Towards a sustainable use of GPUs in Graphics Research

EMILIE YU*, UCSB, USA
ELIE MICHEL*, Adobe Research, France
OCTAVE CRESPEL, Independent Researcher, France
AXEL PARIS, Adobe Research, France
FELIX HÄHNLEIN, University of Washington, USA

ACM Reference Format:
Emilie Yu, Elie Michel*, Octave Crespel, Axel Paris, and Felix Hähnlein. 2025. Towards a sustainable use of GPUs in Graphics Research. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH Talks '25), August 10-14, 2025*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3721239.3734084

1 Introduction

Graphics Processing Units (GPUs) are at the core of Computer Graphics research. Research labs feel compelled to frequently update their hardware, as new GPUs increase the performance of existing methods and as they enable new algorithmic possibilities. And while the wider public does not have access to costly, high-performance hardware, the hegemonic assumption of Graphics researchers is that large-scale tech-hardware renewal is inevitable. We want to question the performative nature of this assumption: aren't we – researchers in Graphics – contributing factors driving the renewal of users' hardware? While we strive to develop more realistic, controllable, and fast ways to create digital content over the years, **what material substrate do we depend on and what are the consequences of that dependency?** We surveyed 888 papers from 2018 to 2024 presented at SIGGRAPH, systematically gathering author-reported GPU models. By contextualizing the hardware reported in graphics papers with publicly available data of consumers' hardware, we demonstrate that graphics research is consistently developed and tested on new high-end devices that do not reflect the state of the consumer-level market (Fig. 1). We find that this pattern of hardware usage is incompatible with the ACM's code of ethics[1] to *"promote environmental sustainability"*. Specifically, developing algorithms and tools that require newer and more powerful hardware than hardware owned by users, may reinforce existing trends of overconsumption caused by the premature replacement of devices [Laurenti et al. 2015], or increase the dependency of individuals and businesses on resource-hungry, industry-owned services in the "cloud" [Ensmenger 2021]. Looking forward, we show that there are exciting avenues for our research community to develop methods that embrace more realistic and sustainable hardware configurations, such as low-end GPUs. We contribute a set of recommendations at multiple institutional levels (authors, reviewing committees, research institutions) to support and incentivize this transition towards more sustainable computer graphics research. We release our data collection code and dataset of 888 papers with GPU models [2].

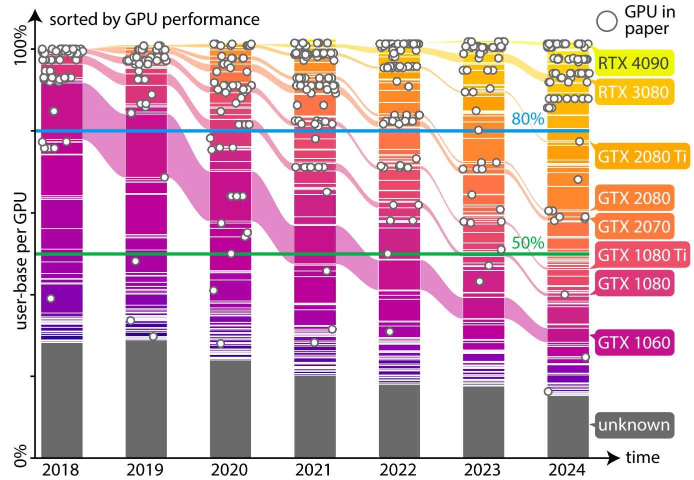

Fig. 1. **Research papers in computer graphics consistently use high-end GPUs:** 87% of GPUs reported in research papers (above the blue line) are available to less than 20% of the consumer-level user base at publication time. We visualize GPUs as colored segments in this graph. The vertical size of a segment represents the proportion of the Steam user base that uses that device, while its vertical placement indicates its performance compared to other devices (higher is more powerful). The curved lines represent the evolution of a GPU in terms of user-base and relative performance. For example, we see the GTX 1060 user-base shrinking, and its relative performance drop as new GPUs like the RTX 4090 are released.

2 Survey Method

GPU database construction. We built a database of GPU devices from an online repository[3]. Each GPU was then associated a relative performance score using the Blender Open Data Benchmark [4] (an open source repository of performance benchmarks by the Blender community) to sort GPU devices with respect to performance.

Research papers survey. We developed a semi-automatic method to search for GPU names in the PDF text of a paper: we detect likely

*Both authors contributed equally to this research.
[1]https://www.acm.org/code-of-ethics
[2]https://eliemichel.github.io/sustainable-gpu-usage
[3]https://www.techpowerup.com/
[4]https://opendata.blender.org/

Authors' Contact Information: Emilie Yu, emyu@ucsb.edu, UCSB, Santa Barbara, USA; Elie Michel*, emichel@adobe.com, Adobe Research, Paris, France; Octave Crespel, octave.crespel@club.fr, Independent Researcher, Paris, France; Axel Paris, aparis@adobe.com, Adobe Research, Paris, France; Felix Hähnlein, fhahnlei@cs.washington.edu, University of Washington, Seattle, USA.

SIGGRAPH Talks '25, August 10-14, 2025, Vancouver, BC, Canada

matches using a conservative automatic string matching algorithm (see supplemental material for the detailed algorithm), which we then manually process to filter out false positives, correct erroneous assignments, and validate each match. We applied this extraction method on ACM Transactions on Graphics papers for all SIGGRAPH proceedings from 2018 to 2024, processing 888 papers in total.

Survey of consumer-level GPU usage. To gain a broader understanding of the hardware usage and adoption patterns beyond academia, we collected usage data from the Steam Hardware Survey[5] – a publicly data base of GPU usage among Steam users across time.

3 Results and Discussion

We successfully detect at least one author-reported GPU for 43% of our corpus (383 papers). We find that a large majority of these papers report one GPU (81%), while some papers report two GPUs (16%) or more (< 3%). By matching the data collected on GPU usage in Computer Graphics papers with the data on GPU usage at a consumer-level, we observe a significant bias towards high-performance GPUs in research papers (see Figure 1, and Supplemental for more results). We acknowledge that the raw data sources we rely on are limited and may present biases, discussed in the supplemental material. In the following paragraphs, we use our quantitative analysis as a basis for broader discussions.

Consumer GPUs renewal: a self-fulfilling prophecy? Our analysis shows that high-end GPUs used in research papers tend to become more wide-spread among the surveyed population after a few years: the GTX 1080 was the most cited GPU in papers in 2018 (11 papers), and similar or better devices become more wide-spread after a 6 years period, owned by 54% of users in 2024. While hardware renewal is a phenomenon driven by wider societal and market forces [Magnier and Mugge 2022], we argue that our community has to question our role in this unsustainable cycle: by developing and disseminating computational methods optimized to work on the newest hardware, are we actively encouraging premature hardware obsolescence among end-users?

Driving the growth of cloud computing. Citations for data center GPUs are increasing across the years we surveyed (the NVIDIA V100 is the 2nd most cited GPU, with 13 papers in 2022-2023, and we see the growing popularity of the A100X in 2023 with 3 papers, and 2024 with 6 papers). Relying on cloud computing services and compute clusters for research is convenient and an unavoidable recourse for large machine learning methods (e.g., training on 256 NVIDIA A800 GPUs). While methods relying on cloud computing do not presuppose or encourage user-level hardware renewal, we do not see this as a satisfactory solution. Data centers are nevertheless material entities [Ensmenger 2021], that already have a negative impact on climate change [Thangam et al. 2024], including through hardware renewal [Wang et al. 2024]. They necessitate large infrastructural investments, which means that they are largely owned and operated by a few private corporations who derive a part of their profits from capturing user data [Zuboff 2023]. By developing methods which depend on cloud computing resources, are we consolidating the dependency of individual users on privatized "cloud" services?

[5]https://store.steampowered.com/hwsurvey/

4 Recommendations

Authors. First, we encourage authors to report what hardware (GPU and CPU) has been used for their work, to facilitate further efforts of analysis and commentary on hardware usage. While it is already common practice in many areas to indicate hardware in the implementation section, we note that our extraction method failed to identify a reported GPU on more than 50% of papers. We compile a list of guidelines to report hardware, which we will make available on our website (see Supplemental). Furthermore, correct reporting could be verified by reviewers, e.g. as part of the standard reproducibility check. Second, we encourage authors to test their method on a variety of different hardware. We found positive examples showcasing a method's performance and portability on multiple GPUS [Keeter 2020; Kenzel et al. 2018], yet this is far from being common practice. The evaluation on multiple devices showcases a method's generalization capability across hardware, which could become a valuable evaluation dimension. We hope that more research will leverage the community's existing expertise in hardware optimization to develop algorithms for mid-to-low-end material, giving rise to a new field of *Low-tech Graphics* [Nardi et al. 2018].

Institutions. We believe that authors will only feel empowered to question dominant practices when they are supported by institutions. First, research labs need access to supra-local testing infrastructures providing heterogeneous, multi-generational hardware, which is currently not the case for computing centers [IDRIS 2024]. Second, to incentivize low-tech support within our community, we need new metrics which provide (i) *relative* performance measures, e.g. rendering performance relative to the particular theoretical hardware computing limit, and (ii) *distributional* performance measures, e.g. performance over the distribution of low-to-high-end hardware. Finally, we believe that the transition towards a sustainable GPUs usage in research can be supported by existing grant programs targeting sustainable computing [6] and through partnerships with film studios encouraging sustainable practices [7].

References

Nathan Ensmenger. 2021. The Cloud Is a Factory. In *Your Computer Is on Fire.*

IDRIS. 2024. Jean Zay. http://www.idris.fr/eng/jean-zay/

Matthew J. Keeter. 2020. Massively Parallel Rendering of Complex Closed-Form Implicit Surfaces. *ACM Transactions on Graphics* (2020).

Michael Kenzel, Bernhard Kerbl, Dieter Schmalstieg, and Markus Steinberger. 2018. A high-performance software graphics pipeline architecture for the GPU. *ACM Transactions on Graphics* (2018).

Rafael Laurenti, Rajib Sinha, Jagdeep Singh, and Björn Frostell. 2015. Some pervasive challenges to sustainability by design of electronic products–a conceptual discussion. *Journal of Cleaner Production* (2015).

L. Magnier and R. Mugge. 2022. Replaced too soon? An exploration of Western European consumers' replacement of electronic products. (2022).

Bonnie Nardi, Bill Tomlinson, Donald J Patterson, Jay Chen, Daniel Pargman, Barath Raghavan, and Birgit Penzenstadler. 2018. Computing within limits. (2018).

Dhanabalan Thangam, Haritha Muniraju, R Ramesh, Ramakrishna Narasimhaiah, N Muddasir, Ahamed Khan, Shabista Booshan, Bharath Booshan, and R Ganesh. 2024. *Impact of Data Centers on Power Consumption, Climate Change, and Sustainability.*

Peng Wang, Ling-Yu Zhang, Asaf Tzachor, and Wei-Qiang Chen. 2024. E-waste challenges of generative artificial intelligence. (2024).

Shoshana Zuboff. 2023. The age of surveillance capitalism. Routledge.

[6]https://www.nsf.gov/funding/opportunities/desc-design-environmental-sustainability-computing/

[7]https://ooolala.fr/en/