Green Computer Architecture

Recap

- Computation requires a lot of power
- When we need more performance than we can achieve in a single platform, computation scales out

Power = Energy / Time

Recap

- Data centers require a lot of energy
- Energy efficiency from innovation alone is insufficient

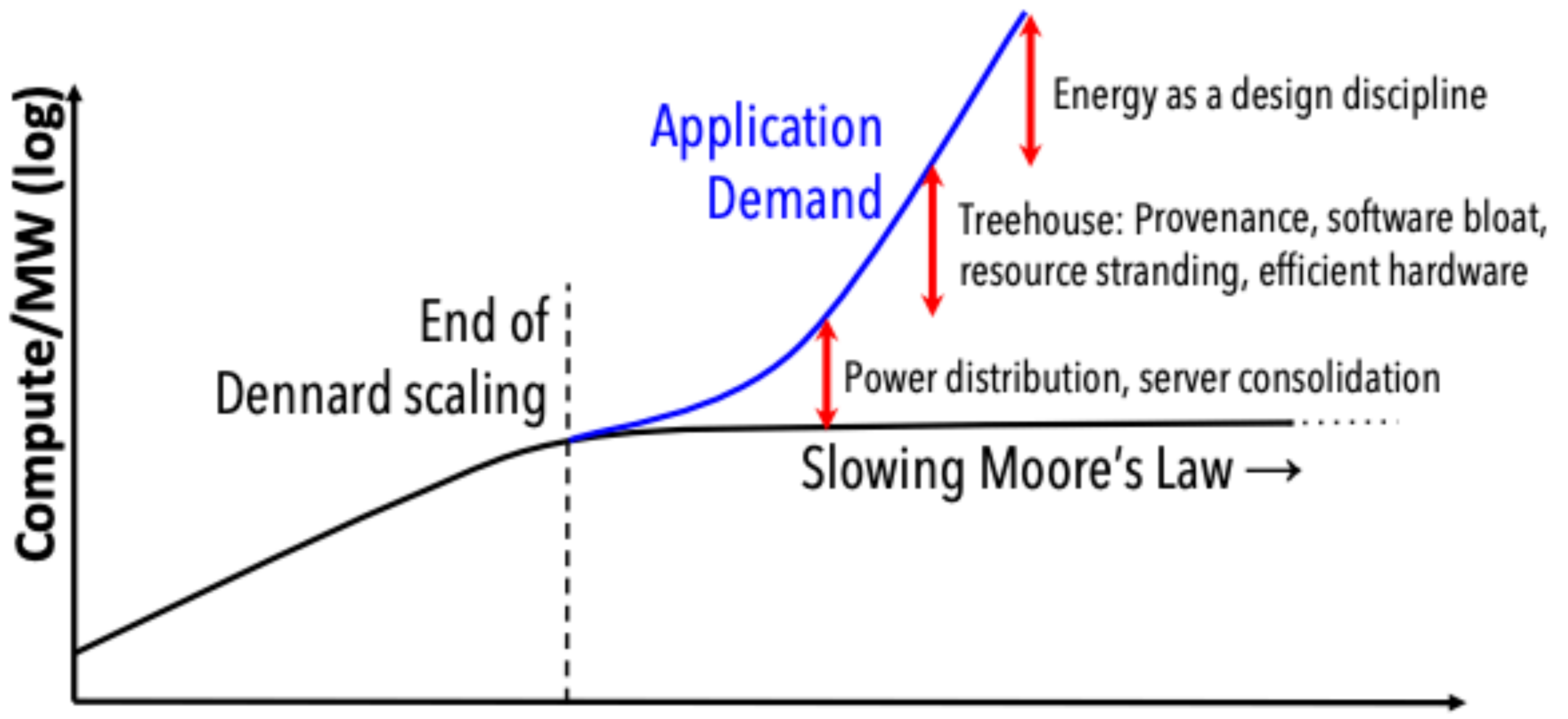
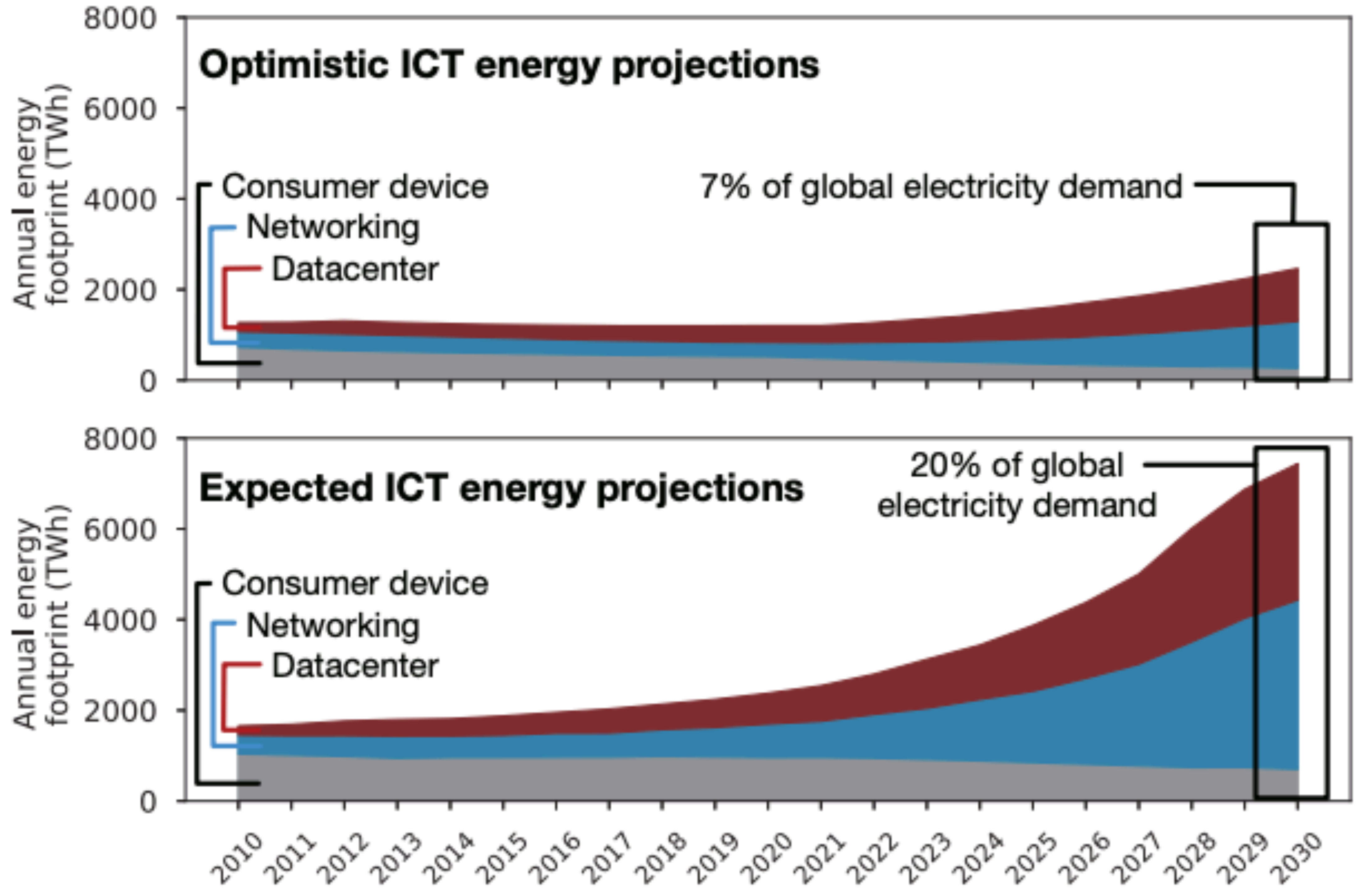


Fig. 1. Application demand for computing is growing faster than circuit-level energy efficiency. Treehouse takes a software-centric approach to reduce this gap.

Anderson, et al. Treehouse, SIGENERGY 2023

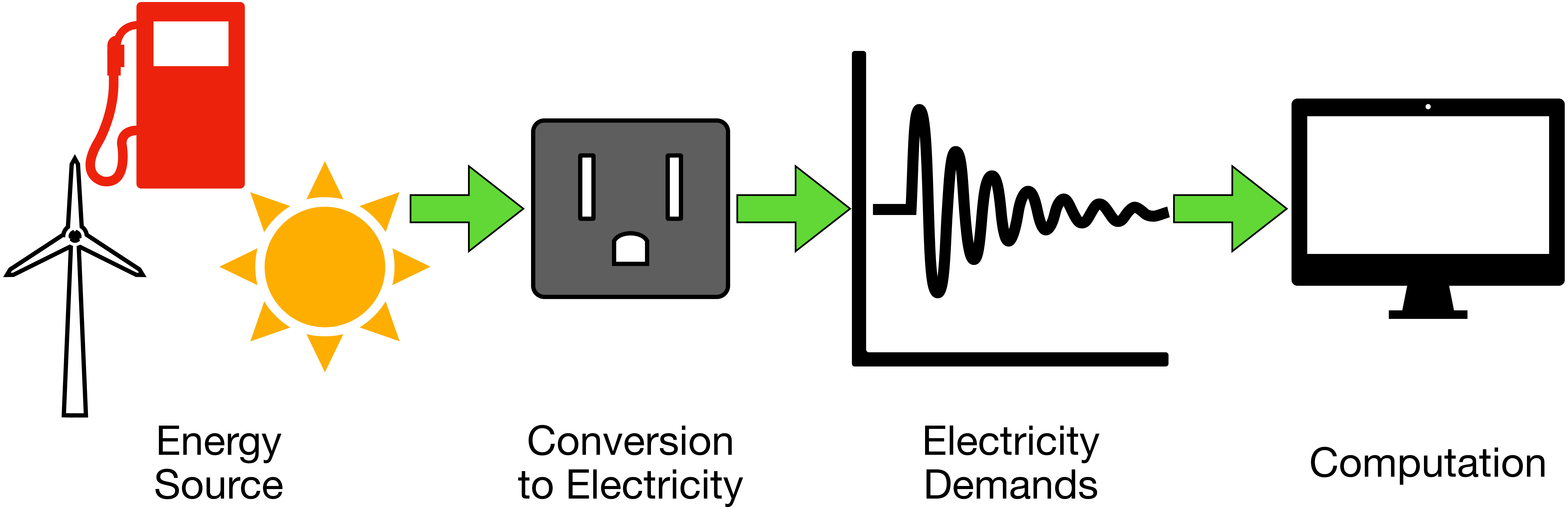


Gupta, et al. Chasing Carbon, HPCA 2021

Outline


From Energy to Carbon
Operational and Embodied Carbon
Profiling Carbon in Data Centers
Other Environmental Considerations

The Energy to Computation Pipeline



Chat with your neighbors!

What are the primary carbon bottlenecks in energy to computation pipeline?

Formalization of “Carbon Emissions”

- Greenhouse Gas Protocol defines an accounting standard followed by many companies to report carbon emissions

Scope 1

Direct Emissions

* Fuel Combustion
* Cooling
* Transportation
* Chemical Emissions

Scope 2

Indirect Emissions

* Purchased Energy Consumed
* Emissions from Converting Energy to Electricity

Scope 3

Upstream and Downstream Emissions

* Hardware Purchasing
* Device Lifetimes
* Transportation

Formalization of "Carbon Emissions"

How long producing energy until the initial energy to produce plant is regenerated

Source	Carbon intensity ($g\ CO_2/kWh$)	Energy-payback time (months)
Coal	820	2 [33]
Gas	490	1 [33]
Biomass	230	∼12 [73]
Solar	41	∼36 [34]
Geothermal	38	72 [74]
Hydropower	24	∼12–36 [33], [75]
Nuclear	12	2 [33]
Wind	11	≤12 [35]

TABLE II
CARBON EFFICIENCY OF VARIOUS RENEWABLE-ENERGY SOURCES.

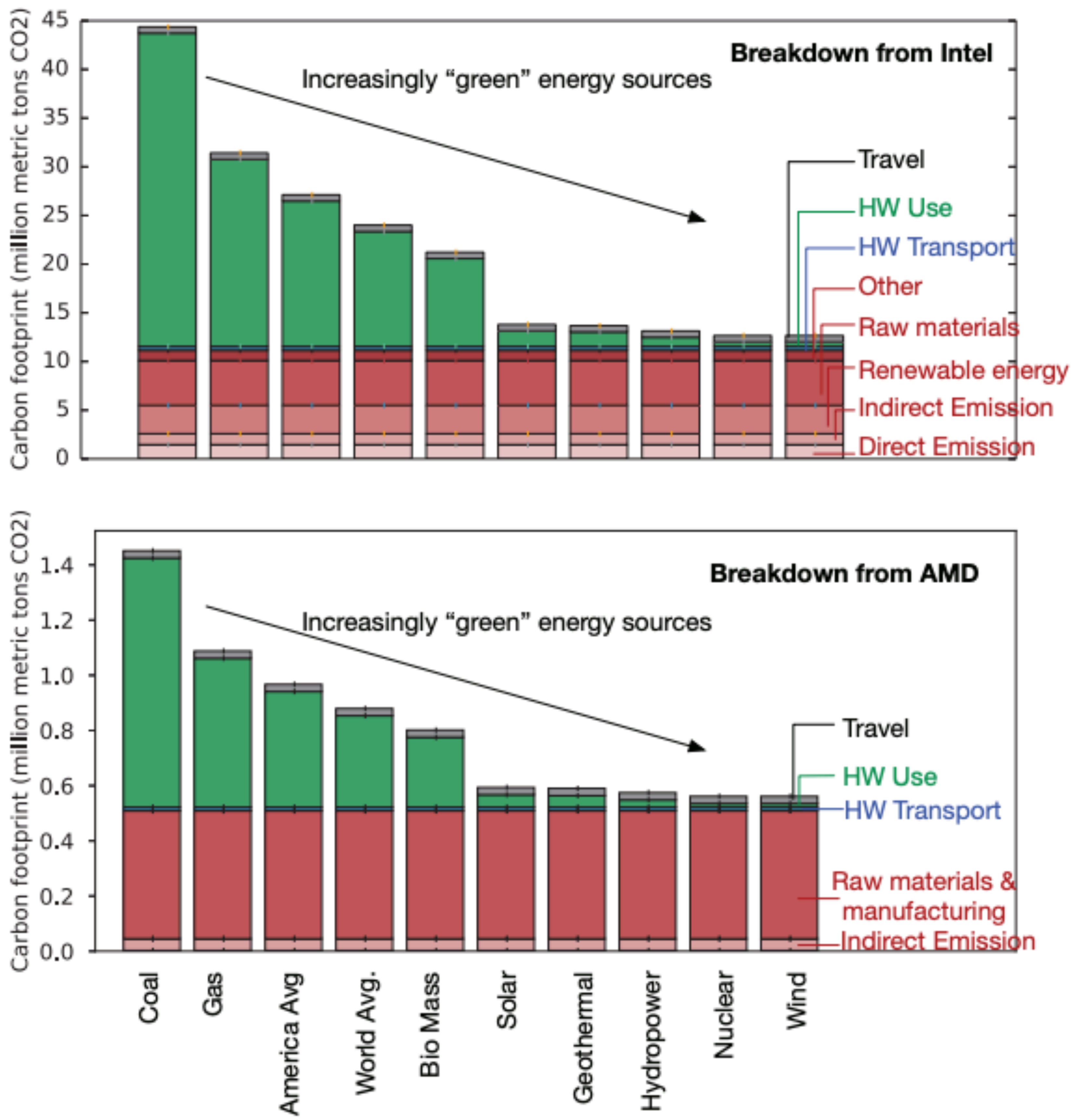


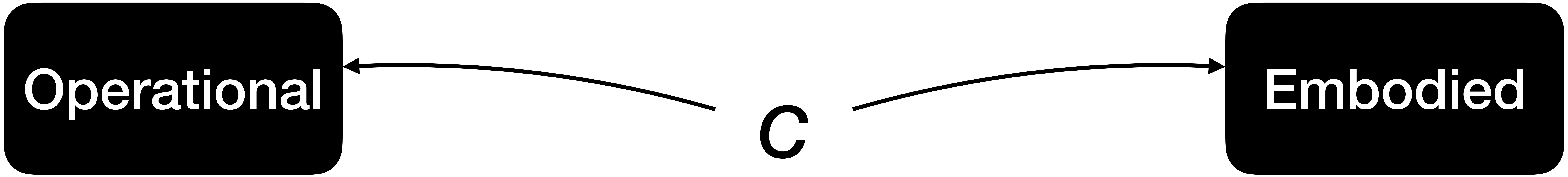
Fig. 13. Reported carbon-footprint breakdown for Intel (top) and AMD (bottom) as renewable energy increasingly (from left to right) powers hardware operation. The use of renewable energy reduces carbon emissions dramatically; most of the remaining emissions are from manufacturing.

Gupta, et al. Chasing Carbon, HPCA 2021

Takeaways

- Direct relation between computing energy and carbon emissions
- Emissions can be further characterized based on when they are produced
- Renewables reduce the overall carbon footprint of computation
- Producing renewable energy is not "free"

Characterizing Computational Carbon



Carbon cost at runtime

Carbon cost of manufacturing devices

Chat with your neighbors!

Come up with an argument for why embodied or operational carbon is a bigger overhead!

Operational Carbon

- Carbon Footprint =

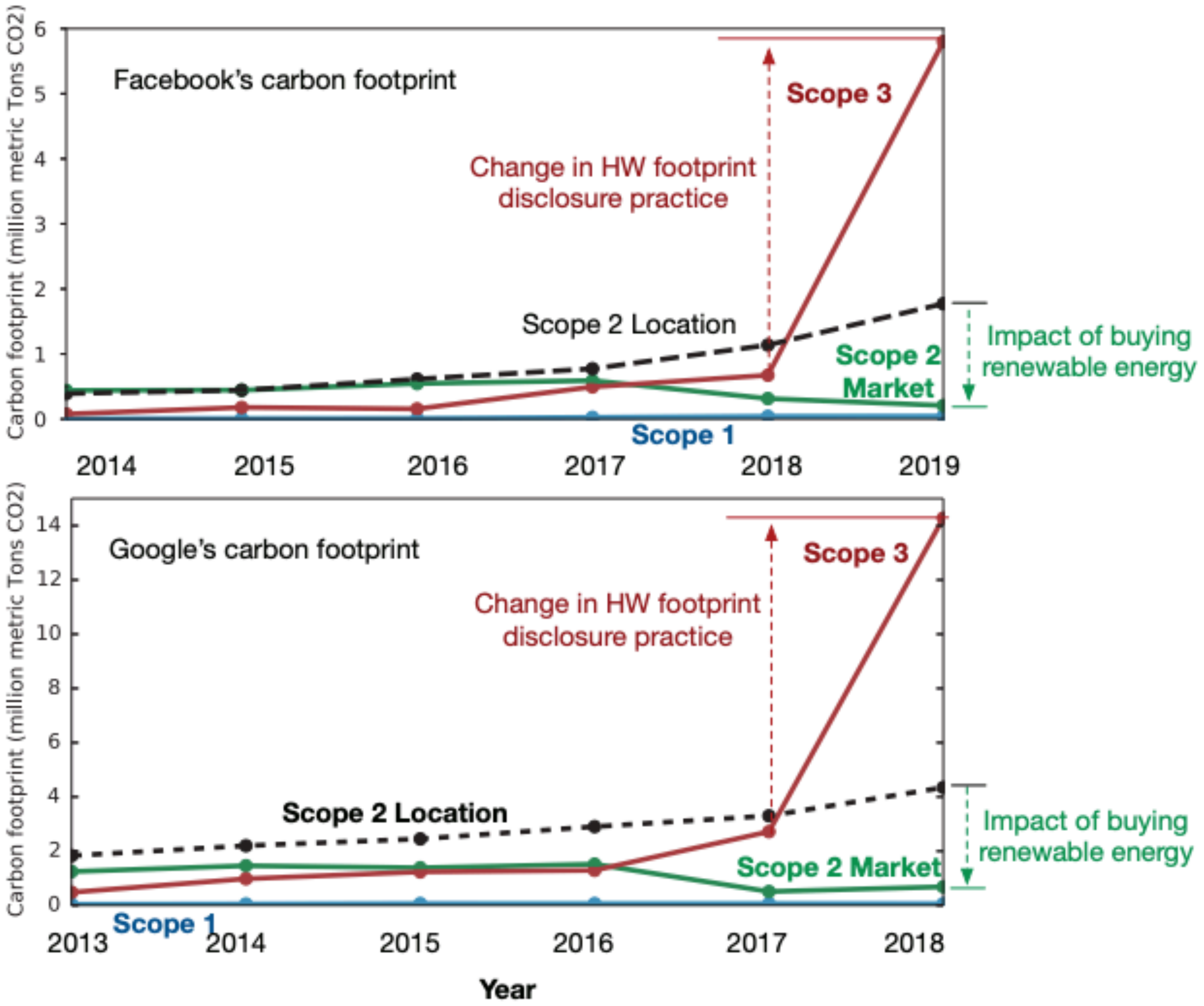
  Operational Carbon Footprint + (Embodied Carbon / System Lifetime)

- Operational Carbon Footprint = Carbon Intensity * Energy Source

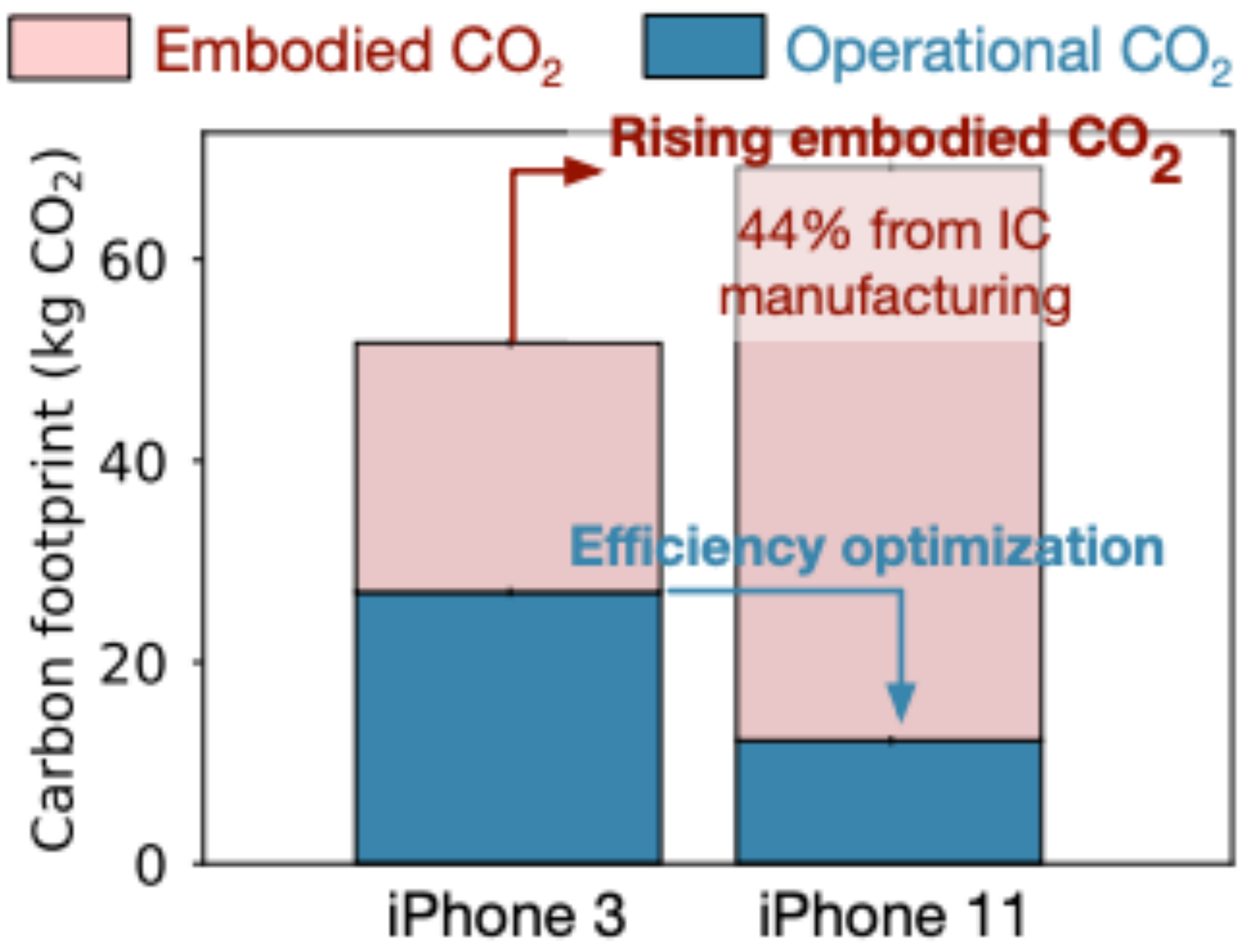
Conventional Thinking…

- Optimize for operational carbon
- Embodied carbon cost is amortized over a device’s lifetime
- Lo[illegible]s the “effective” embodied cost is lower
- Se[illegible] years

In practice…

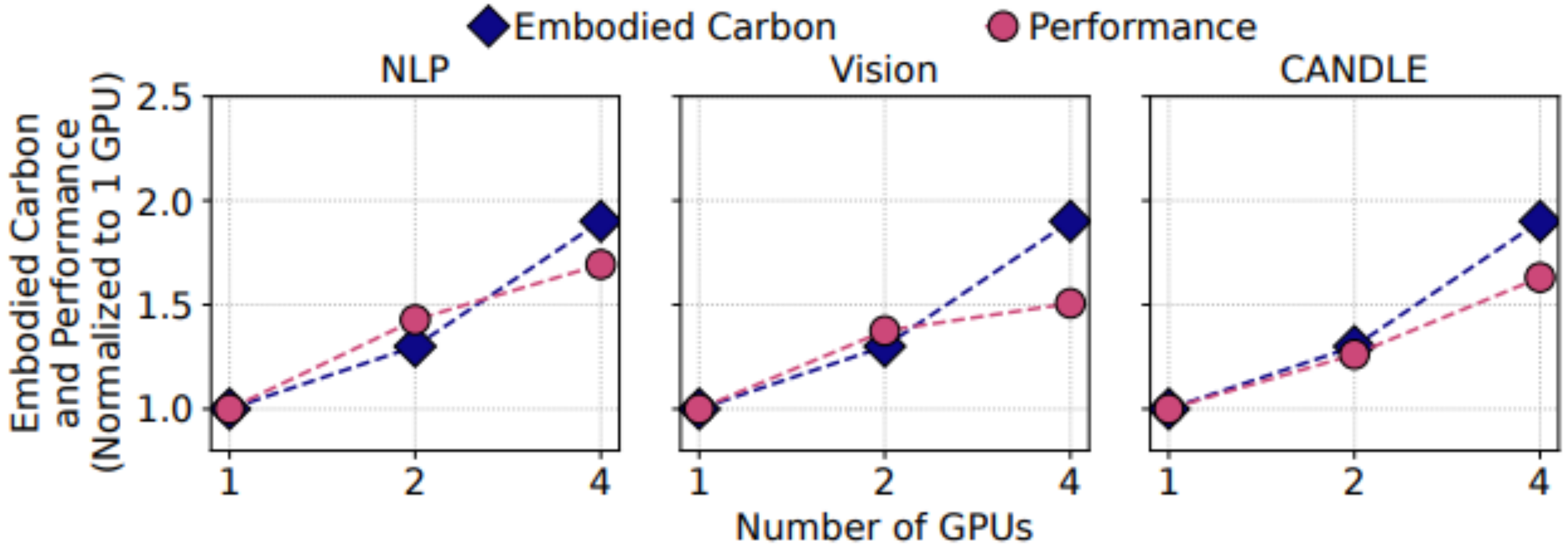


Gupta, et al. Chasing Carbon, HPCA 2021



Gupta, et al. ACT, ISCA 2022

Why Embodied Carbon?



Li, et al. Toward Sustainable HPC. SC 2023

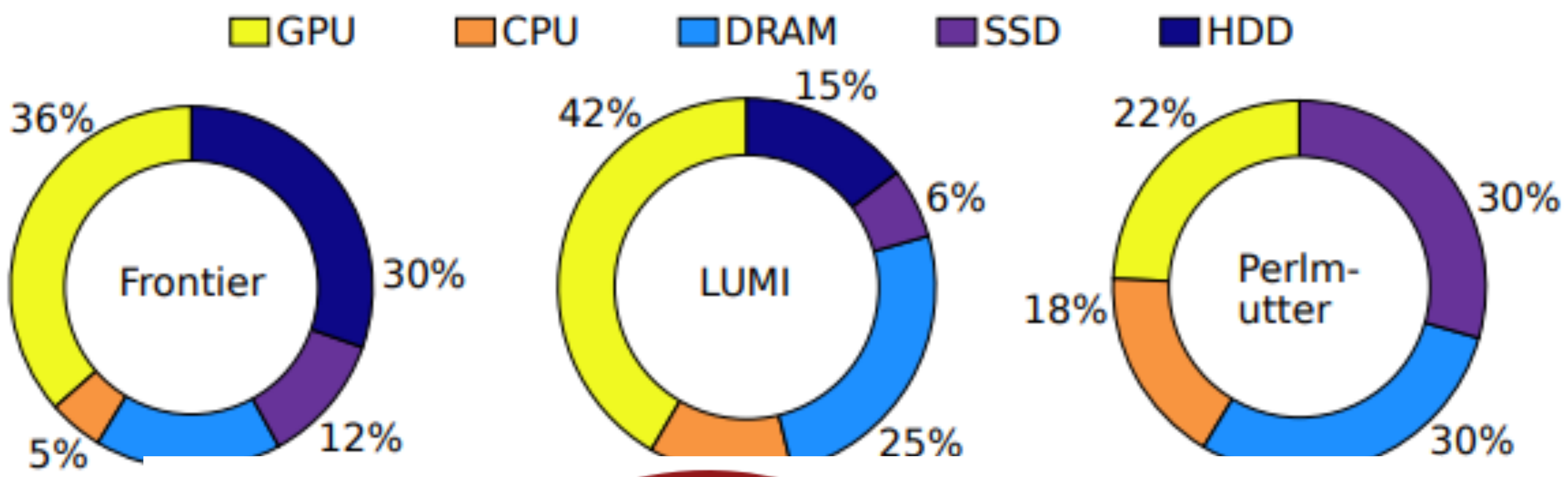


Figure 5: ... nt com-
ponents ... , LUMI,
and Perl...

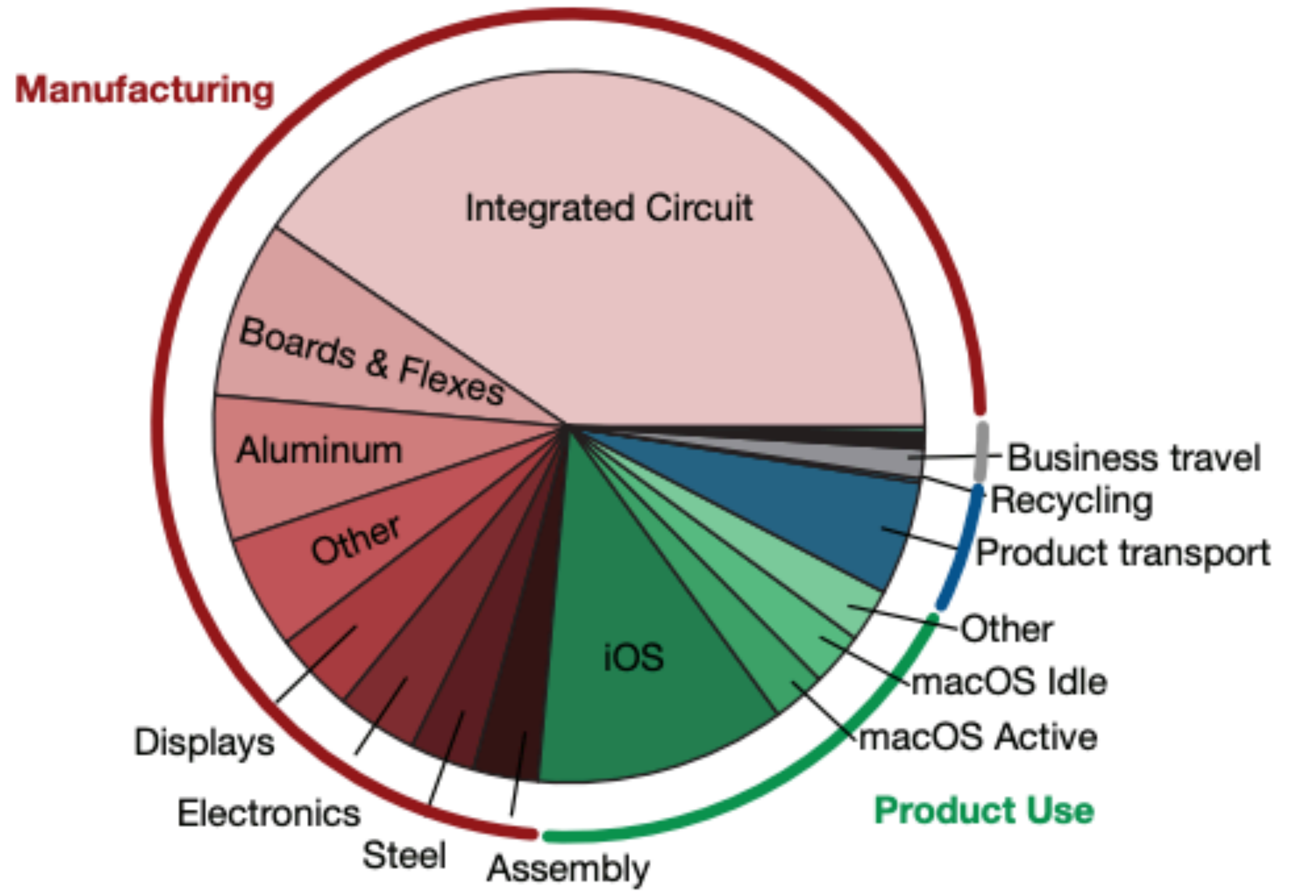


Fig. 5. Apple's carbon-emission breakdown. In aggregate, the hardware life cycle (i.e., manufacturing, transport, use, and recycling) comprises over 98% of Apple's total emissions. Manufacturing accounts for 74% of total emissions, and hardware use accounts for 19%. Carbon output from manufacturing integrated circuits (i.e., SoCs, DRAM, and NAND flash memory) is higher than that from hardware use.

More GPUs improves performance to a point, but more hardware requires more embodied carbon

The Embodied Cost

- Carbon emissions are a function of integrated circuitry
- For CPU and GPU (kg CO_2 per cm^2)
  - 0.1-0.4 kCO_2/cm^2
- For memory and storage (kg CO_2 per GB):
  - DRAM: 0-.6 kCO_2/GB, SSD: 0-.3 kCO_2/GB, HDD: 0-.12 kCO_2/GB

Takeaways

- Embodied carbon can often exceed operational carbon costs
- Larger components require more carbon

Operational Carbon in Data Centers

- Improved device efficiency
- Using renewable energy sources
  - Intermittent reliability
  - Geographic implications
  - Batteries
- Inter-Data Center Scheduling

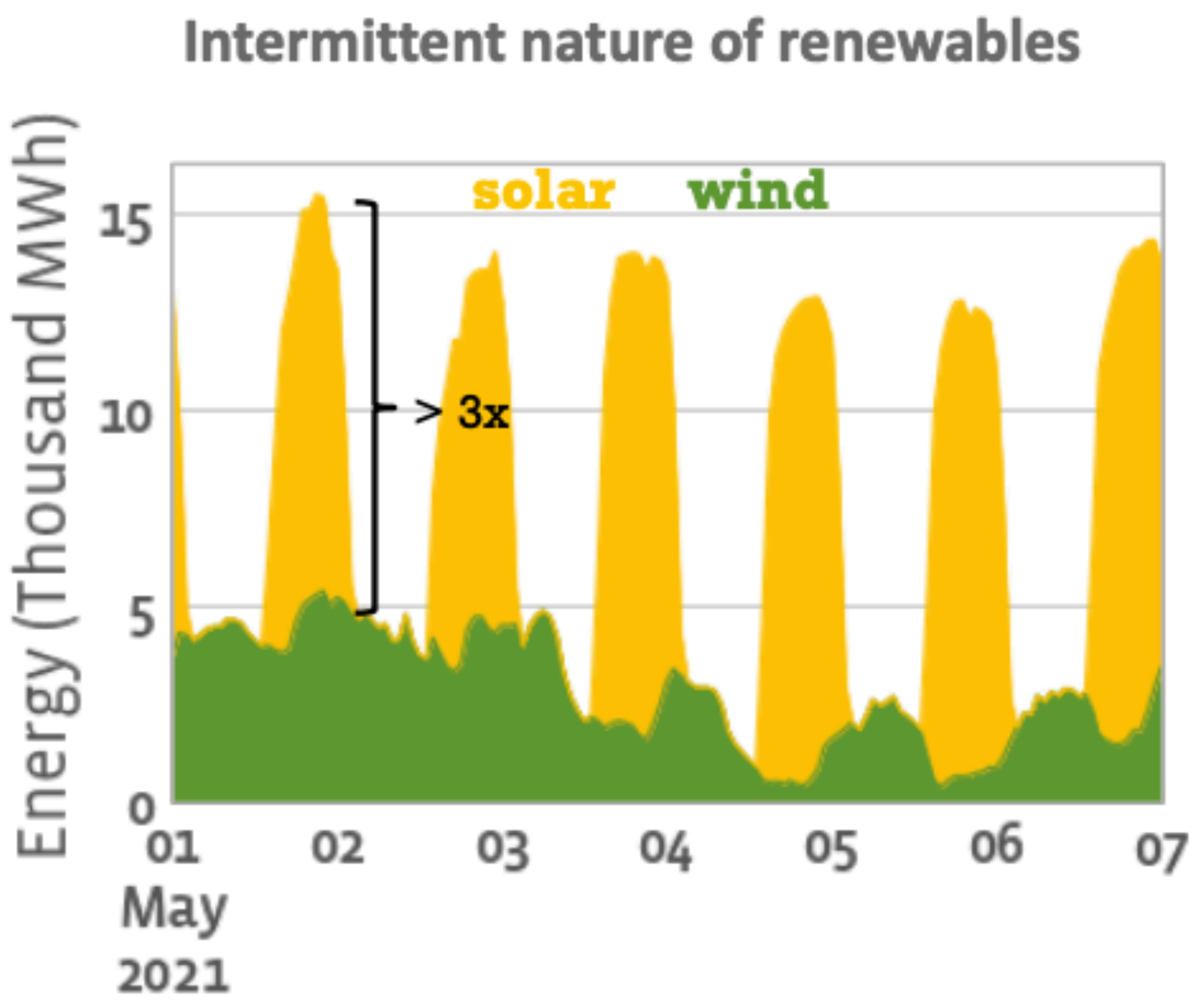


Figure 1: Hourly wind and solar energy generation in California grid during a week of time-frame.

Acun, et al. Carbon Explorer. ASPLOS 2023

Heterogeneous Components

- Components wear at different rates
  - Compute lifetime 3-5 years
  - Memory lifetime 5-7 years
- Reintegrate memory devices with newer compute components
- See also, “Junkyard Computing”

Wang, et al. Designing Cloud Servers for Lower Carbon. ISCA 2024

Fig. 2. Moving average (black) of raw (gray) normalized failure rates vs. DDR4 DIMMs’ deployment time in production. Failure rates tend to stay constant over a 7-year period.

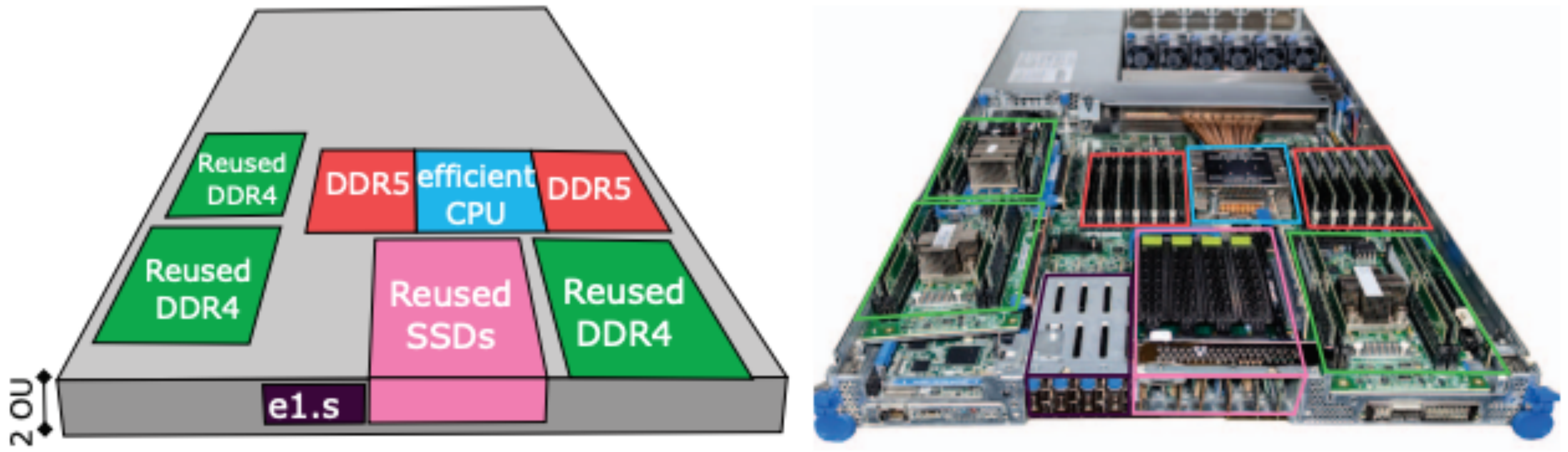


Fig. 5. Our *GreenSKU-Full* design with AMD’s efficient CPU, reused DDR4 DRAM (via CXL), and reused m.2 SSDs (via e1.s and PCIe adapters).

Chat with your neighbors!

What other considerations go into green computing?

Why just carbon?

- Forever chemicals
- Water cooling of data centers
- Electronic waste

https://sustainability.atmeta.com/wp-content/uploads/2020/12/FB_Sustainability-Data-Disclosure-2019.pdf

Further Reading!

- Chasing Carbon: The Elusive Environmental Footprint of Computing
- ACT: Designing Sustainable Computer Systems with an Architectural Carbon Modeling Tools
- Treehouse: A Case for Carbon-Aware Datacenter Software
- Carbon Explorer: A Holistic Framework for Designing Carbon Aware Datacenters
- Toward Sustainable HPC: Carbon Footprint Estimation and Environmental Implications of HPC Systems
- Designing Cloud Servers for Lower Carbon