EVIDEN x MIAI Grenoble Alpes
Multidisciplinary Institute In Artificial Intelligence

GREEN HPC

An analysis of the domain based on Top500

Link to the article: https://hal.science/hal-04519645

Authors : BENHARI Abdessalam, Yves Denneulin, Fanny Dufossé, Frederic Desprez, Denis Trystram

May 2023

an atos business

1. Introduction

Introduction :

Context

- The demand for computing power has continuously increased over the years.
- Today, the performance of the most powerful systems surpasses the *exascale*, and the number of *petascale* systems continues to rise.
- However, this growth is accompanied by escalating energy costs, leading to a significant carbon footprint.

Introduction :

Some interesting numbers …

- A recent survey of studies focusing on methodologies and tools for estimating the carbon footprint of the ICT sector indicates a value ranging **between 2.1% and 3.9% of global greenhouse gas (GHG) emissions** (1).
- Among the various electronic and computing systems, data centers alone contribute to approximately **one percent of the world's electricity consumption** (2).
- In France, it's estimated that the ICT sector will **represent 10% of electricity consumption by 2030** (3).

1. Freitag et al. : The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations.
2. Masanet et al. : Recalibrating global data center energy-use estimates
3. ADEME: Evaluation de l'impact environmental du numérique en france et analyse prospective.

Introduction :

Contributions

- analyze the empirical laws governing the evolution of HPC computing systems both from the performance and energy perspectives,
- Analyze the most relevant data to study the performance and energy efficiency of large-scale computing systems,
- Put these analyses into perspective with effects and impacts (lifespan of HPC systems),
- Derive an estimation of GHG emissions for the HPC domain within the horizon 2030.

EVIDEN x MIAI Grenoble Alpes
Multidisciplinary Institute In Artificial Intelligence

2. Experimentations

Experimentations :
Data, methodology & limitations

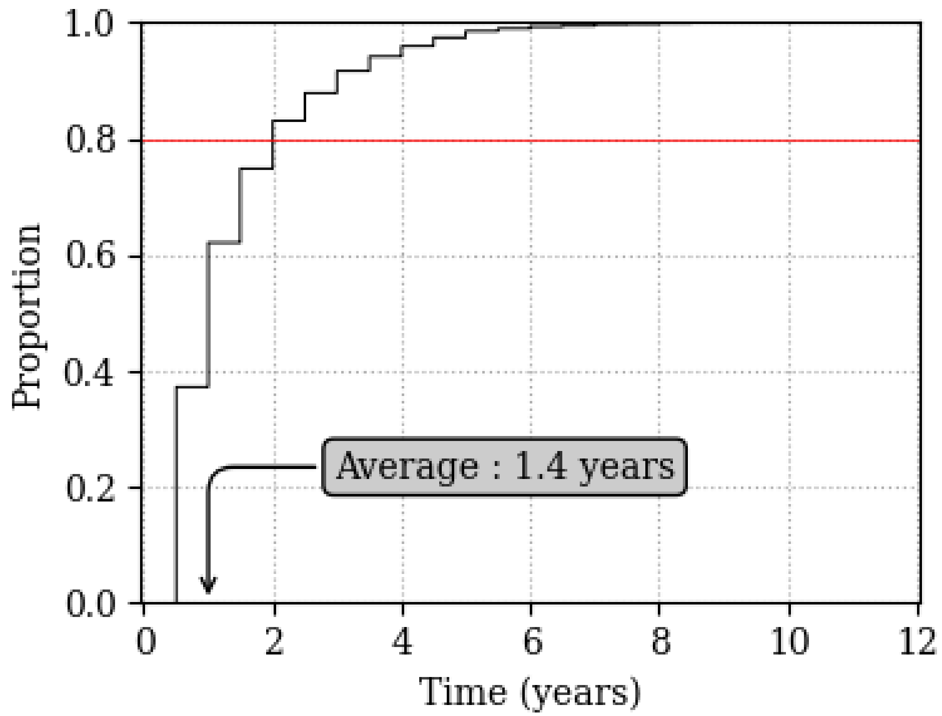
- **Data ->** The Data used to run the experiments (and thus produce the different figures) comes from the Top500, Green500 and Graph500 Lists.
- **Limitations ->** Although these lists offer a consistent perspective on collected features spanning numerous years, making them a valuable resource for tracking the progression of HPC architectures. They present two main drawbacks : Their **declarative character** and **orientation towards large scale HPC systems.**

Experimentations :

First experiments (HPC Lifespan) : The Top500 apparition time of HPC machines.

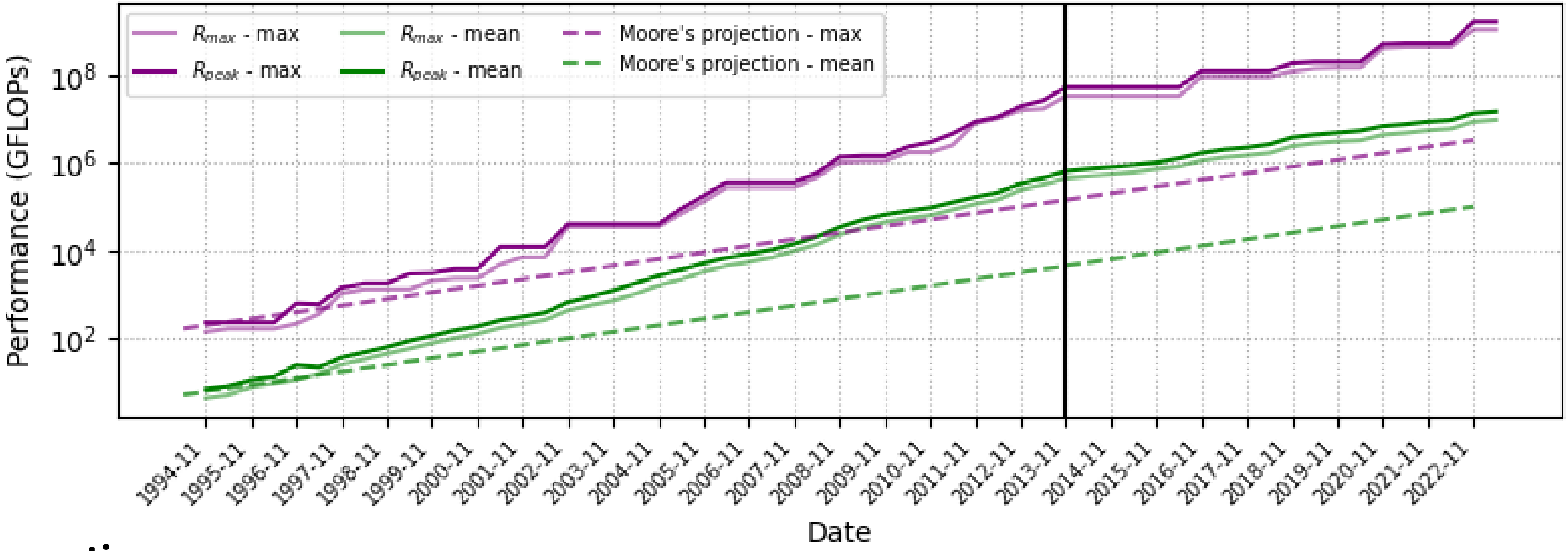
Observation: Only a select few of the highest-ranking supercomputers on the Top500 list maintain their position for up to 8 years, with 80% of them appearing for no more than 2 years, resulting in an average appearance duration of 1.4 years, indicating the rapid evolution of large-scale HPC systems.

N.B: this duration does not truly reflect the total lifespan of these systems (HPC).



Experimentations :

Second experiments (Performance): The evolution of the Top500 performance metrics.

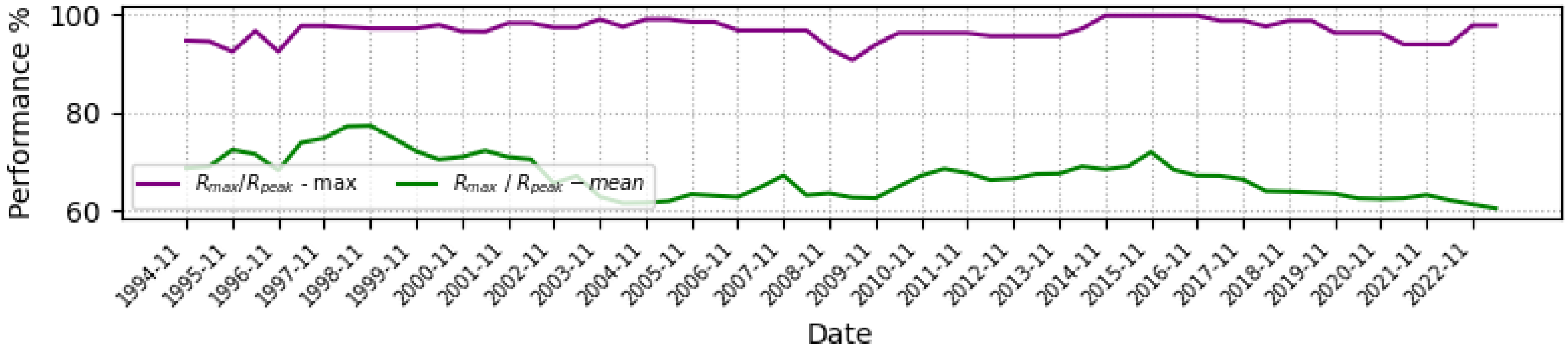


Observations:

- The graph depicts a consistent performance climb among listed supercomputers, especially those ranking highest.
- Both Rmax and Rpeak exhibit robust growth since the inception of the Top500.
- **However,** this exponential increase began to decelerate notably from the second half of 2013, indicating a clear inflection point.

Experimentations :

Second experiments (Performance): The evolution of the Top500 performance ratio.

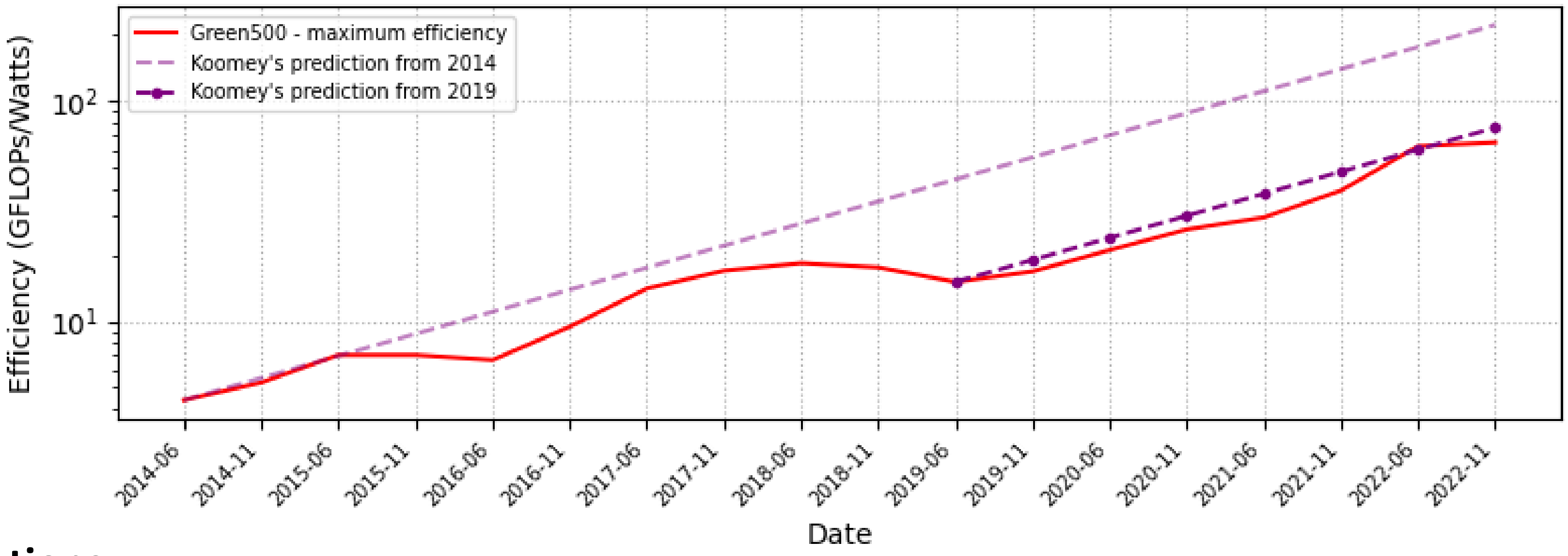


Observations:

- This graph shows a clear performance decrease in terms of percentage between the first Top500 supercomputer and the latter ones.
- Apart from constraints related to energy or budget, the decline in growth and performance efficiency can be attributed to the increasing architectural complexity of high-performance computing systems

Experimentations :

Third experiments (Energy & CO_2 footprint): The evolution of the Green500 efficiency metric.

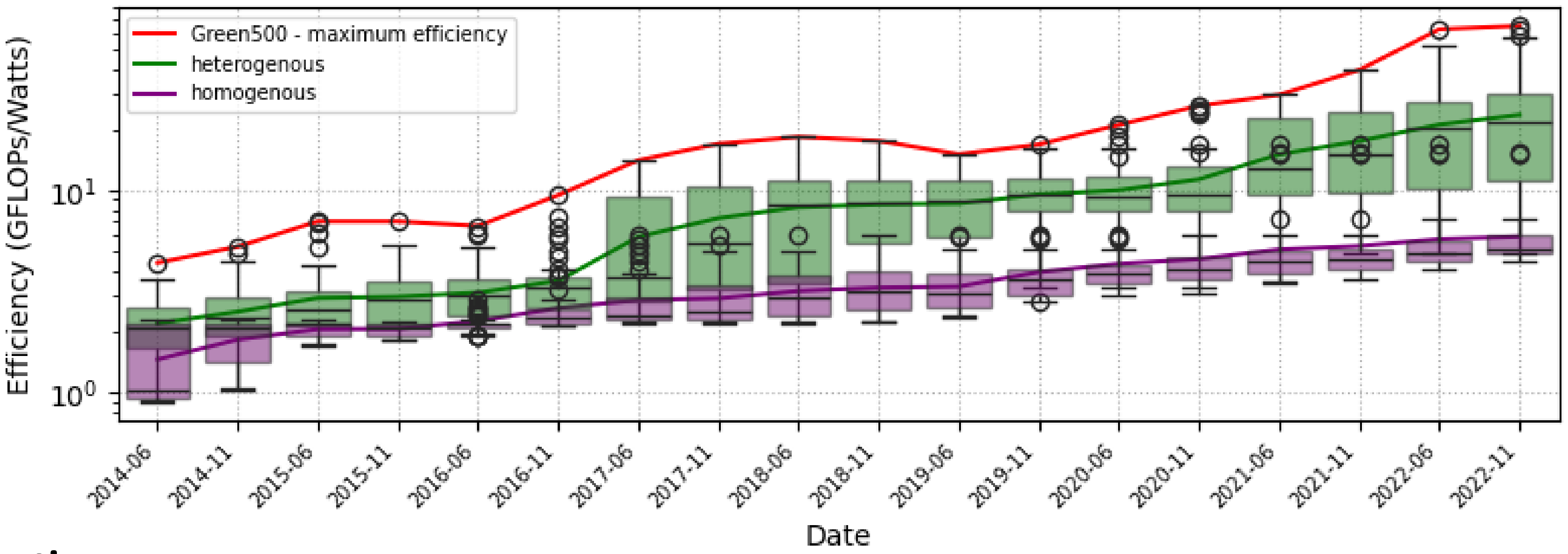


Observations:

- There's a noticeable increase in energy efficiency, albeit at a slower pace compared to performance improvements.
- Significant progress has been achieved, with the maximum value rising from 4.5 GFLOPS/Watt in late 2013 to 65.39 GFLOPS/Watt in 2023.
- Although this progress does not align with *Koomey's law*, recent years show signs of improvement, attributed to the growing interest in energy efficiency within the HPC community and the emergence of heterogeneous supercomputers, which appear to be more energy efficient.

Experimentations :

Third experiments (Energy & CO2 footprint): The evolution of the Green500 efficiency metric.



Observations:

- The figure clearly demonstrates that heterogeneous systems outperform homogeneous ones by an order of magnitude, with the gap between the two categories widening.
- This highlights the effectiveness of employing dedicated architectures for specific operations, particularly vectorial tasks, as a means to enhance efficiency.

Experimentations :

Third experiments (Energy & CO_2 footprint): The evolution of the increase ratio between graph500 & Green500.

			2010-2014	2014-2018	2018-2022
Top500	R_{max} (FLOPS)	top 1	13.20	4.24	7.68
		top 5	11.19	4.72	5.26
		top 15	8.50	4.53	4.75
		top 500	7.07	4.58	3.44
	HPCG (FLOPS)	top 1	-	-	5.47
		top 5	-	-	6.11
		top 15	-	-	5.03
		top 500	-	-	4.99
Graph500	BFS (TEPS)	top 1	-	1.32	3.29
		top 5	-	1.50	1.79
		top 15	-	1.57	1.76
	SSSP (TEPS)	top 1	-	-	11.81
		top 5	-	-	54.18
		top 15	-	-	186.98

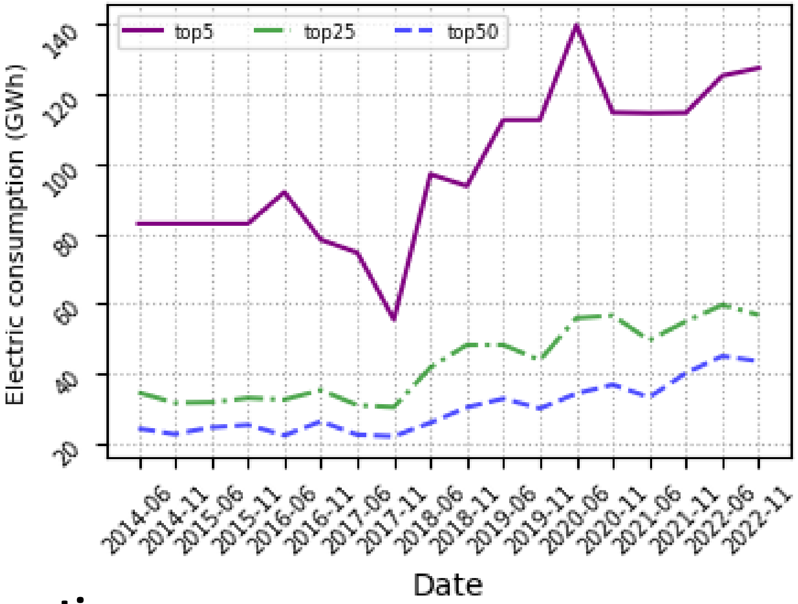
Observations:

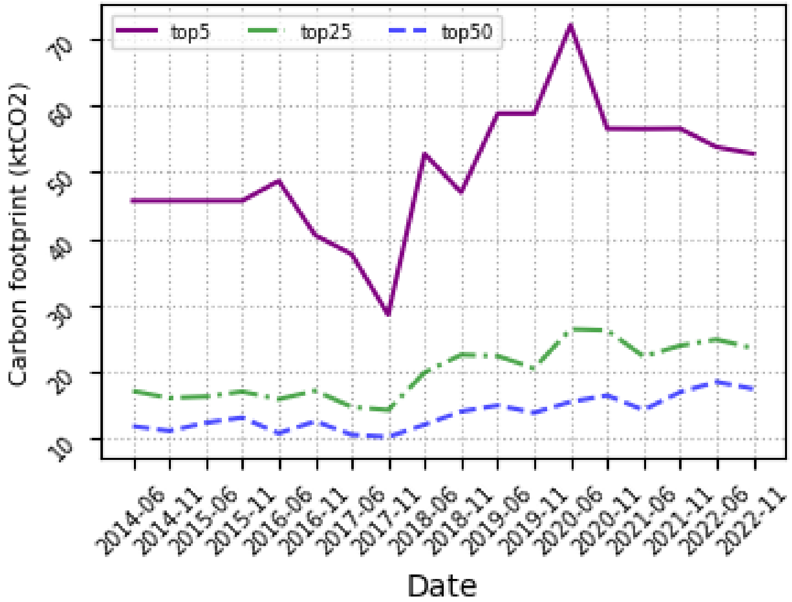
- A large difference can be observed between the different benchmarks : On the period 2018-2022 for example, the increase ratio for vary between 1.76 for the top 15 of BFS benchmark and 186.98 for the top 15 of SSSP benchmark.
- BFS and Rmax also behave differently: Top 1 Rmax performance increased around 7.5 times during the same period, while BFS increased around 3.3 times only.
- The focus on period 2014-2022 shows an acceleration on top systems after 2018: This highlights the impact of accelerators arrival in HPC.

2. Conclusion - Projection models

Conclusion - projection models :

Footprint efficiency.



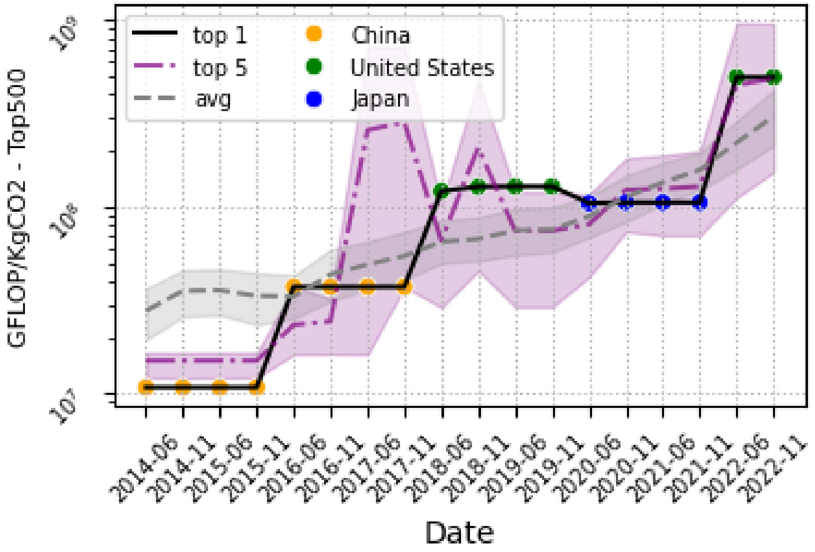


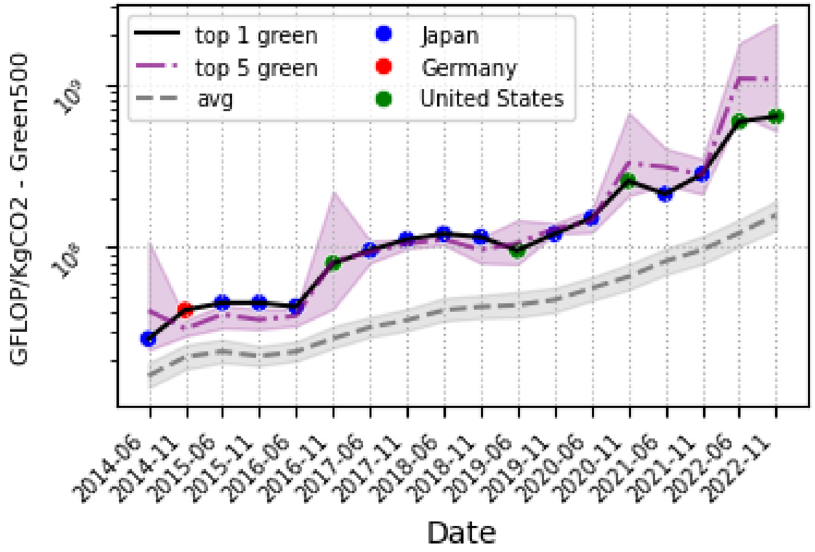
Observations:

- The top 5 systems significantly differ from the rest.
- Consumption remained nearly constant before 2018 but doubled by 2022.
- Despite the increase, the footprint growth was limited due to reduced electricity footprint in many countries during the same period.
- Both curves closely mirror each other, except for the most recent period, influenced by changes in the energy mix.

Conclusion - projection models :

Footprint efficiency.





Observations:

- There's a consistent rise in the mean efficiency since 2014, with an average doubling period of **2.83** years.
- The initial Top500 systems align closely with the mean efficiency, while the efficiency of Green500 systems demonstrates a more rapid increase.

Conclusion - projection models :

Scenarios.

- The European Union Green Deal aims to achieve a 55% reduction in greenhouse gas (GHG) emissions by 2030, making this the targeted horizon for our analysis.
- Extrapolating from carbon efficiency data presented in previous figures: the number of GFLOPs per kilogram of CO_2 is **projected to reach $1.64 * 10^9$ in 2030 compared to $3.05 * 10^8$ in 2022**.
- This represents a **537%** increase between 2022 and 2030, with a slope of **11.87% per issue and 24.99% per year relative to the 2022 value**.
- Although it is a rough estimate, the analysis indicates considerable stability in this improvement over an extended period. However, the disparity between this increase and the expected reduction in emissions is substantial.

Thank you!

For more information, please contact:

Abdessalam.benhari@eviden.com

Article available here :

https://hal.science/hal-04519645