

3 CONCLUSION AND FUTURE WORK

Our ongoing work highlights the potential for revenue-neutral improvements in the operational energy and carbon performance of research computing data centers. By utilizing the HEP score high-energy physics benchmark suite, we can quantitatively map the landscape of heterogeneous fleets of servers of different ages. These practices, based on the approach described in this article, are expected to yield exclusively positive outcomes. Operational costs associated with energy purchases and environmental emissions could be reduced, while research productivity could be increased. If our ongoing experiment hypothesis is validated, we advocate for the implementation of proactive life-cycle management practices on computer equipment.

REFERENCES

U.S. Environmental Protection Agency. 2024. *Pounds of coal burned.* Retrieved May 6, 2024 from https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references#lbscoal

Srini Bangalore and Bhargs Srivathsan. 2023. Investing in the rising data center economy. (Jan. 2023). Retrieved May 6, 2024 from https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy

The HEPiX Benchmarking Working Group. 2023. *HEPScore23 (HS23).* Retrieved April 20, 2024 from https://w3.hepix.org/benchmarking/scores_HS23.html

Andrew Schaap. 2024. *Five Trends Driving The Booming Data Center Economy In 2024.* Retrieved May 6, 2024 from https://www.forbes.com/sites/forbestechcouncil/2024/01/22/five-trends-driving-the-booming-data-center-economy-in-2024-and-why-investors-are-taking-notice/?sh=69a15ebd6fa5

A CALCULATED FIELDS AND PARAMETERS

A.1 Processor Details

- `Score`: HEPiX CPU Benchmark for describing experiment requirements, lab commitments, existing compute resources, and procurement of new hardware.
- `TDP (W)`: Thermal Design Power, referring to the power consumption under maximum theoretical load.
- `Electricity Cost ($) / Year`: Annual electricity cost per processor.
- `Electricity Use (kWh) / Year`: The amount of electricity used by a processor each year.

- `Carbon Operational Cost (kg) / Year`: Carbon emissions released annually. item `Equivalent Pou` Equivalent pounds of coal burned to produce the same amount of carbon emissions by a processor annually.

A.2 Per Score Measurements

- `Carbon Operational Cost (kg) / Score`: Per-score measurement of carbon operational cost.
- `Electricity Cost ($) / Score`: Per-score measurement of electricity cost.
- `Electricity Use (kWh) / Score`: Per-score measurement of electricity use.

A.3 Single Processor Estimation

- `Single Model - # Processor Needed`: The number of a particular type of processor needed to reach a desired amount of performance score, assuming we only use a single type of processor.
- `Single Model - Carbon Operational Cost (kg) / Year`: The carbon emissions cost each year, assuming using one type of processors.
- `Single Model - Electricity Cost ($) / Year`: The electricity expenses each year, assuming using one type of processors.
- `Single Model - Pounds of Coal Burned / Year`: The equivalent pound of coal burned to produce the same amount of carbon emissions by the processors annually.

A.4 Baseline Comparison

- `# New Processors Needed`: Difference of number needed between any chosen processor and the baseline processor to reach a desired amount of performance score.
- `Saved Carbon Operational Cost (kg) / Year`: Difference of carbon operational between any chosen processor and the baseline processor.
- `Saved Electricity Cost ($) / Year`: Difference of electricity cost between any chosen processor and the baseline processor.
- `Saved Pounds of Coal Burned / Year`: Difference of equivalent amount of carbon emission produced by coal burned between any chosen processor and the baseline processor.

A.5 Financial Payback Time

- `Purchase Price ($)`: The purchase price of a model of processor.
- `Financial Payback Time (Year)`: The time it takes to make it worth it to purchase a new processor. By using the money it spends to buy the processor that meets the requirement, and divide by the saved electricity cost over a year. Purchase Price of Replaced processor ($): Purchase Price of Replaced processor.
- `Yearly Savings ($)`: The saved electricity cost each year after the processor renewal.
- `Electricity Cost of the New processor ($)`: Electricity cost of the new processor.
- `Electricity Cost of the Old processor ($)`: Electricity cost of the old processor.
- `Score of the Old Processor`: Performance score of the old processor.

A.6 Parameters (User Inputs)

- `# Old Processors`: The model of the old processor that will be replaced.
- `# Old Processors to Replace`: The number of the old processors that will be replaced.
- `Performance Score Required`: The desired computational power of the data center, measured by the HEPiX score.

- `Baseline Processor`: The baseline processor to be compared with in the baseline comparison section.
- `Carbon Emission (kg CO2 / kWh)`: The amount of carbon generated per kWh electricity used in this data center.
- `Electricity Cost ($/kWh)`: The monetary cost of each kWh of electricity.
- `Length of Day`: The length of time the data center runs each day.
- `PUE`: Power usage effectiveness of the data center
- `Utilization Percentage`: How much percent a processor is utilized. The range is between 0 and 1, with a step size of 0.05.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009