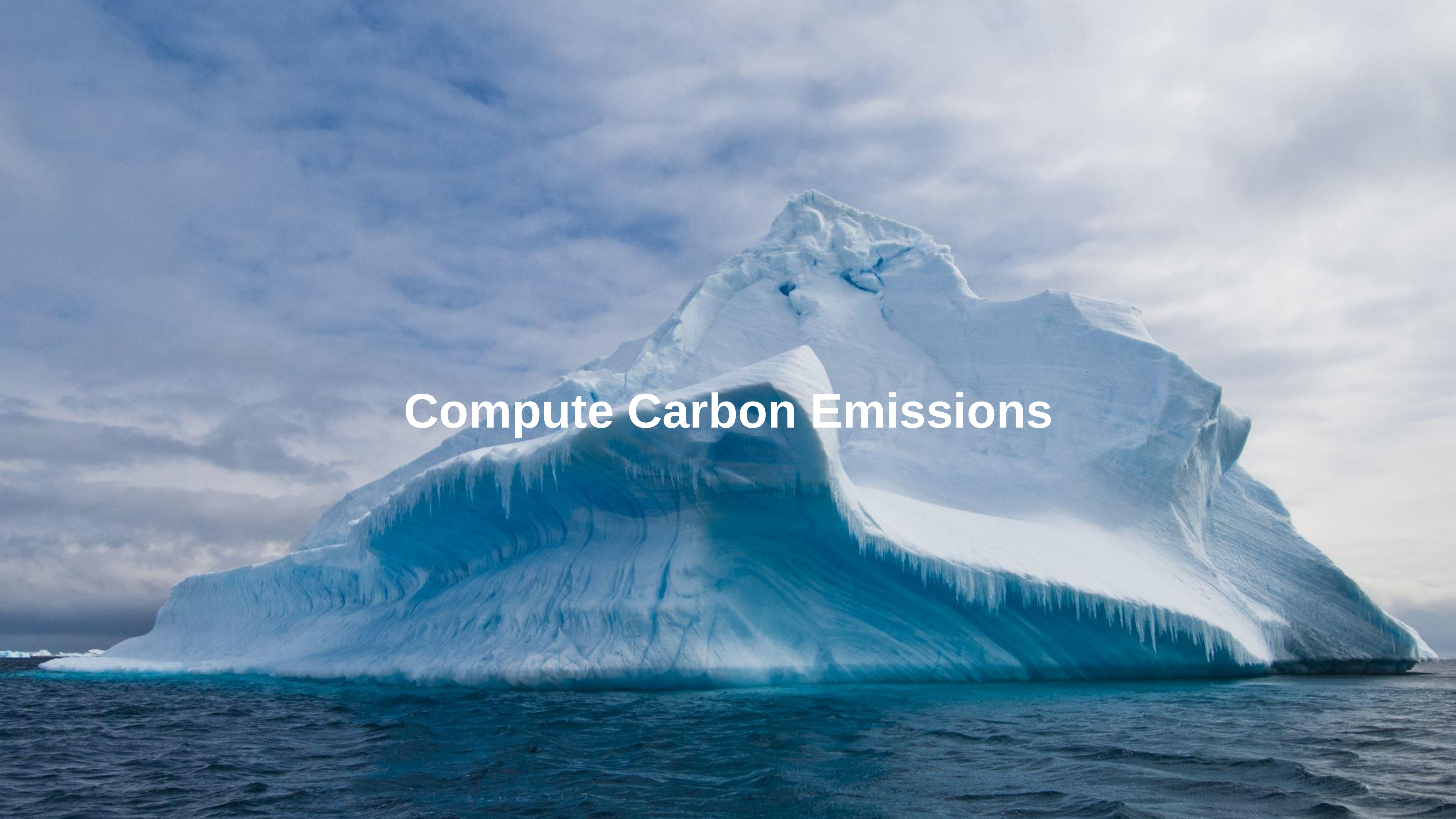
Sustainable Computing at Scale

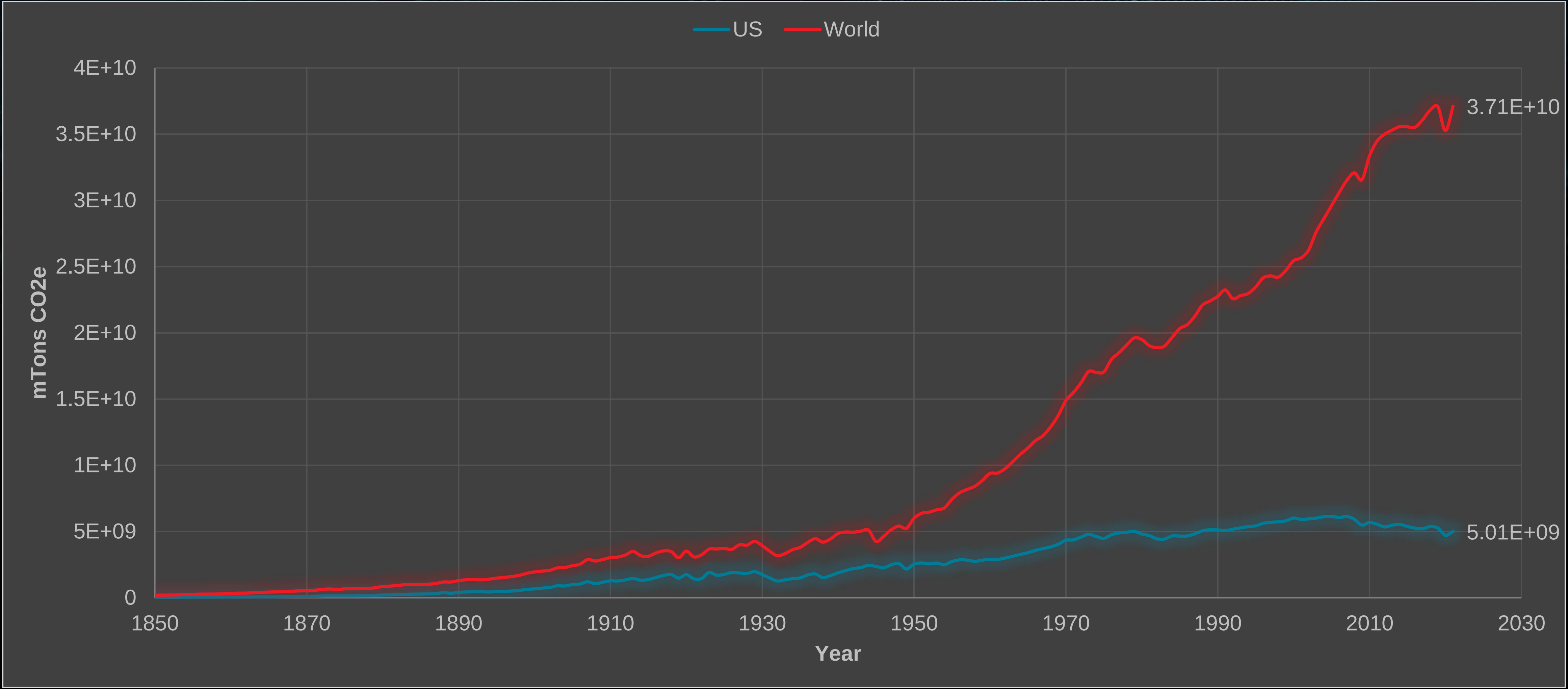
Srilatha (Bobbie) Manne
Senior Fellow, AMD

AMD
together we advance_

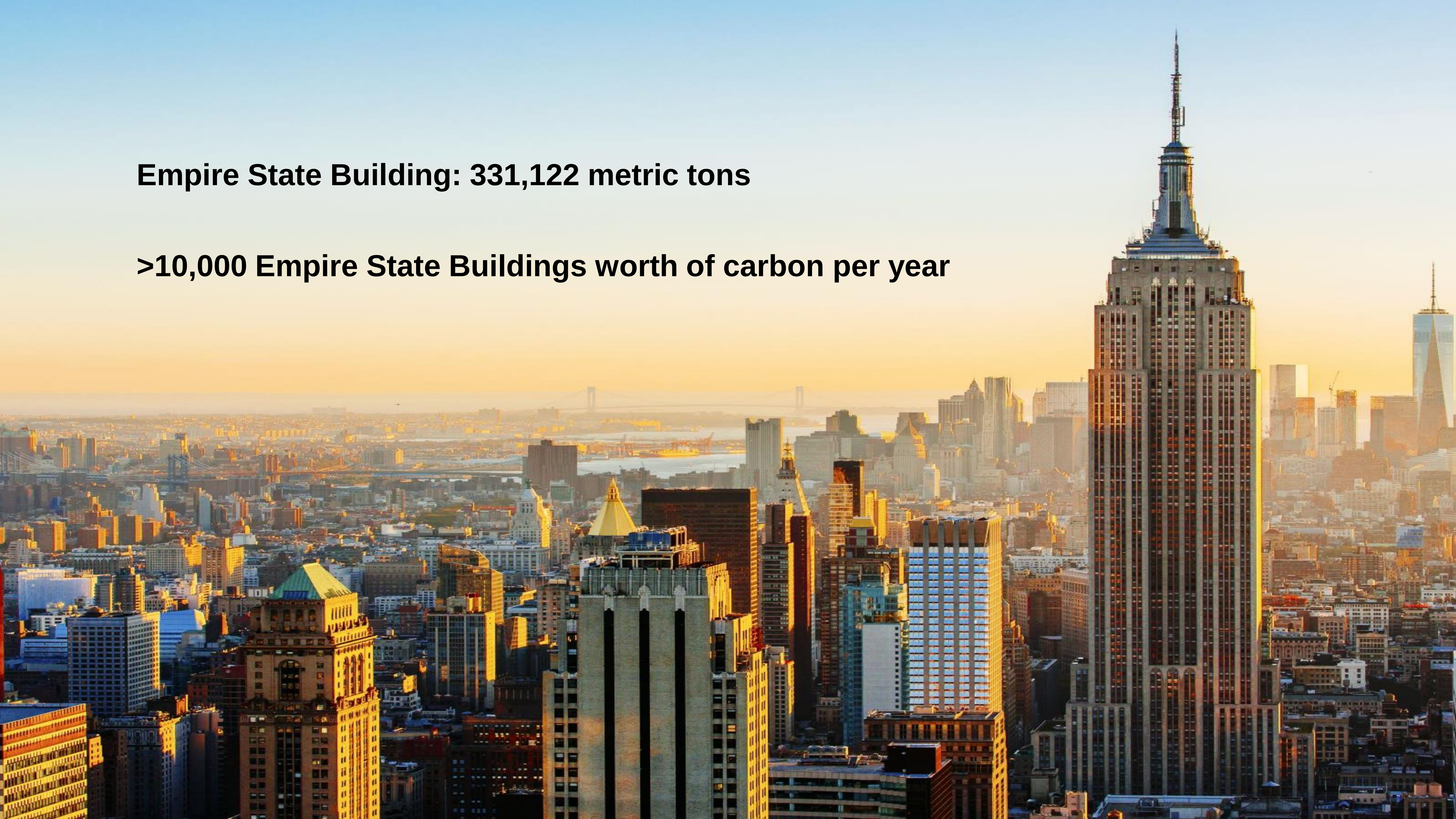
Compute Carbon Emissions



Carbon Emissions Over Time



Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "CO_2 and Greenhouse Gas Emissions". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/co2-and-greenhouse-gas-emissions' [Online Resource]


Empire State Building: 331,122 metric tons
>10,000 Empire State Buildings worth of carbon per year

Data Center Power Use

- Up to 460 TWh of electricity/year in 2022 [†]
- 2% of global energy demand [†]

Equivalent to 1 year of *

46M cars

50,000 wind turbines

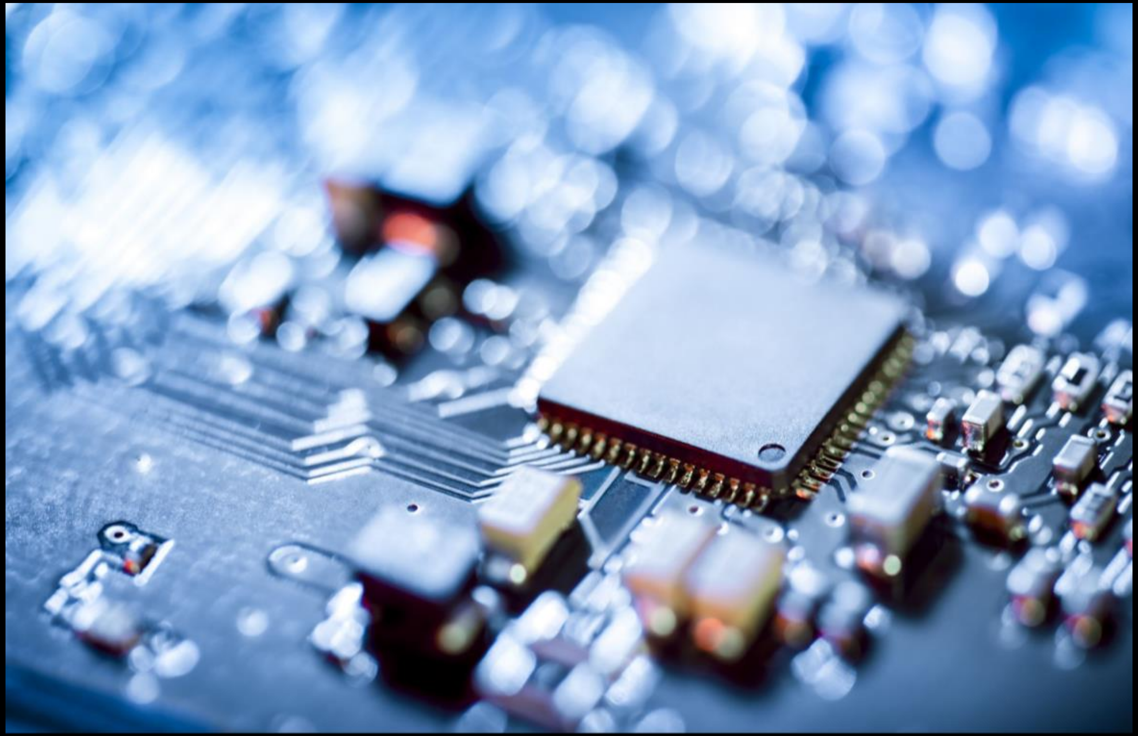
224M acres of forestland (Texas is ~172M acres)

192M metric tons CO_2e

[†] IEA (2024), Electricity 2024, IEA, Paris https://www.iea.org/reports/electricity-2024, Licence: CC BY 4.0
* Based on EPA Greenhouse Gas Equivalences Calculator calculated on 3/20/2024

Hardware Innovations Over Last 5 Decades

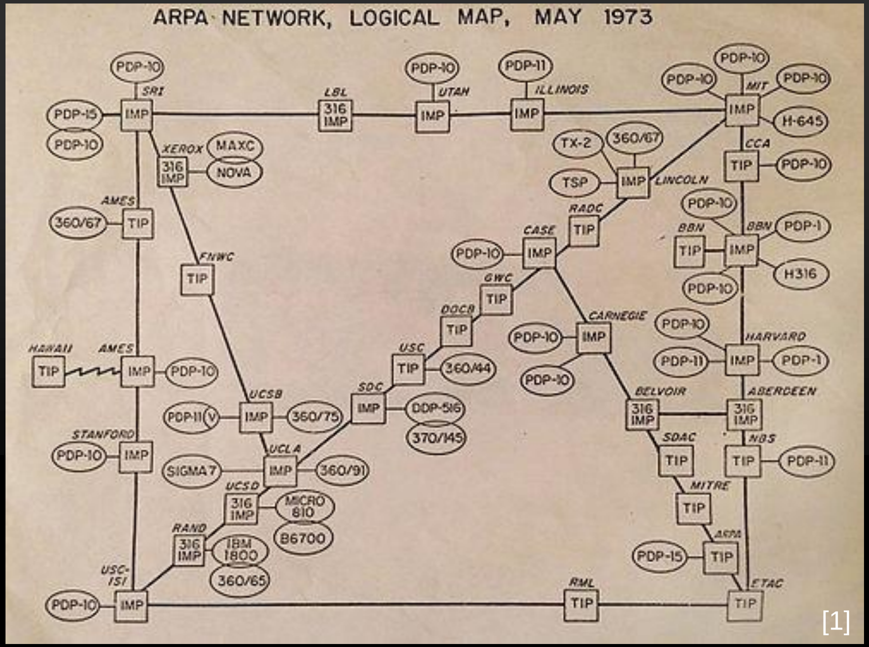
>2000 times faster



>3000 times smaller

Analysis based on comparing an AMD AM9080 processor running at 2MHz to an AMD Ryzen 9 7950X with base clock frequency of 4.5GHz.

The Past 50 Years

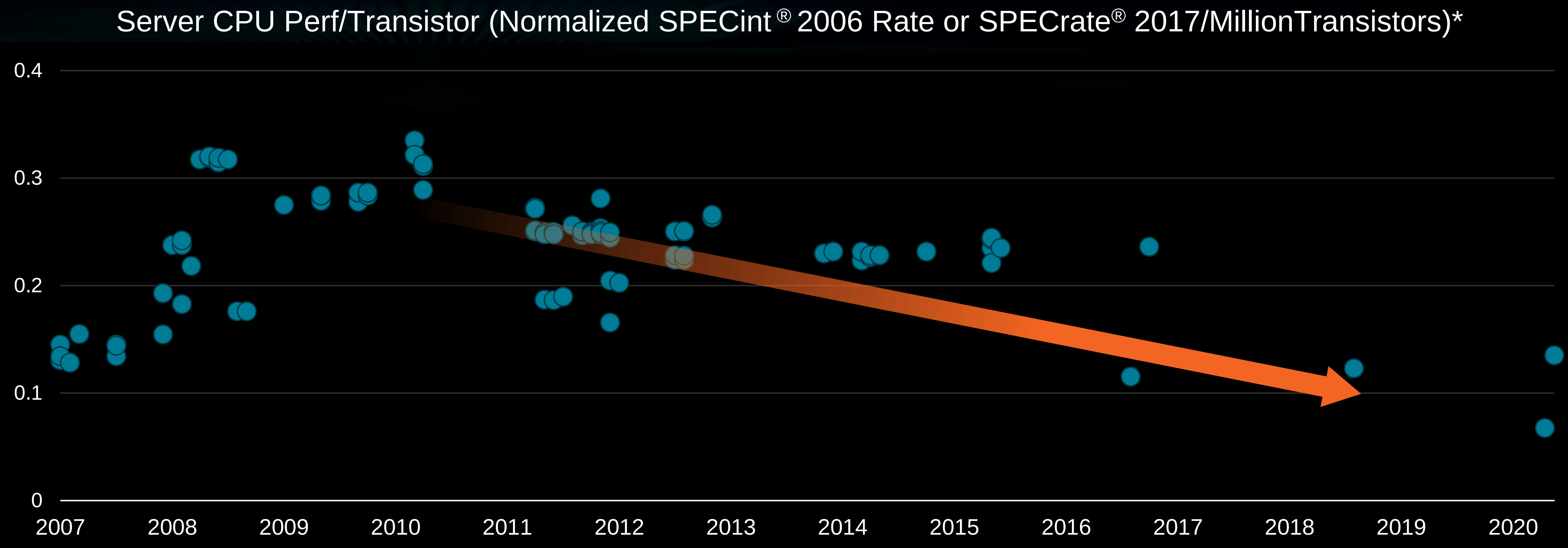


1973 ARPANET: 42 computers

2023: Over 42 Billion connected devices[2]

[1] https://commons.wikimedia.org/wiki/File:Arpanet_map_1973.jpg, Public Domain, Wiesoweshalbwarum
[2] https://www.forbes.com/sites/bernardmarr/2022/11/07/the-top-4-internet-of-things-trends-in-2023/?sh=67f33bf62aea

Transistors and Performance



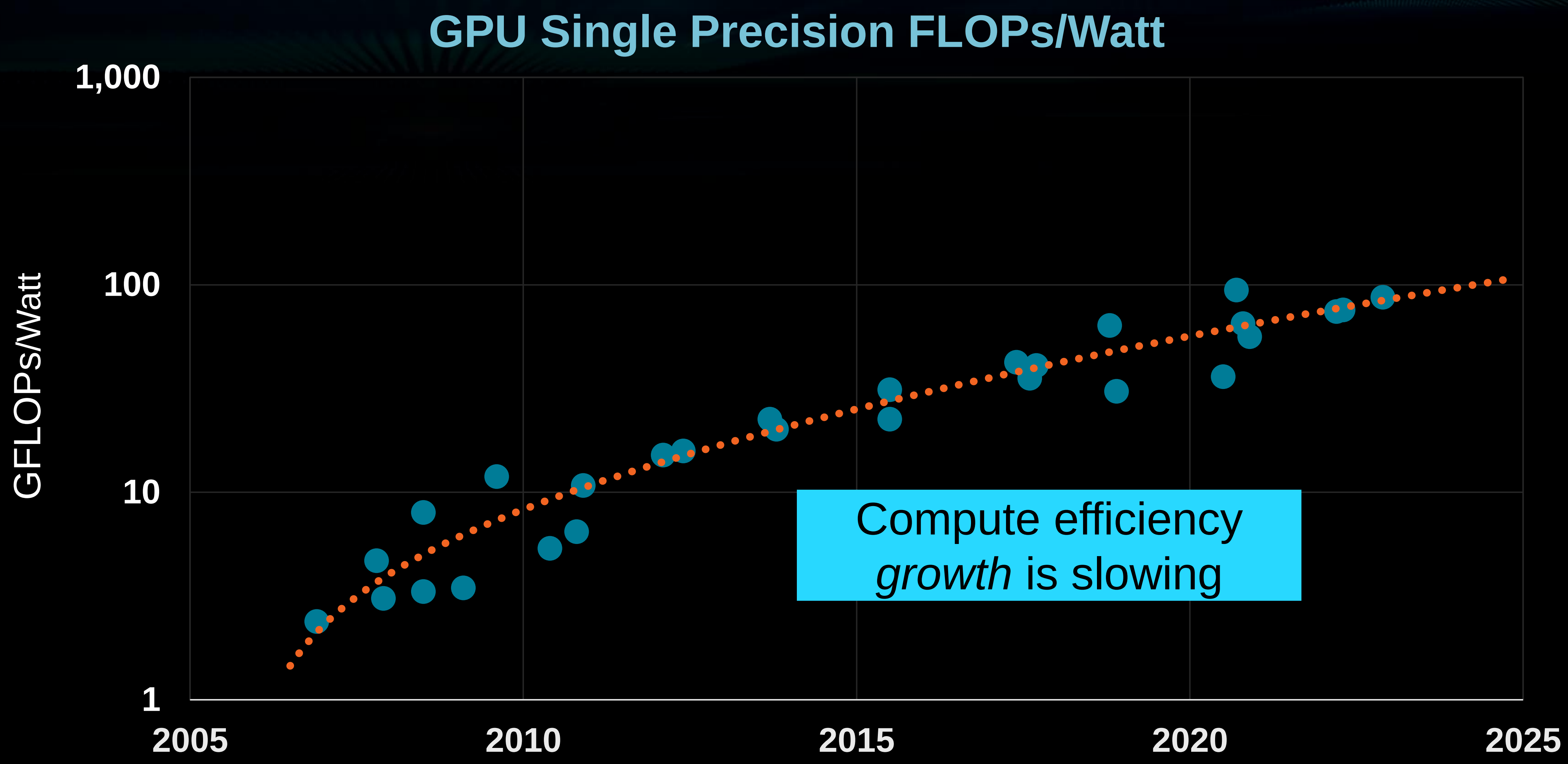
Another way to look at this is through the lens of performance/transistor

We're using more of them to extract a unit of performance

*Normalized SPECint® 2006 Rate and SPECrate® 2017 published results from www.spec.org as of 12/31/20 divided by total CPU transistors. SPEC®, SPEC CPU®, SPECrate®, and the benchmark name SPECint® are registered trademarks of the Standard Performance Evaluation Corporation (www.spec.org). SPEC CPU® 2006 is a retired benchmark and SPEC ® is no longer reviewing or publishing SPEC ® CPU 2006 results.

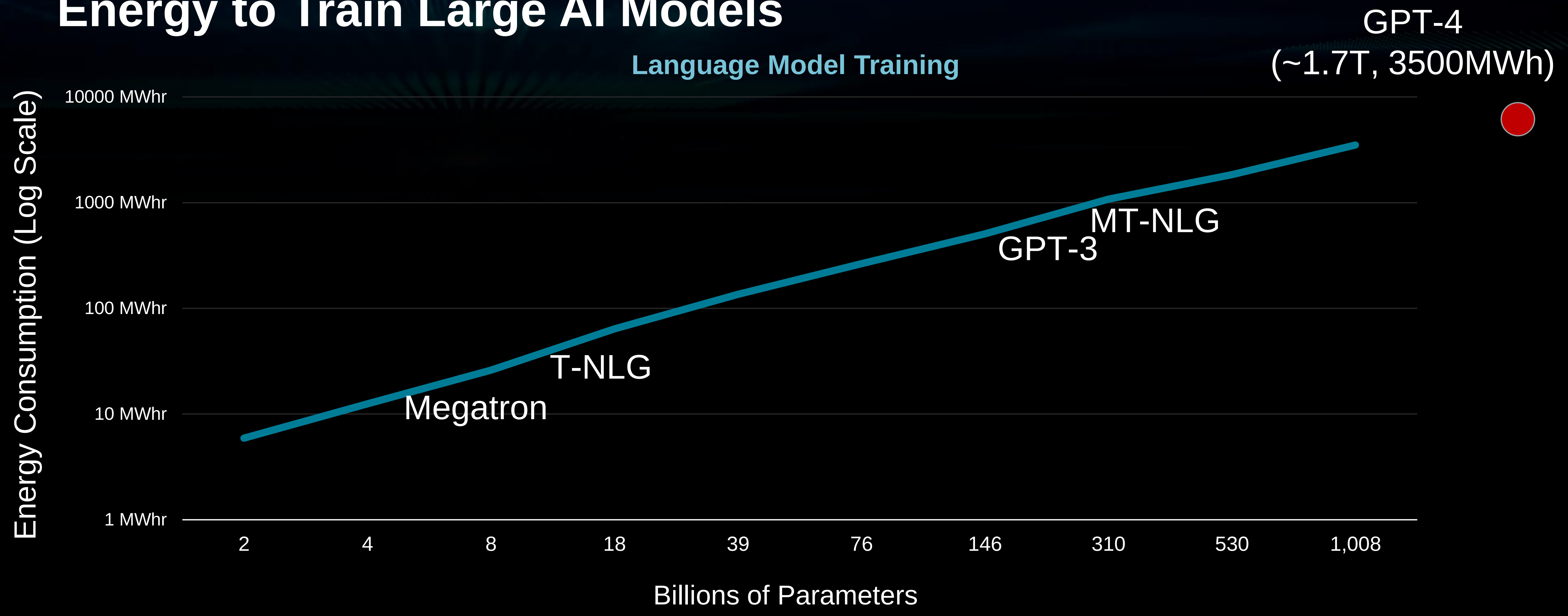
AMD
together we advance_

Compute Efficiency Trends



AMD
together we advance_

Energy to Train Large AI Models

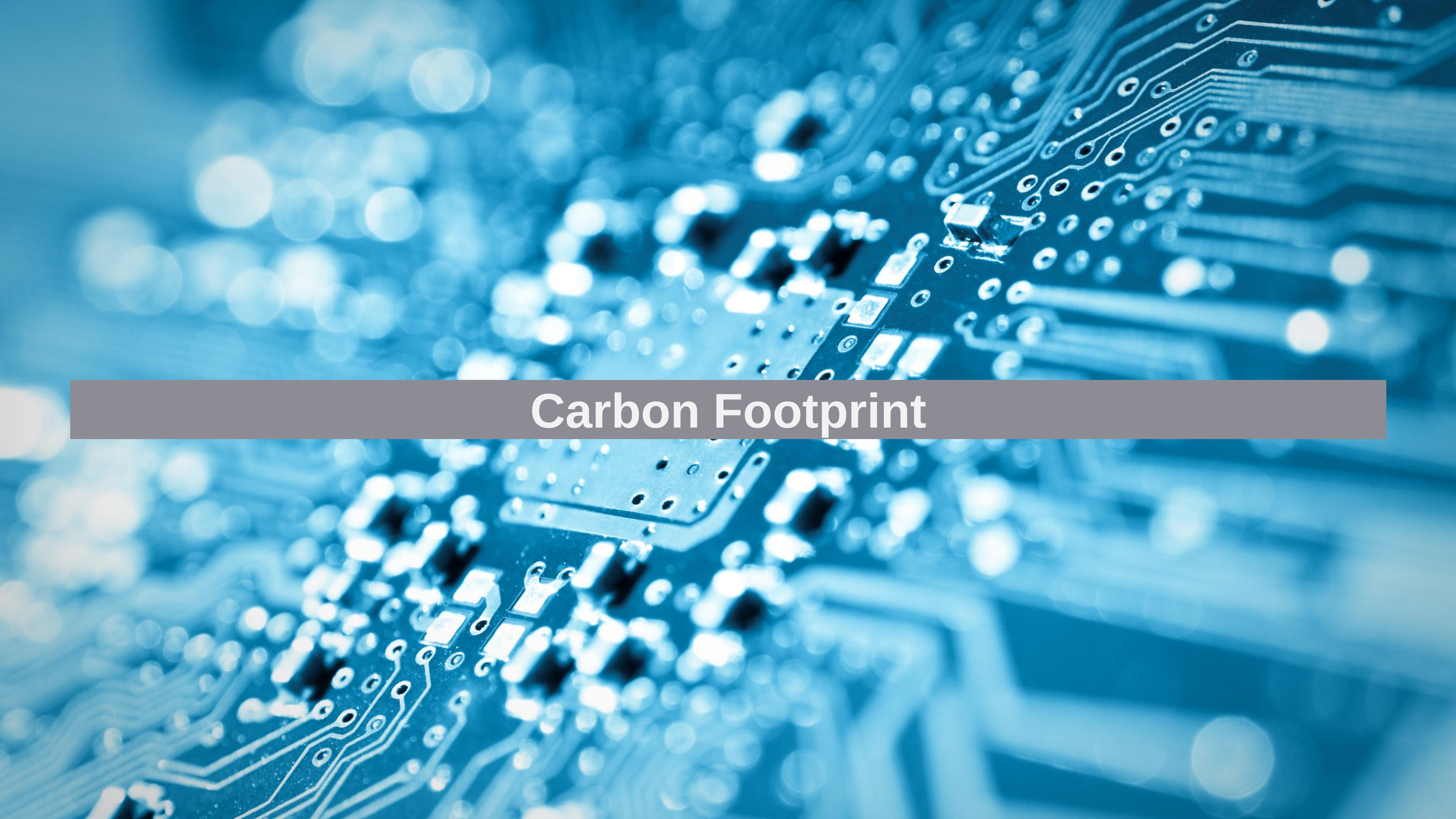


Exponentially growing model sizes drive immense growth in energy for training.

The upper bound on training requirements is yet to be determined.

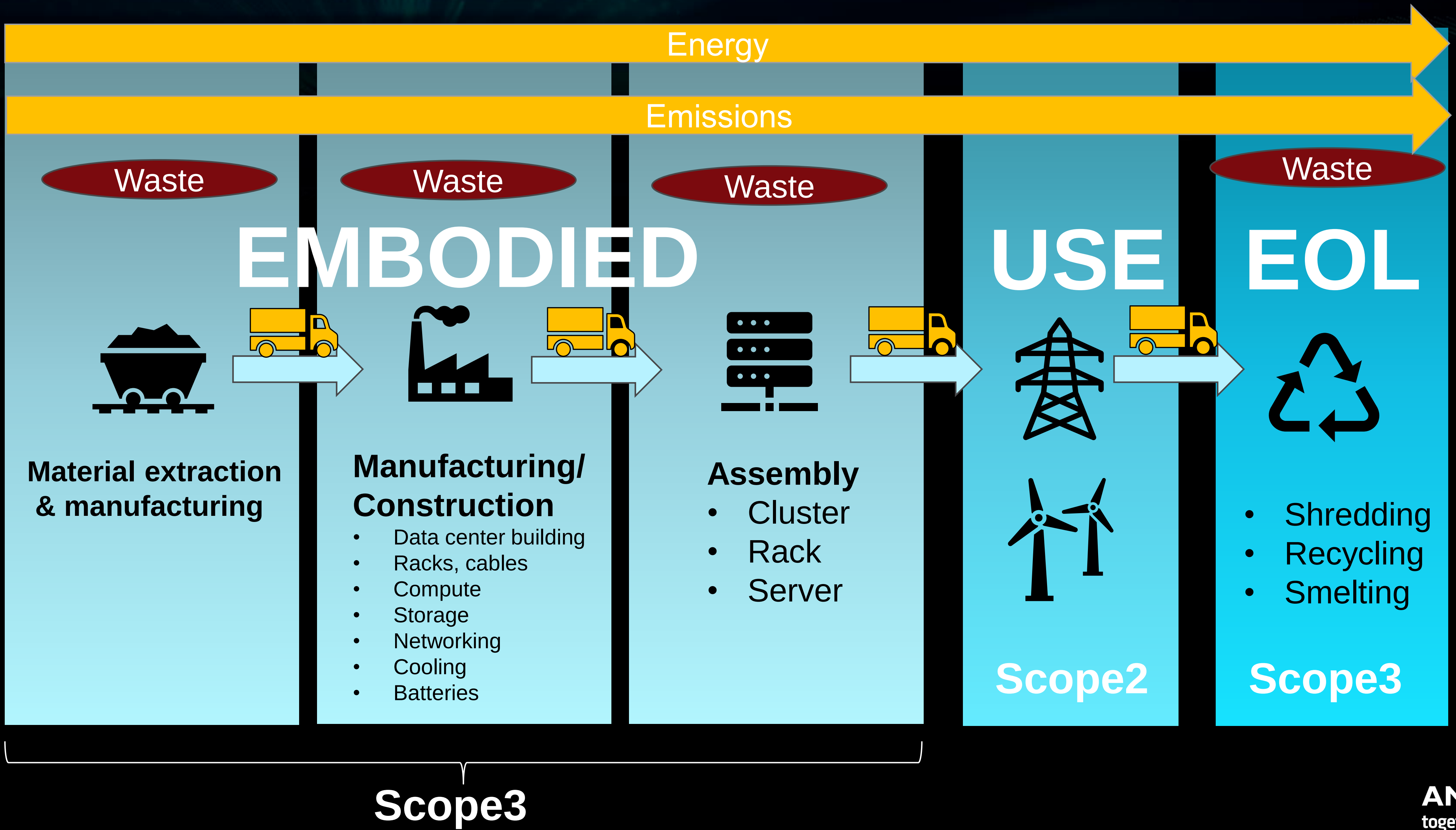
Based on published parameter counts of leading training models; and AMD internal calculations

AMD together we advance_



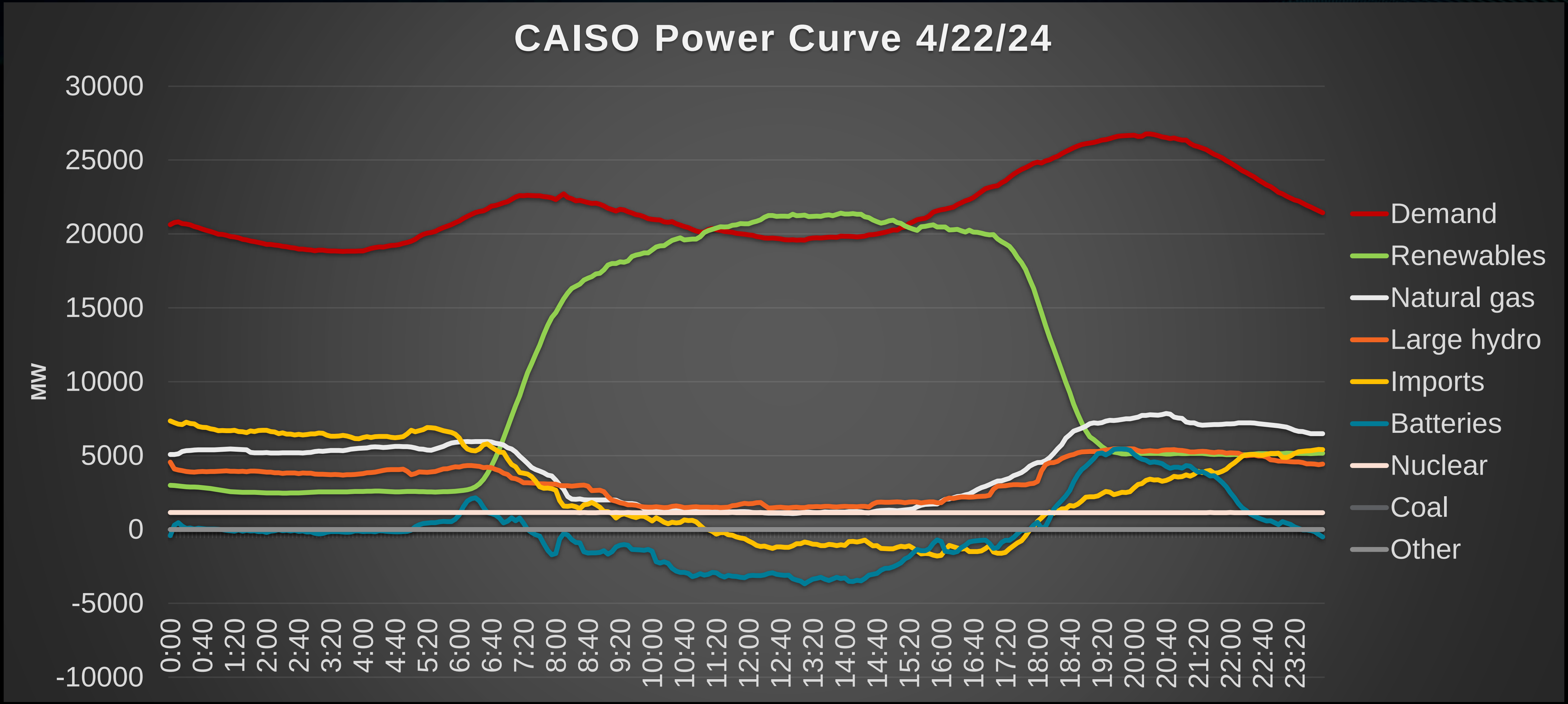
Carbon Footprint

Data Center Carbon Footprint



AMD
together we advance_

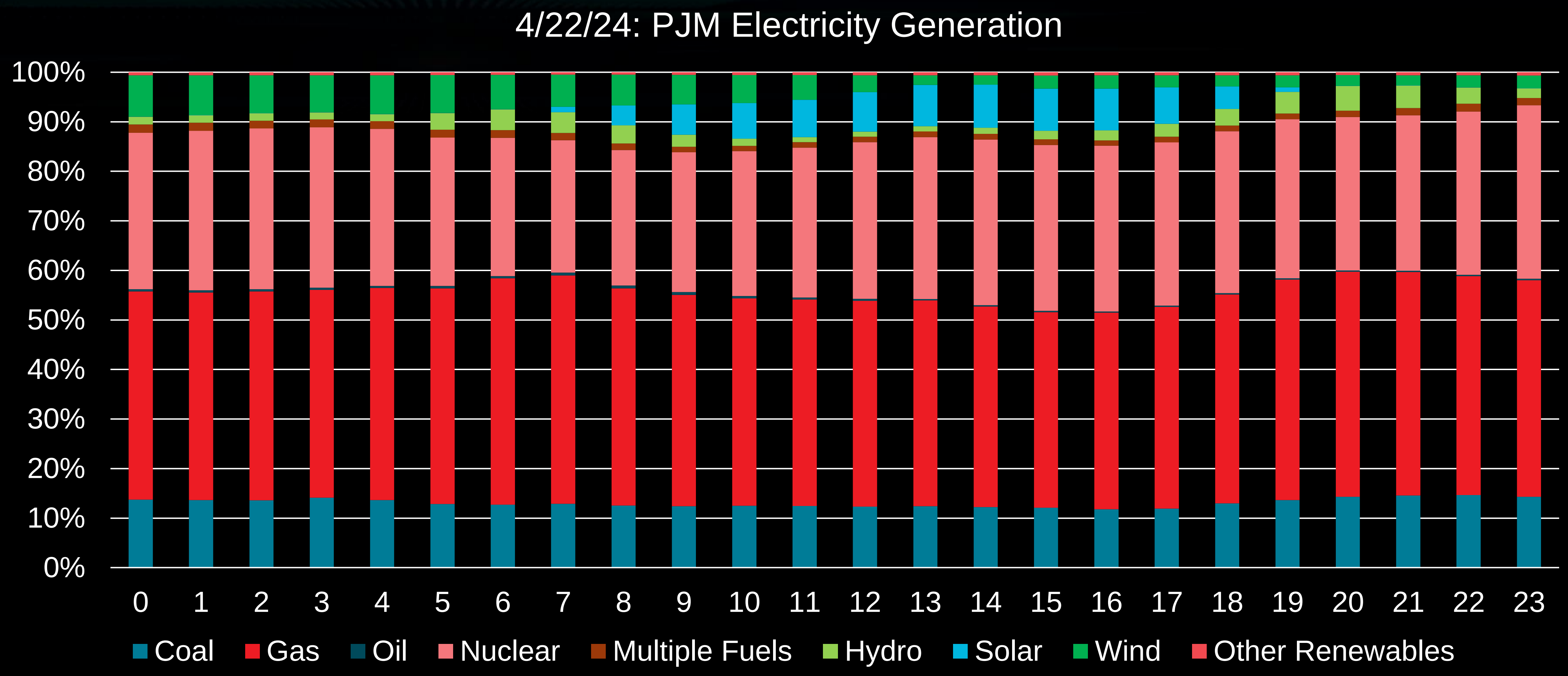
Renewable Energy in CA



Data from California Independent System Operator

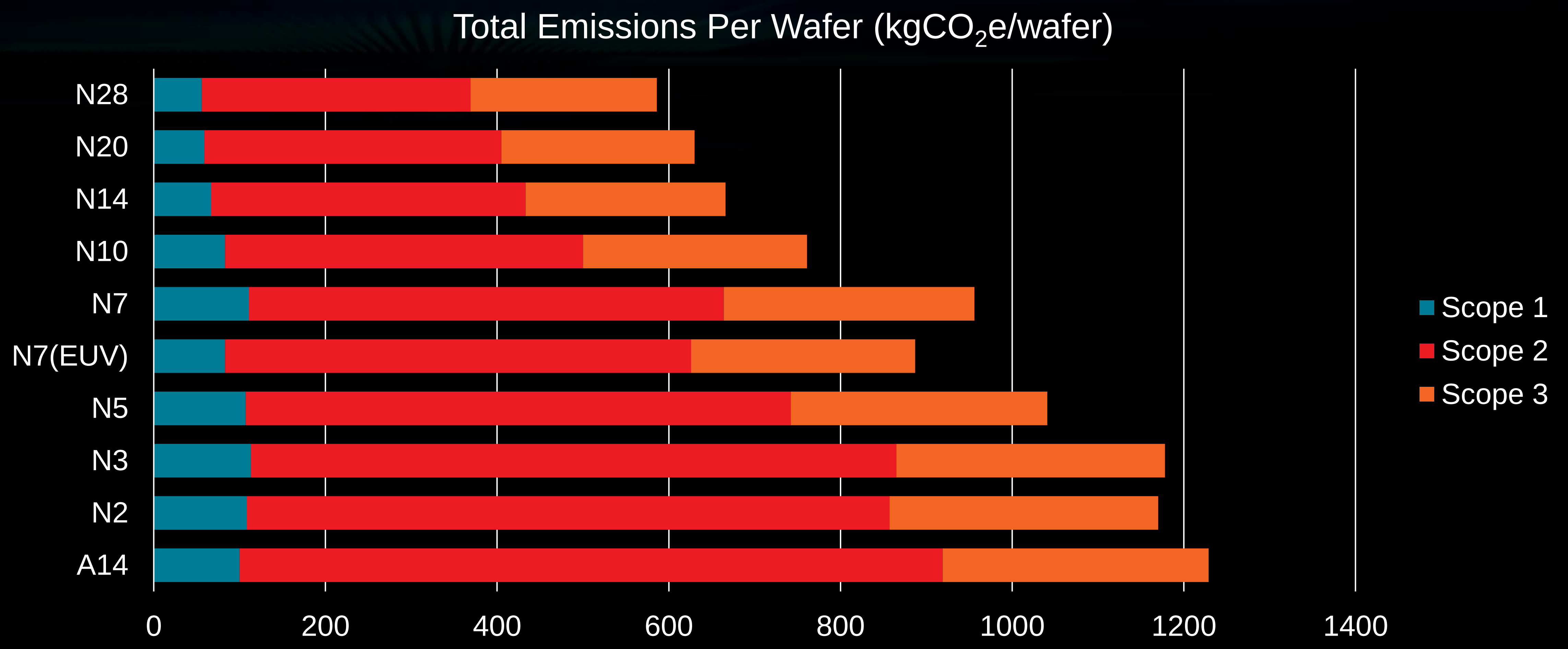
AMD together we advance_

Fuel Mix in PJM Grid



Data from PJM generation by fuel type: https://dataminer2.pjm.com/feed/gen_by_fuel

Estimated Silicon Manufacturing Emissions

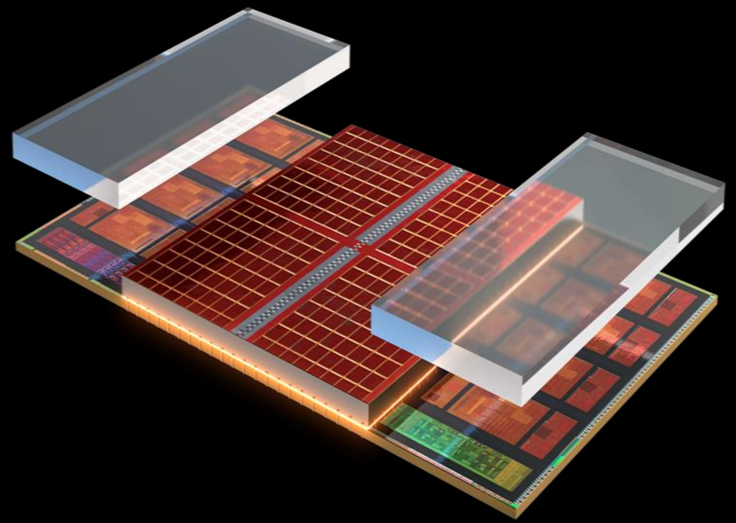


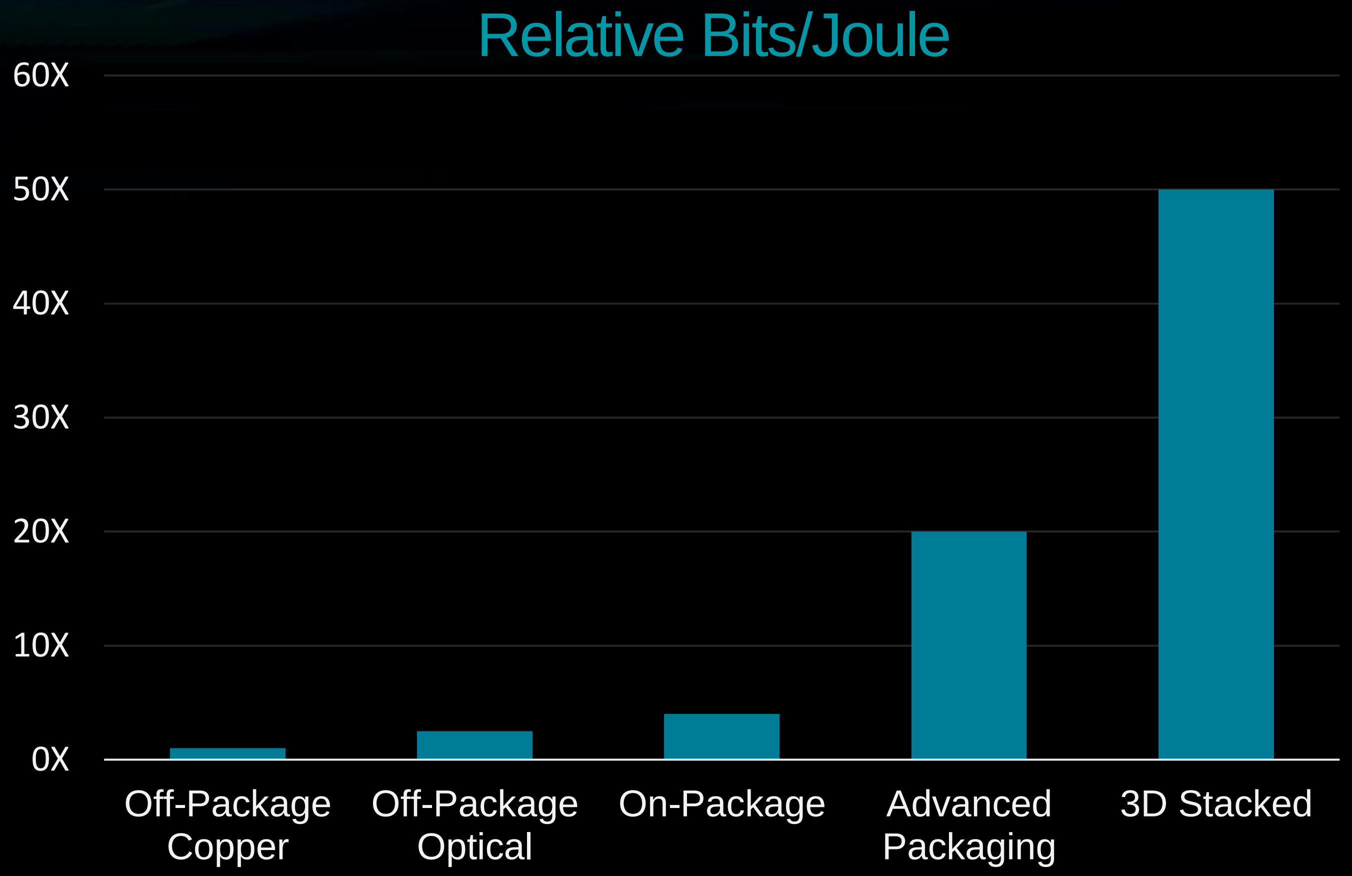
Data from “Cradle-to-gate Life Cycle Assessment of CMOS Logic Technologies”, L Boakes et al., IEDM 2023

AMD
together we advance_

Compute Efficiency Examples

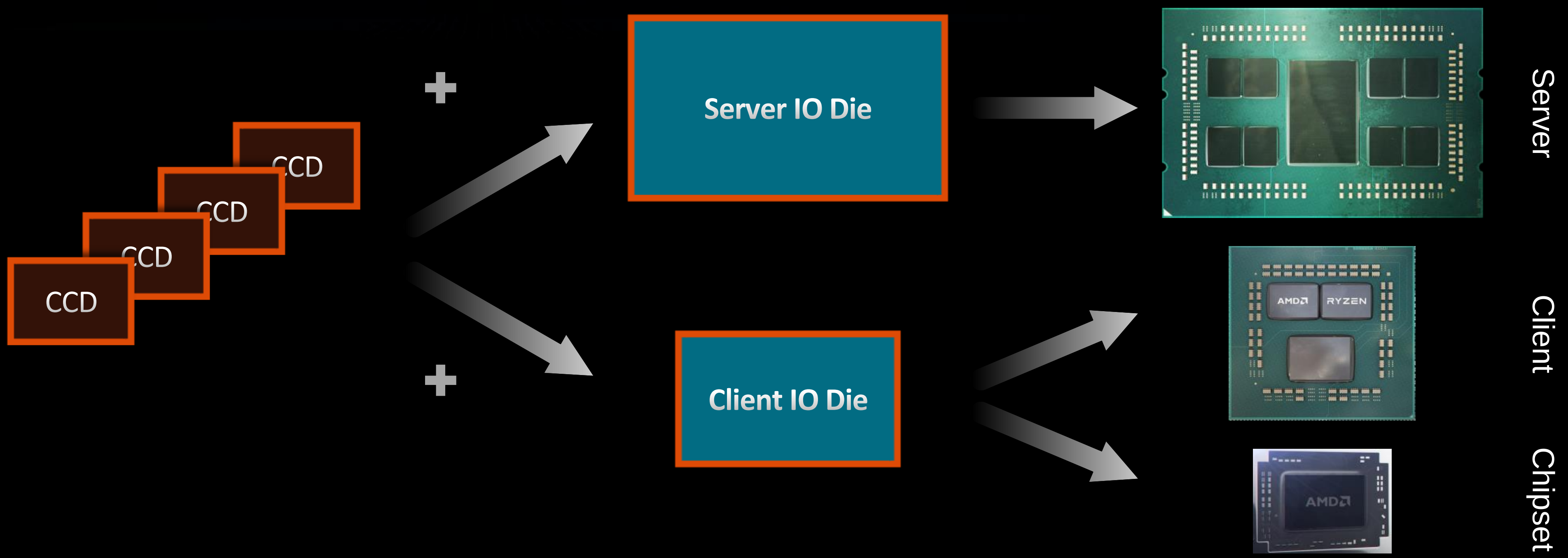
3D Chiplets and Communication Power



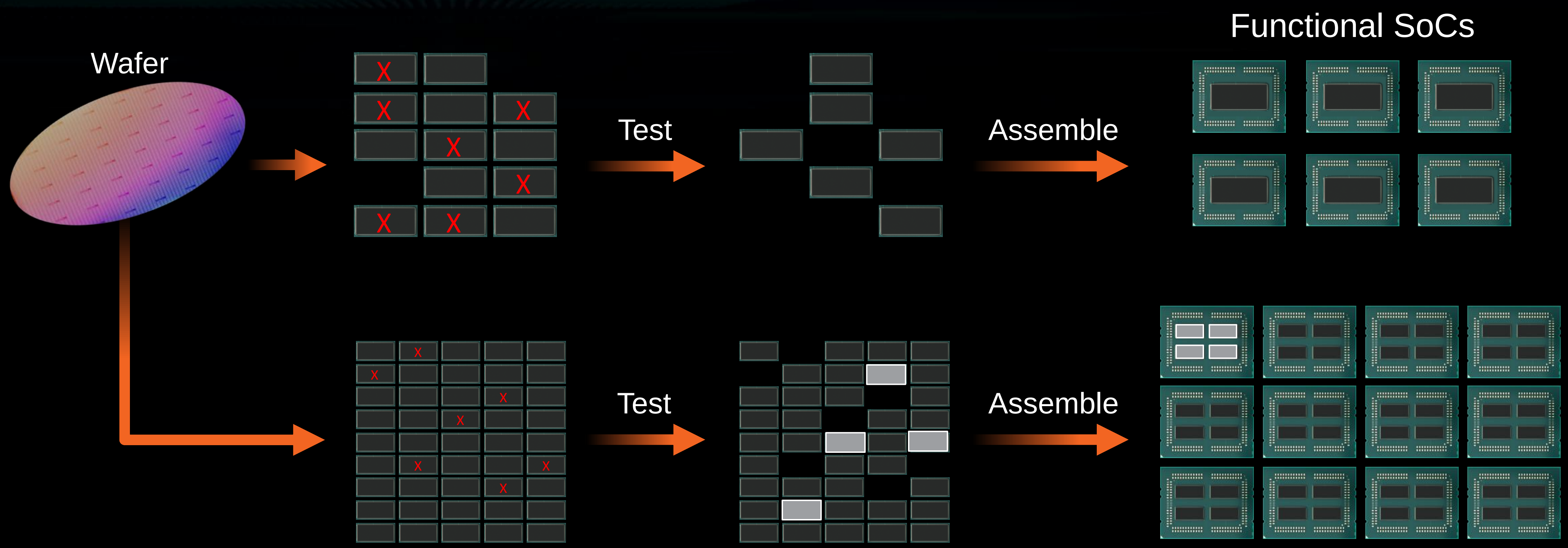


Advanced Packaging Provides up to a 50x Reduction in Communication Power

Chiplets and Reuse



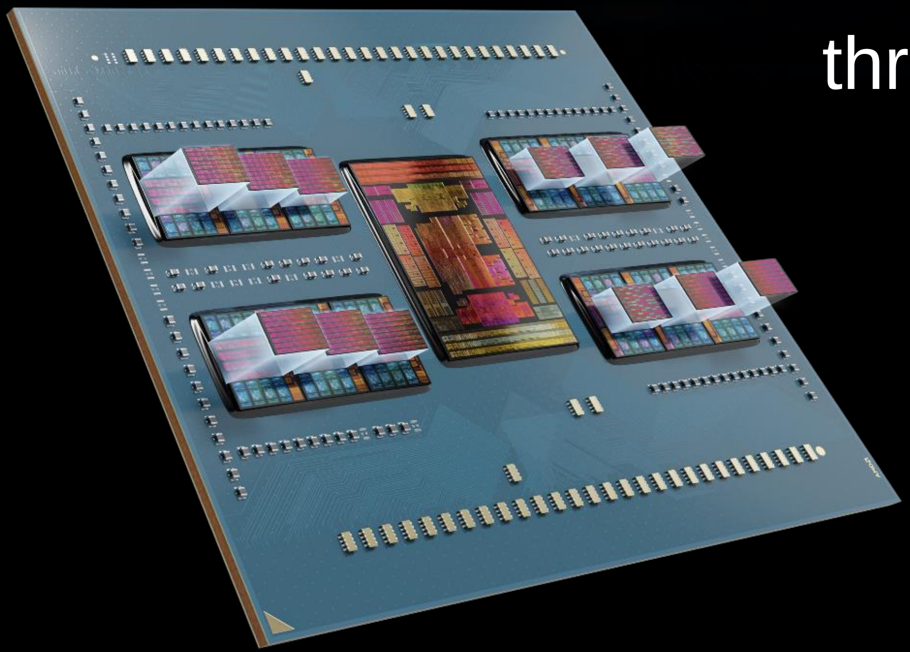
Chiplets and Yield



Many More Functional SoCs

Ability to mix and match at a finer grained level

EPYC™ Embodied Carbon Savings with Chiplets



~132,000 metric tons CO2e saved in 2023 through avoidance of wafers manufactured [1]

(Up to 12 CCD chiplets versus 1 monolithic die)

Equivalent to:

- 2.8X the operational CO2e of AMD in 2023
- 14.9 million gallons of gas [2]
- Yearly electricity for ~26,000 homes [2]

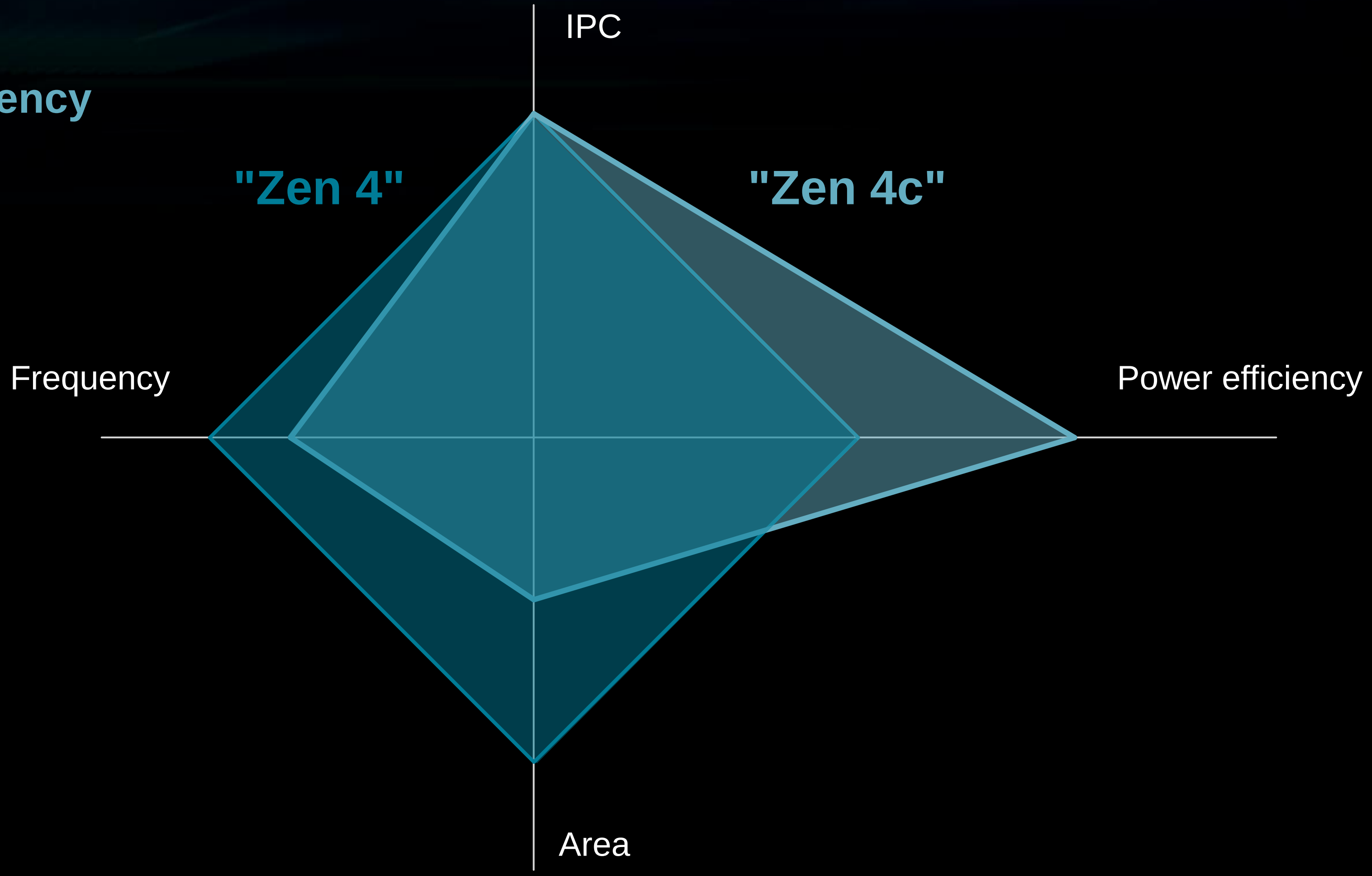
[1] AMD estimation based on defect density (defects per unit area on the wafer), chip area and n-factor (manufacturing complexity factor) to estimate the number of wafers avoided in one year. Yield = (1 + A*D0)^(-n) where A is the chip area, D0 is the defect density and n is the complexity factor. The area is known from our design, D0 is known based our manufacturing yield data, and n is a number provided by a foundry partner for a given technology. The calculations are not meant to be precise, since chip design can have a large influence on yield, but it estimates the area impact on yield. The carbon emission estimates of 132,000 mtCO2e were calculated using the estimated number of 5 nm wafers saved in one year, based on the TechInsights' Semiconductor Manufacturing Carbon Model. Comparison to AMD corporate footprint is based on AMD reported scope 1 and 2 market-based GHG emissions in 2023: 46,605 tCO2e.

[2] Data generated from the EPA Greenhouse Gas Equivalency Calculator on 7/31/2024: https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator

"Zen 4c" – a Compact "Zen 4"

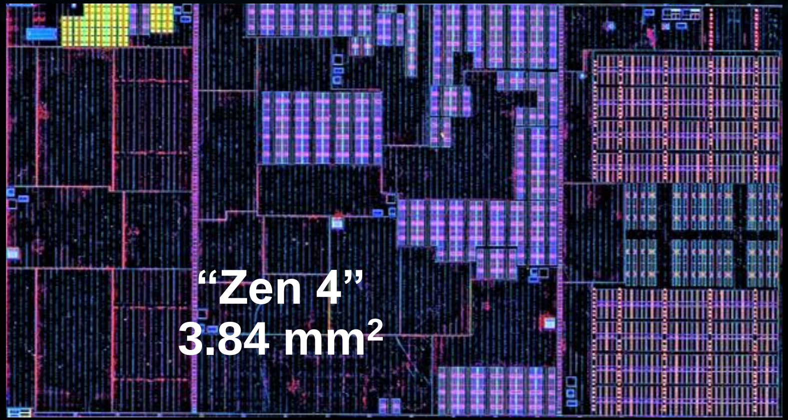
Optimized for density and power efficiency

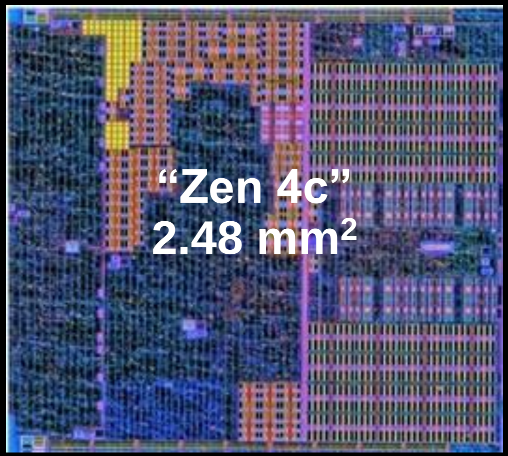
- Same IPC and features
- Lower max frequency
- Smaller area
- Increased power efficiency

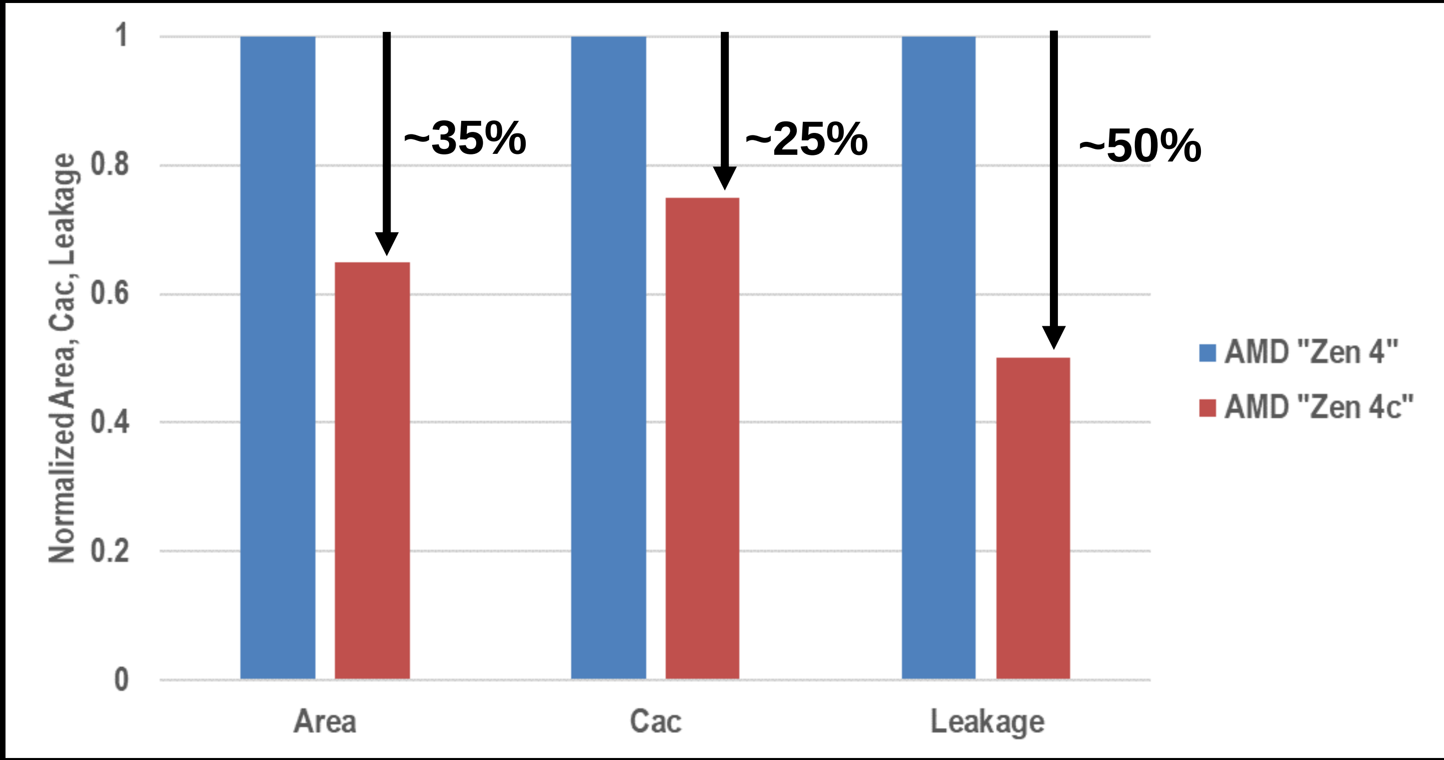


AMD
together we advance_

Benefits of Specialization[1]



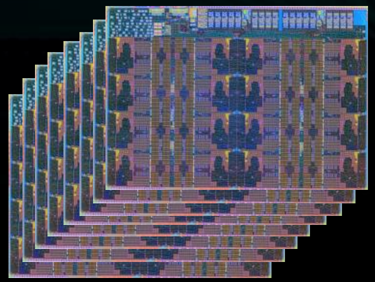




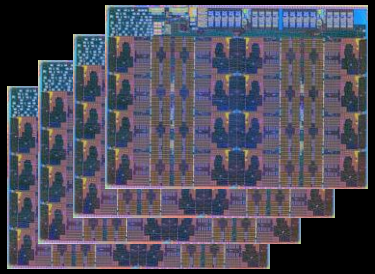
[1]Figures and data from "Zen 4c": The AMD 5nm Area-Optimized x86-64 Microprocessor Core, T. Burd et. Al., ISSCC 2024

AMD together we advance_

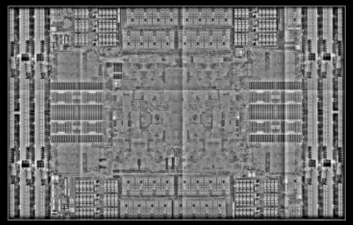
Zen4C Server Configurations



4-8 CCDs (5 nm)



2-4 CCDs (5 nm)



IO Die (6 nm)
~11B Transistors, 387 mm^2

200-400W TDP

4^{th} Gen AMD EPYC^{TM} "Bergamo"
82B Transistors

70-225W TDP

4^{th} Gen AMD EPYC^{TM} "Siena"
47B Transistors

Same IO die used in all "Zen 4" and "Zen 4c" server products

AMD together we advance_

Other Sustainable Solutions

SMT

- Increases utilization of wide pipelines

Virtualization

- Better utilization of SOC

Disaggregation

- Less fragmentation and waste through pooling of resources
- Reuse of older technology

Reconfigurable Hardware

- Multi-purpose use

Component Redundancy

- Improves yield and increases lifetime

Power Oversubscription

- Improves data center capacity

NEW CHALLENGES

The Power Problem: Transmission Issues Slow Data Center Growth Data Center Frontier

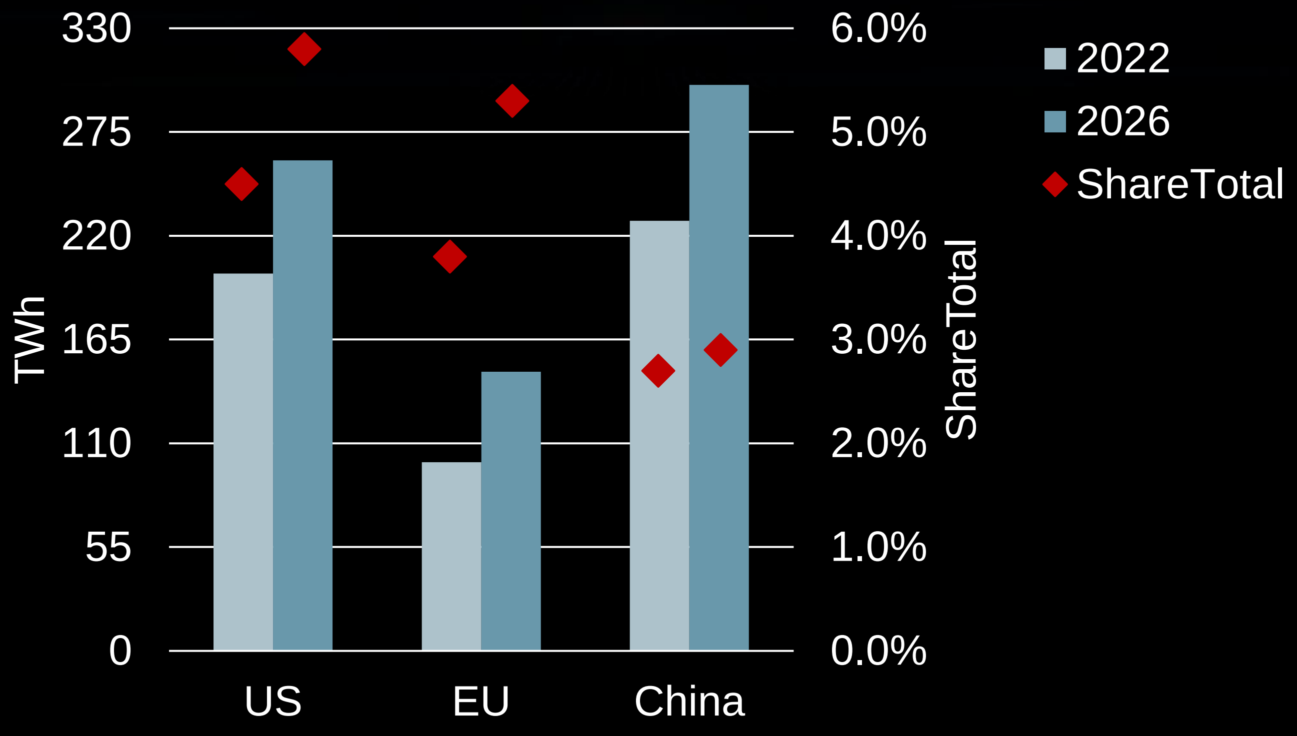
Power Shortages Are Turning More Data Centers Into Their Own Utilities Bisnow

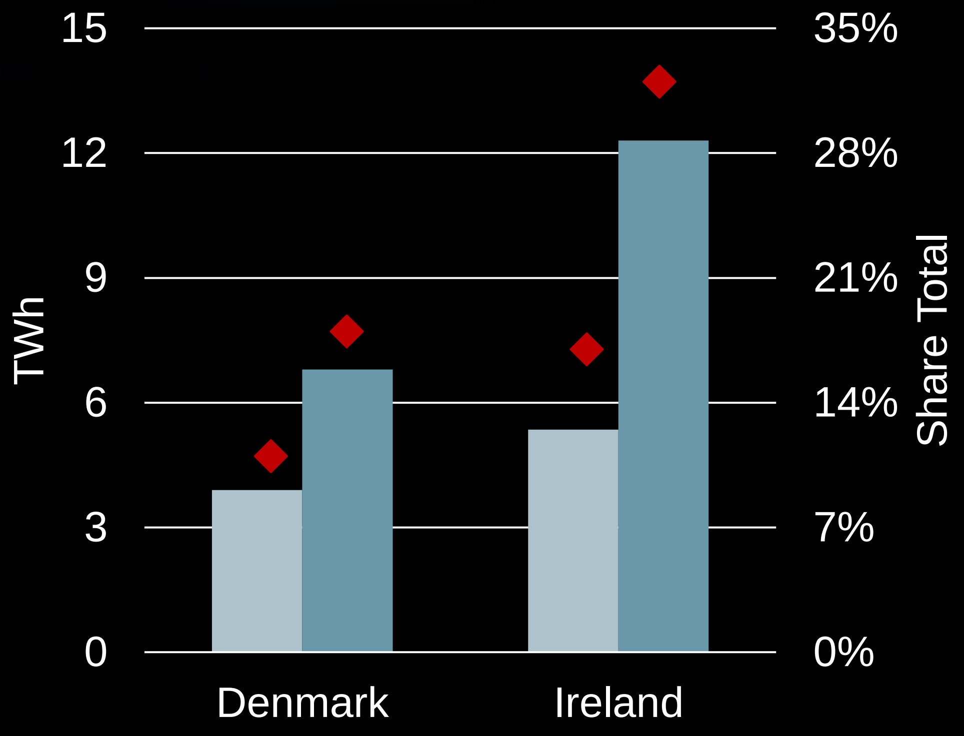
AI Is Changing the Future of Energy. What to Know. Barrons

Amid record high energy demand, America is running out of electricity Washington Post

Amazon goes nuclear, acquires atomic datacenter for $650M The Register

Data Center Demand Projections





Data source: IEA (2024), Electricity 2024, IEA, Paris https://www.iea.org/reports/electricity-2024, License: CC BY 4.0

AMD
together we advance_

Takeaways

DESIGN FOR EFFICIENCY

DESIGN FOR DISRUPTION

DESIGN FOR ALL

Disclaimer

The information contained herein is for informational purposes only, and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD's products are as set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale. GD-18

AMD