How Carbon Metrics Impact Device Selection

Debajyoti Halder, Deboparna Banerjee, Akash Mani, Anshul Gandhi, Erez Zadok
Stony Brook University, NY, USA
{dhalder, debbanerjee, akmani, anshul, ezk}@cs.stonybrook.edu

ABSTRACT

As computing systems increasingly contribute to carbon emissions, understanding how comprehensive carbon metrics influence device selection is crucial for sustainable computing. We investigate how considering embodied carbon alongside operational carbon affects optimal device choice for AI inference workloads. Our results show that including embodied carbon changes the optimal device choice for up to 58% of workloads, with the impact being more pronounced in low carbon intensity regions. This demonstrates that operational carbon alone is insufficient for sustainable device selection, highlighting the need for comprehensive carbon-aware metrics.

1 Introduction

The increasing environmental impact of computing systems has spurred a shift toward more sustainable computing practices. Traditionally, the computing community assessed a system's carbon footprint via its operational energy (e.g., PUE [5]) or via its operational carbon emissions—those produced during active workload execution [6]. However, recent studies [8, 19] have highlighted the importance of also accounting for (i) embodied carbon (emissions incurred during hardware manufacturing and deployment) and (ii) the carbon intensity of the energy source (which depends on, for example, the amount of renewable energy employed). For example, recent works on temporal and spatial shifting of workloads have focused on exploiting lower carbon intensity to reduce the total carbon footprint of execution [26, 21, 15].

This broader perspective has led to the development of holistic metrics such as Software Carbon Intensity (SCI) [12], defined as:

$$SCI = (O \times \text{carbon intensity} + M) \text{ per functional unit } R \tag{1}$$

Here, O and M are the operational and embodied carbon emissions. A functional unit R can be a user, an API-call, or an inference run.

In recent years, several works advocated using one carbon-related metric (or metrics [3, 13, 11]) over the other [22, 7, 14]. However, it is not always clear ***what the impact of the carbon metric is on the underlying system design choice***. That is, would using a different carbon metric result in a significantly different system design that is optimal for procurement and use? The Carbon Explorer work [2] investigated this question for the specific use case of employing renewable energy infrastructure (e.g., solar panels, batteries) for data centers. The authors highlight the trade-off between the operational carbon savings that such infrastructure can provide and the embodied carbon cost of procuring that infrastructure. The authors also observe that the location of the data center affects the total carbon footprint as the cost of setting up renewable energy farms factors in.

Inspired by the aforementioned works, in this paper we consider the impact of the carbon metric on the system design choice for the specific problem of device procurement for edge-inference tasks—*given an inference job that is to be run on the edge, which edge device should be procured and employed to execute the job in the most "carbon" efficient manner?* Recent works have considered this question [10, 25], but only using operational energy as the metric. While we do not endorse any specific sustainability metric, the goal of this work is to highlight the impact of the metric (and associated factors such as carbon intensity, job runtime, and embodied carbon) on the optimal device choice for an edge-inference task.

We investigate this problem using AI-inference workloads on various edge devices under traditional and contemporary carbon metrics. Our results indicate that, for about 58% of the workloads, the optimal device choice differs between the metrics, and these discrepancies are more pronounced for low carbon intensity regions (75% of the workloads see a change in optimal device).

2 "Carbon"-aware edge device selection

We experimentally investigated the edge device selection question.

Edge devices.

We experimented with four edge devices, listed here in ascending order of their compute capability: (i) Jetson Nano (4 ARM cores, 128 CUDA cores), (ii) Xavier NX (6 ARM, 384 CUDA, 48 Tensor cores), (iii) Orin Nano (6 ARM, 1024 CUDA, 32 Tensor cores), and (iv) Orin NX (8 ARM, 1024 CUDA, 32 Tensor cores). These are NVIDIA's Jetson series devices [1], capable of running Deep Learning (DL) models due to their integrated GPUs with CUDA support. They also consume low power; for example, the Xavier NX consumes at most 20W of power.

AI-inference workloads. We consider 5 PyTorch inference models across 4 different batch sizes: (i) AlexNet [18], (ii) VGG16 [24], (iii) ResNet50 [16], (iv) DenseNet121 [17], and (v) MobileNetV2 [23]. Due to the limited computational capabilities of the edge devices, we restrict our workload choice to these models (as opposed to LLMs or training).

"Carbon" metrics. For evaluation, we consider:

1. Operational energy (measured from device)
2. SCI—Software Carbon Intensity (defined in Eq. (1))

CarbonMetrics 2025 Stony Brook, NY

3. EDP—Energy Delay Product (operational energy × runtime)
4. CDP—Carbon Delay Product (SCI × runtime)

We use EDP as it captures the trade-off between energy and performance in a single metric and is widely used in system evaluation studies [32, 4]. We introduce CDP to contrast it from EDP and to study the impact of energy vs. carbon on edge device selection. Prior works have suggested various carbon metrics for different purposes [13, 14]. To illustrate the impact of metrics on system design, we consider the above four metrics that account for factors such as carbon intensity, manufacturing emissions, and hardware performance. The evaluation could be expanded to include other metrics as well, such as throughput per watt or total SCI [7].

To obtain embodied carbon for the edge devices (say, M_{device}), we use estimates derived using ACT [14], a recent framework that models manufacturing emissions of computing hardware based on process node, chip area, fabrication location, RAM size, and other parameters. For each device, we estimate the embodied carbon using the processor die and memory specifications from TechPowerUp GPU Specs Database [27]:

- Jetson Nano 4GB [28]: 54.60 $kgCO_2e$
- Xavier NX 8GB [31]: 132.39 $kgCO_2e$
- Orin Nano 8GB [29]: 168.26 $kgCO_2e$
- Orin NX 16GB [30]: 230.32 $kgCO_2e$

To estimate the embodied cost of each job (M_{job}), we amortize the device embodied carbon over its operational lifetime of five years [9] and multiply this result with the job's measured runtime:

$$M_{job} = \frac{M_{\text{device}}}{\text{device lifetime}} \times \text{job runtime} \qquad (2)$$

Carbon intensity. We use regional average values based on data from Electricity Maps [20], which provides hourly carbon intensity metrics (gCO_2e/kWh) across global regions. We chose three regions—France ($22gCO_2e/kWh$), California ($138gCO_2e/kWh$), and Italy ($325gCO_2e/kWh$)—whose intensities are representative of low, medium, and high carbon-intensity regions, respectively.

Methodology. For each device, we disabled DVFS and set them to MAXN performance mode. Each configuration was tested for 10 iterations, measuring runtime and energy consumption. We used the `jetson-stats` package to get the edge device's total energy consumption. Averaged results for each model were then sorted per the target "carbon" metric across the four devices.

Results and insights

Figure 1(a) illustrates the comparative sustainability performance of the four Jetson devices across low, medium, and high carbon intensity regions for both operational energy and SCI metrics when running VGG16 with a batch size of 4. For each metric in each graph, the value is normalized by that under Jetson Nano. The order of devices is then sorted based on increasing operational energy.

Starting with Figure 1(a) for the medium carbon intensity region (California), we see that the sorted order of devices changed noticeably between operational energy and SCI metrics. All four devices had a different sorted rank under the SCI metric compared to their rank under the energy metric. We quantify this as "4 changes" in sorted order (e.g., Orin Nano has the lowest energy but has the second-lowest SCI). In particular, Orin NX consumes more energy than Orin Nano, and Jetson Nano consumes more energy than Xavier NX. However, SCI prefers Jetson Nano over Xavier NX because Jetson Nano has very low embodied carbon—so despite higher operational energy, the SCI is lower. Interestingly, *even though Orin NX has the highest embodied carbon among all devices, it has the lowest SCI.* This is because of the shorter runtime under Orin NX, which reduced the job's embodied carbon share (M_{job}, see Eq. (2)). This shows how SCI can favor faster devices, even if they have a higher embodied footprint.

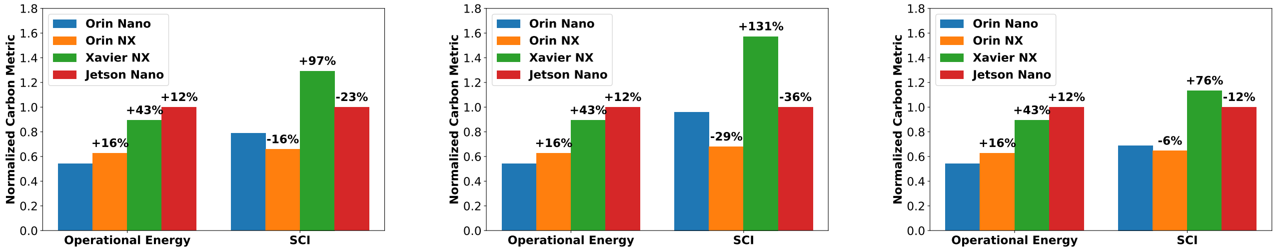
If we picked the most "carbon"-efficient device per energy rankings, we would choose Orin Nano. However, using SCI as the metric, we would choose Orin NX, which results in 16% lower SCI compared to the energy-minimizing Orin Nano device. This shows that *the "carbon" metric greatly impacts the device choice.*

A similar shift in sorted order, with 4 changes, is noticed under the low carbon intensity (Figure 1(b)) and high carbon intensity (Figure 1(c)) regions. However, we observe a more pronounced change for the low-intensity region (France) as the SCI for Orin NX drops even further and the gain in SCI for Xavier NX is higher.

Across all 20 workloads we experimented with, for the low-intensity region, 17 of them experienced some change in sorted order, with 4 of them experiencing a complete change in order (i.e., 4 changes). For medium and high-intensity regions, we still observed 16 and 13 workloads experiencing some change in sorted order, respectively. However, the biggest shifts happened in the low carbon intensity region, highlighting the *importance of the location when making the most carbon-efficient decision.* Overall, for 58% of the workloads, the optimal device (the one with the lowest "carbon" metric value) changes under SCI when compared with operational energy. Table 1 ("Energy vs. SCI") provides further details.

Next, we also consider performance as a deciding factor. EDP and CDP capture this by combining operational energy and total carbon with runtime, respectively. In Figure 2, when ResNet50 is run with a batch size of 8 in a low carbon intensity region, the device order changes significantly (with 4 changes) for both energy vs. SCI and for EDP vs. CDP. Using EDP as the metric, we would choose Orin Nano as the best device (11% better than the next-best Orin NX), whereas under CDP, we would choose Orin NX, which provides 8% lower CDP than Orin Nano. It is worth noting that although Jetson Nano has a lower SCI, its CDP is much higher compared to Orin NX and Orin Nano. This is because Orin devices, being faster, complete inference sooner, and this advantage shifts the order under CDP where runtime is taken into account explicitly. This highlights the trade-off between total carbon and runtime, which becomes important in performance-sensitive applications.

Across all workloads, under low, medium, and high intensity, we observed a change in order for 8, 6, and 5 workloads, respectively. We find that the optimal device



(a) Medium Carbon Intensity – California

(b) Low Carbon Intensity – France

(c) High Carbon Intensity – Italy

Figure 1: Change in device order for VGG16 (batch size 4) when sorted on distinct carbon metrics in different locations.

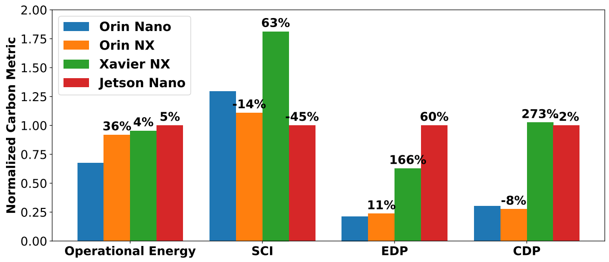


Figure 2: Device order under all 4 carbon metrics for ResNet50 with batch size 8.

	Energy vs SCI			EDP vs CDP		
# changes	2	3	4	2	3	4
Low CI	5	8	4	7	0	1
Mid CI	9	2	5	6	0	0
High CI	9	2	2	5	0	0

Table 1: Number of workloads (out of 20) experiencing various degrees of sorted-order changes for edge devices under different carbon-intensity regions.

changes under CDP for 12% of the workloads. Table 1 ("EDP vs CDP") provides further details.

3 Conclusion

Our study shows that carbon metrics significantly impact edge device selection for AI inference workloads. We found that considering various factors, like embodied emissions of the devices and carbon intensity of data center location, changes the optimal device for procurement and use, with low carbon intensity regions showing bigger differences. Based on the above results, we conclude that ***the choice of the "carbon" metric can impact the system design choices being made for sustainable computing***, especially based on how pronounced some factors (carbon intensity, job runtime, embodied carbon) are in the chosen metric.

Acknowledgment

This work was supported by NSF grants 2230078, 2106263, CCF-2324859, CNS-2214980, CNS-2106434, and CNS-1750109.

4 References

[1] NVIDIA Embedded Systems for Next-Gen Autonomous Machines — nvidia.com. `https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/`. [Accessed 05-07-2024].

[2] B. Acun, B. Lee, F. Kazhamiaka, K. Maeng, U. Gupta, M. Chakkaravarthy, D. Brooks, and C.-J. Wu. Carbon explorer: A holistic framework for designing carbon aware datacenters. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, pages 118–132, 2023.

[3] F. AL-Hazemi, A. F. Yousif Mohammed, L. Isaac Yoseke Laku, and R. Alanazi. PUE or GPUE: A Carbon-Aware Metric for Data Centers. In *2019 21st International Conference on Advanced Communication Technology (ICACT)*, pages 38–41, PyeongChang, South Korea, 2019.

[4] G. Ali, S. Bhalachandra, N. J. Wright, M. Side, and Y. Chen. Optimal gpu frequency selection using multi-objective approaches for hpc systems. In *2022 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–7, 2022.

[5] V. Avelar, D. Azevedo, and A. French. PUE: A comprehensive examination of the metric. *White Paper WP-49, The Green Grid*, 2012.

[6] D. Azevedo, M. Patterson, J. Pouchet, and R. Tipley. Carbon usage effectiveness (CUE): A green grid data center sustainability metric. *White Paper WP-32, The Green Grid*, 2010.

[7] N. Bashir, V. Gohil, A. B. Subramanya, M. Shahrad, D. Irwin, E. Olivetti, and C. Delimitrou. The Sunk Carbon Fallacy: Rethinking Carbon Footprint Metrics for Effective Carbon-Aware Scheduling. In *Proceedings of the 2024 ACM Symposium on Cloud Computing*, SoCC '24, page 542–551, Redmond, WA, USA, 2024.

[8] A. Bhagavathula, L. Han, and U. Gupta. Understanding the implications of uncertainty in embodied carbon models for sustainable computing. In *HotCarbon'2024*. 2024.

[9] N. Corporation. Jetson Modules Data Sheet. `https://developer.nvidia.com/embedded/downloads`, 2025.

[10] A. Dutt, S. P. Rachuri, A. Lobo, N. Shaik, A. Gandhi, and Z. Liu. Evaluating the energy impact of device parameters for dnn inference on edge. In *Proceedings of the 14th International Green and Sustainable Computing Conference (IGSC'23)*, Toronto, Canada, 2023.

[11] T. Eilam, P. Bello-Maldonado, B. Bhattacharjee, C. Costa, E. K. Lee, and A. Tantawi. Towards a Methodology and Framework for AI Sustainability Metrics. In *Proceedings of the 2nd Workshop on Sustainable Computer Systems*, HotCarbon '23, Boston, MA, USA, 2023.

[12] G. S. Foundation. Software Carbon Intensity (SCI) Specification — sci.greensoftware.foundation. `https://sci.greensoftware.foundation/`.

[13] A. Gandhi, K. Ghose, K. Gopalan, S. Hussain, D. Lee, D. Liu, Z. Liu, P. McDaniel, S. Mu, and E. Zadok. Metrics for sustainability in data centers. In *Proceedings of the 1st Workshop on Sustainable Computer Systems Design and Implementation (HotCarbon'22)*, San Diego, CA, USA, July 2022. USENIX.

[14] U. Gupta, M. Elgamal, G. Hills, G.-Y. Wei, H.-H. S. Lee, D. Brooks, and C.-J. Wu. ACT: designing sustainable computer systems with an architectural carbon modeling tool. In *Proceedings of the 49th Annual International Symposium on Computer Architecture*, ISCA '22, page 784–799, New York, NY, USA, 2022.

[15] W. A. Hanafy, Q. Liang, N. Bashir, D. Irwin, and P. Shenoy. Carbonscaler: Leveraging cloud workload elasticity for optimizing carbon-efficiency. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 7(3):1–28, 2023.

[16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[19] B. Li, R. Basu Roy, D. Wang, S. Samsi, V. Gadepally, and D. Tiwari. Toward sustainable hpc: Carbon footprint estimation and environmental implications of hpc systems. In *Proceedings of the international conference for high performance computing, networking, storage and analysis*, pages 1–15, 2023.

[20] E. Maps. Live 24/7 CO_2 emissions of electricity consumption — app.electricitymaps.com. `https://app.electricitymaps.com/map`.

[21] J. Murillo, W. A. Hanafy, D. Irwin, R. Sitaraman, and P. Shenoy. Cdn-shifter: Leveraging spatial workload shifting to decarbonize content delivery networks. In *Proceedings of the 2024 ACM Symposium on Cloud Computing*, pages 505–521, 2024.

[22] V. D. Reddy, B. Setz, G. S. V. R. K. Rao, G. R. Gangadharan, and M. Aiello. Metrics for sustainable data centers. *IEEE Transactions on Sustainable Computing*, 2(3):290–303, 2017.

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018.

[24] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[25] P. S.K, S. A. Kesanapalli, and Y. Simmhan. Characterizing the Performance of Accelerated Jetson Edge Devices for Training Deep Learning Models. *Proc. ACM Meas. Anal. Comput. Syst.*, 6(3), 2022.

[26] A. Souza, S. Jasoria, B. Chakrabarty, A. Bridgwater, A. Lundberg, F. Skogh, A. Ali-Eldin, D. Irwin, and P. Shenoy. Casper: Carbon-aware scheduling and provisioning for distributed web services. In *Proceedings of the 14th International Green and Sustainable Computing Conference*, pages 67–73, 2023.

[27] TechPowerUp. GPU Database: Graphics Card Specifications. `https://www.techpowerup.com/gpu-specs/`, 2025.

[28] TechPowerUp. Jetson Nano Specifications. `https://www.techpowerup.com/gpu-specs/jetson-nano.c3643`, 2025.

[29] TechPowerUp. Jetson Orin Nano 8 GB Specifications. `https://www.techpowerup.com/gpu-specs/jetson-orin-nano-8-gb.c4082`, 2025.

[30] TechPowerUp. Jetson Orin NX 16 GB Specifications. `https://www.techpowerup.com/gpu-specs/jetson-orin-nx-16-gb.c4086`, 2025.

[31] TechPowerUp. Jetson Xavier NX 8 GB Specifications. `https://www.techpowerup.com/gpu-specs/jetson-xavier-nx-8-gb.c3642`, 2025.

[32] Y. Zhang, Q. Wang, Z. Lin, P. Xu, and B. Wang. Improving GPU Energy Efficiency through an Application-transparent Frequency Scaling Policy with Performance Assurance. In *Proceedings of the Nineteenth European Conference on Computer Systems*, EuroSys '24, pages 769–785, Athens, Greece, 2024.