Towards a sustainable use of GPUs in Graphics Research

SIGGRAPH Talk 2025

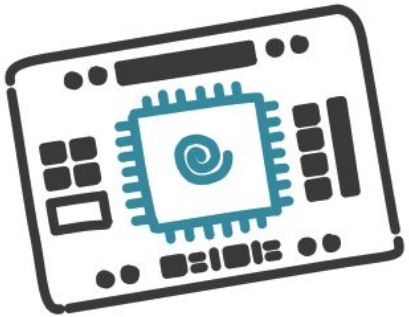
Emilie Yu, Élie Michel, Octave Crespel, Axel Paris, **Felix Hähnlein**

UC SANTA BARBARA

UNIVERSITY of WASHINGTON

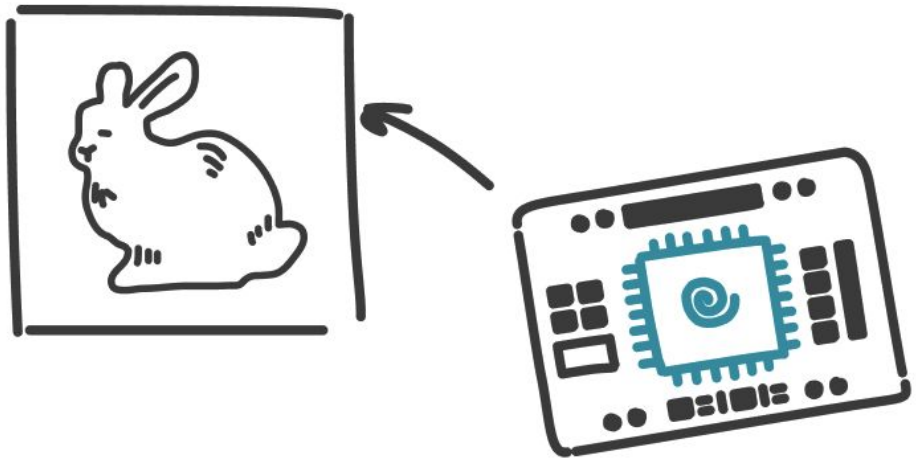
Motivation

GPUs & Graphics Research: historical ties…



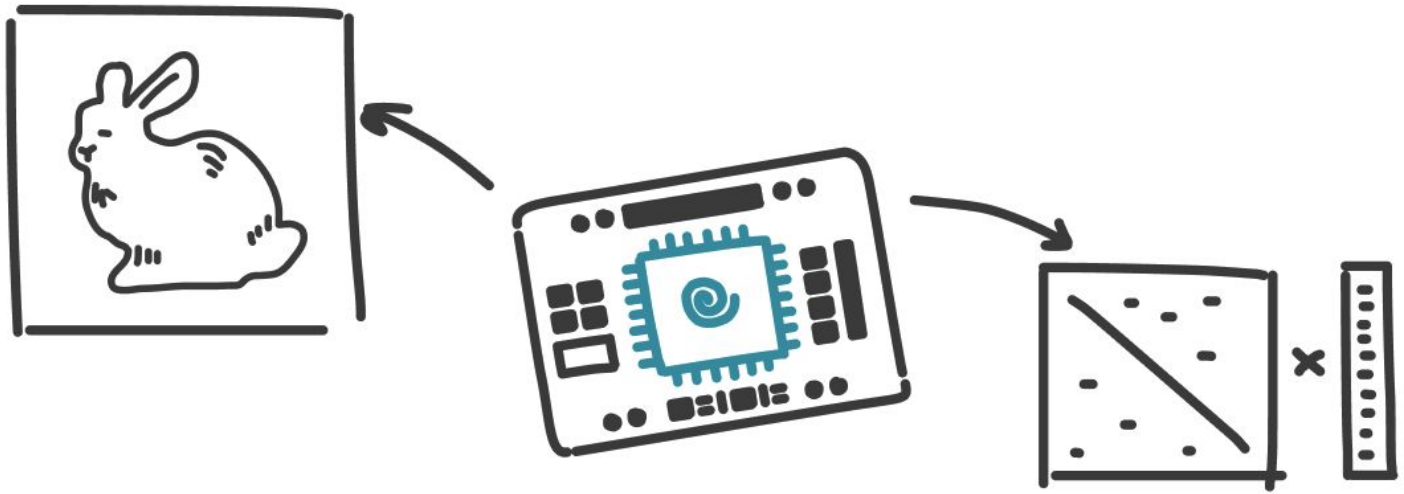
Motivation

GPUs & Graphics Research: historical ties…



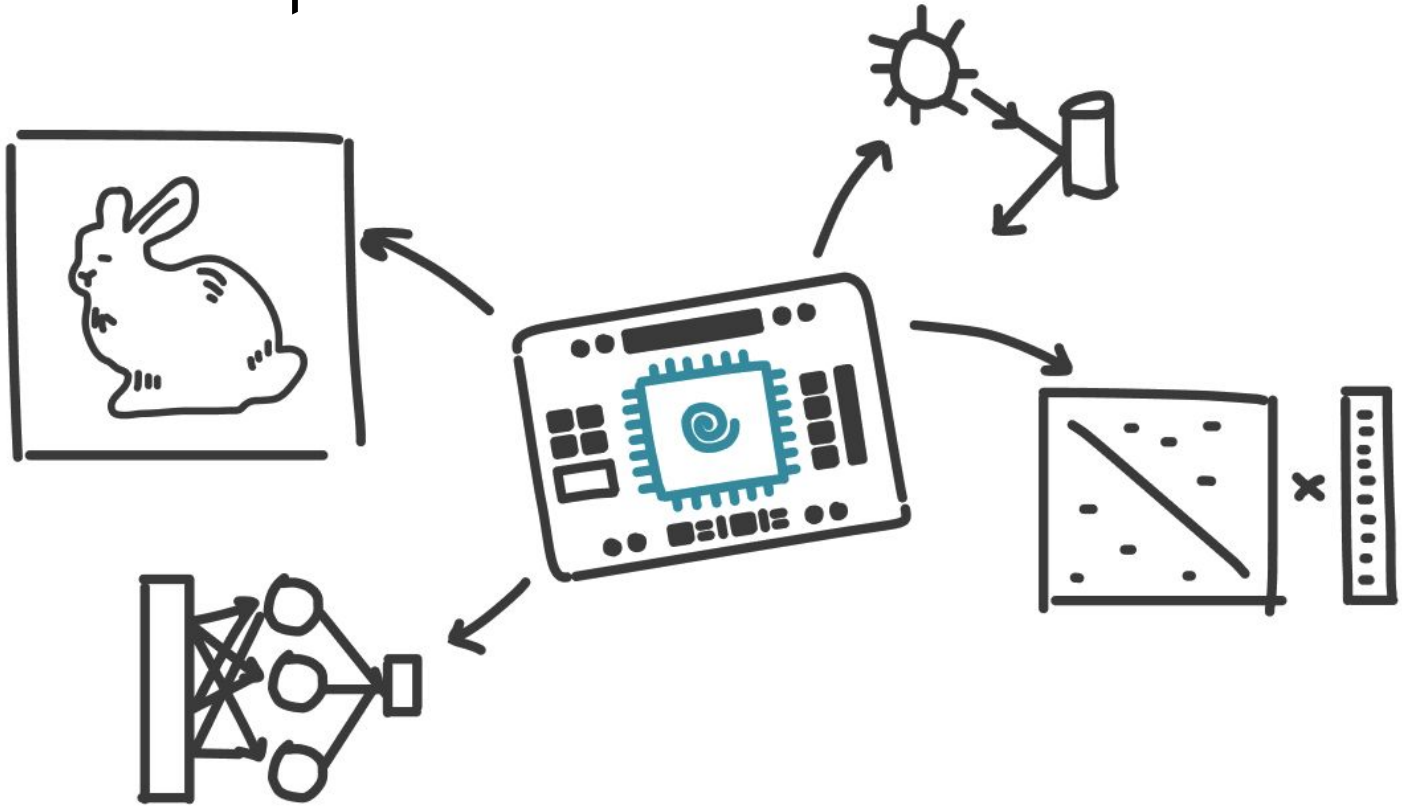
Motivation

GPUs & Graphics Research: historical ties…



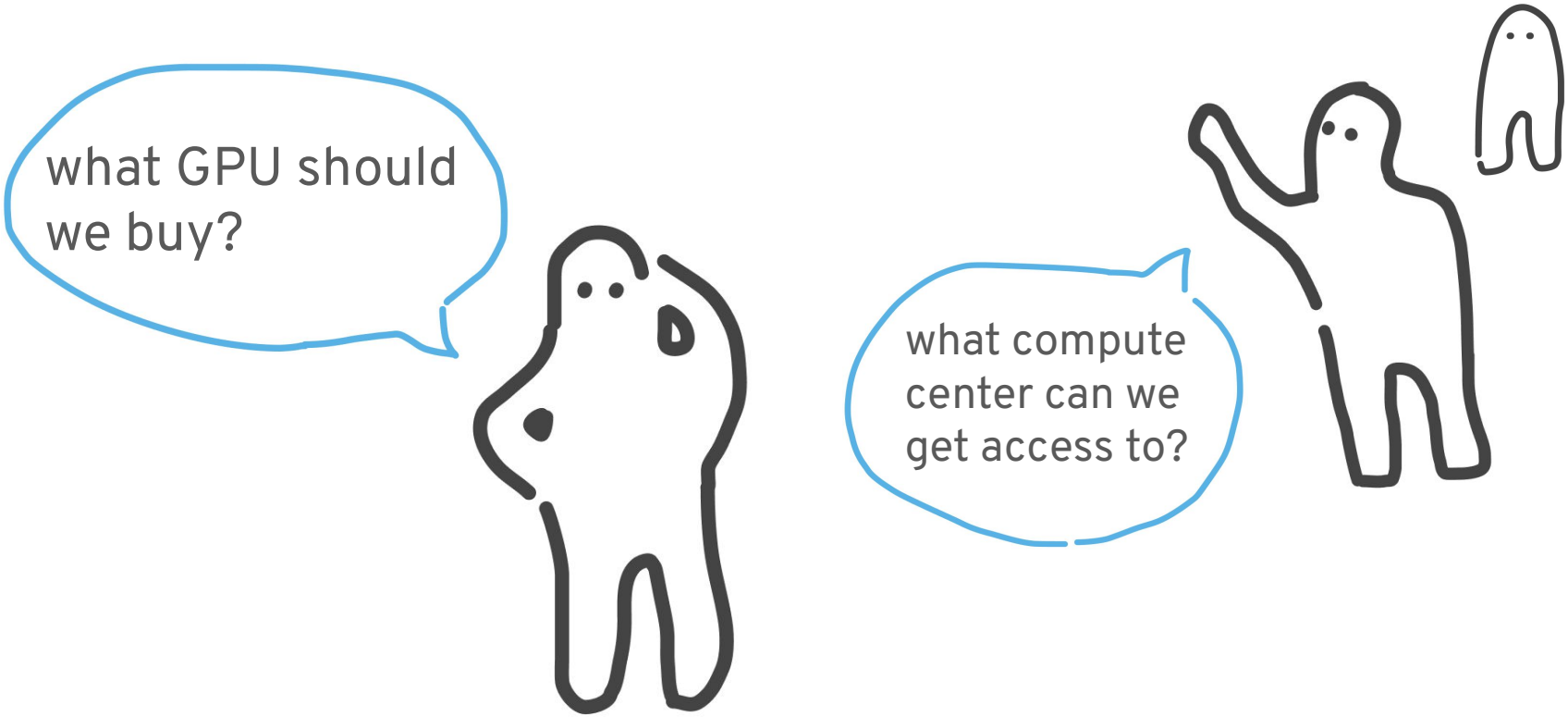
Motivation

GPUs & Graphics Research: historical ties…



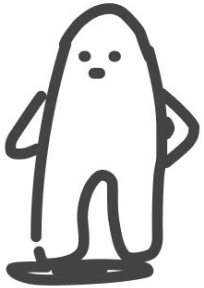
Motivation

… that persist today



Motivation

A systemic issue

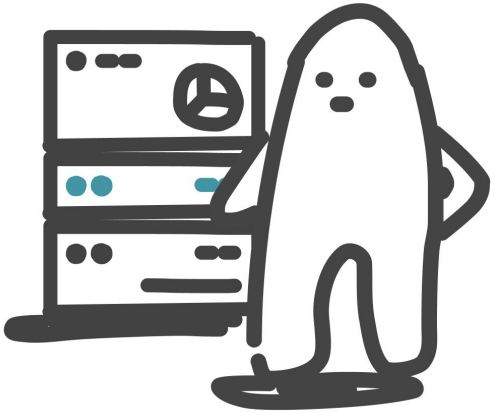


individual labs and researchers behave according to **systemic incentives**

we consider **trends** and implications at the scale of the graphics research community

Motivation

Broader implications

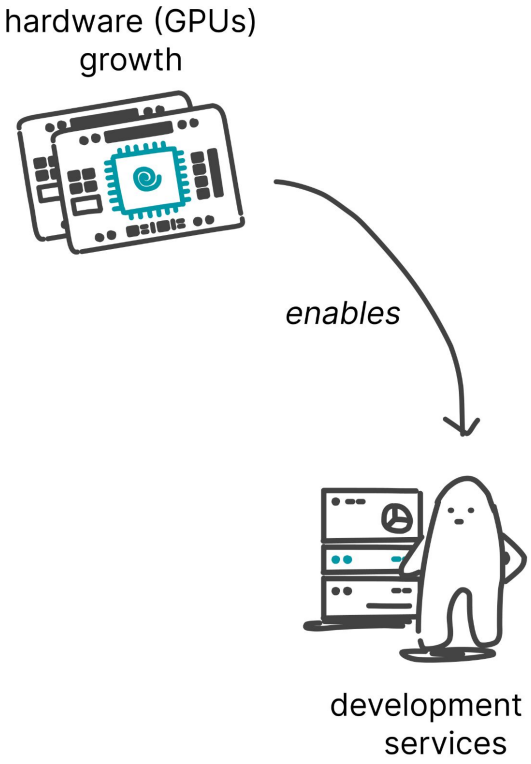


Our techniques are often meant to reach end users, who have access to different hardware than researchers.

Our research has an impact on **material infrastructure.**

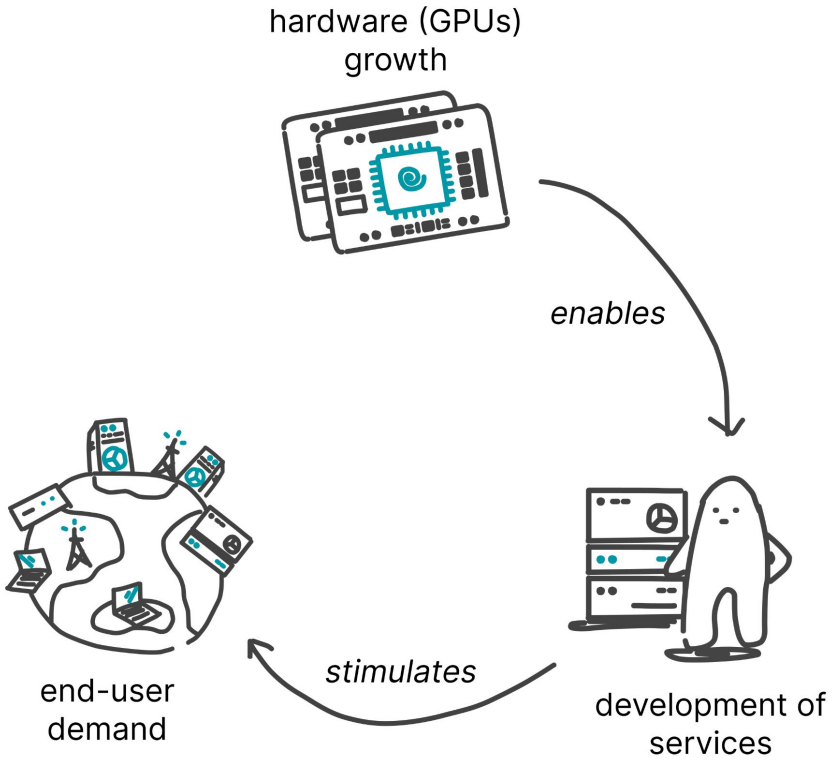
Motivation

Our part in a Cornucopian feedback loop



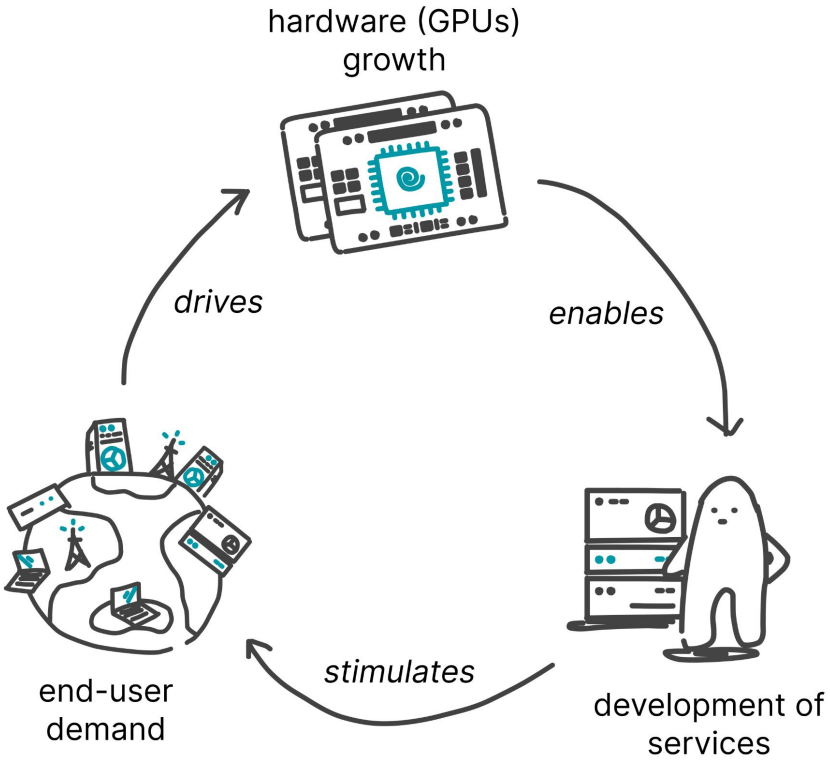
Motivation

Our part in a Cornucopian feedback loop



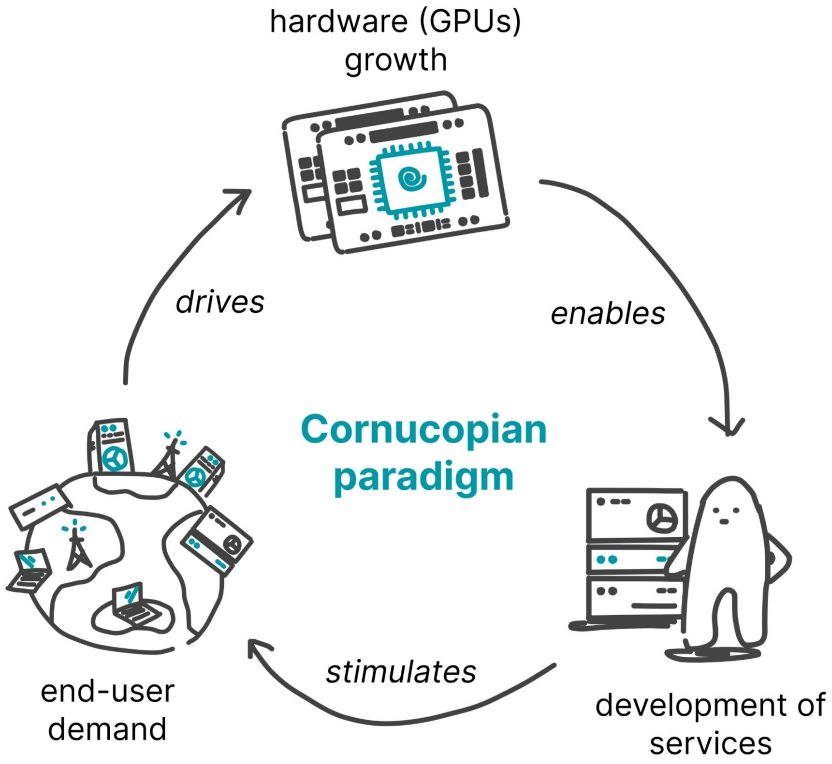
Motivation

Our part in a Cornucopian feedback loop



Motivation

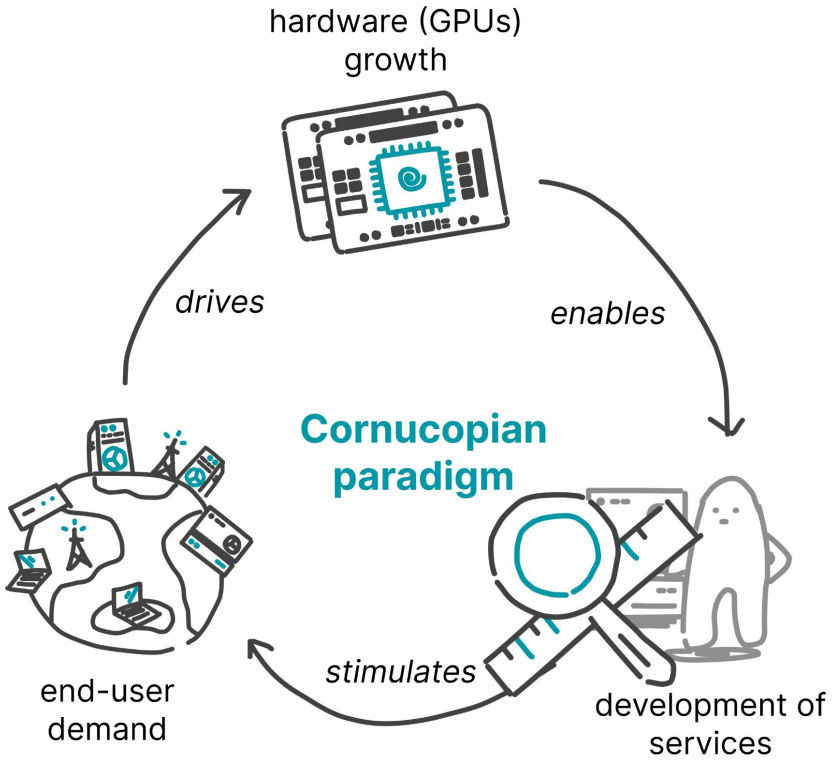
Our part in a Cornucopian feedback loop



Understanding and Mitigating the Effects of Device and Cloud Service Design Decisions on the Environmental Footprint of Digital Infrastructure [Preist et al. CHI 2016]

Motivation

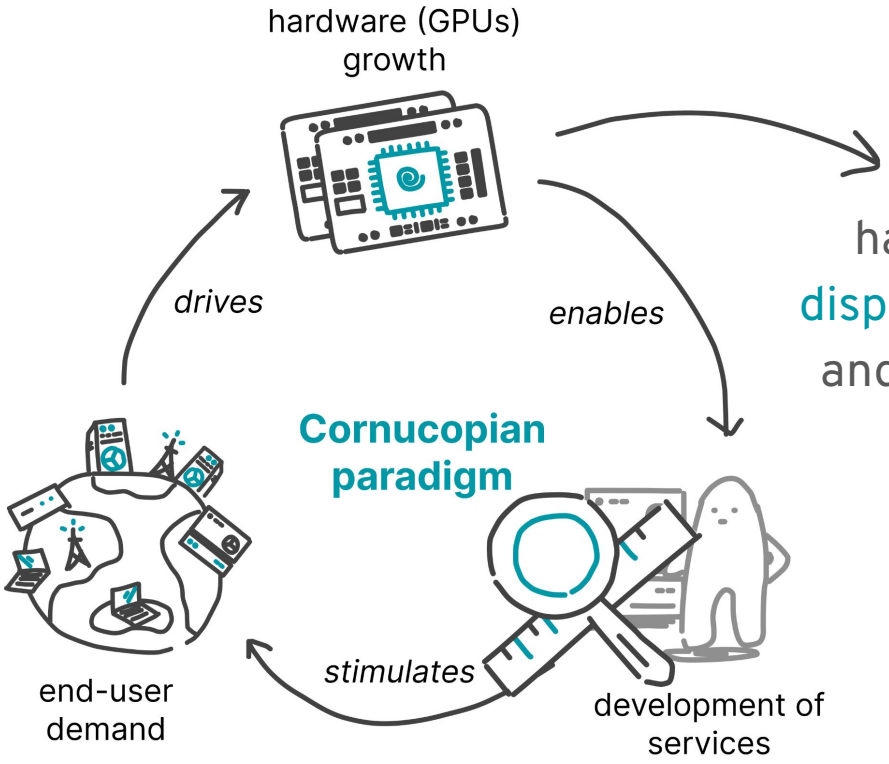
Our part in a Cornucopian feedback loop



Understanding and Mitigating the Effects of Device and Cloud Service Design Decisions on the Environmental Footprint of Digital Infrastructure [Preist et al. CHI 2016]

Motivation

Our part in a Cornucopian feedback loop

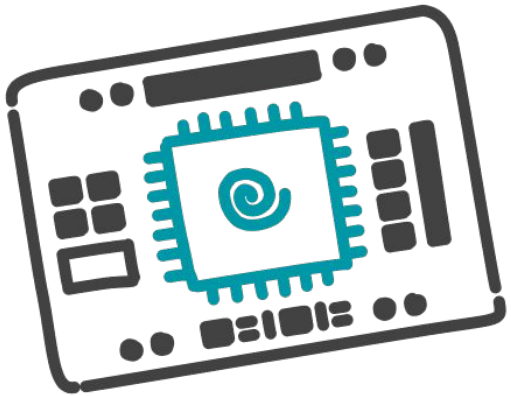


harmful impact from **disposing** of old hardware and **manufacturing** new hardware

Understanding and Mitigating the Effects of Device and Cloud Service Design Decisions on the Environmental Footprint of Digital Infrastructure [Preist et al. CHI 2016]

Motivation

The impact of hardware renewal

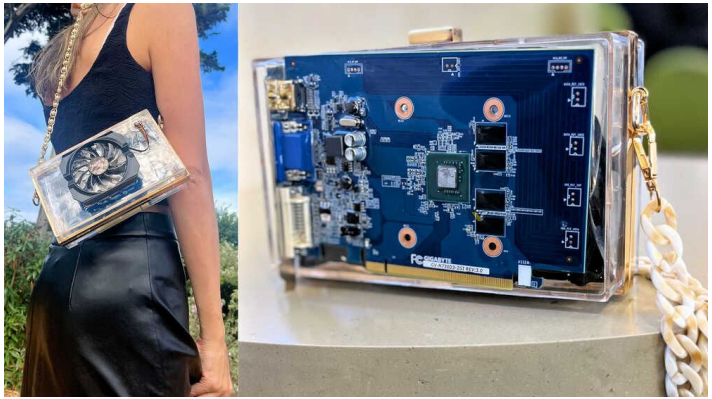


manufacturing a RTX 4090
costs around 140 kg CO2-eq

Source: Gigabyte GPU Carbon Footprints

Motivation

The impact of hardware disposal



Motivation

The impact of hardware disposal

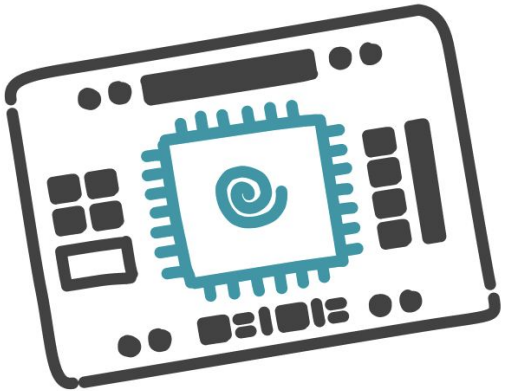
the **e-waste** problem

- **62 billion kg** produced in 2022
- only **22%** formerly collected & recycled
- the rest causes **harm**: soil, water, human health

Source: Global E-waste monitor

Motivation

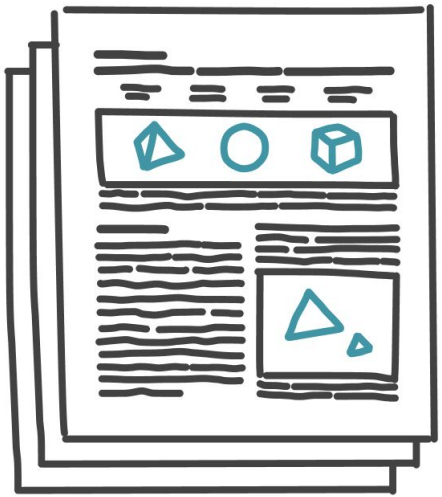
Research questions



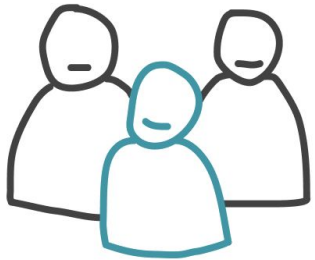
What GPUs do we use in computer graphics **research**?
What GPUs do **users** have?
How do those **compare** to one another?

Method

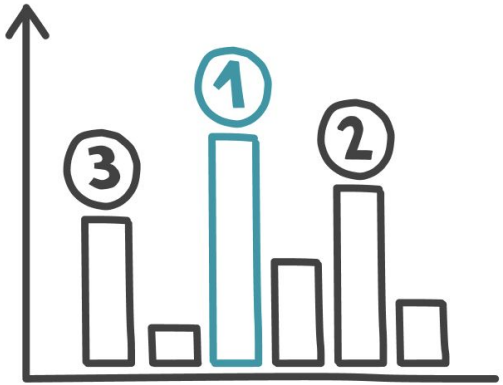
Three data sources



x



x



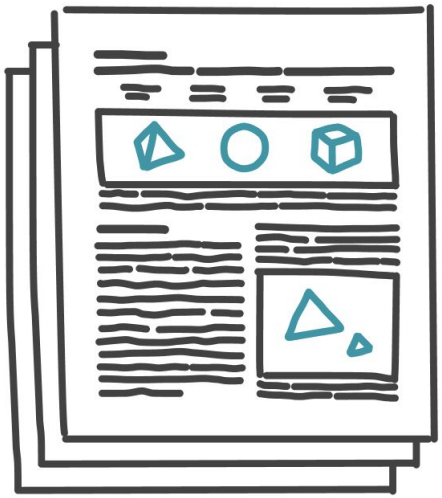
research papers

usage data

performance ranking

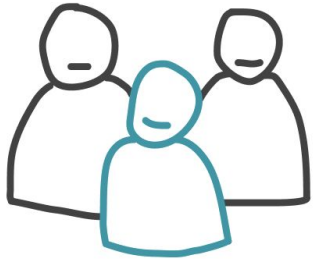
Method

Three data sources



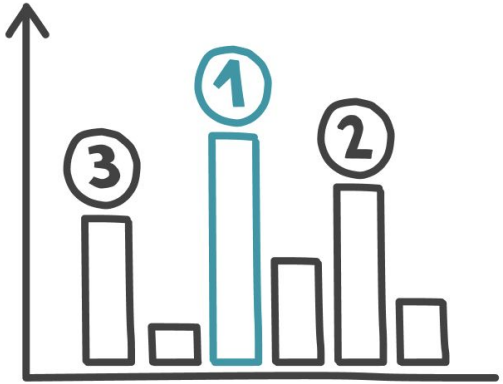
research papers

x



usage data

x



performance ranking

Method

Author-reported GPUs in papers

SIGGRAPH journal papers 2018-2024: 888 papers

4.2 Implementation

In general, let k be the number of eigenvectors/eigenvalues in use, we recommend to use the number of root nodes $m > k \times 2$. In Fig. 21 we show that if m is too small, the degrees of freedom are insufficient to capture the eigenfunctions with higher frequencies.

Our serial C++ implementation is built on top of LIBIGL [Jacobson et al. 2018] and SPECTRA [Qiu 2018]. We test our implementation on a Linux workstation with an Intel Xeon 3.5GHz CPU, 64GB of RAM, and an NVIDIA GeForce GTX 1080 GPU. We evaluate our runtime using the mesh from Fig. 4 in three different cases: (1) varying the size of input operators n, (2) varying the size of output operators m,

Liu et al. 2019

5.1 Performance

All scenes were processed using a PC with 3.4 GHz 6-core Intel Xeon E5-2643 CPU and a NVIDIA Titan X GPU and 64 GB of memory. Our implementation mostly consists of unoptimized CPU code. The GPU is currently only (insignificantly) used in the warping stage. We ran our system also on a slower 14" Razer Blade laptop with a 3.3 GHz 4-core Intel i7-7700HQ CPU and a NVIDIA GTX 1060 GPU. Interestingly, the warping stage performs faster on the laptop, most likely because CPU computation and CPU/GPU transfers dominate the runtime. Table 1 breaks out the timings for the various algo-

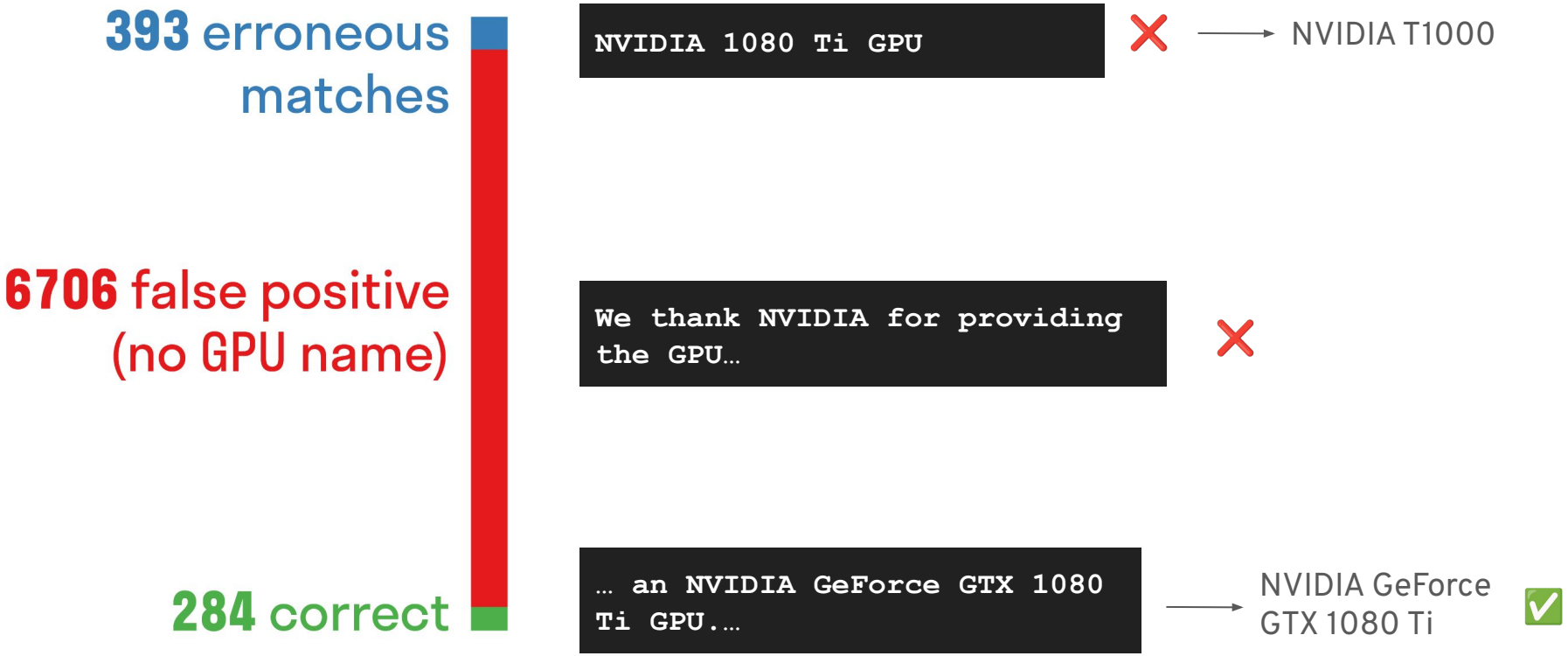
Hedman and Kopf, 2018

We train our CNNs using Tensorflow [Abadi et al. 2015] using a single NVIDIA Titan RTX; for each of our CNNs, an overnight run is typically enough to converge to good results using the Adam optimizer. After our CNNs have been trained, we run inference on a PC with an i7-9700K CPU and an NVIDIA GTX 1080 GPU. With a 4000 × 2160 input image, an inference pass of *CornerdetNet*

Chen et al. 2021

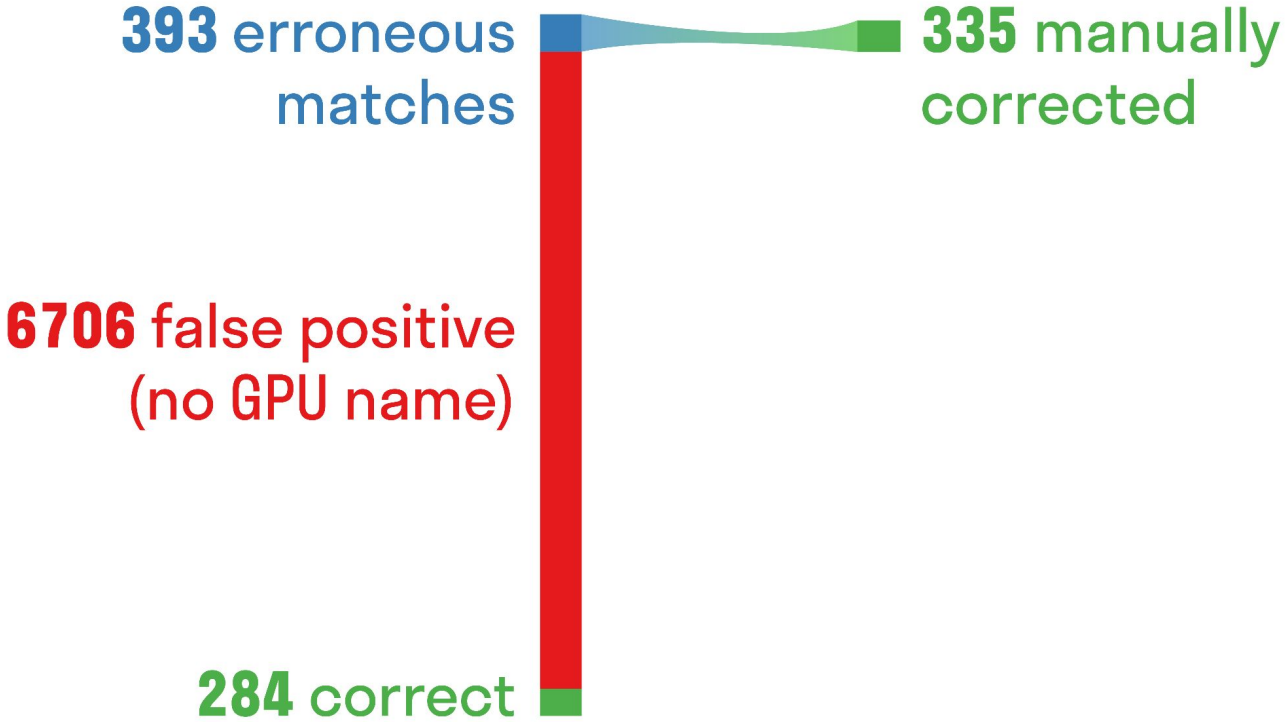
Method

“Automatic” GPU detection in papers



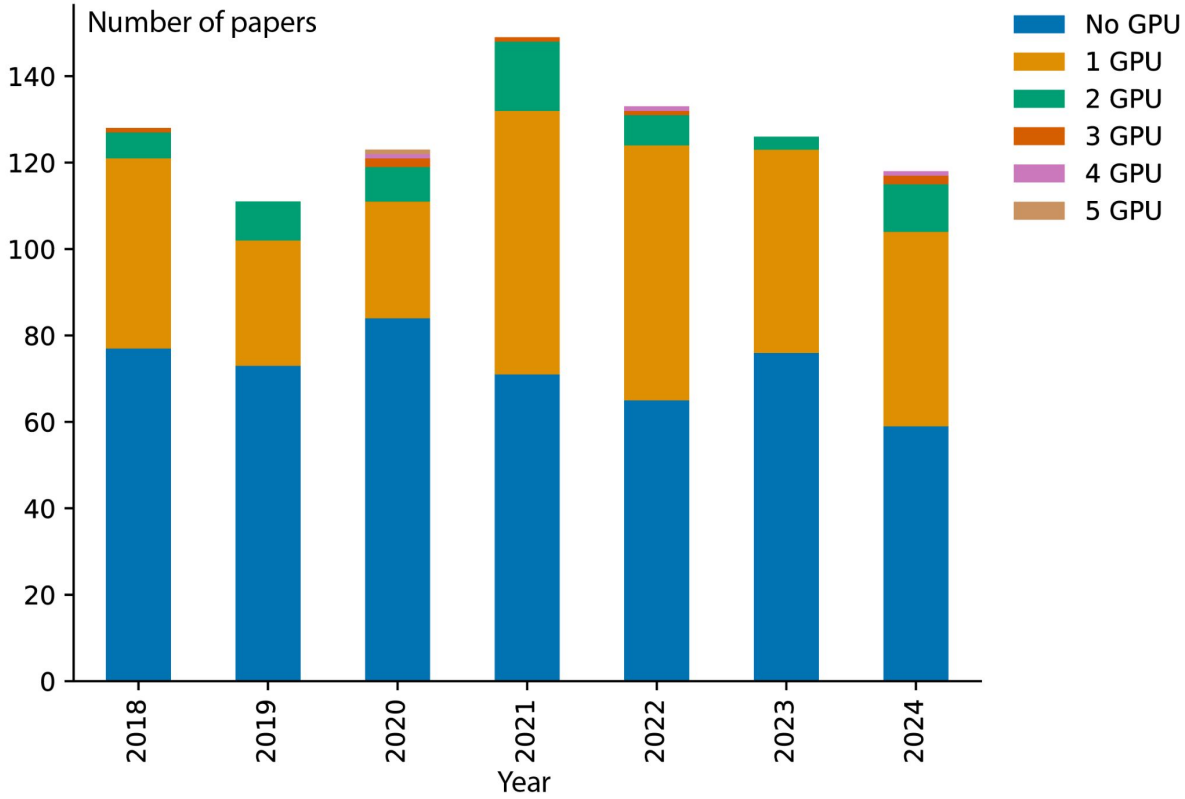
Method

Manual correction



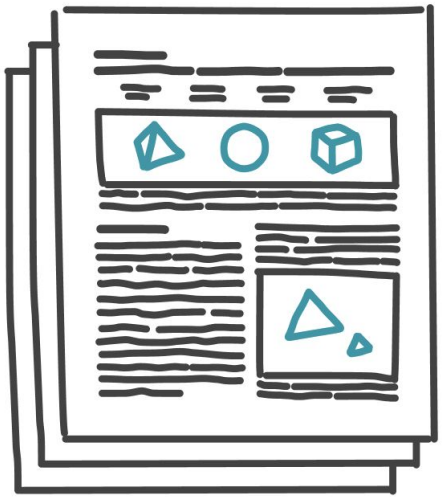
Results

GPU reports in papers



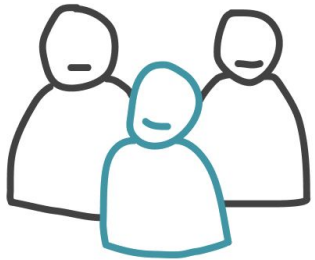
Method

Three data sources



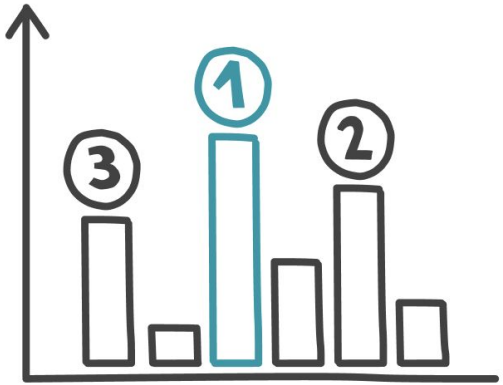
research papers

x



usage data

x



performance ranking

Method

GPU usage

STEAM®

Steam Hardware & Software Survey: April 2025

PC VIDEO CARD USAGE DETAILS

Sort By: Percent Share Sort

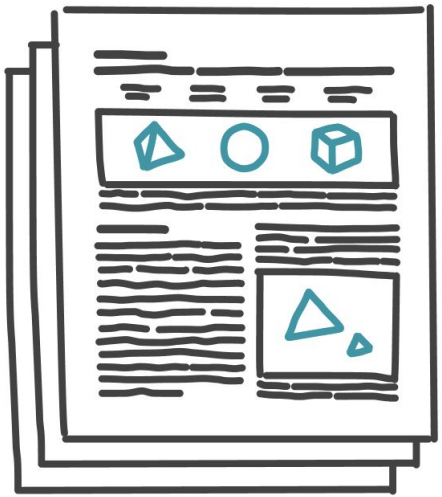
OVERALL DISTRIBUTION OF CARDS	DEC	JAN	MAR	APR	
DirectX 12 GPUs	92.71%	93.17%	92.01%	**90.96%**	-1.05%
DirectX 11 GPUs	0.63%	0.60%	0.59%	**0.55%**	-0.04%
DirectX 10 GPUs	0.41%	0.38%	0.39%	**0.36%**	-0.03%
DirectX 9 Shader Model 2b and 3.0 GPUs	0.00%	0.00%	0.00%	**0.00%**	0.00%
DirectX 9 Shader Model 2.0 GPUs	0.00%	0.00%	0.00%	**0.00%**	0.00%
DirectX 8 GPUs and below	6.25%	5.85%	7.01%	**8.13%**	+1.12%

ALL VIDEO CARDS	DEC	JAN	MAR	APR	
NVIDIA GeForce RTX 4060 Laptop GPU	4.19%	4.61%	4.48%	**4.99%**	+0.51%
NVIDIA GeForce RTX 3060	5.01%	5.20%	5.10%	**4.72%**	-0.38%
NVIDIA GeForce RTX 4060	4.04%	4.60%	4.77%	**4.51%**	-0.26%
NVIDIA GeForce GTX 1650	3.75%	3.56%	3.54%	**3.51%**	-0.03%
NVIDIA GeForce RTX 4060 Ti	3.26%	3.45%	3.15%	**3.25%**	+0.10%
NVIDIA GeForce RTX 3050	2.87%	2.93%	2.96%	**3.12%**	+0.16%
NVIDIA GeForce RTX 3060 Ti	3.07%	3.15%	3.05%	**2.92%**	-0.13%
NVIDIA GeForce RTX 3060 Laptop GPU	2.80%	2.81%	2.64%	**2.72%**	+0.08%
NVIDIA GeForce RTX 3070	2.99%	2.94%	2.87%	**2.68%**	-0.19%
NVIDIA GeForce RTX 2060	2.75%	2.76%	2.68%	**2.43%**	-0.25%
NVIDIA GeForce RTX 4070	2.74%	2.89%	2.49%	**2.43%**	-0.06%
NVIDIA GeForce GTX 1060	2.60%	2.54%	2.40%	**2.34%**	-0.06%
AMD Radeon Graphics	2.25%	2.07%	2.17%	**2.10%**	-0.07%
AMD Radeon(TM) Graphics	1.60%	1.66%	1.92%	**2.01%**	+0.09%

monthly reports of Steam users' hardware
(2008 - today)

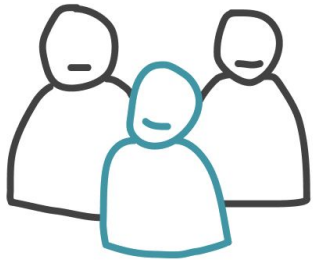
Method

Three data sources



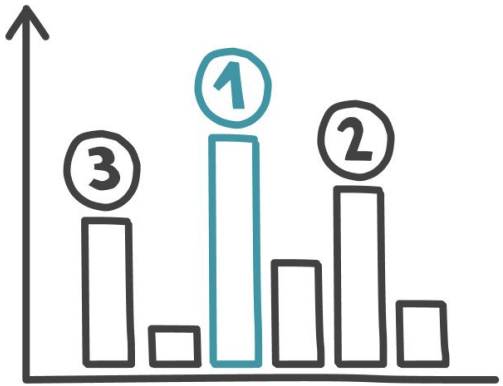
research papers

x



usage data

x

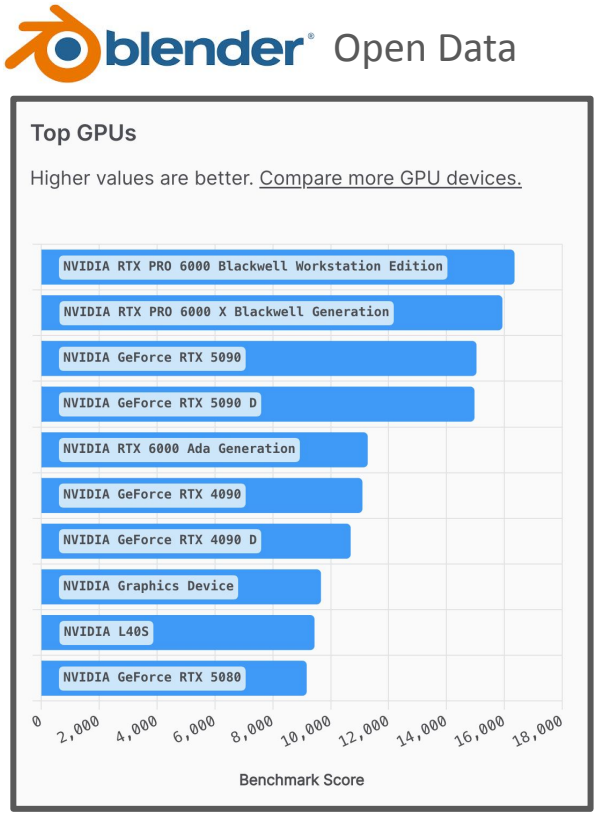


performance ranking

Method

GPU ranking

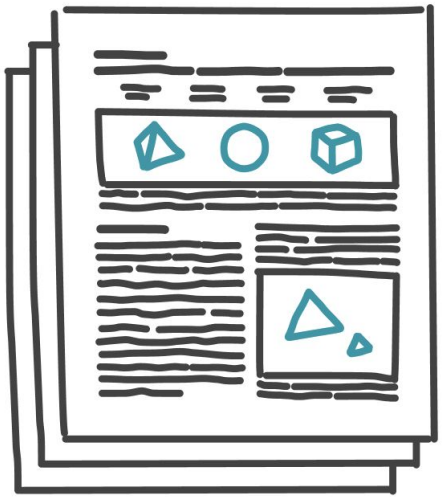
- “The Blender Benchmark Score is a measure of how quickly Cycles can render path tracing samples…”
- “The higher the number, the better.”



Source: [Blender Open Data]

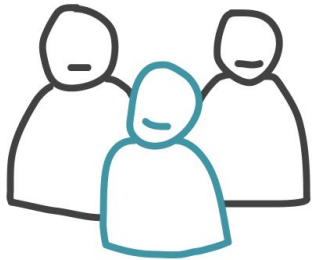
Method

Three data sources



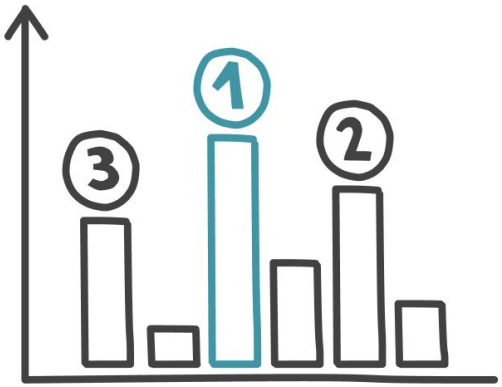
research papers

x



usage data

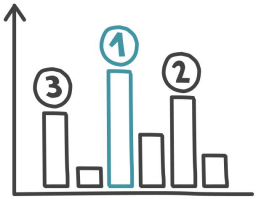
x

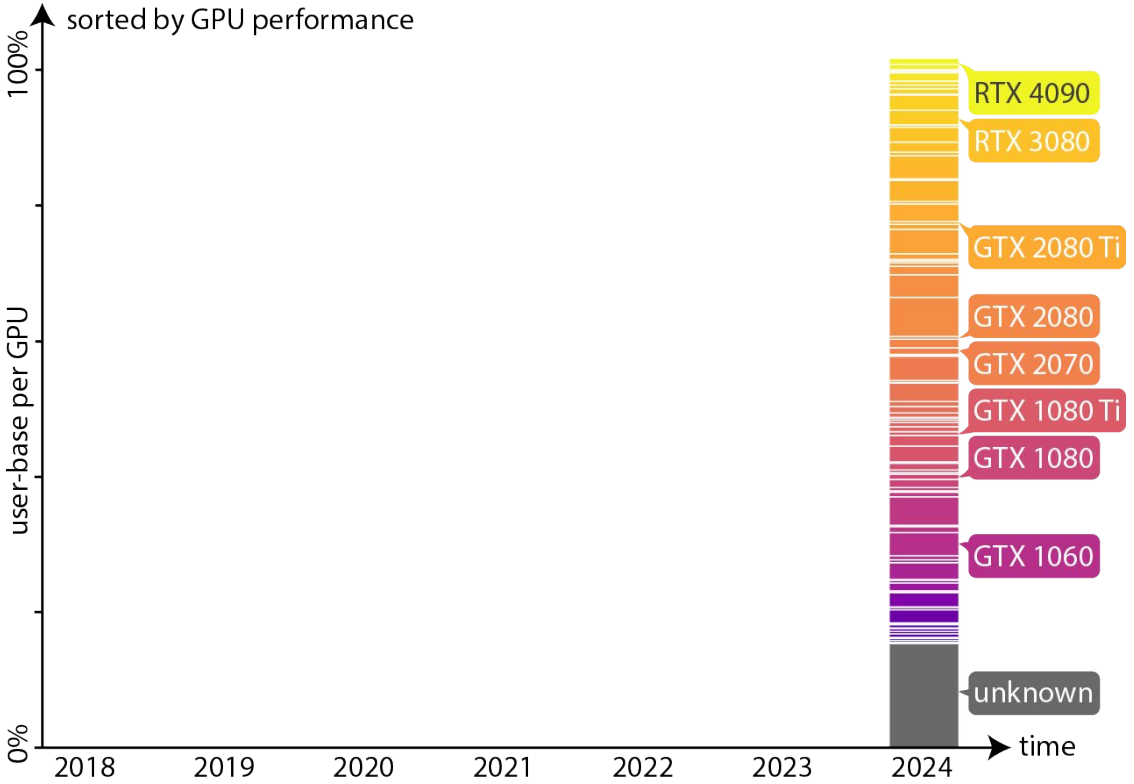


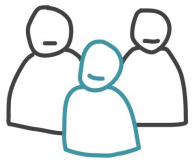
performance ranking

Results

Which GPU do people use?

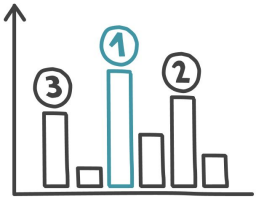


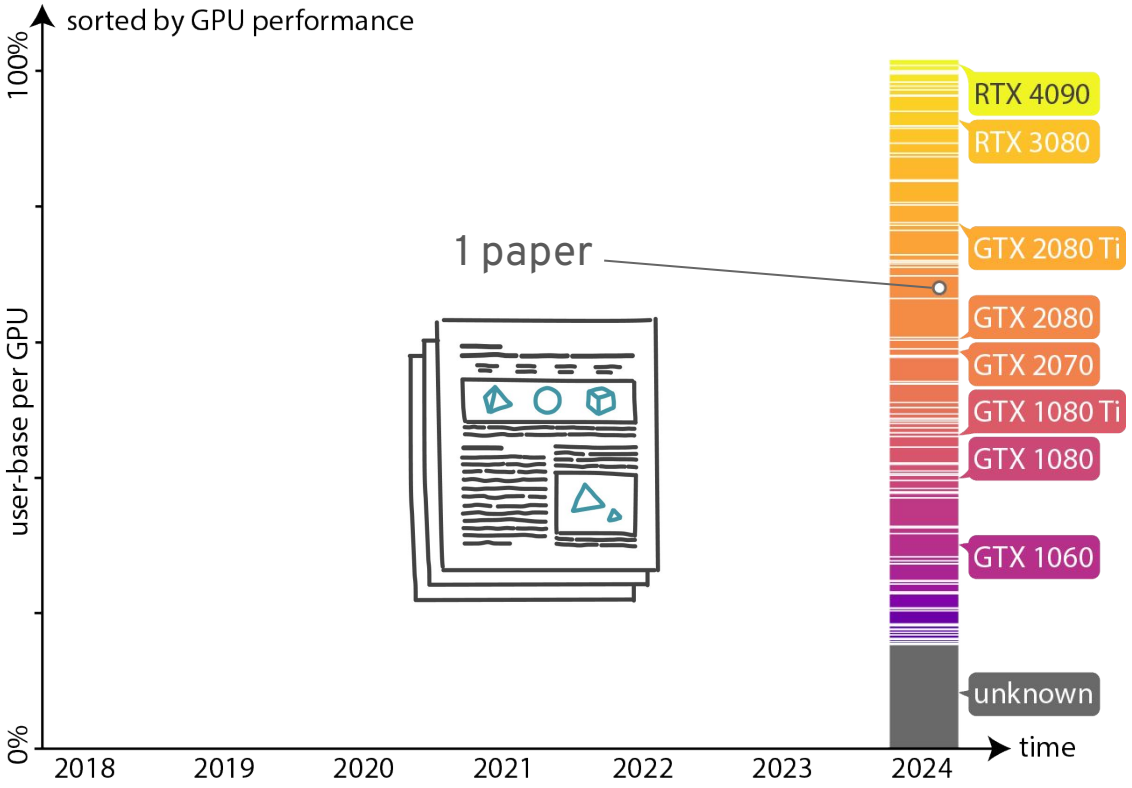


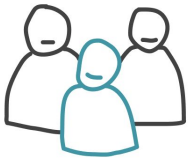


Results

Which GPU do papers report?

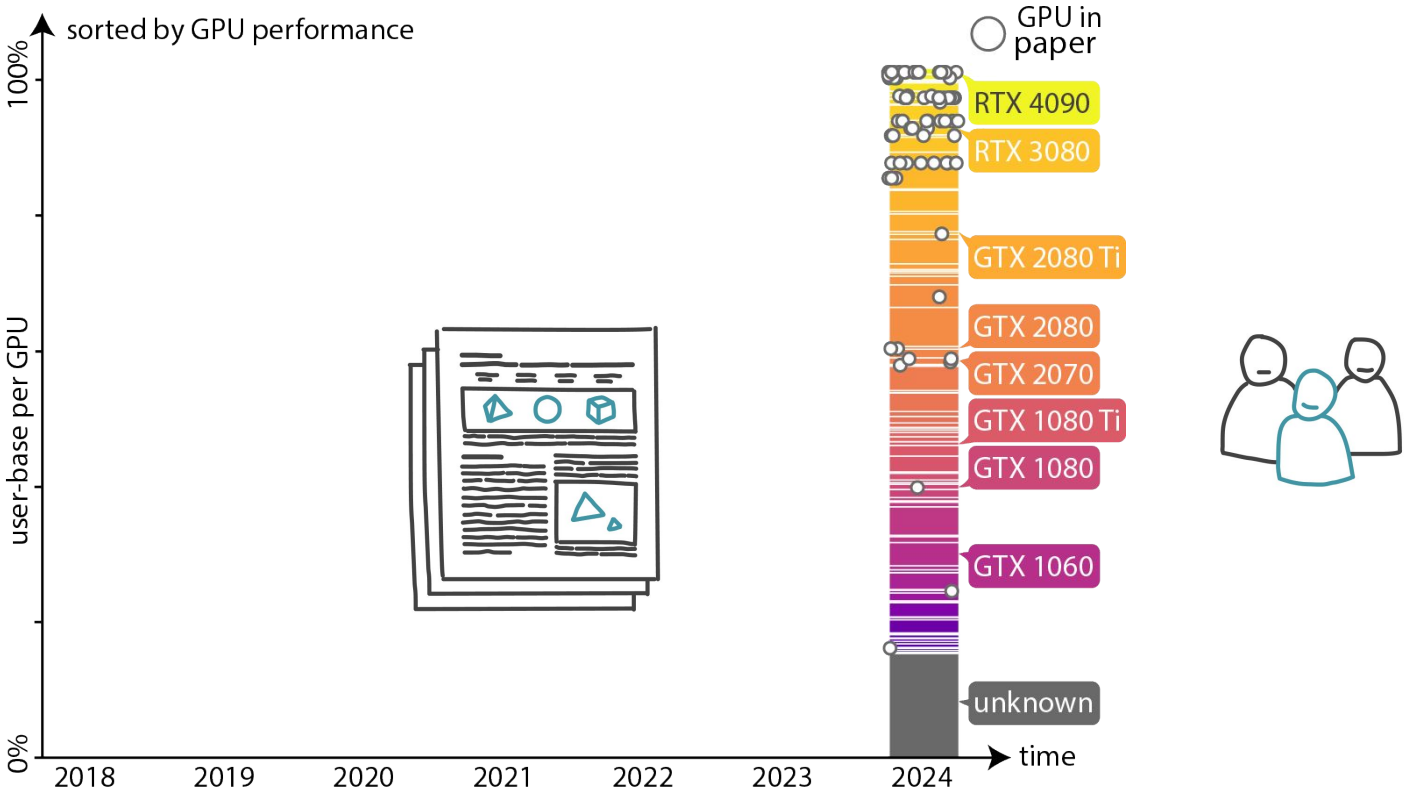






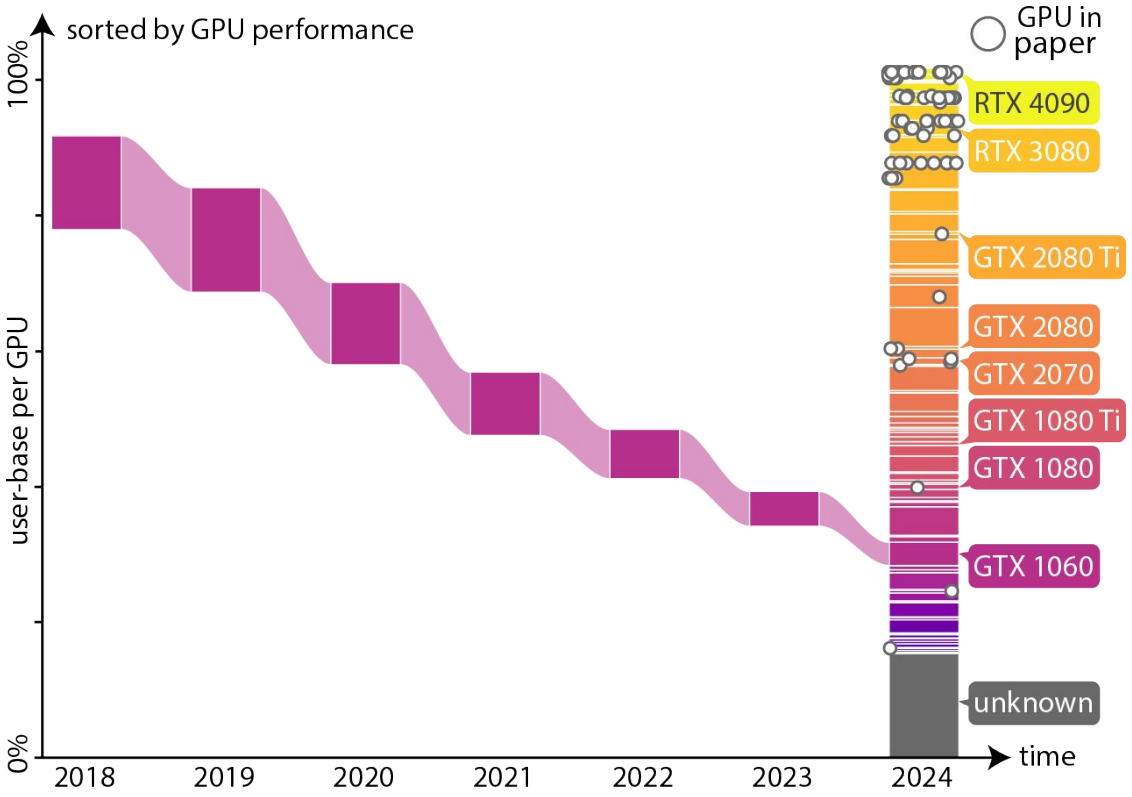
Results

Which GPU do papers report?



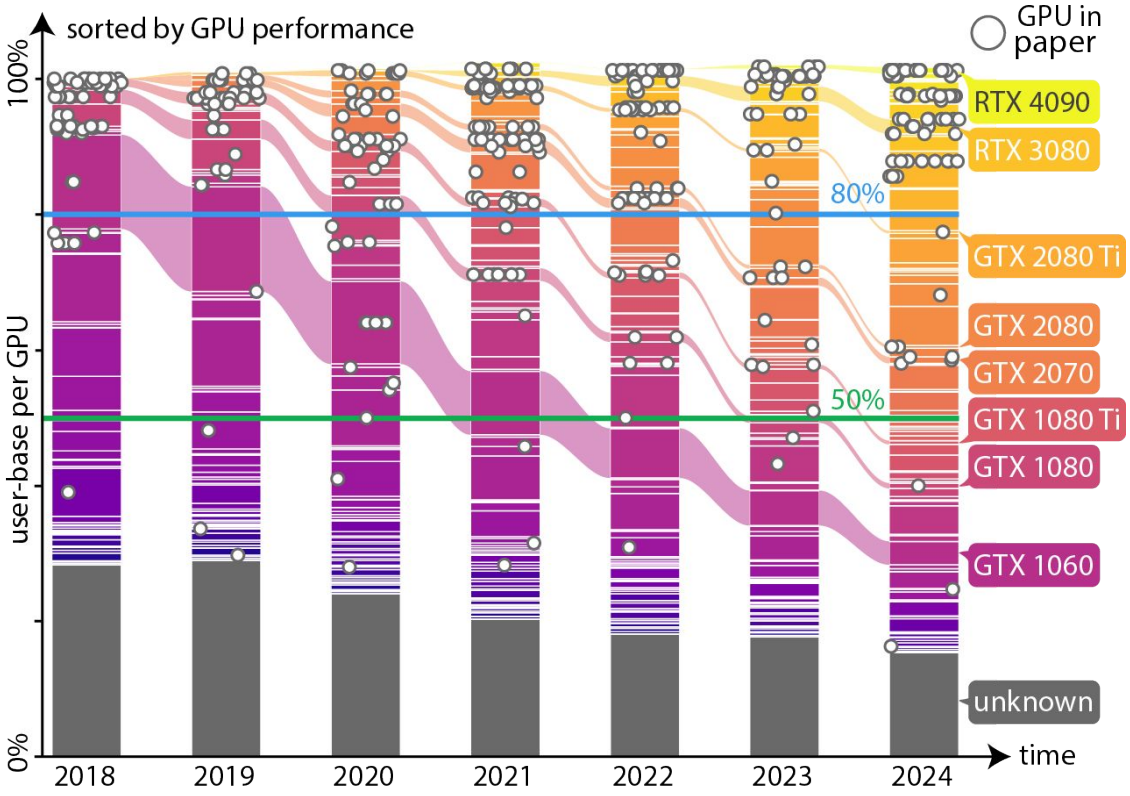
Results

How does usage evolve over the years?



Results

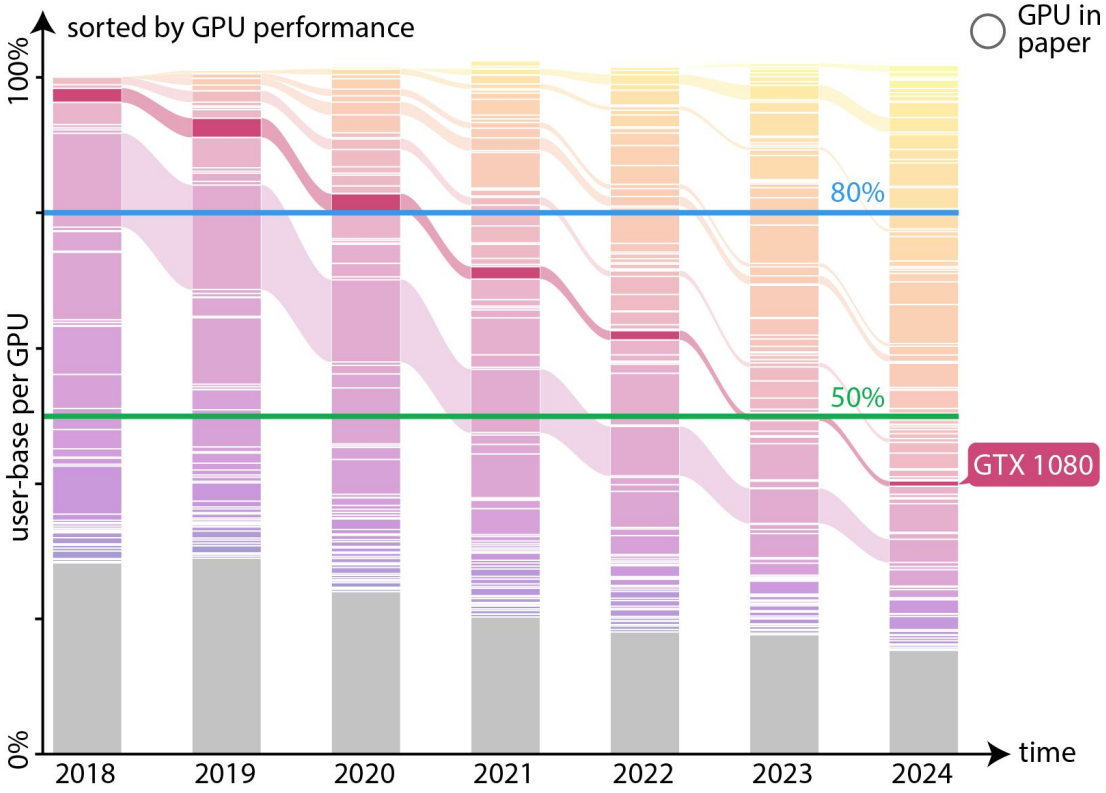
Putting it all together…



Results

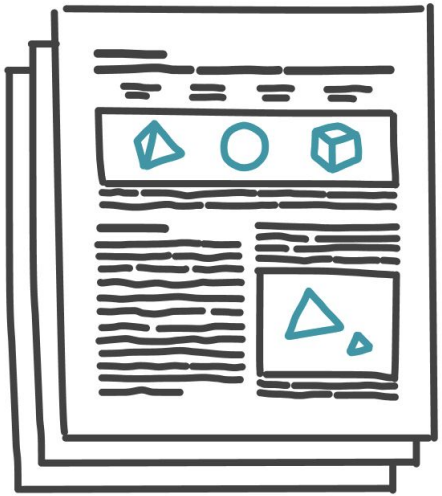
Key findings

- Consistent bias towards **high-end GPUs**
  - 87% available to less than 20%
- High-end GPUs become more **wide-spread after a few years**
- Increasing adoption of **cluster-GPUs** in papers (e.g. A100X, A800)

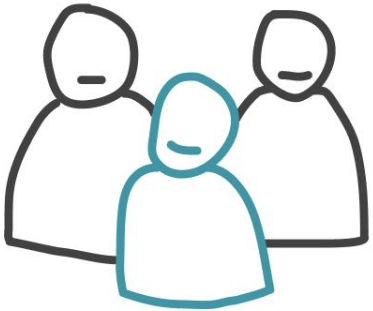


Limitations

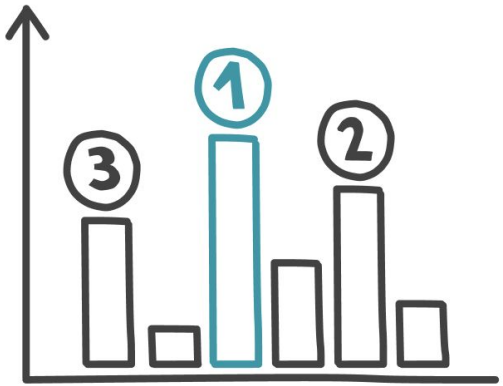
Potential biases in data sources



GPU names found in paper text



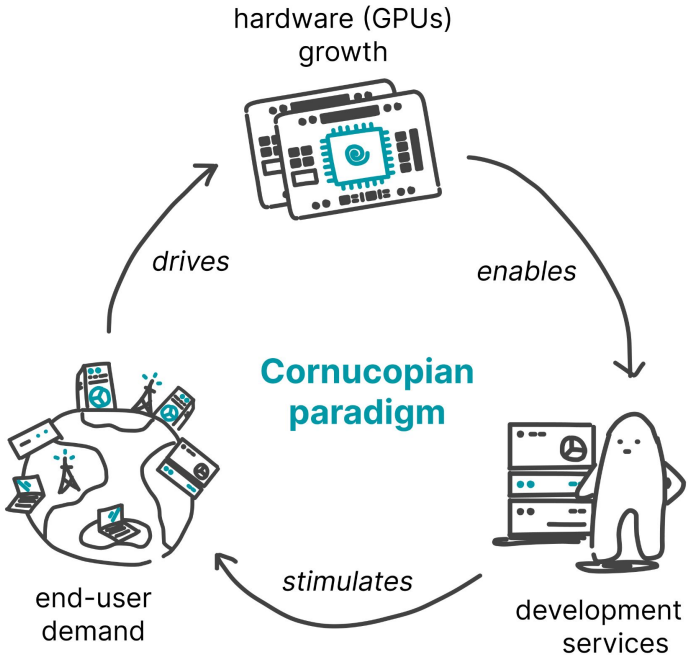
usage data among gamers population

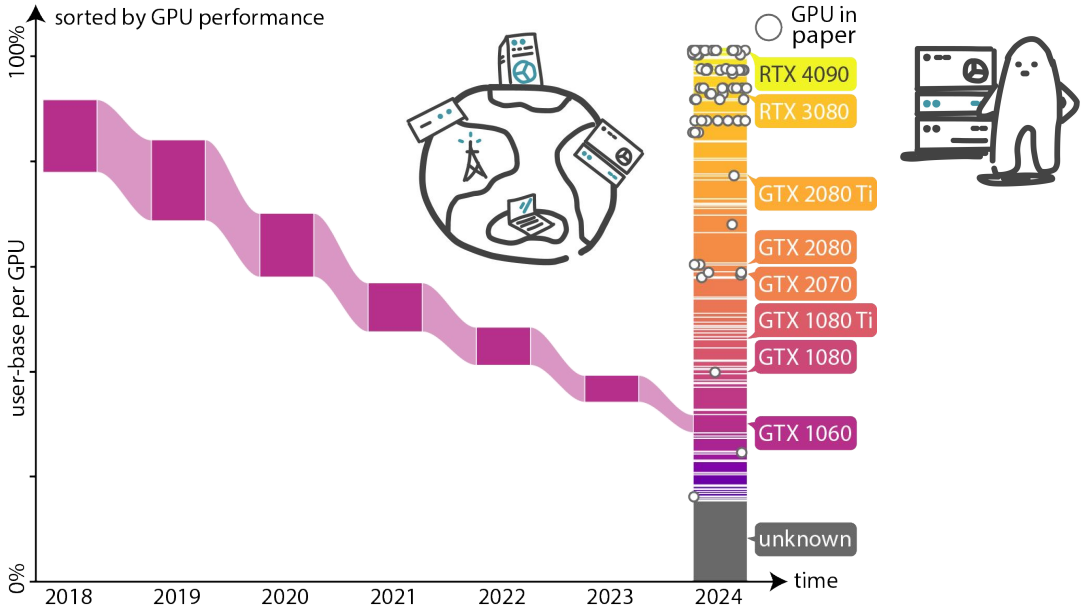


benchmark on a specific rendering task

Discussion

Confirmation of the Cornucopian paradigm?

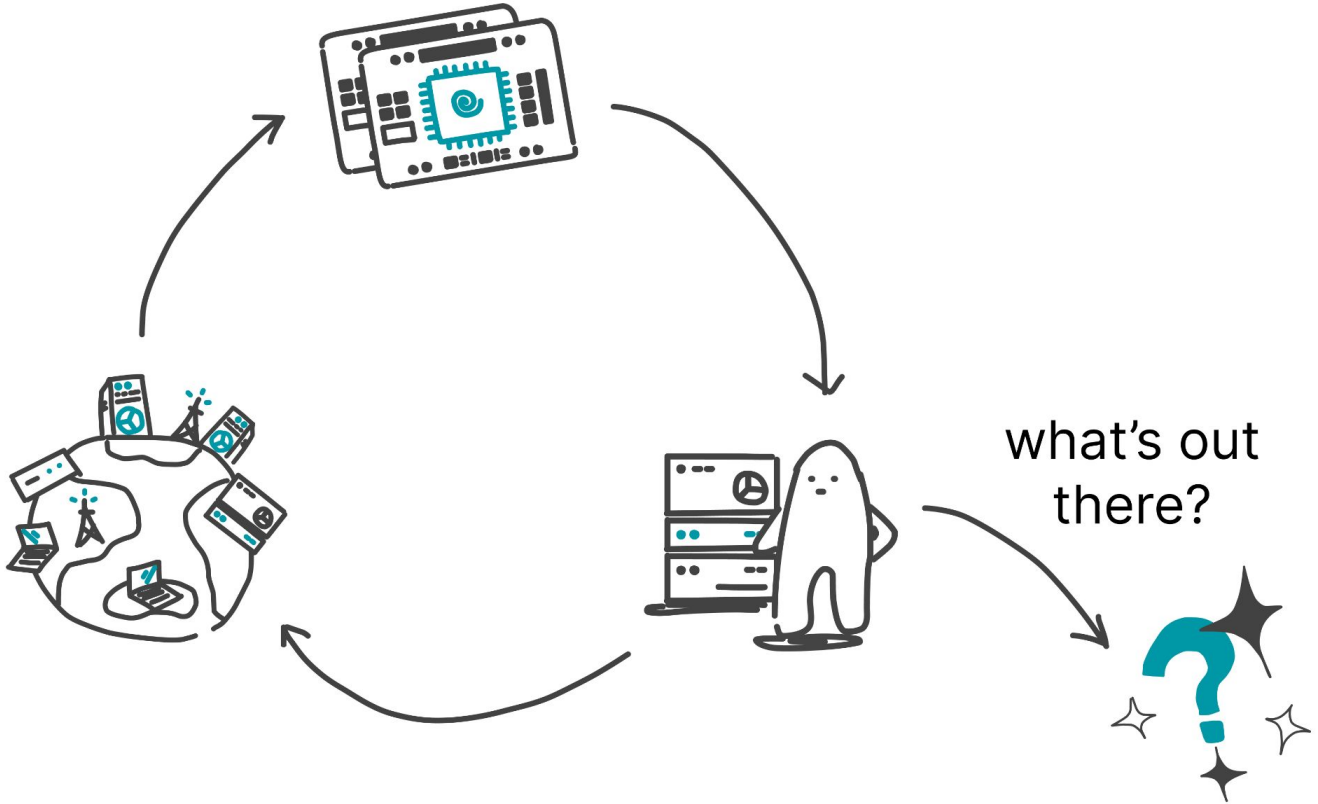




we do not prove causality between GPU use in graphics research and GPU renewal,
yet current GPU use in research **takes part in** the Cornucopian paradigm

Discussion

We can break out from the cycle



Opportunities for change

Levers for change at multiple levels

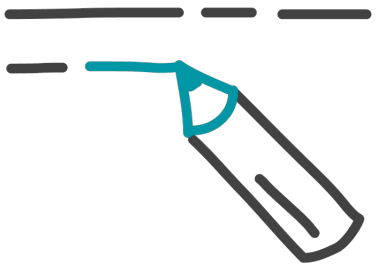
there are opportunities for change among different organizations and institutions

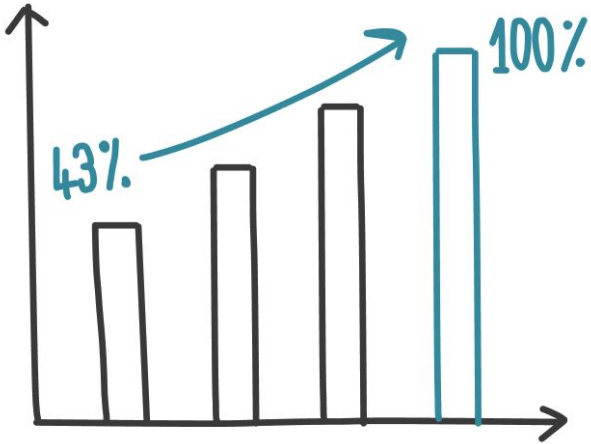
we present ideas as provocations, and prompts for further thinking, rather than definitive guidelines

Opportunities for change

Reviewing practices

Ensure adequate hardware reporting





enforce hardware reporting in reviews, adapt reviewing guidelines & forms

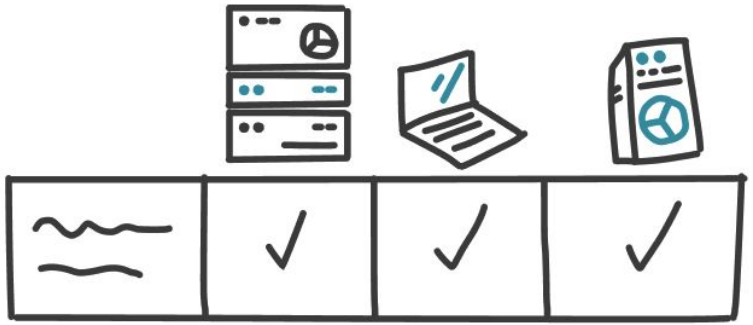
can we reach 100% of graphics papers that report their GPU?

Opportunities for change

Reviewing practices

Promote the reporting of metrics that are hardware agnostic

assess performance relatively to a baseline running on the same hardware, instead of absolute values



encourage reporting performance on a diversity of devices, when possible

Opportunities for change

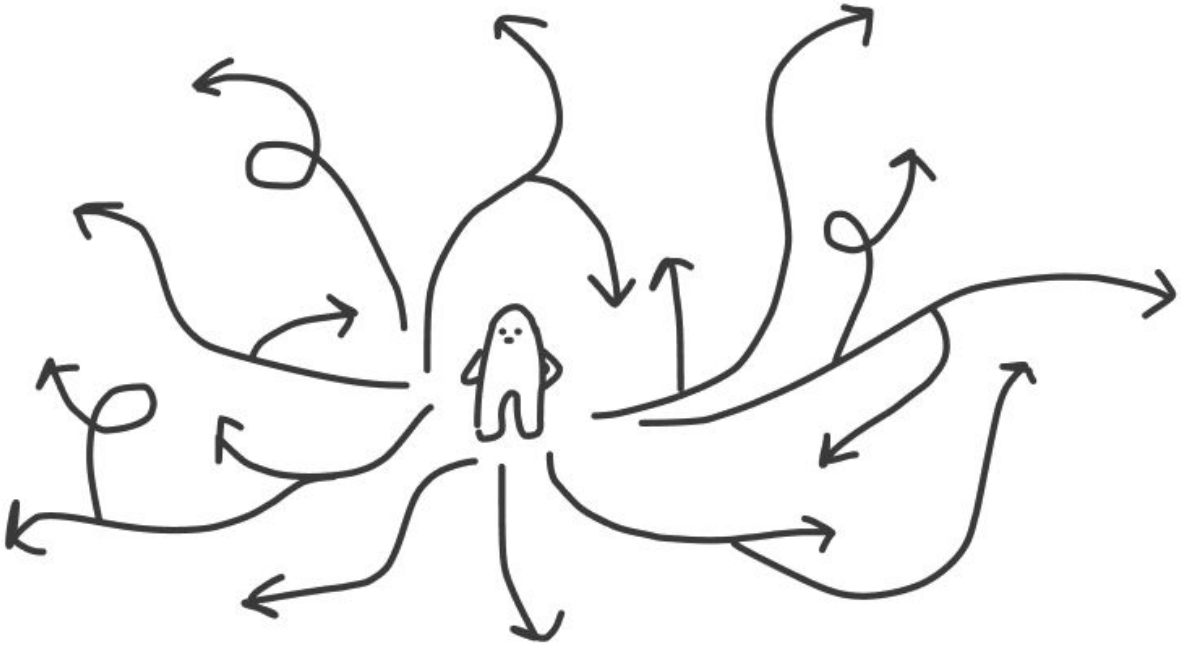
Reviewing practices

`Assess hardware accessibility as acceptance criteria`

expensive compute as an obstacle to adoption and reproducibility

Opportunities for change

Research directions

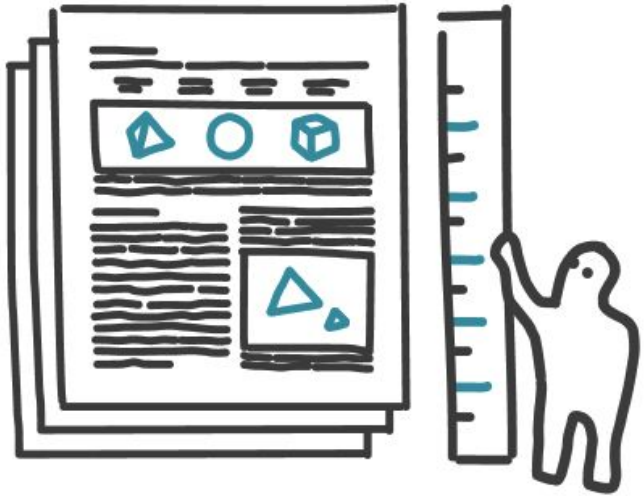


researchers have some* lee-way in choosing what to work on
there are exciting research directions outside of the cornucopian paradigm

Opportunities for change

Research directions

`Develop ways to measure performance effectively`

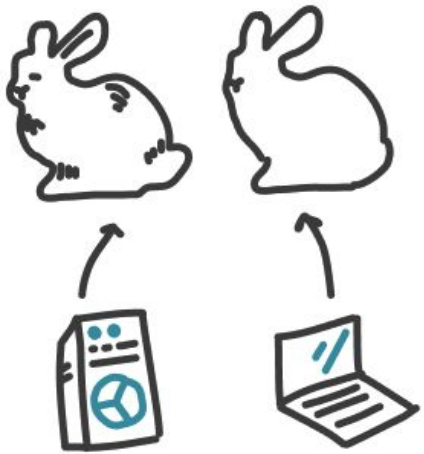


benchmarks and metrics to characterize performance
across a distribution of low-to-high-end hardware

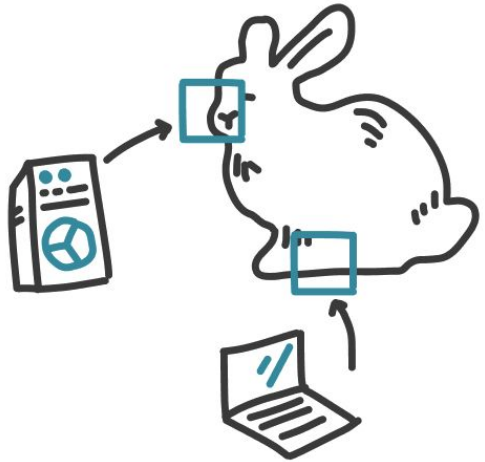
Opportunities for change

Research directions

Consider constrained hardware setups as opportunities for innovative solutions



adaptive techniques for diverse hardware

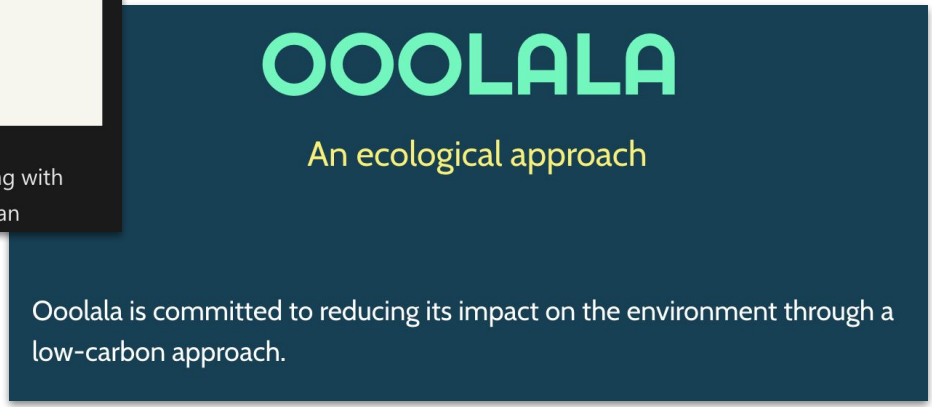


distributed techniques to split the workload on low-end machines

Opportunities for change

Graphics industry actors

Source: [Microsoft Gaming Sustainability]



Source: [Ooolala – 2D animation studio]

Conclusion

Takeaways

Research papers at Siggraph quasi-systematically use **high-end GPUs**.

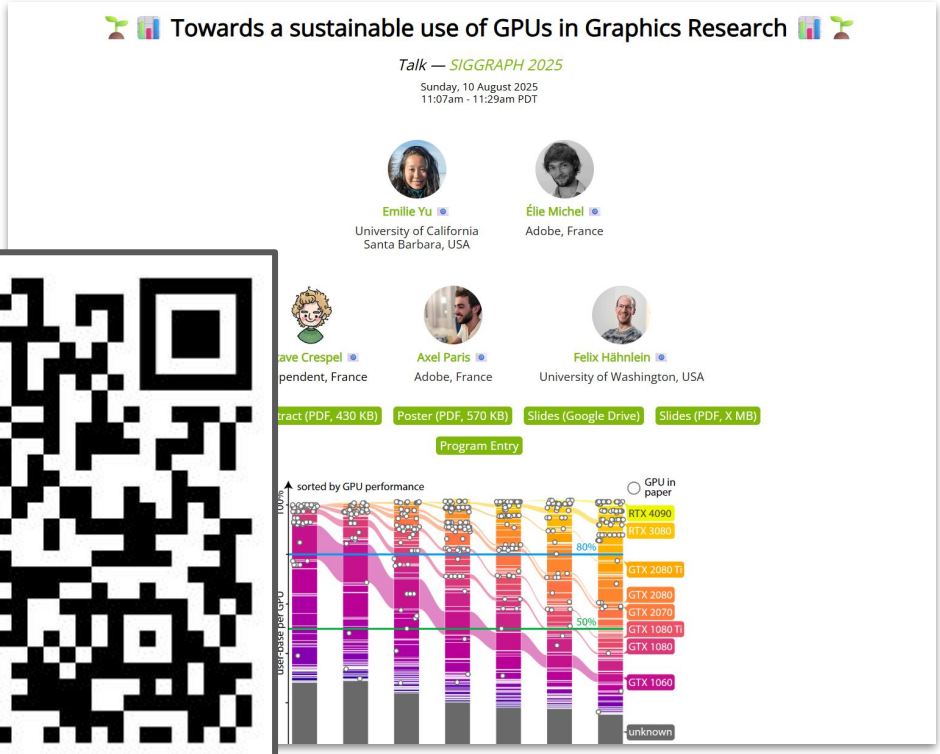
This behavior contributes to a cycle of premature **end-user device renewal**.

This is not inevitable.

Reviewers, researchers, and industry practitioners can **change their practices** to tend towards a more sustainable use of GPUs.

Thank you for your attention!

Going further



https://eliemichel.github.io/sustainable-gpu-usage

Continuing the discussion

Birds of a Feather
Sustainable Research in Computer Graphics

Tuesday 4-5PM
East Building room 9

End of slides