Cloud Computing -- The Carbon Footprint of Datacenters

Thomas Ropars

Email: thomas.ropars@univ-grenoble-alpes.fr

Website: tropars.github.io

Warning

Studying the carbon footprint of Datacenters is complex:

- A rather recent research topic
- It is difficult to collect accurate data
- The data presented in these slides might not be 100% correct

References

The following references were used to prepare these slides:

- Reports from the Shift project: https://theshiftproject.org/en/lean-ict-2/
- L. A. Barroso, U. Hölzle, and P. Ranganathan. The datacenter as a computer: Designing warehouse-scale machines. Synthesis Lectures on Computer Architecture, 13(3):i–189, 2018.
- Other papers are cited directly in the slides

Carbon footprint

Definition

We are trying to evaluate:

- The quantity of Greenhouse Gases (GHGs) emitted into the atmosphere by some activities
  - Direct emissions (made directly during the progress of a process)
  - Embodied emissions (manufacturing, transportation, etc.)

We express it in terms of CO2 equivalent mass (CO2-e):

- There is more than CO2 in Greenhouse Gases (see the 6 Kyoto gasses)
- CO2-e is used to express the global warming potential of all these gasses
  - A common unit that allows making comparisons
  - The most important gas: 70% of the emissions in France

Pandey, Divya, Madhoolika Agrawal, and Jai Shanker Pandey. "Carbon footprint: current methods of estimation." Environmental monitoring and assessment 178 (2011): 135-160.

Impact of datacenters on the environment

Not only about the carbon footprint

- Use of different metals
    - Rare earth elements (soil pollution)
- Soil artificialization
- Water consumption

IT Carbon footprint

Energy consumption of the IT domain

A constant growth and an increased percentage in the total energy consumption

- 6.2% of energy increase per year between 2015 and 2019
    - Doubling in 12 years
- More that 5% of the energy consumption due to IT

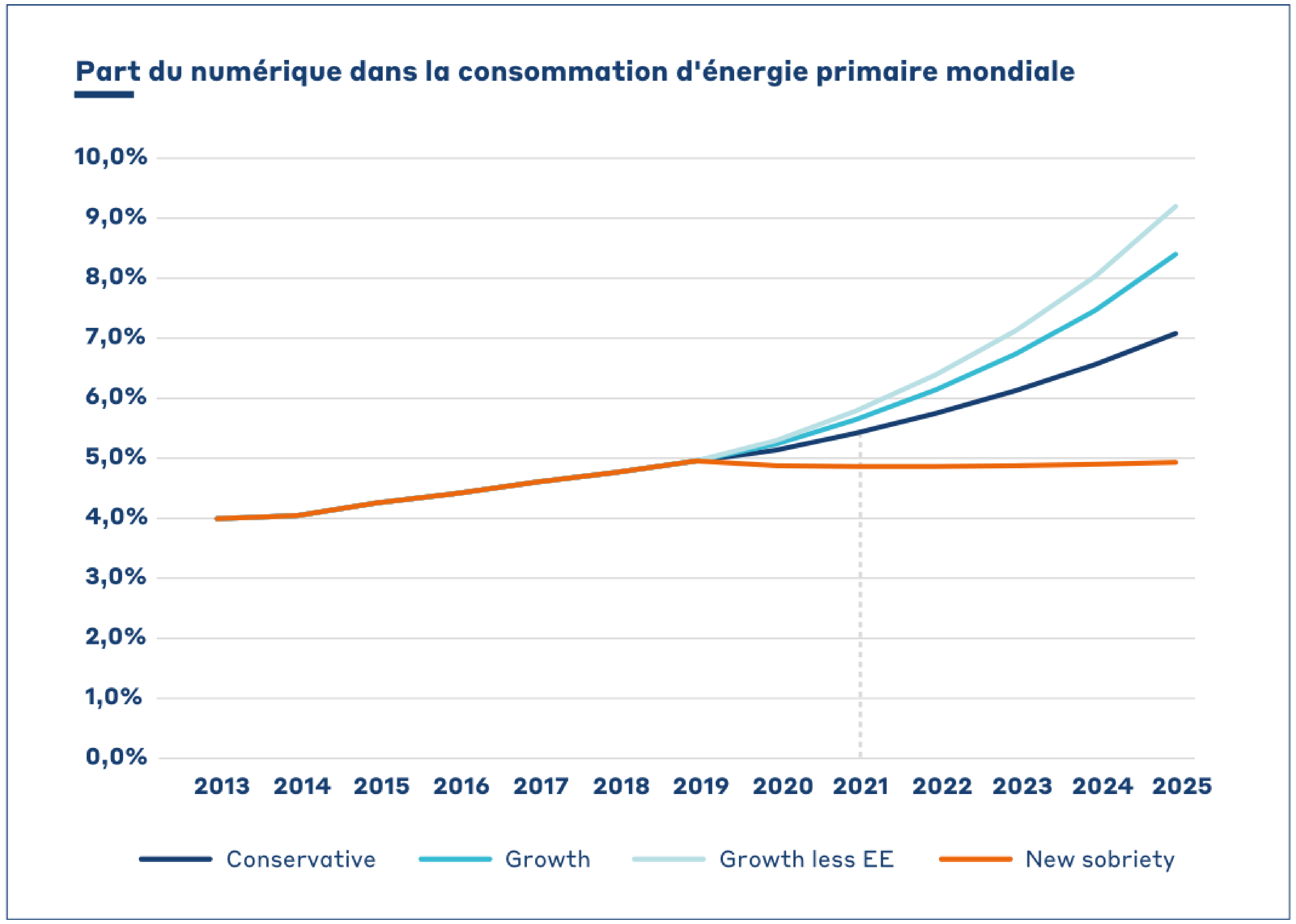


Figure 1 : Évolution 2013-2025 de la part du numérique dans la consommation d'énergie primaire mondiale
(The Shift Project - Forecast Model 2021)

Reasons for this huge growth

Huge traffic increase

- Mobile traffic has increased by 20% in 2025

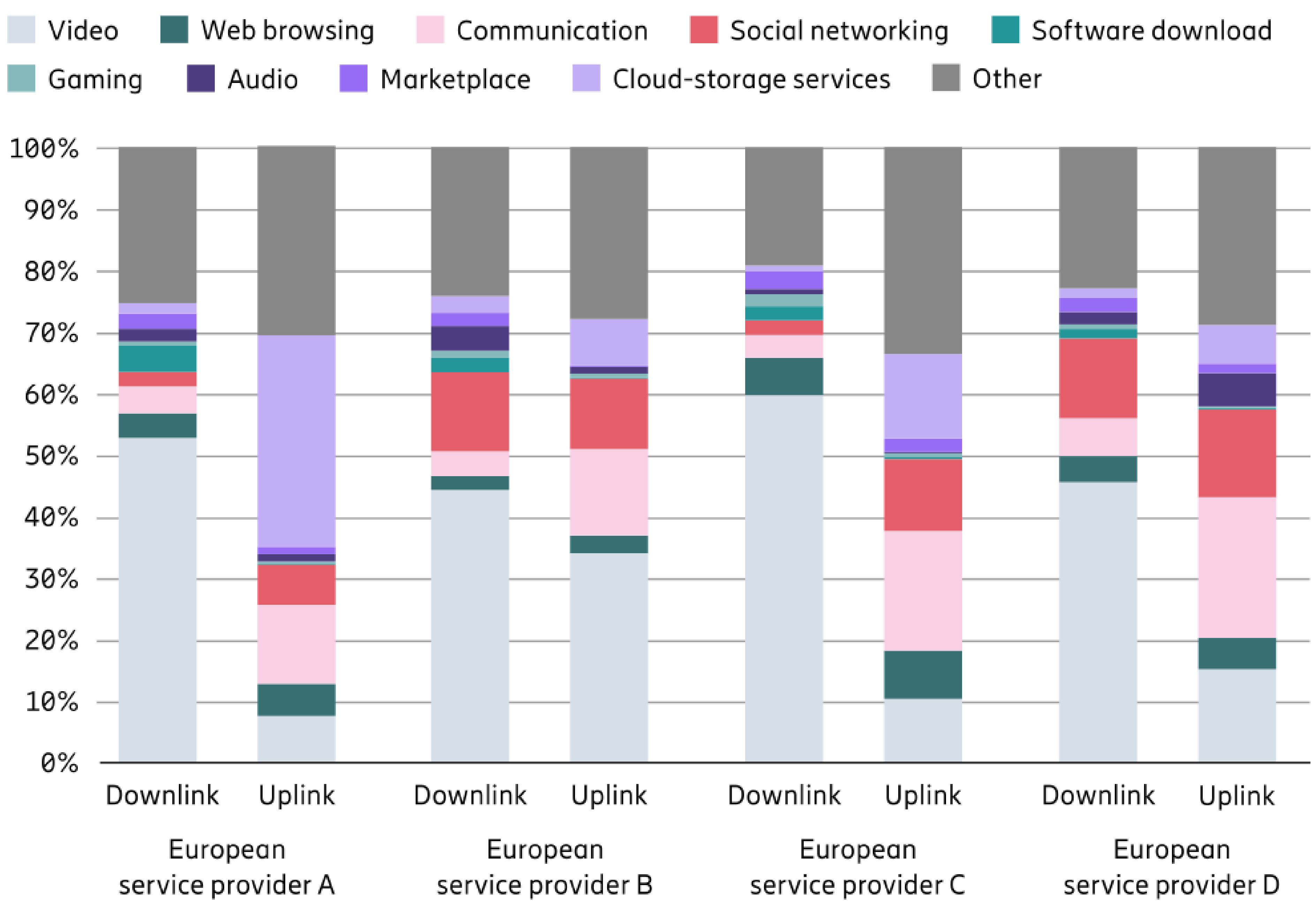
Source: Ericsson Mobility Report, November 2025

Reasons for this huge growth

Huge traffic increase

- The main data are videos (50% of the traffic)

Figure 21: Share of traffic volume in downlink and uplink per application category

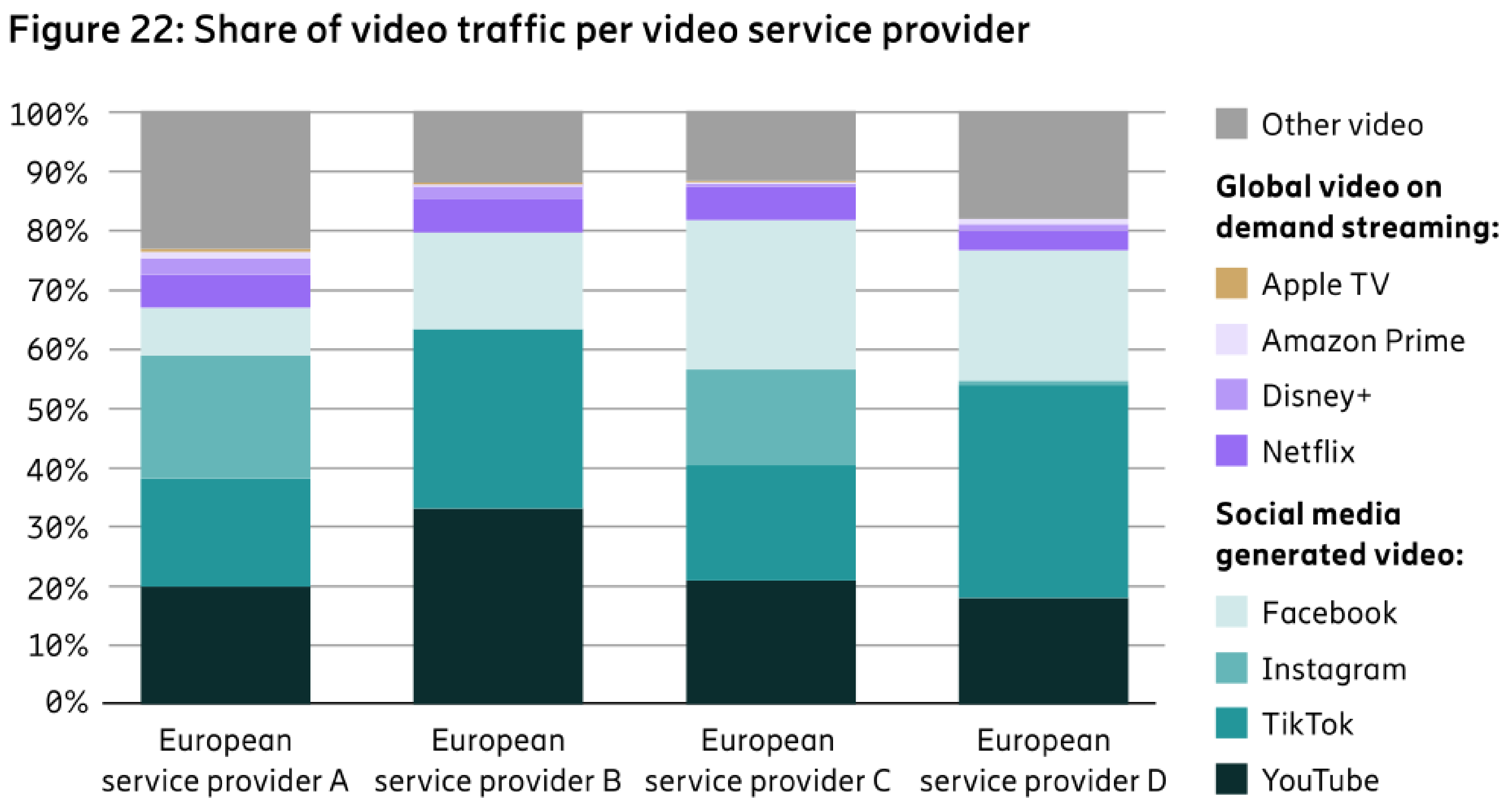


Source: Ericsson Mobility Report, November 2025

Reasons for this huge growth

Huge traffic increase

- Short videos from the social media are the majority (at least 70% of the traffic)



Source: Ericsson Mobility Report, November 2025

Huge traffic increase

More information

- For some providers: 80% of the traffic comes from the GAFAMs

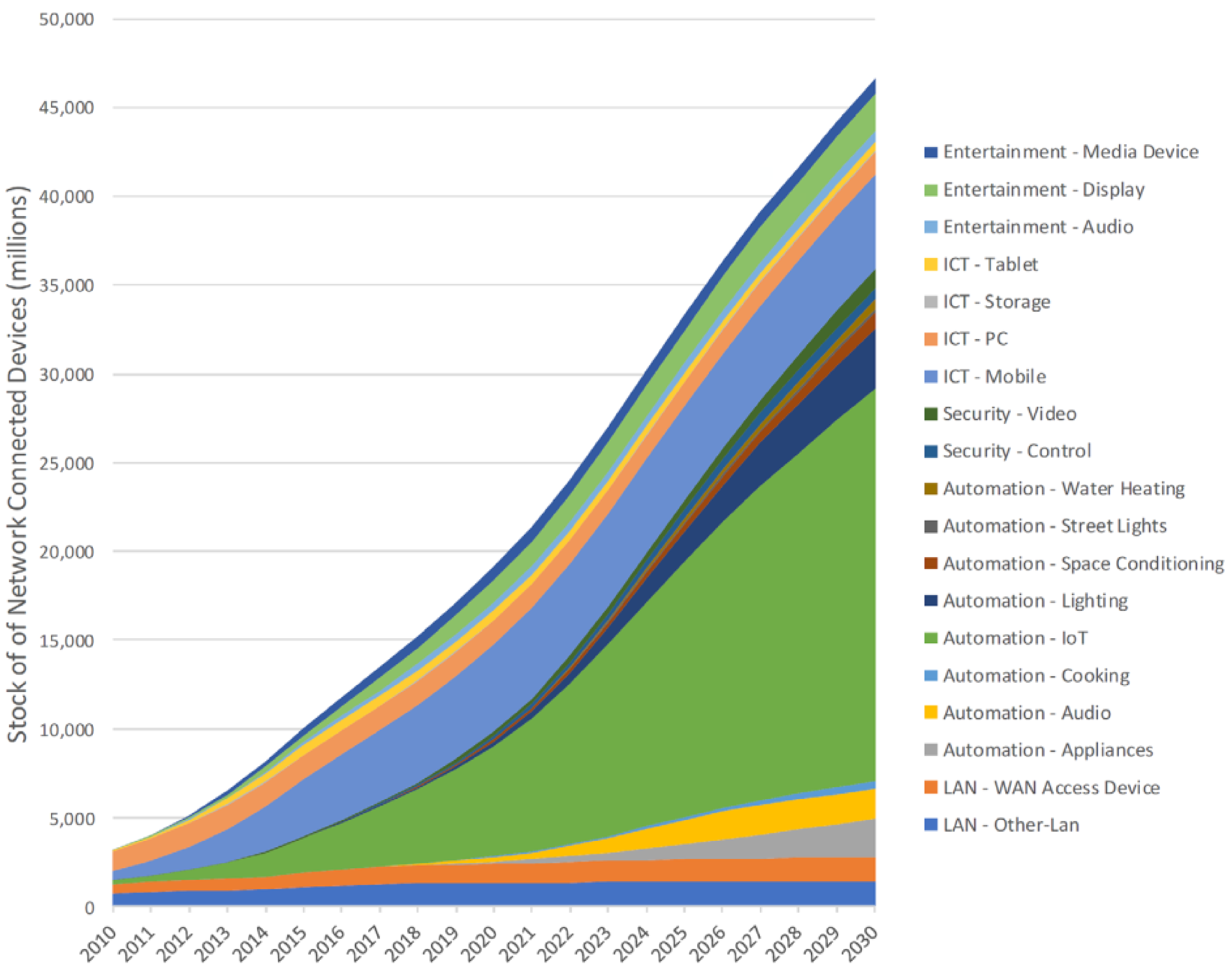
Impact on the Cloud

- Increase of total amount of data stored by data centers: +40% per year
- Data stored in datacenters represent 20% of the total data

Source: The Shift project -- Environmental impacts of digital technology : 5-year trends and 5G governance (2021)

Reasons for this huge growth (2)

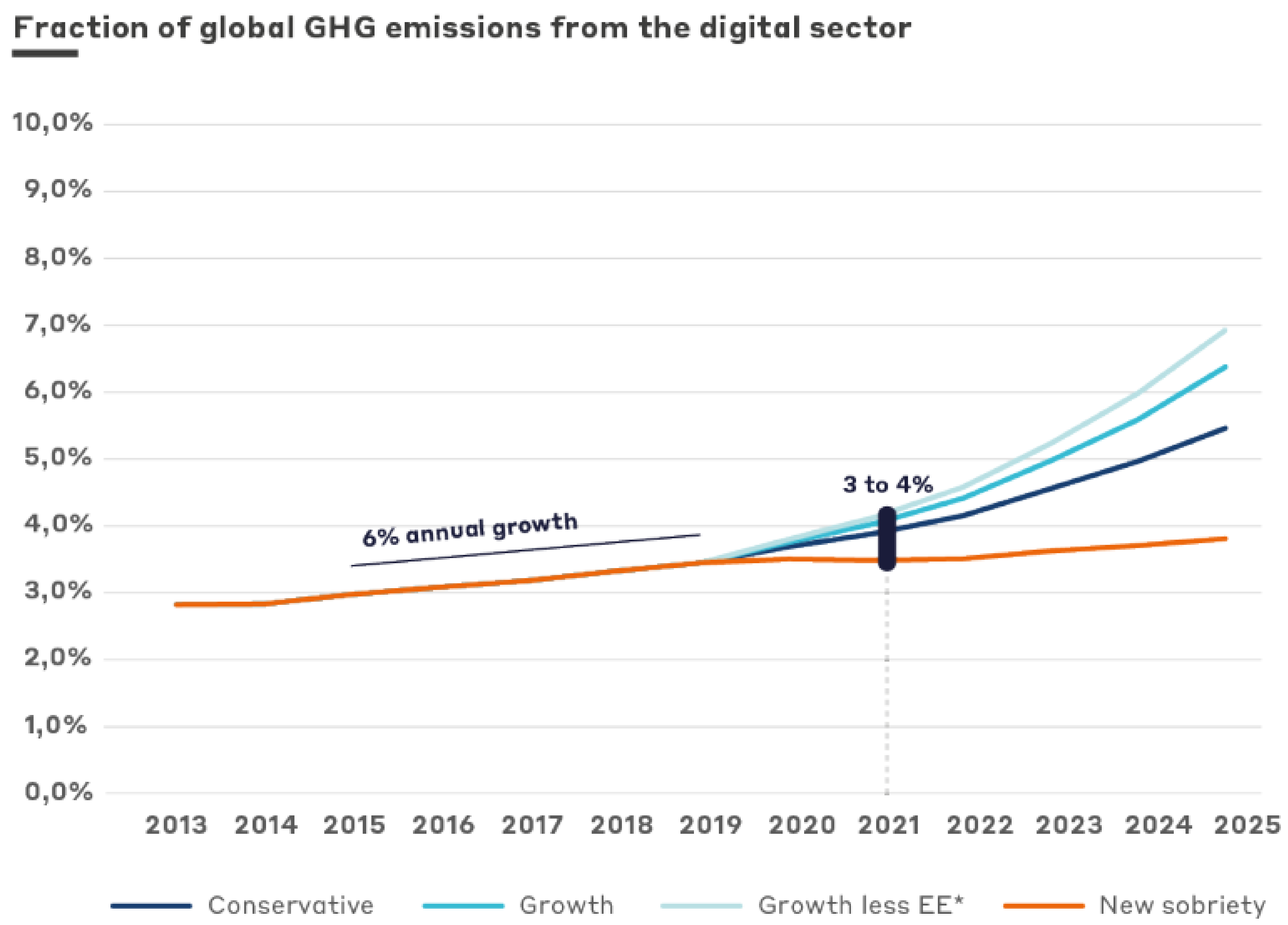
Huge increase in the number of connected devices



Source: The Shift project -- Environmental impacts of digital technology : 5-year trends and 5G governance (2021)

Carbon footprint of the IT domain

- Between 3% and 4%
    - Note that energy is often used as a proxy for evaluating the carbon footprint

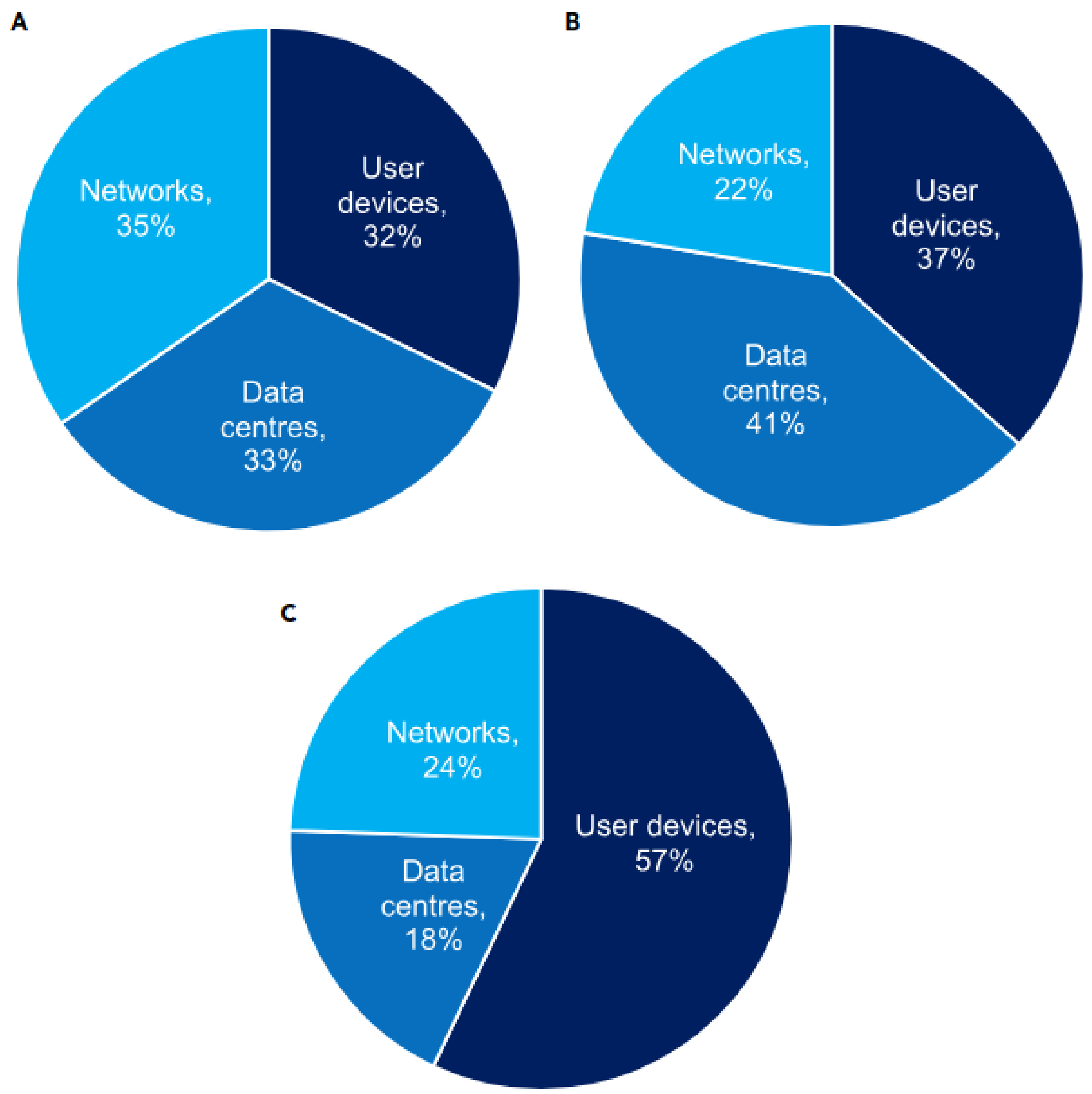


Source: The Shift project -- Environmental impacts of digital technology : 5-year trends and 5G governance (2021)

Carbon footprint of datacenters

The impact of datacenters cannot be ignored

- Breakdown of the contributors to IT carbon footprint (excluding TV)
  - Summary of 3 studies from 2015 to 2020
  - A complex problem: Numbers vary a lot between the studies



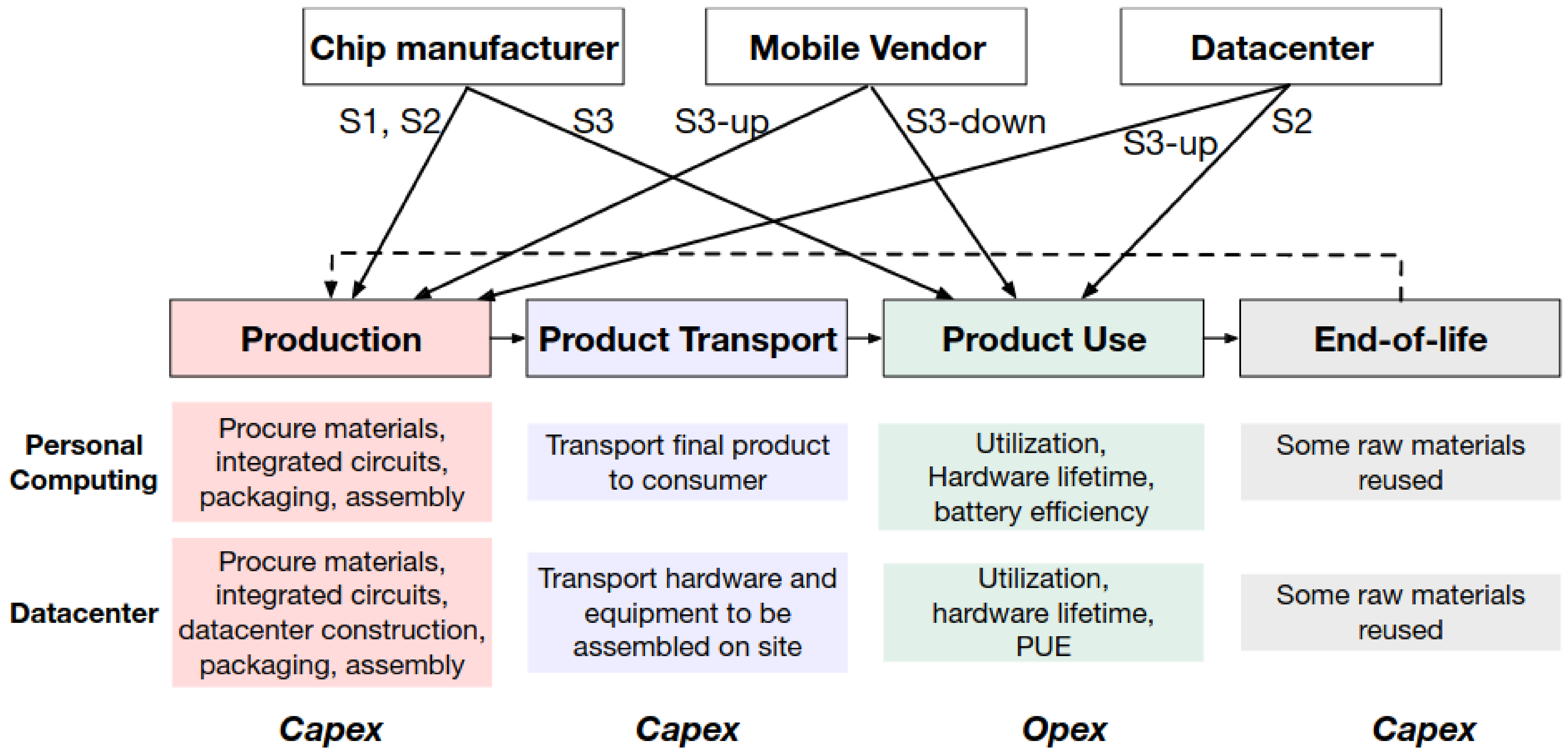
- Datacenters: 1.8-2.8% of worldwide GHG emissions

Source: Freitag, Charlotte, et al. "The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations." Patterns 2.9 (2021).

Which parts of the datacenter emit carbon?

Embodied vs operational emissions

- Embodied emissions = Emissions for the production of the hardware (Capex)
- Operational emissions = Emissions for the operation (use) of the hardware (Opex)
- They need to be evaluated through life-cycle analyses



Source: Gupta, Udit, et al. "Chasing carbon: The elusive environmental footprint of computing." HPCA 2021.

What is the main contributor to carbon emissions?

- Traditionally for a server, it was considered that:
  - 50% of the footprint is the embodied
  - 50% of the footprint is the operational

Significant improvements have been made in recent years

What does is change?

Current trends in datacenters

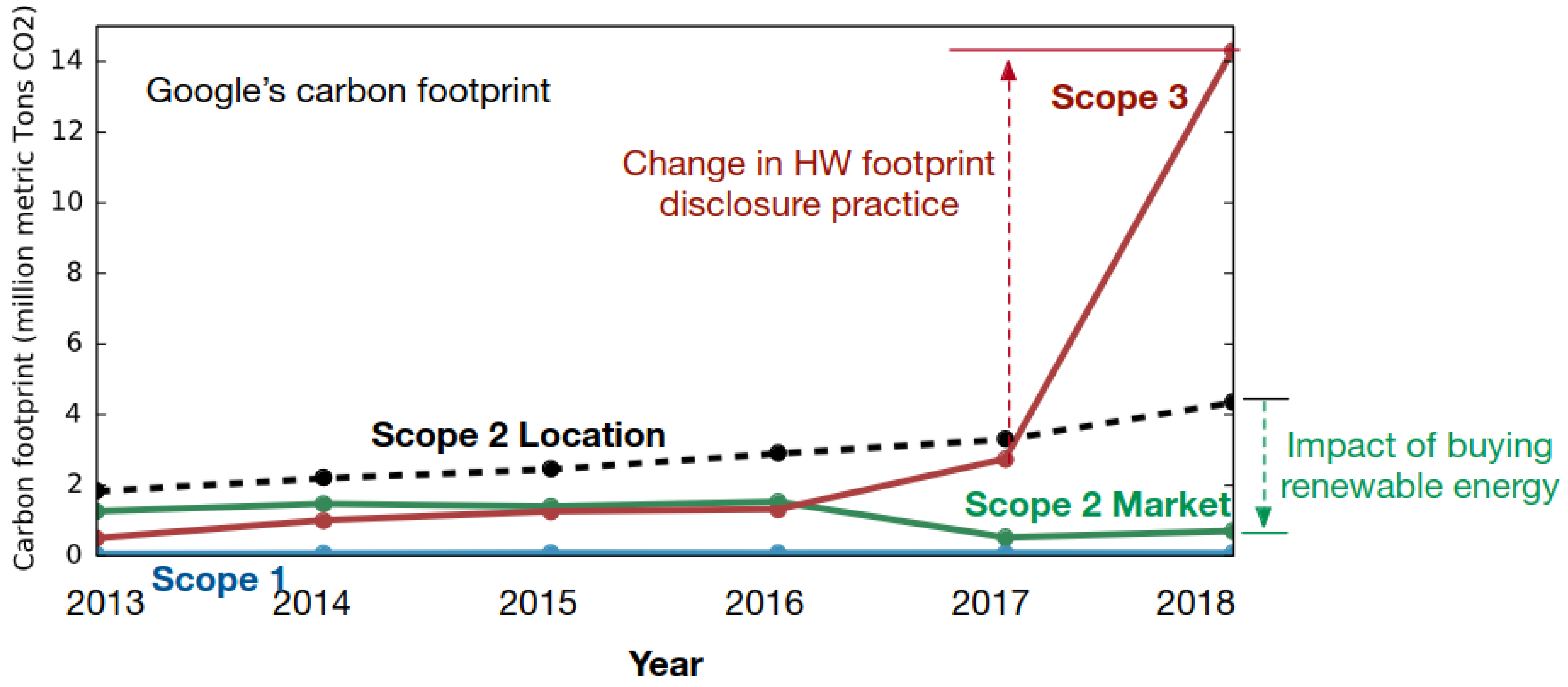
- Use of greener energy
- Use of more energy-efficient hardware
- Improve the energy efficiency of the infrastructure

Study of the operational energy of datacenters

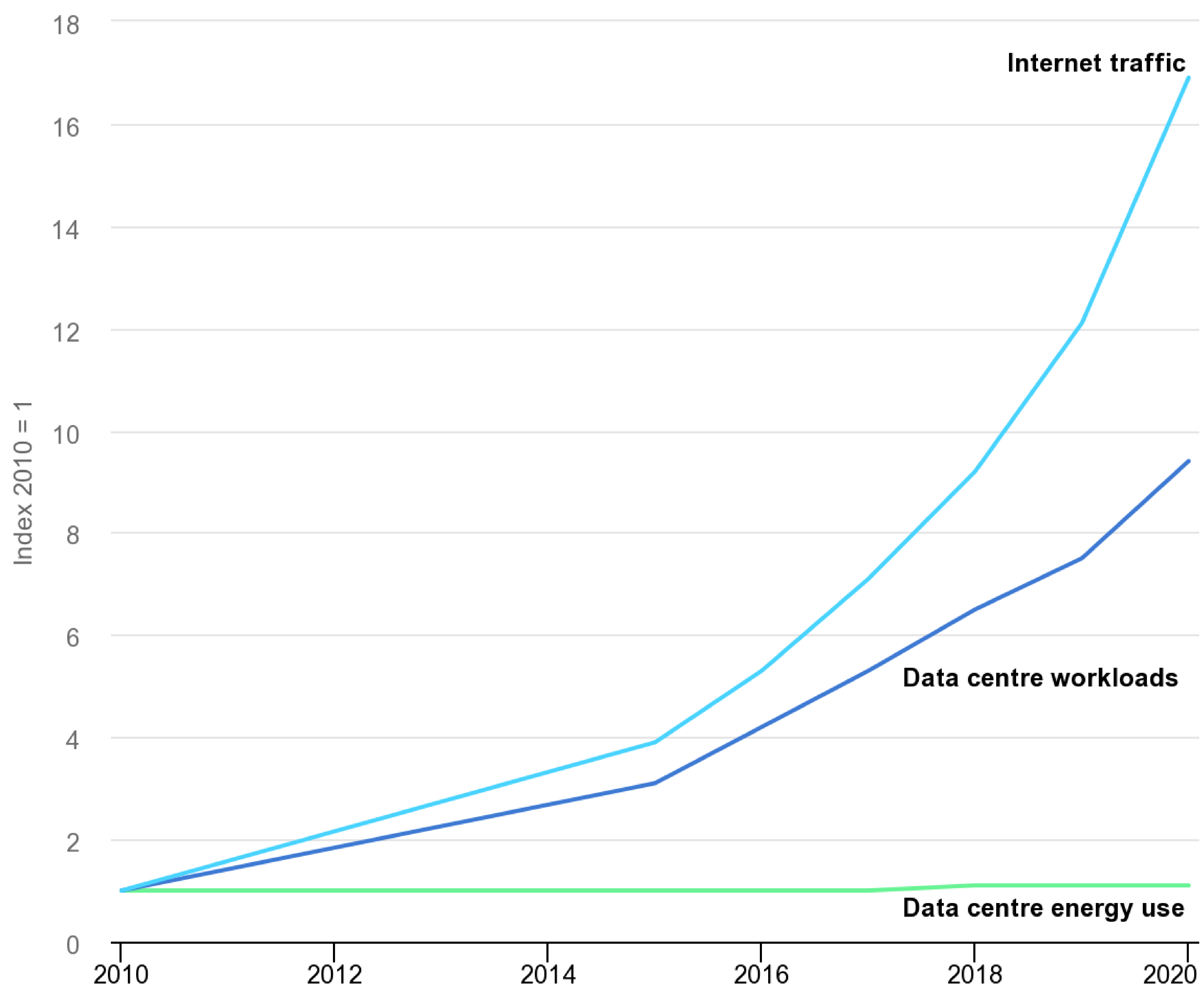
Using green energy

Easy way to improve operational carbon footprint

- Change the energy source -- renewable energy
  - Example of a google data center (scope 3 = manufacturing)



Improving the Energy consumption of data centers



Source: https://www.iea.org/data-and-statistics/charts/global-trends-in-internet-traffic-data-centres-workloads-and-data-centre-energy-use-2010-2020

Energy consumption of data centers

Good news or bad news

Energy consumption of data centers

Good news or bad news

- Positive side
  - Datacenters energy efficiency has improved dramatically
- Negative side
  - The increase in workload is so big that it nullifies efficiency improvements

The goal should be to decrease the absolute energy consumption

The rebound effect

Problem with the optimization of energy

- It almost always leads to an increase in the usage
  - If I have a more efficient car, I use it more
- Improving the energy efficiency implies that:
  - Reduction of the costs of the goods
  - Possibility to improve the service
- Does it apply to all domains of CS? (see: Woodruff, Jackson, et al. "When Does Saving Power Save the Planet?." Proceedings of the 2nd Workshop on Sustainable Computer Systems. 2023.)

Jevons Paradox

The rebound effect

Problem with the optimization of energy

- It almost always leads to an increase in the usage
    - If I have a more efficient car, I use it more
- Improving the energy efficiency implies that:
    - Reduction of the costs of the goods
    - Possibility to improve the service
- Does it apply to all domains of CS? (see: Woodruff, Jackson, et al. "When Does Saving Power Save the Planet?." Proceedings of the 2nd Workshop on Sustainable Computer Systems. 2023.)

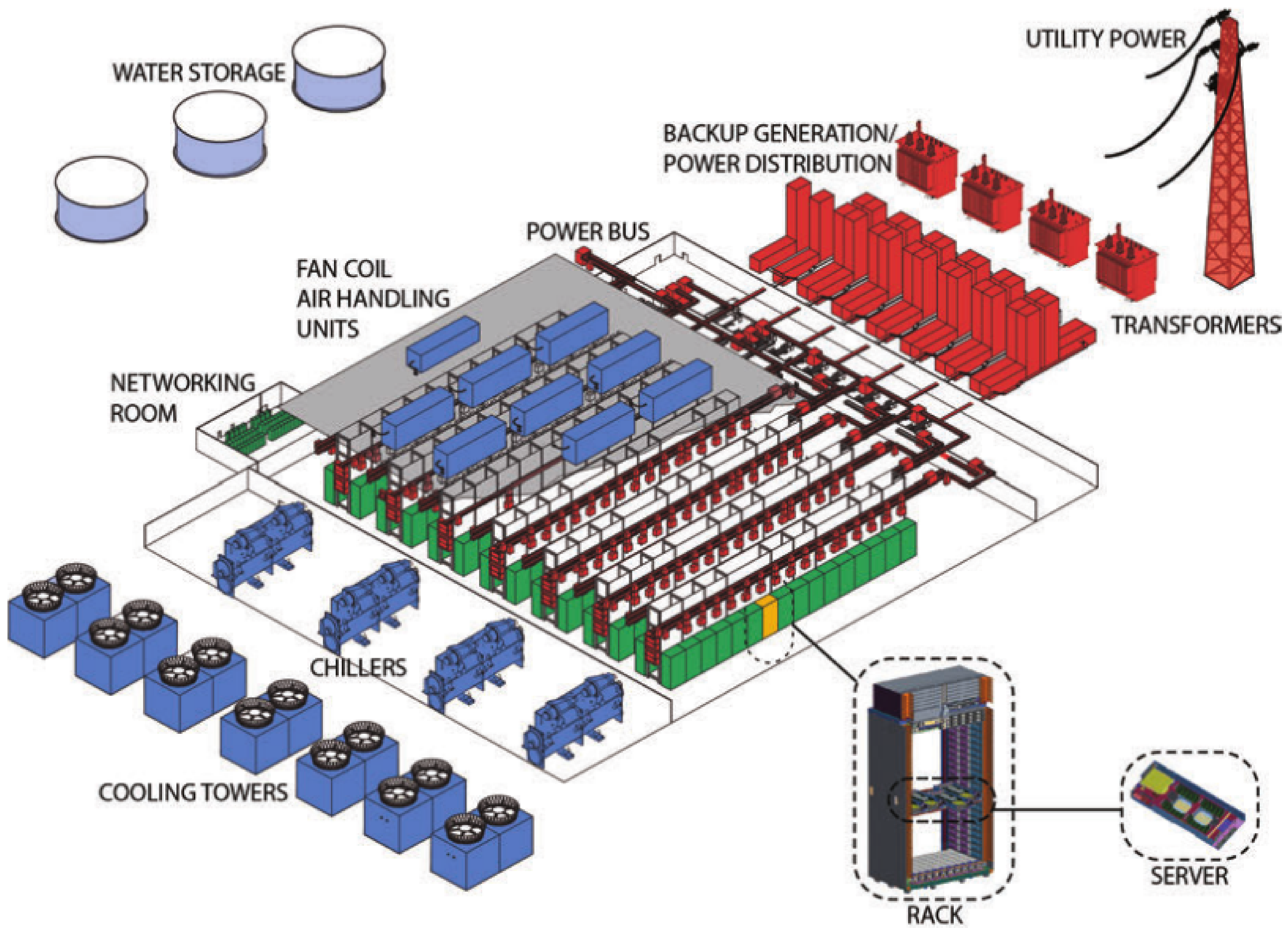
Jevons Paradox

- First discussed the the 19's century about coal

The rebound effect might be so important that it leads to an increase in energy usage

Optimizing the energy consumption of data centers

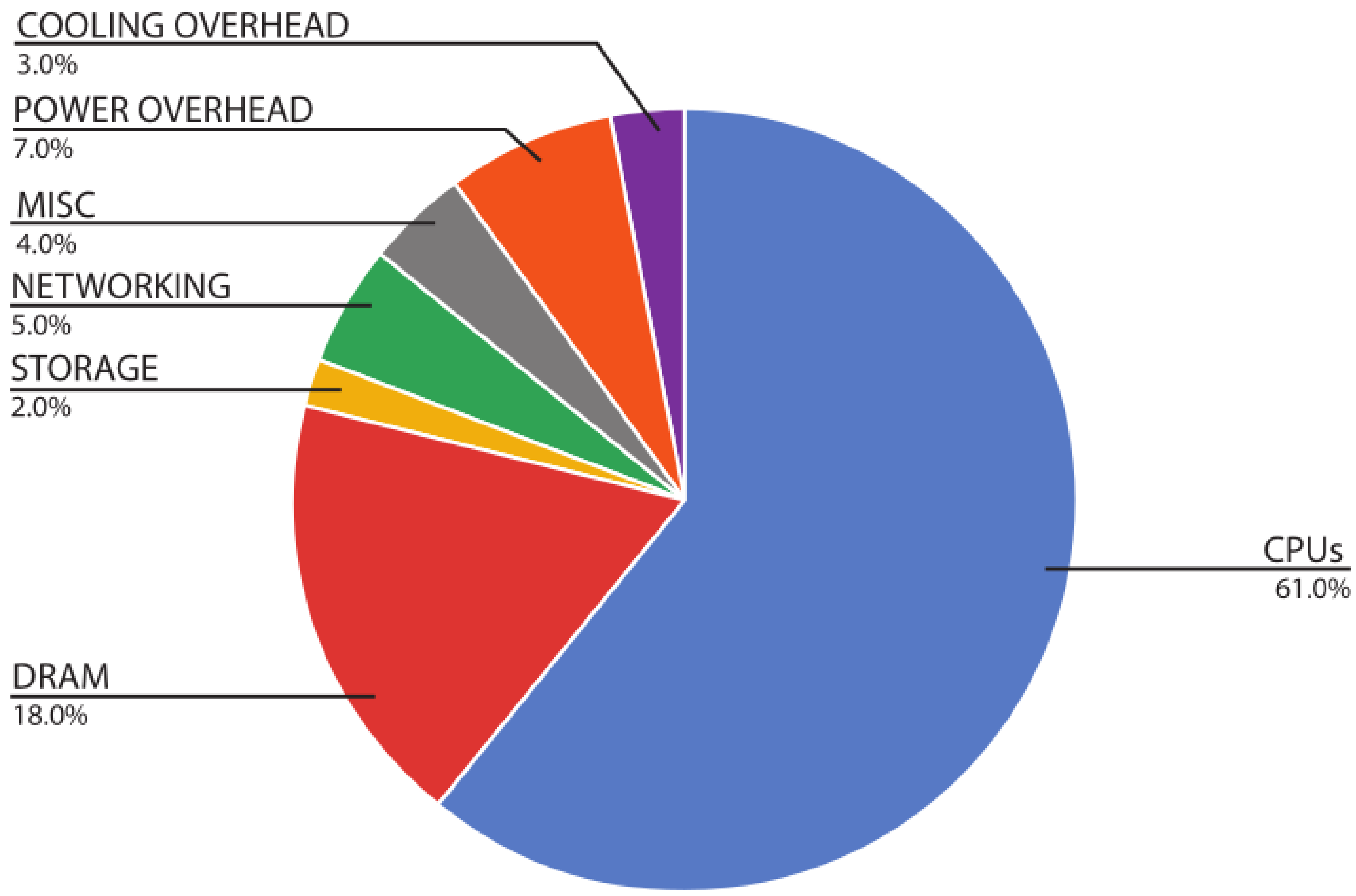
What consumes energy?



Source: The datacenter as a computer

What consumes energy?

Peak power usage for a 2-socket server at 80% of max utilization (2017)



Power Usage Efficiency (PUE)

We want to optimize energy efficiency:

$$\text{Energy Efficiency} = \frac{\text{Work}}{\text{Used energy}}$$

PUE

- Power Usage Efficiency captures the quality of the datacenter building
    - How much energy is used for something else than *computing*?
- SPUE is the PUE at the level of a server

$$\text{Energy Efficiency} = \frac{1}{\text{PUE}} \times \frac{1}{\text{SPUE}} \times \frac{\text{Work}}{\text{Energy used by electronic components}}$$

Power Usage Efficiency

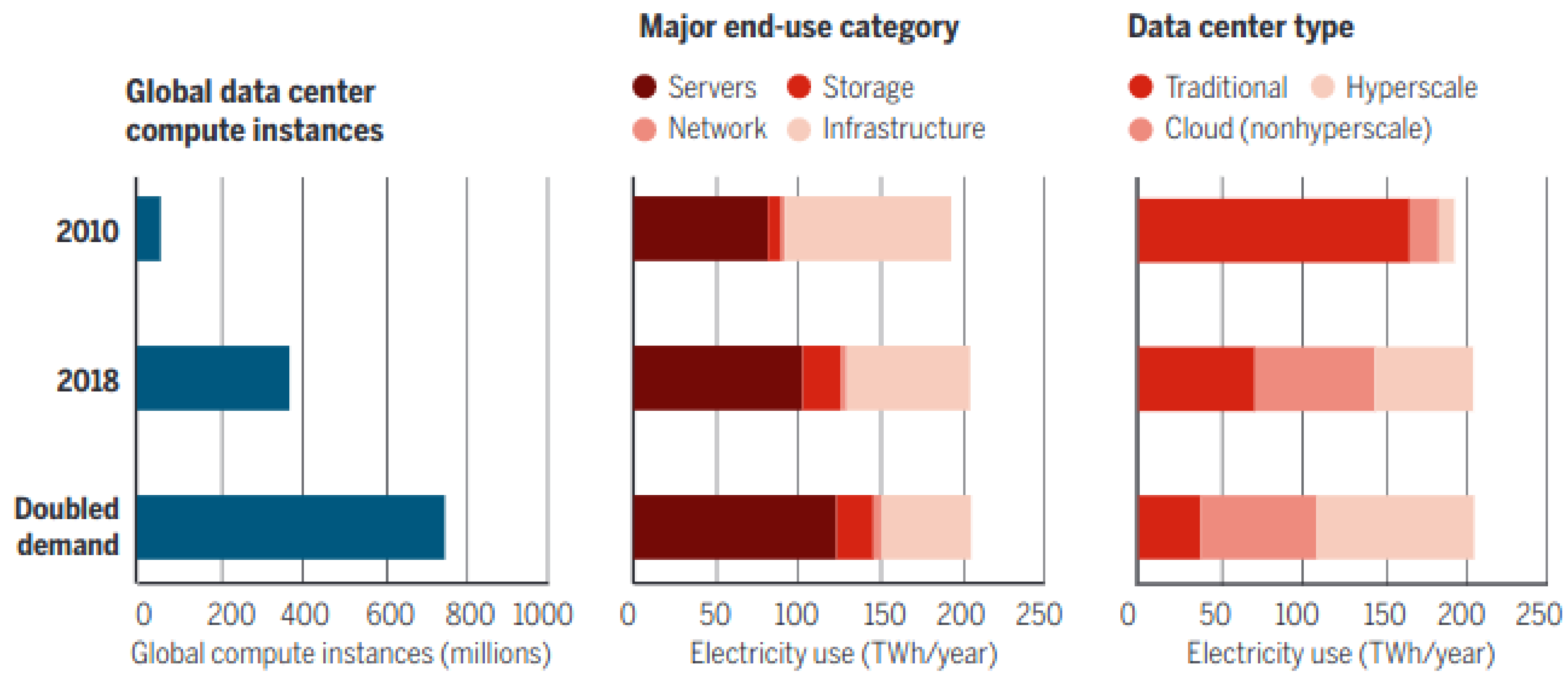
Evolution of the PUE

- Before 2006, the PUE of most datacenters was above 3

- Today:
    - Traditional DC have a PUE between 1.6 and 2.5
    - Hyperscalers have a PUE below 1.2
        - Google DC average PUE is below 1.1

About hyperscalers

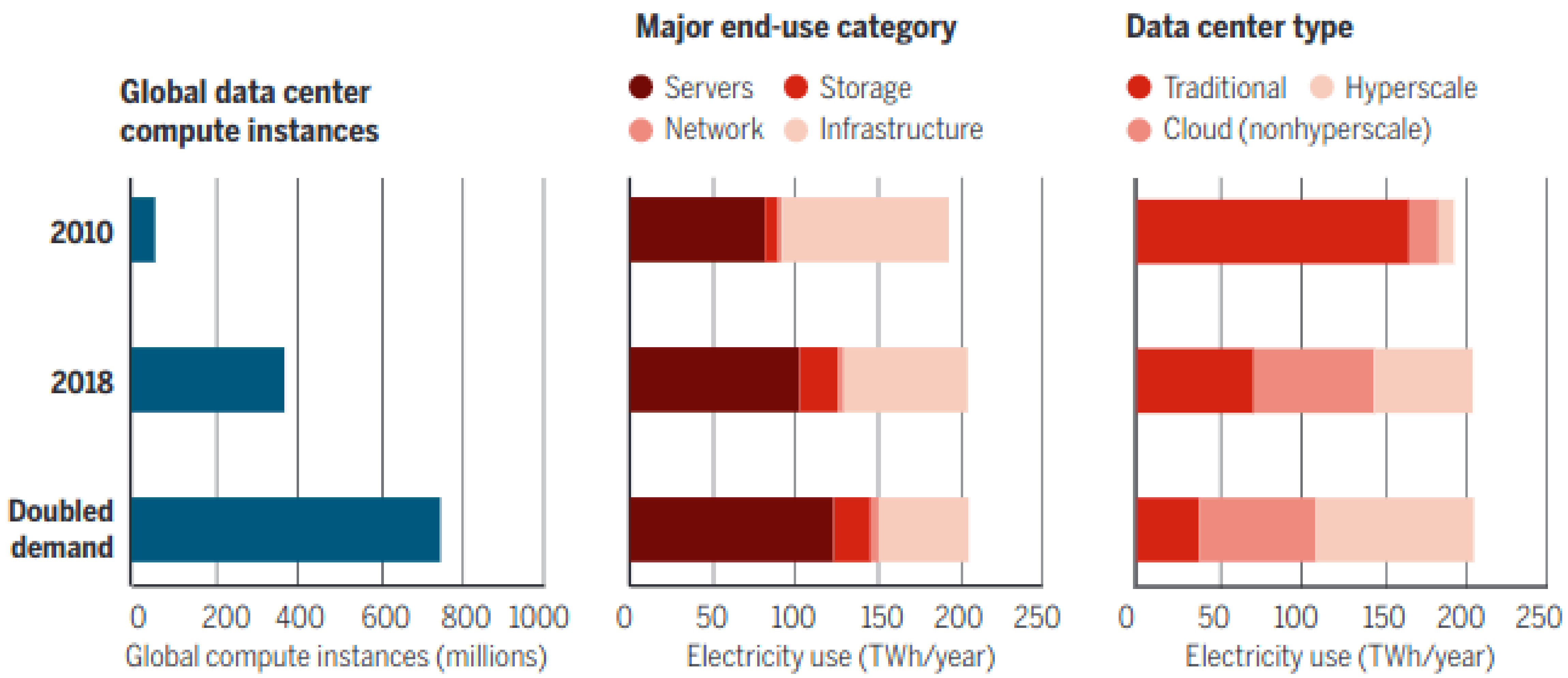
Trend: More and more applications are hosted by hyperscalers



Good news?

About hyperscalers

Trend: More and more applications are hosted by hyperscalers



Good news?

- Yes because better PUE
- No because more data movements

It is a complex question

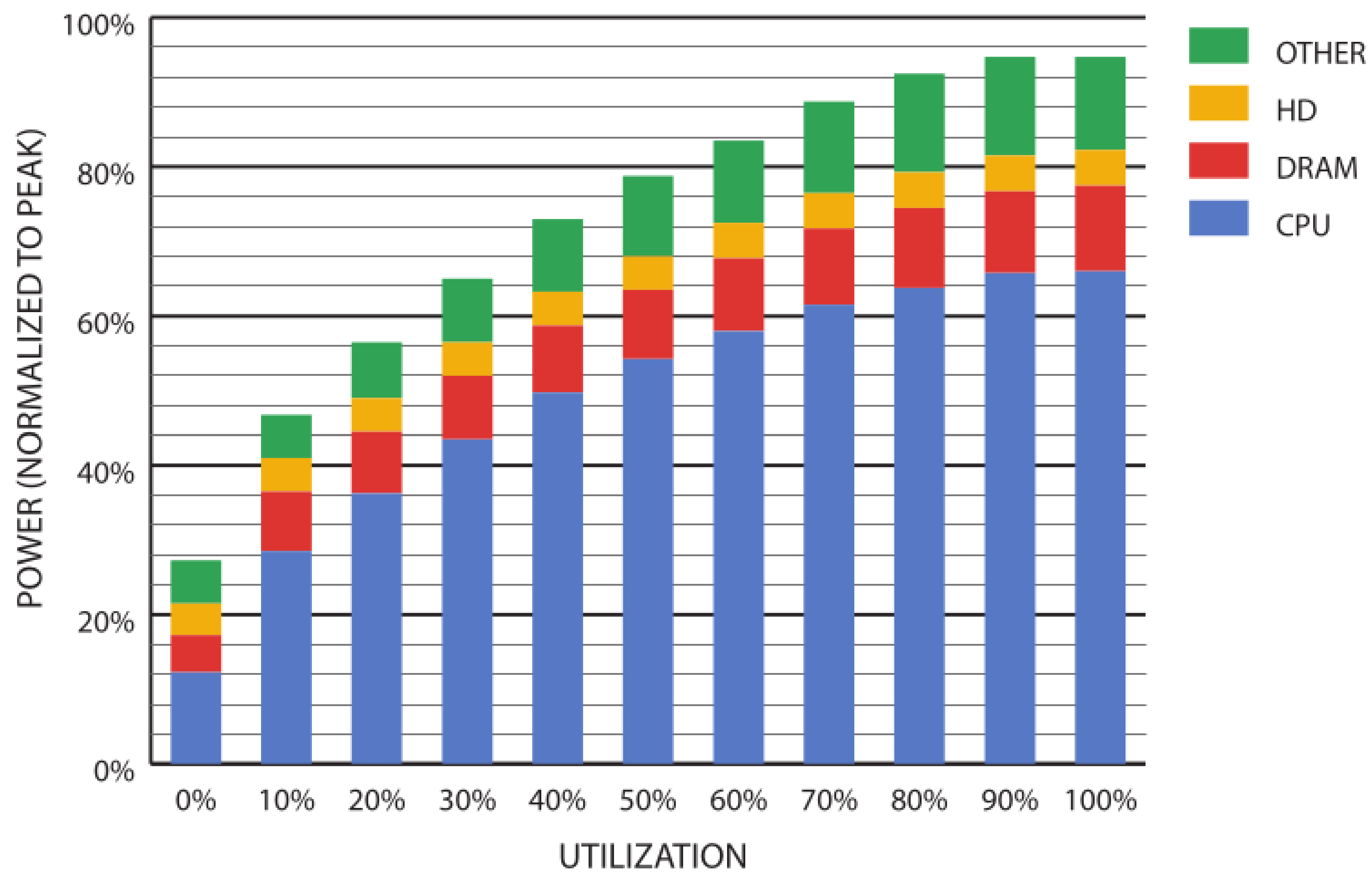
Energy proportionality

Definitions

- **An energy proportional system is one in which the energy consumed by the system is directly proportional to the activity**
- Energy consumed by a device = static energy + dynamic energy
    - Static energy: Energy consumed when the activity is null
    - Dynamic energy: Energy variations depending on the activity

Energy proportionality

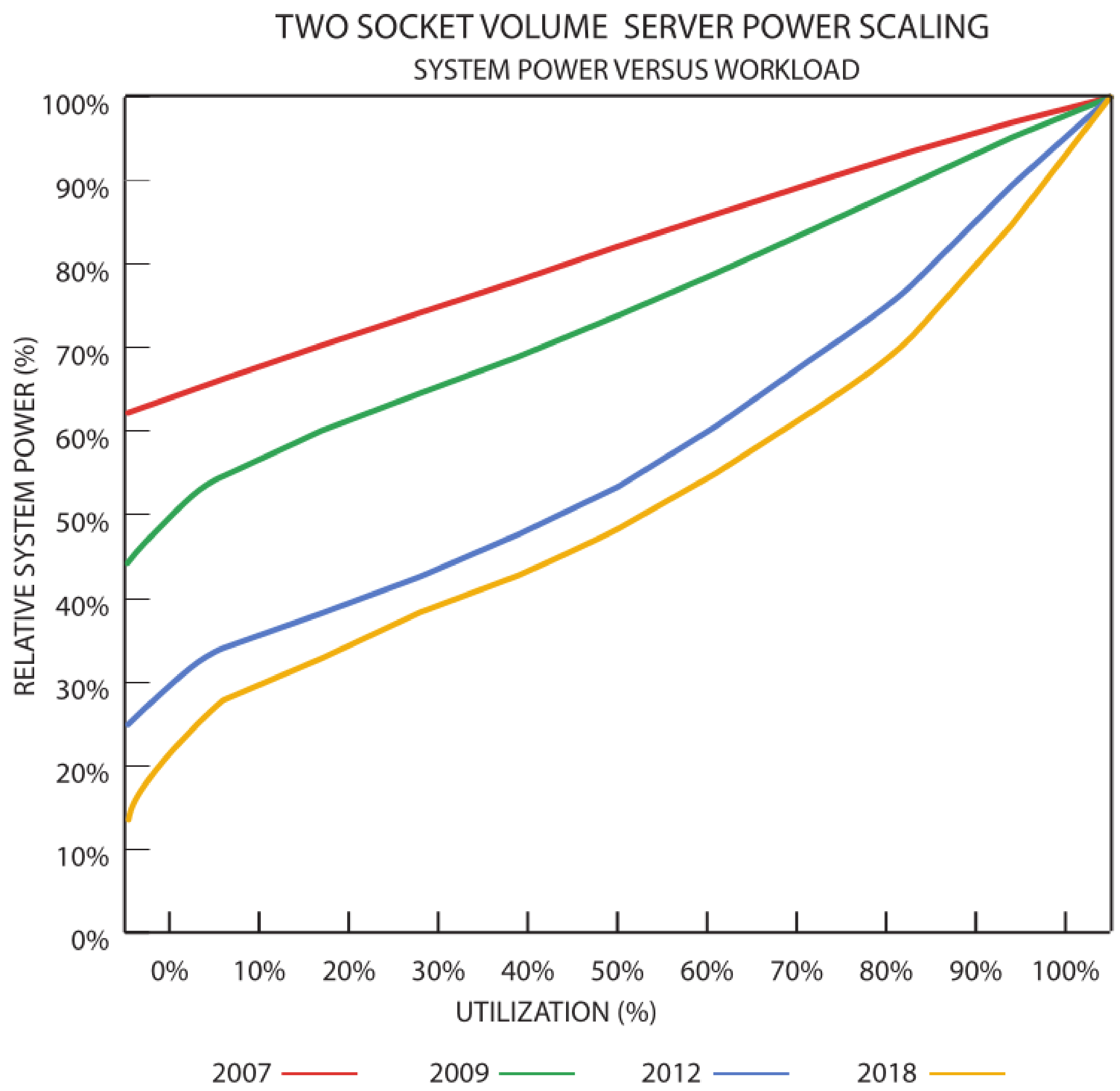
Energy proportionality of a x86 server



- The CPU is not the only thing to focus on
- The network is an example of device with a high static energy consumption

Energy proportionality

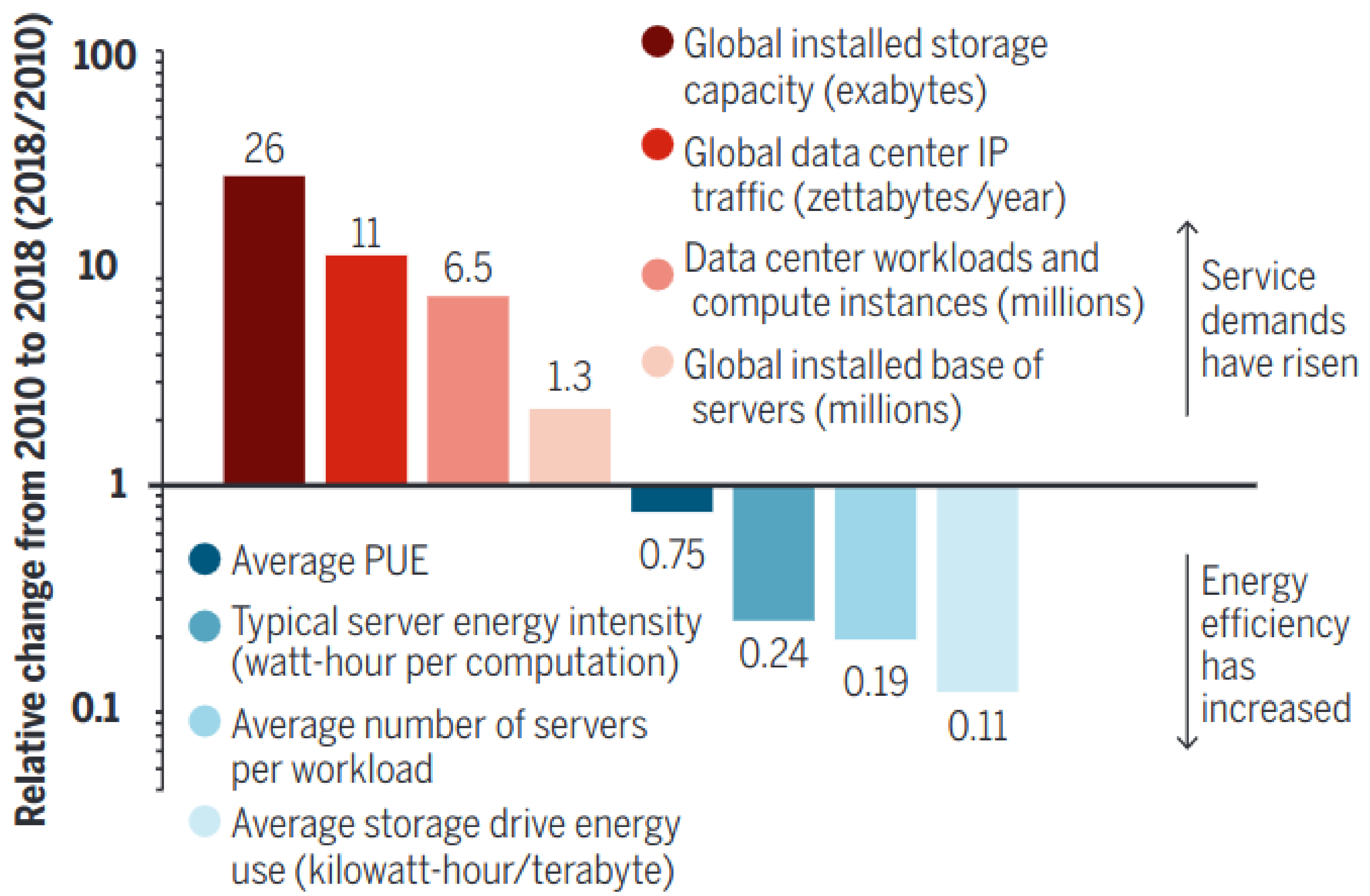
At the level of the CPU



Huge improvements have been made:

- Dynamic Voltage and Frequency Scaling (DVFS)
- Sleep states

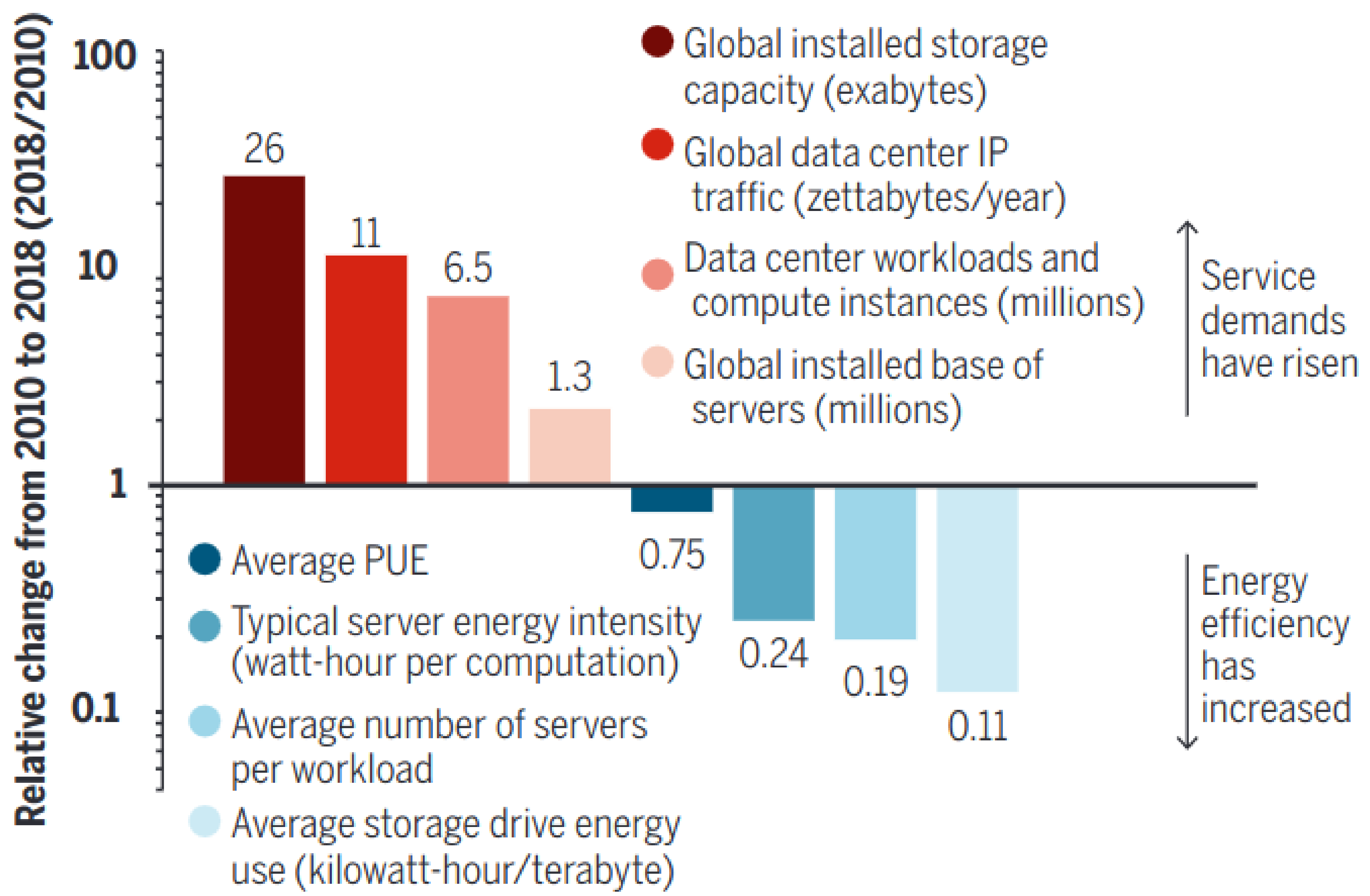
A summary



Is it a good result?

Masanet, Eric, et al. "Recalibrating global data center energy-use estimates." Science 367.6481 (2020): 984-986.

A summary



Is it a good result?

Not really! Energy consumption did not decrease

Masanet, Eric, et al. "Recalibrating global data center energy-use estimates." Science 367.6481 (2020): 984-986.

Additional comments about the use of green energy (and the associated carbon footprint)

The numbers provided by cloud providers have to be taken with care

- Cloud providers use different strategies to hide the fact that they use not-so-clean energy
    - Buying carbon credits (Strategy to compensate carbon emissions)

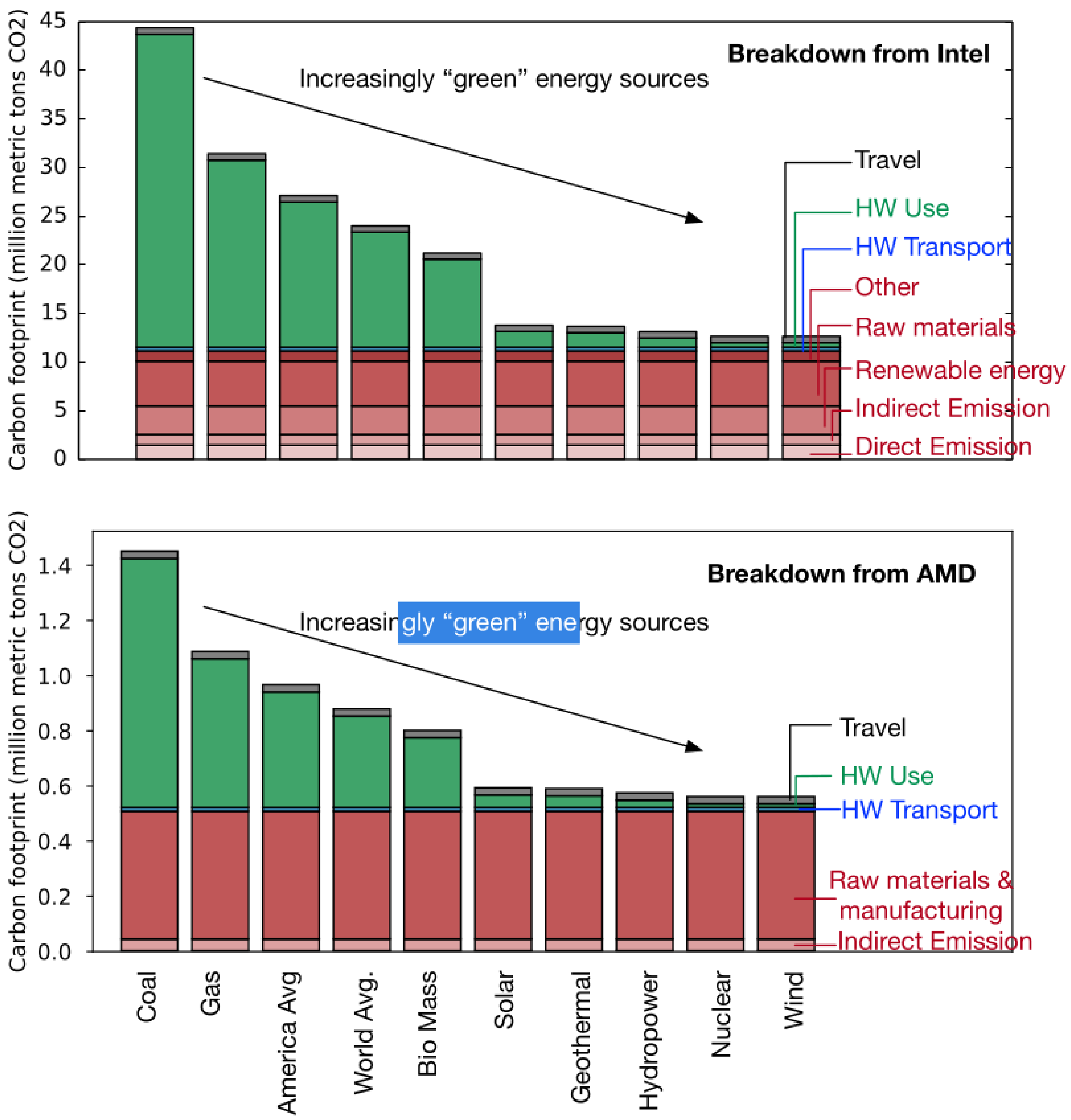
Green energy is not infinite

- The green energy consumed by data centers is not available for other usages

Improving the embodied carbon footprint of datacenters

Embodied carbon footprint has become the main concern

- Using green energy to power datacenters implies that embodied footprint becomes the main part of the carbon footprint



Source: Gupta, Udit, et al. "Chasing carbon: The elusive environmental footprint of computing." HPCA 2021.

Reducing the embodied footprint is difficult

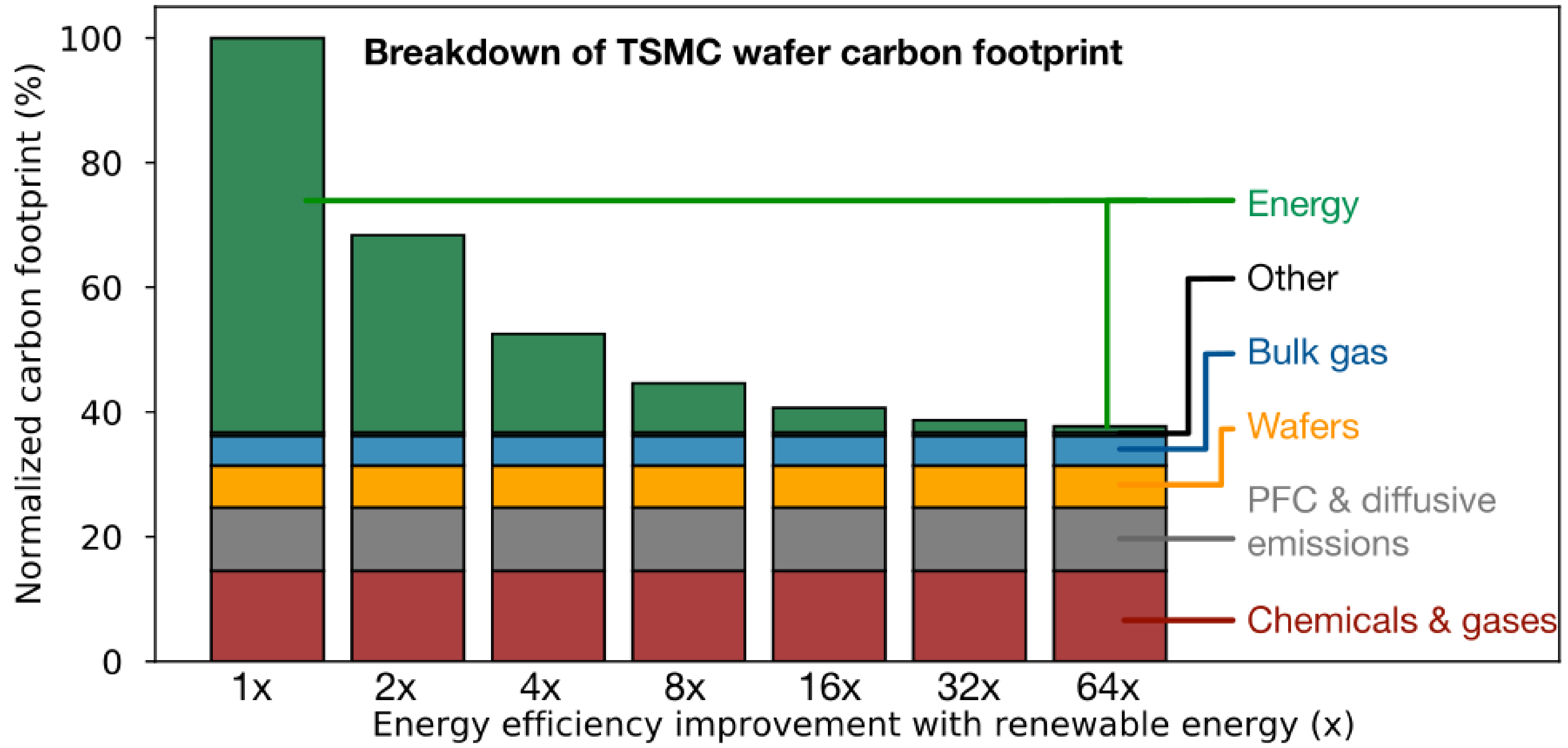
Relying on green energy to produce hardware

- Depends on decisions of the constructors
    - Are they willing to use green energy?
    - Is it even possible?

Reducing the embodied footprint is difficult

Relying on green energy to produce hardware

- The expected improvements in terms of carbon footprint are limited



Source: Gupta, Udit, et al. "Chasing carbon: The elusive environmental footprint of computing." HPCA 2021.

Contributors to the embodied footprint

Contributors to the embodied footprint

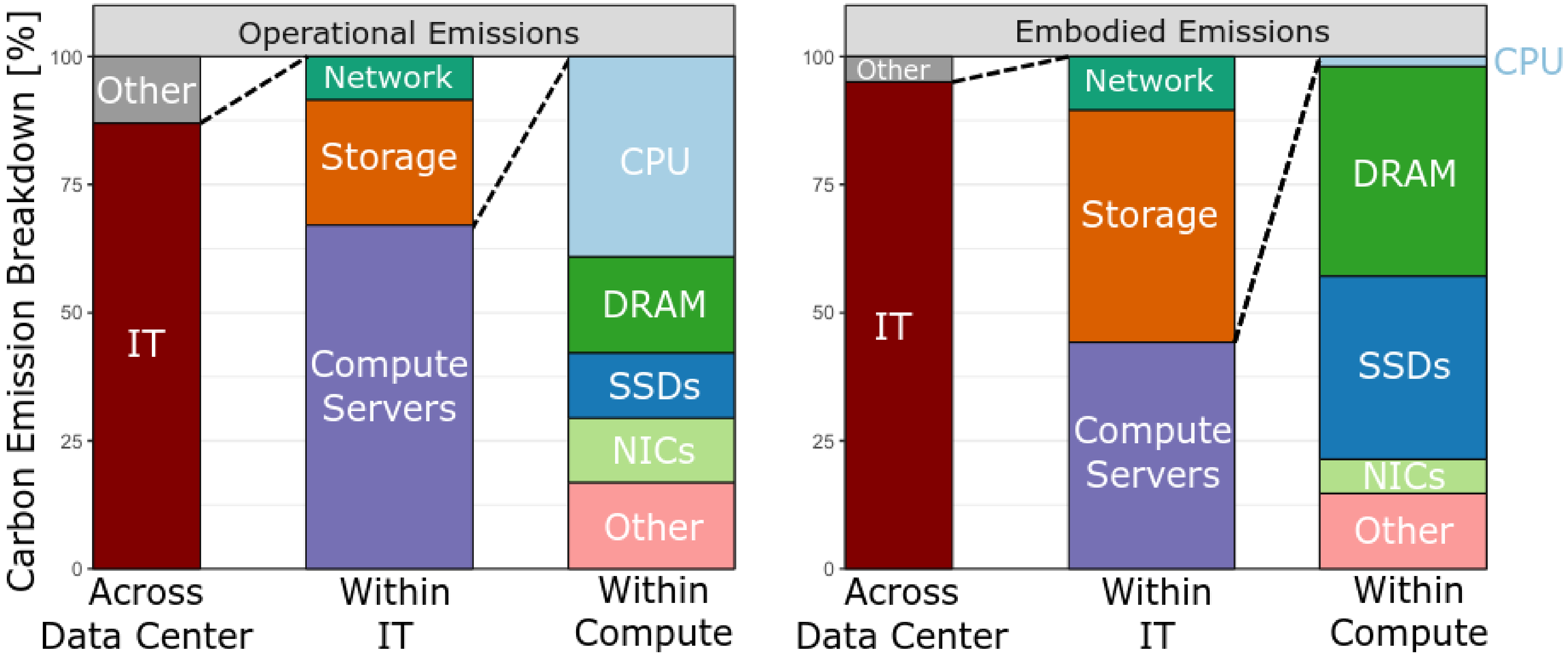


Fig. 1. Carbon breakdown of general-purpose data centers at Azure.

- The CPUs are not an important contributor
- Storage is a major concern

Source: Wang et al. "Designing cloud servers for lower carbon". ISCA, 2024.

How to improve the embodied footprint?

- Not so many factors we can have an impact on

How to improve the embodied footprint?

- Not so many factors we can have an impact on
- Main directions
    - Improve resource usage
    - Improve algorithms and software to use less resources
    - Increase the lifetime of servers
        - More generally of hardware components
    - Select the hardware carefully
        - Do we always need to most efficient hardware?

Problem: Evaluating/Collecting information about the embodied carbon footprint can be difficult

About improving resource usage

About improving resource usage

- On hyperscalers
  - Already very aggressive resource usage optimization is applied (see "Prequal" paper)
- On small/private datacenters
  - Might be difficult to have always enough load to keep all servers busy
    - Use resource consolidation strategies to be able to switch-off some servers
    - Reduces the operational footprint but not the embodied
  - Move the applications to an hyperscaler?
    - Also has drawbacks

Improve algorithms and software to use less resources

Improve algorithms and software to use less resources

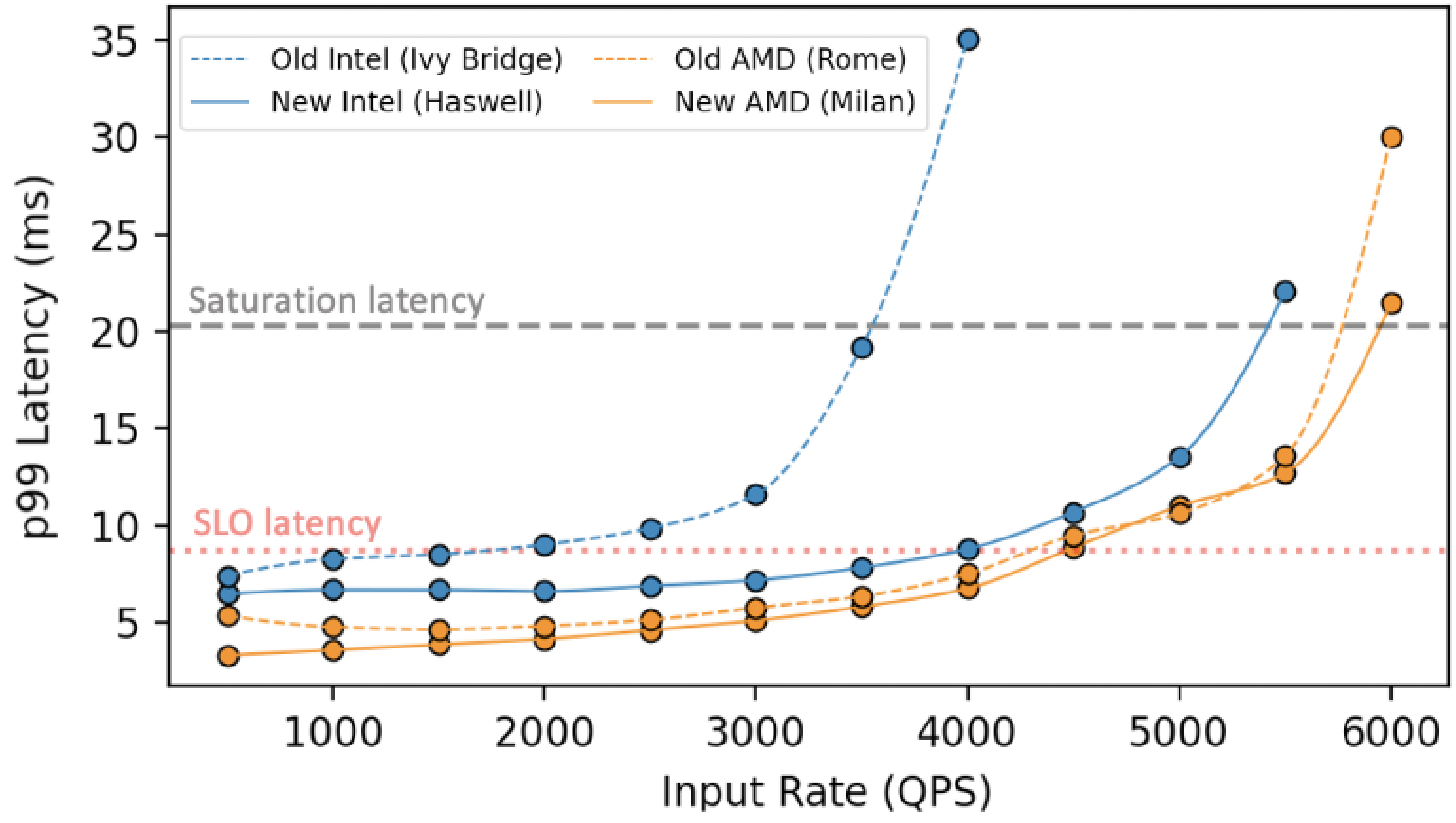
Beware of rebound effects!

Increasing the lifespan of servers

- Using the servers for longer period allows amortizing the embodied carbon footprint
  - What about the reliability of the hardware?
  - What about performance?

Increasing the lifespan of servers

Performance impact

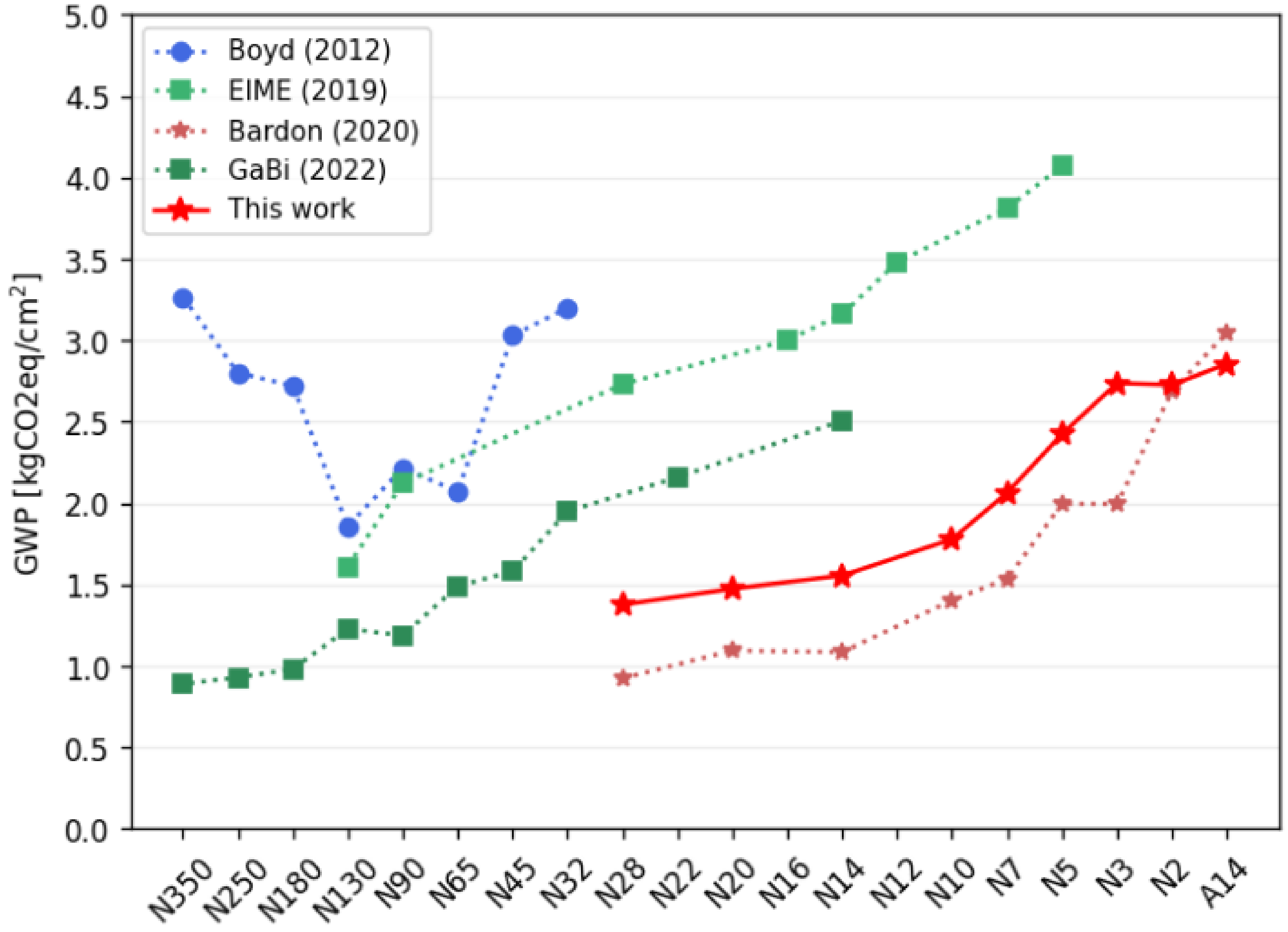


- DeathStarBench application deployed over 15 servers of a given type
    - Intel servers (2012, 2013)
    - AMD servers (2019, 2021)
- SLO = 75% of saturation for the best performing server
    - For low load, old servers can work

See: Wang, Jaylen, Udit Gupta, and Akshitha Sriraman. "Peeling Back the Carbon Curtain: Carbon Optimization Challenges in Cloud Computing." Proceedings of the 2nd Workshop on Sustainable Computer Systems. 2023.

Select the hardware carefully

- Newer semiconductor manufacturing processes have a higher carbon footprint per cm2 (technology node)



All studies show the same trends

- We could argue that with smaller transistors, we have more computing power per cm2

See: Boakes et al. "Cradle-to-gate life cycle assessment of CMOS logic technologies." International Electron Devices Meeting, 2023.

Select the hardware carefully

Select the hardware carefully

Beware of rebound effects!

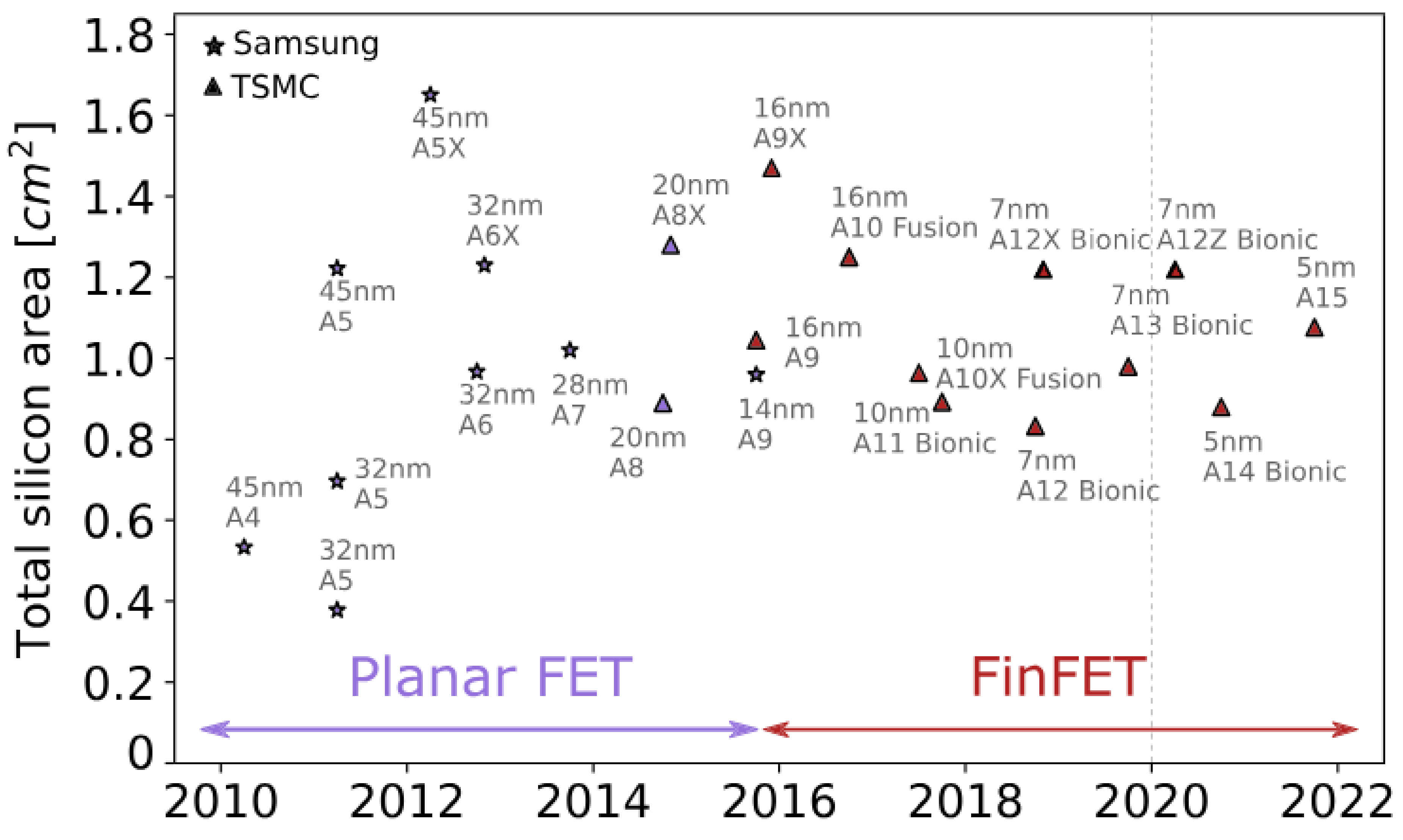


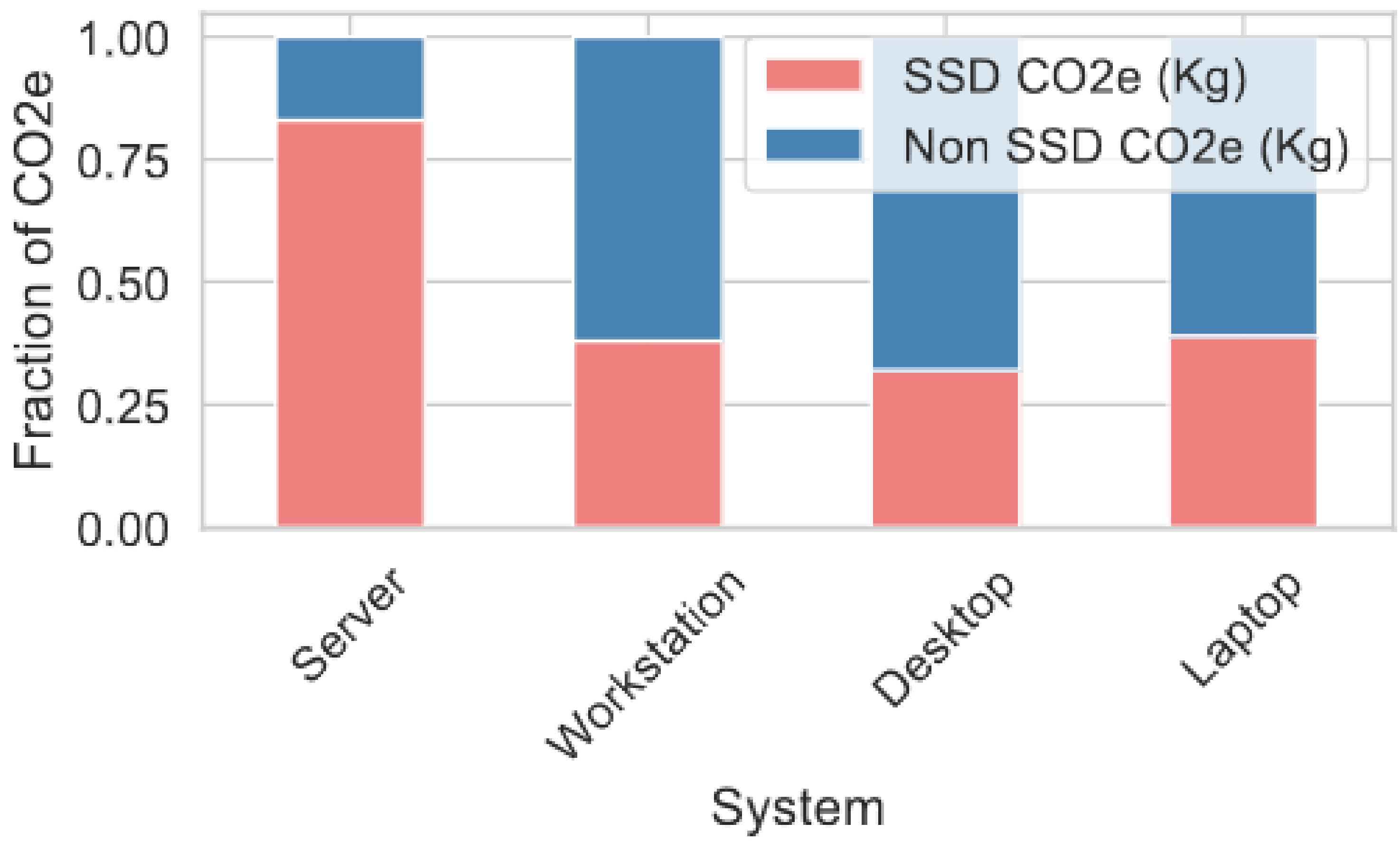
Fig. 7. Evolution of the average chip area of Apple's application processors.

- The area size of chips did not decrease

See: Pirson et al. "The environmental footprint of IC production: Review, analysis, and lessons from historical trends." IEEE Transactions on Semiconductor Manufacturing, 2022.

Select the hardware carefully

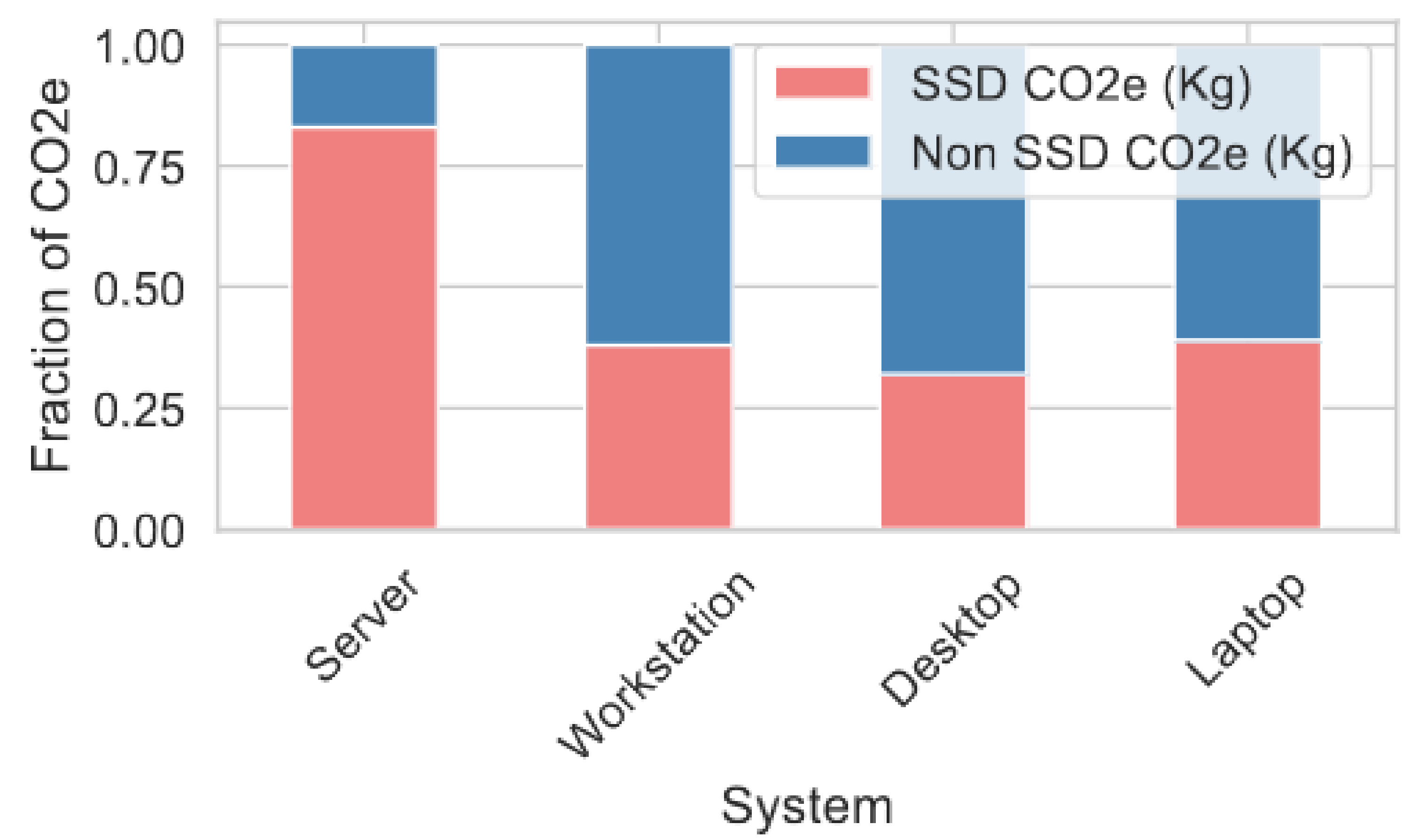
- SSDs are much more efficient than hard drives but:
    - The embodied Carbon Footprint of SDDs is 8x higher



See: Tannu, Swamit, and Prashant J. Nair. "The dirty secret of ssds: Embodied carbon." ACM SIGENERGY Energy Informatics Review 3.3 (2023): 4-9.

Select the hardware carefully

- SSDs are much more efficient than hard drives but:
  - The embodied Carbon Footprint of SDDs is 8x higher



- Could there be usecases where:
  - Users accept lower performance
  - The capabilities of SSDs cannot be fully utilized

See: Tannu, Swamit, and Prashant J. Nair. "The dirty secret of ssds: Embodied carbon." ACM SIGENERGY Energy Informatics Review 3.3 (2023): 4-9.