



Sustaining sustainability in a power hungry (HPC) world

Kirk W. Cameron, Ph.D.

Computer Science

Virginia Tech

“It's tough to make predictions, especially about the future.”

Yogi Berra



“If you don't know where you've come from, you don't know where you're going.”

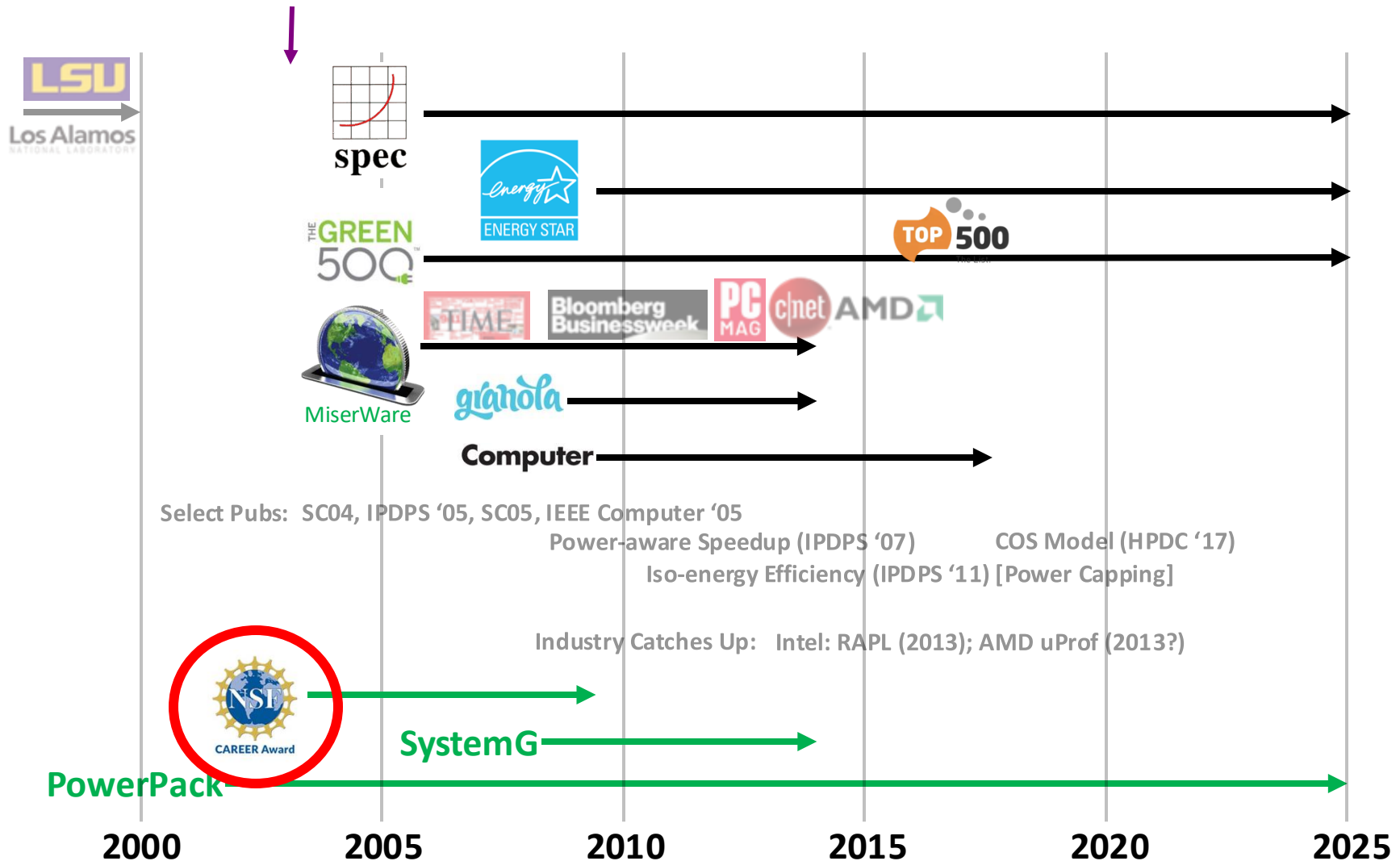
Maya Angelou, 1928-2014

Professors Privilege

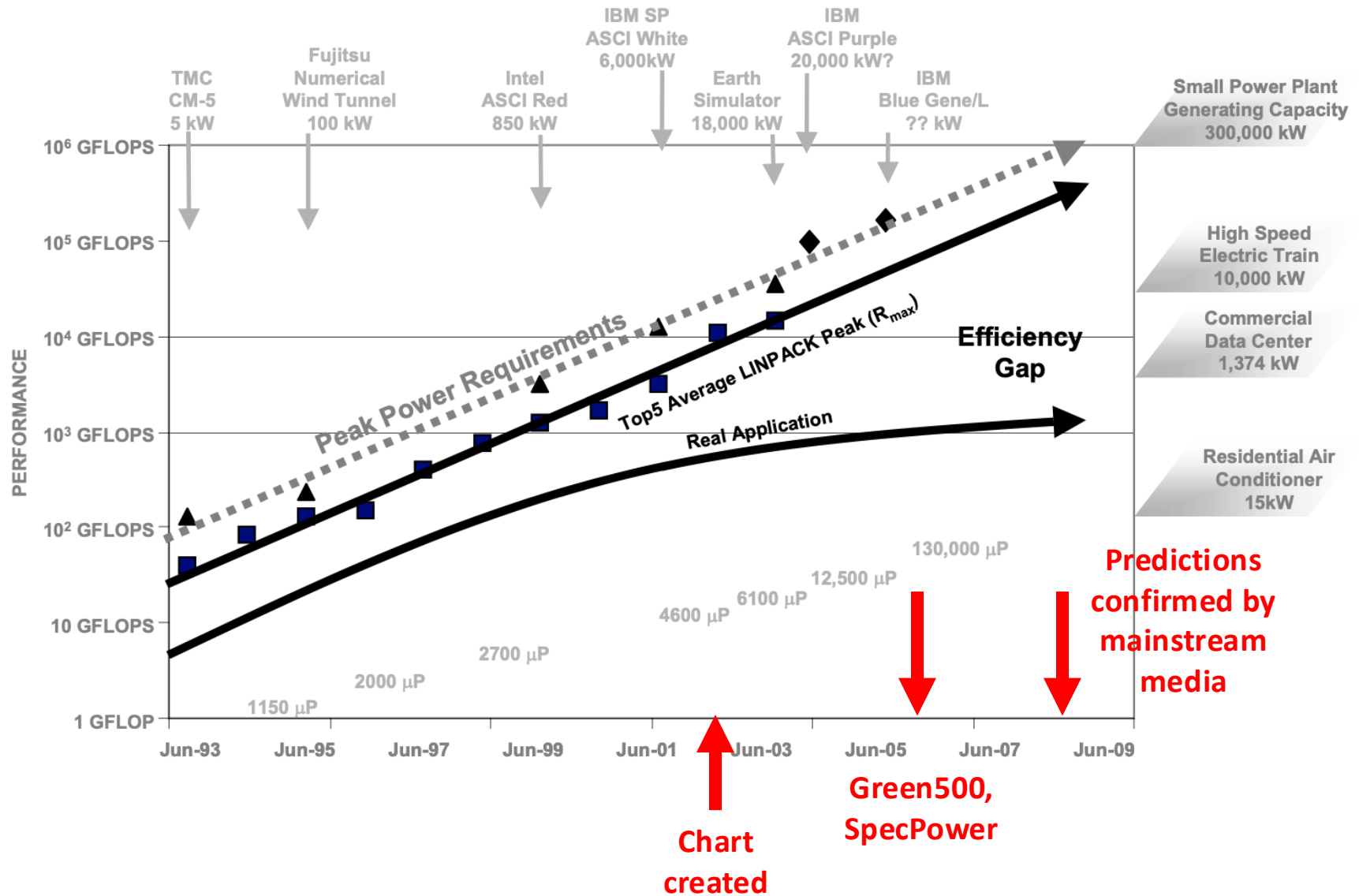


“People don’t invite me to these things for quiet platitudes. I speak my mind. One of the benefits of being a professor. I encourage you to disagree and I hope it pushes you in the RIGHT direction.” -me

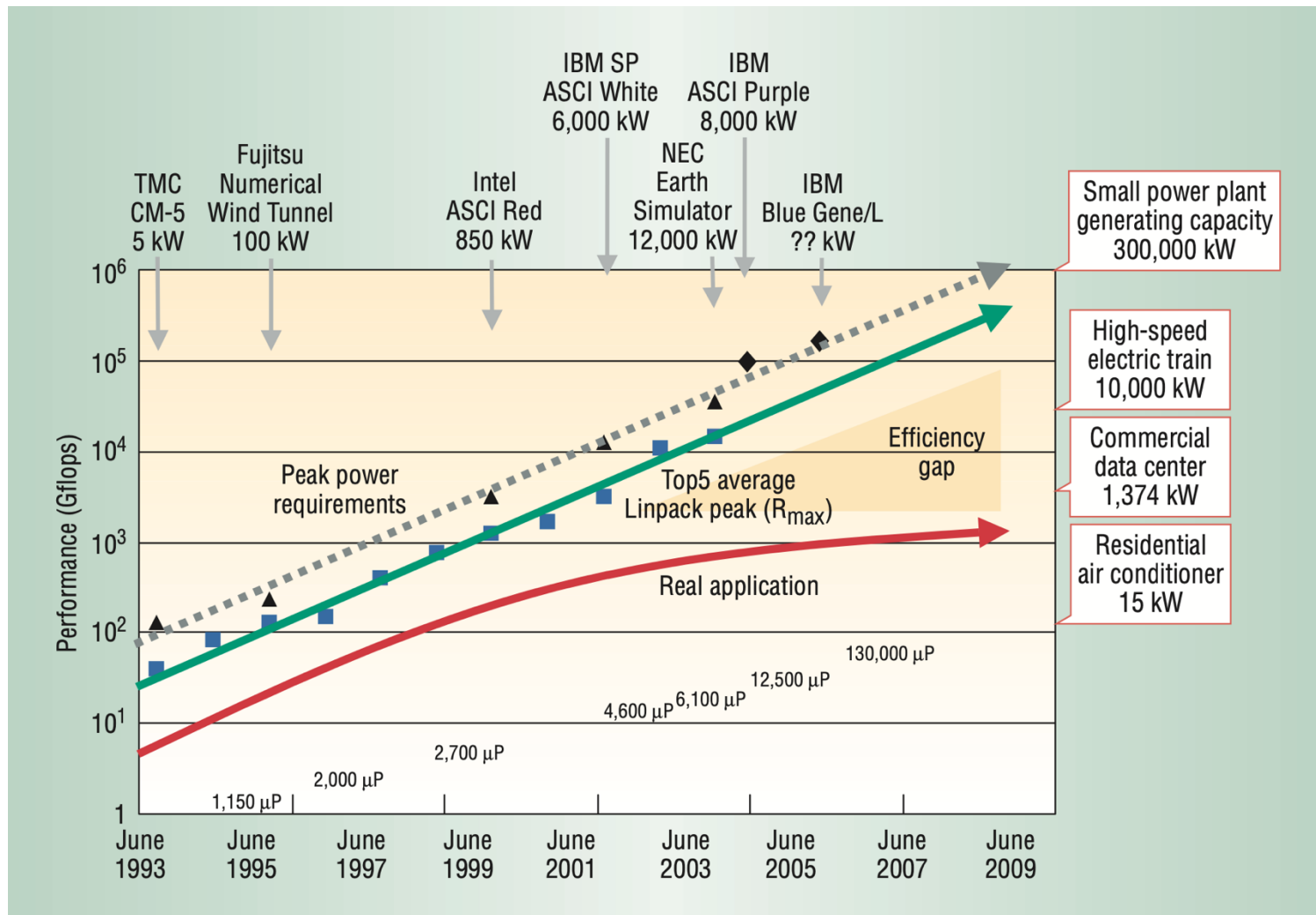
My Green HPC Timeline



Data-driven Predictions (circa 2002)

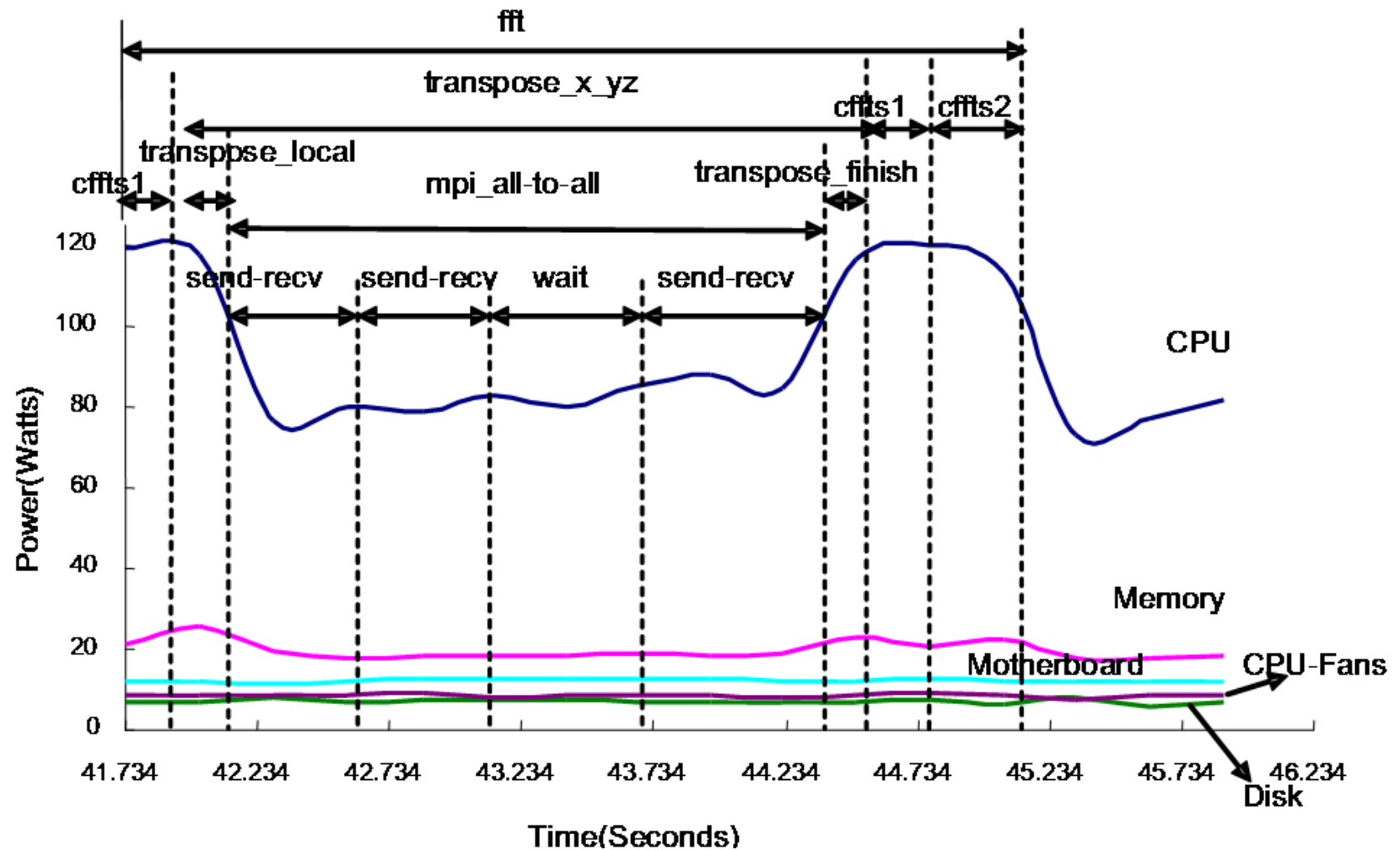


For reference...



Kirk W. Cameron, Rong Ge, Xizhou Feng: High-Performance, Power-Aware Distributed Computing for Scientific Applications. Computer 38(11): 40-47 (2005)

PowerPack (System level power)



Rong Ge, Xizhou Feng, Shuaiwen Song, Hung-Ching Chang, Dong Li, Kirk W. Cameron:
PowerPack: Energy Profiling and Analysis of High-Performance Systems and Applications. IEEE
Trans. Parallel Distributed Syst. 21(5): 658-671 (2010)

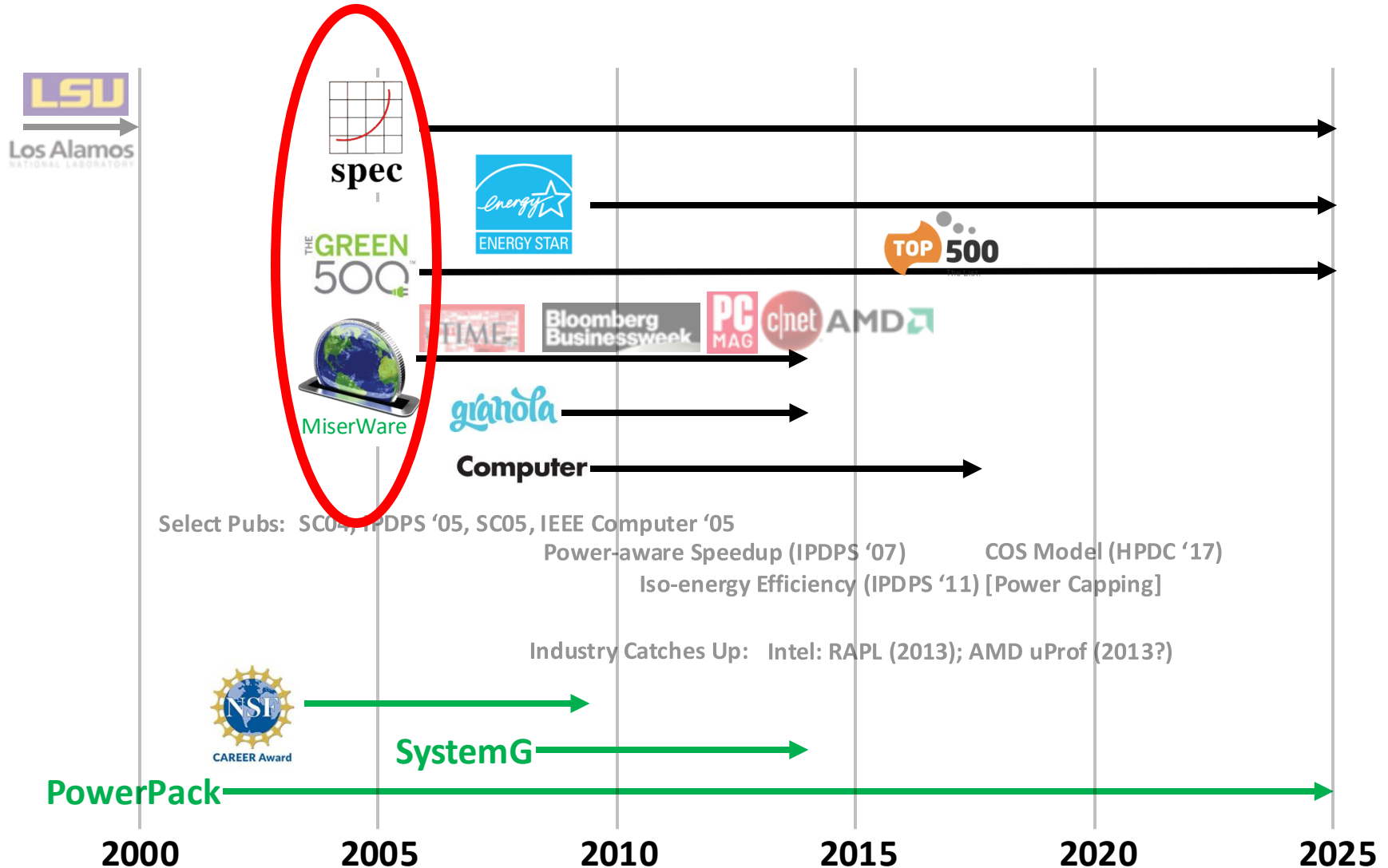
ASCR EE Thought Experiment

- Apply same principles from 2002 to predictions using current data
 - Data-driven predictions
 - Market-driven prediction and analyses
- Find a path forward...

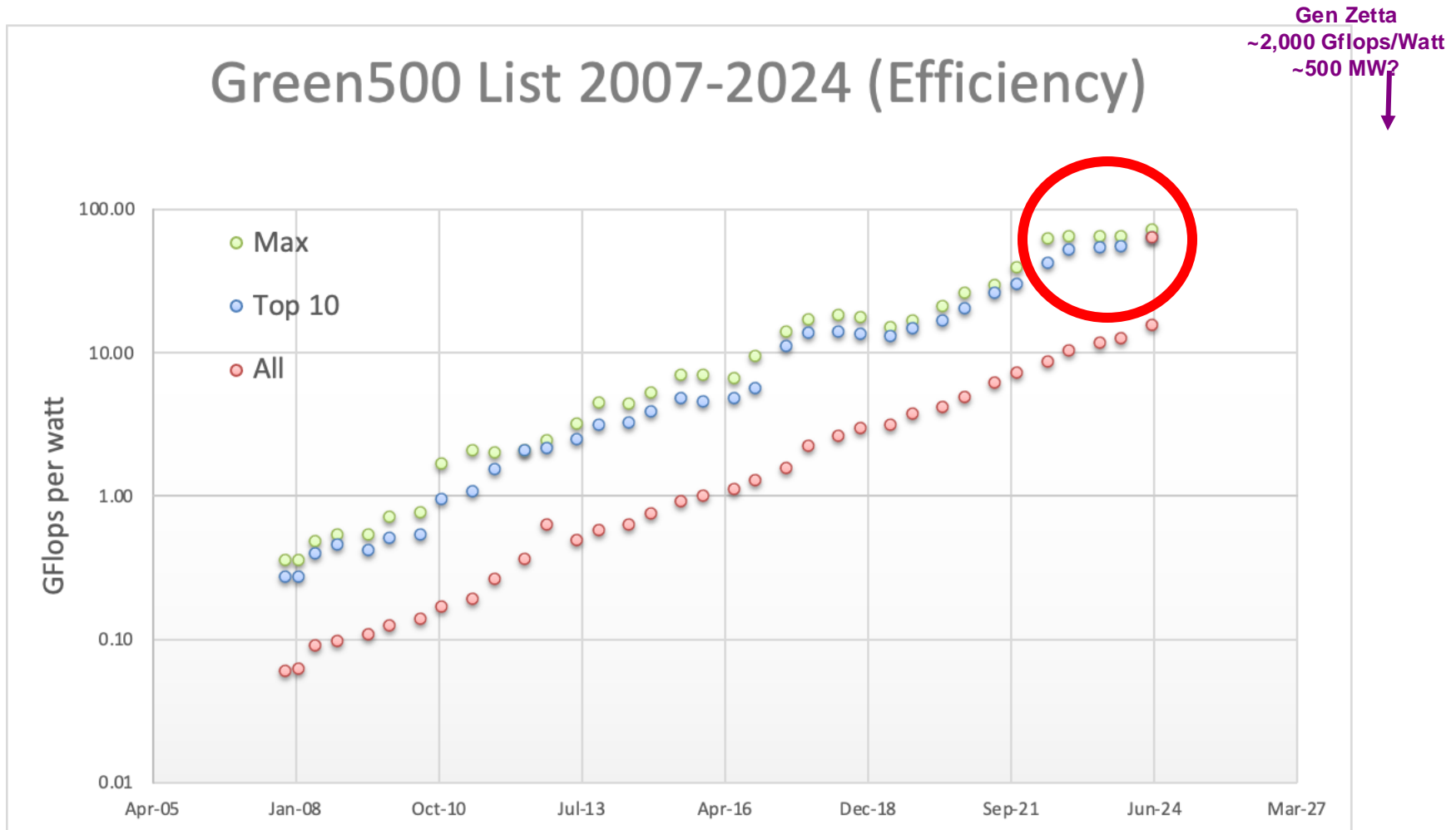
Data-driven Predictions (circa 2002)

- Perf efficiency in SC gets worse with scale
- \$700k/MW/year IS disruptive to HPC
- Laptops outselling PC's
 - Laptop features will migrate to servers
 - Power-aware computing will increase HPC efficiency
- High-performance, power-aware computing
 - Title of my 2004 NSF Career Award

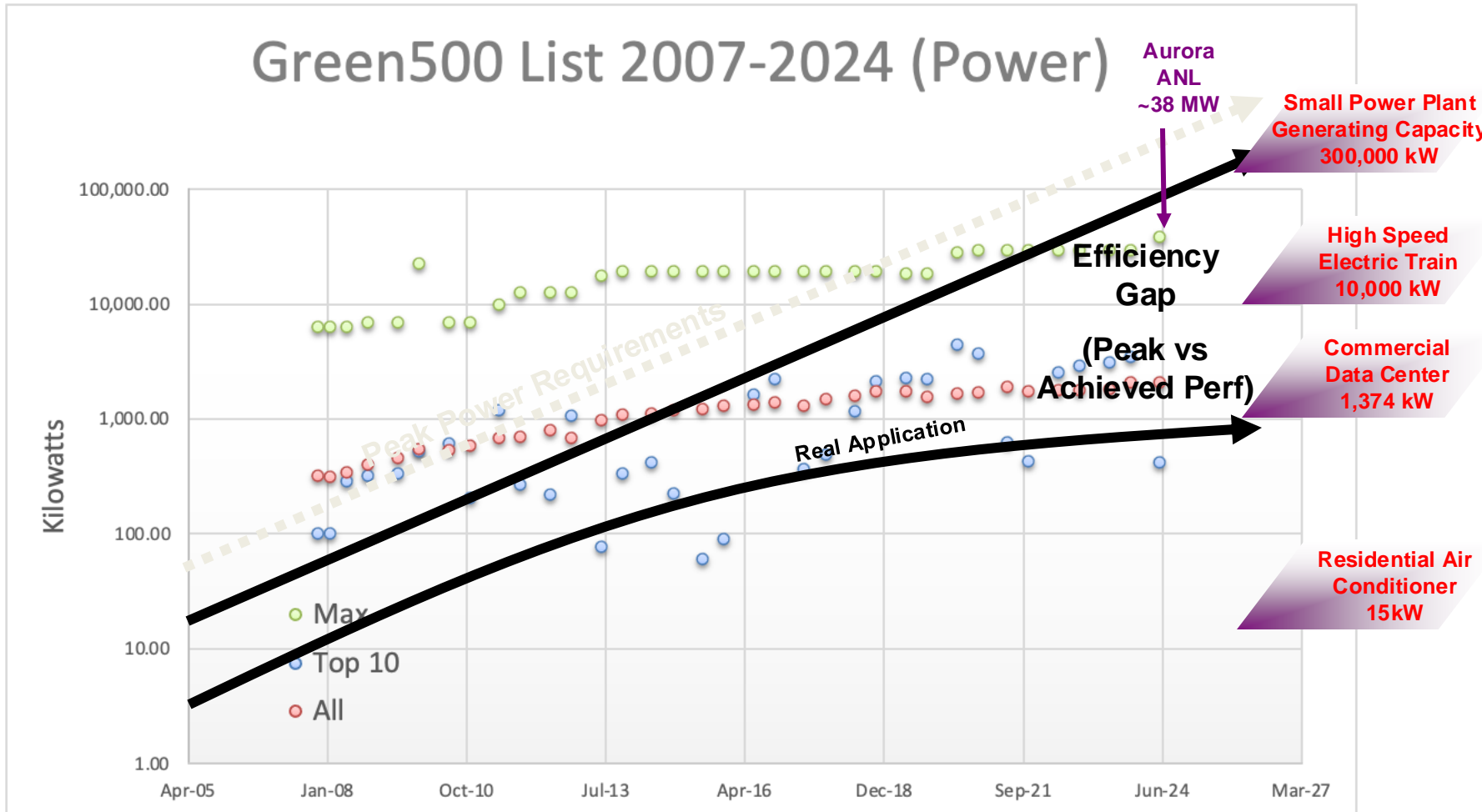
My Green HPC Timeline



Energy Efficiency (June 2024)



Power Consumption (June 2024)



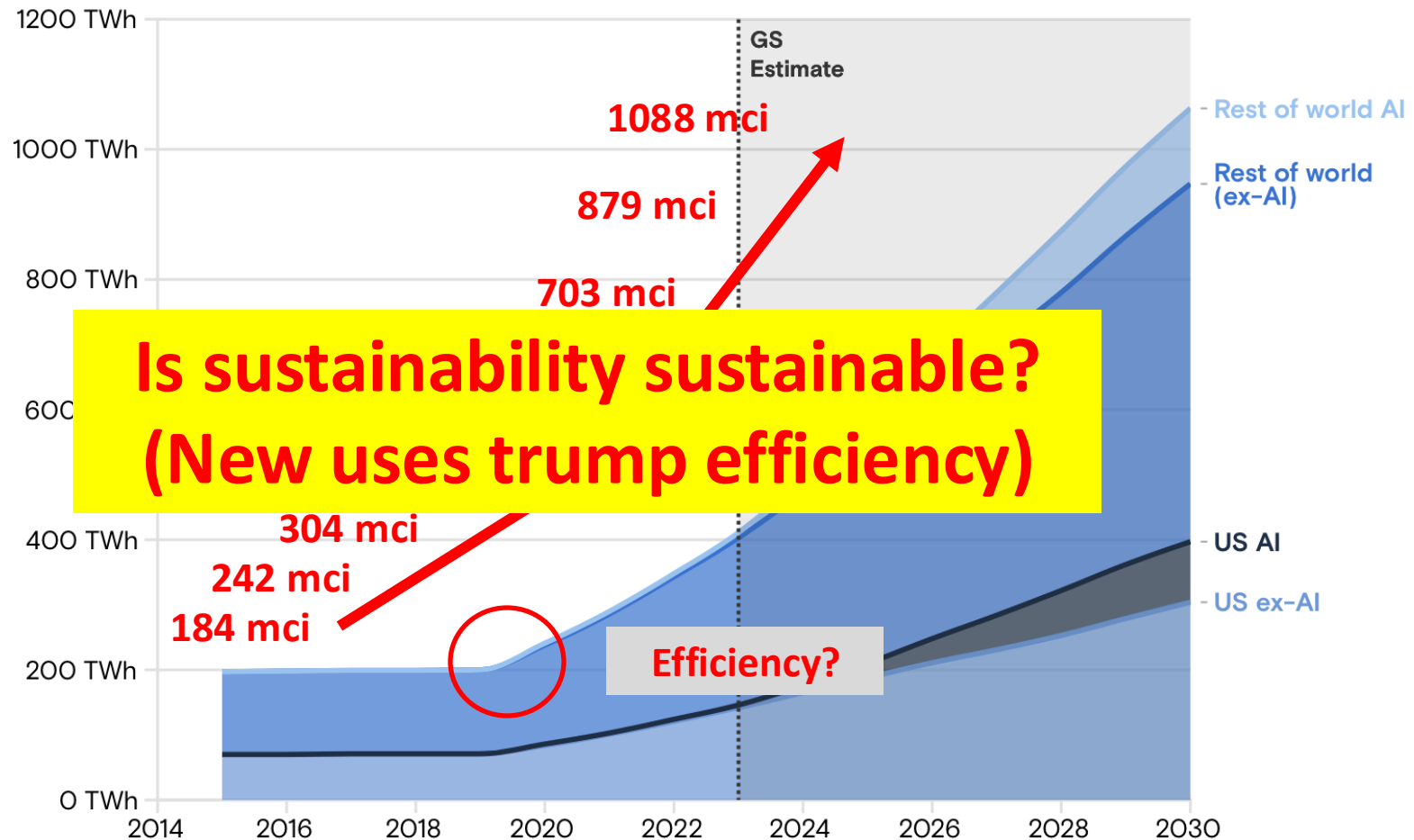
At full throttle, at about 8 cents/kWh, Aurora could cost > \$20M/year to operate.

2024 Market Predictions

- Consider trending data (HW, SW, eff)
- Consider market shifts
 - Must acknowledge HPC not a driver
 - HPC can “innovate”; be an “early adopter”
 - Mainstream apps drive market adoption
 - CAGR and market penetration ARE indicators
 - Which markets matter to HPC?
 - How will they drive technology?

The AI Power Surge (DC Market)

Data center power demand



Source: Masanet et al. (2020), Cisco, IEA, Goldman Sachs Research

mci = million compute instances

NVIDIA (GPU) Market Adoption

1. Consider how long it takes to adapt/refactor and the risks of adoption for a disruptive technology.



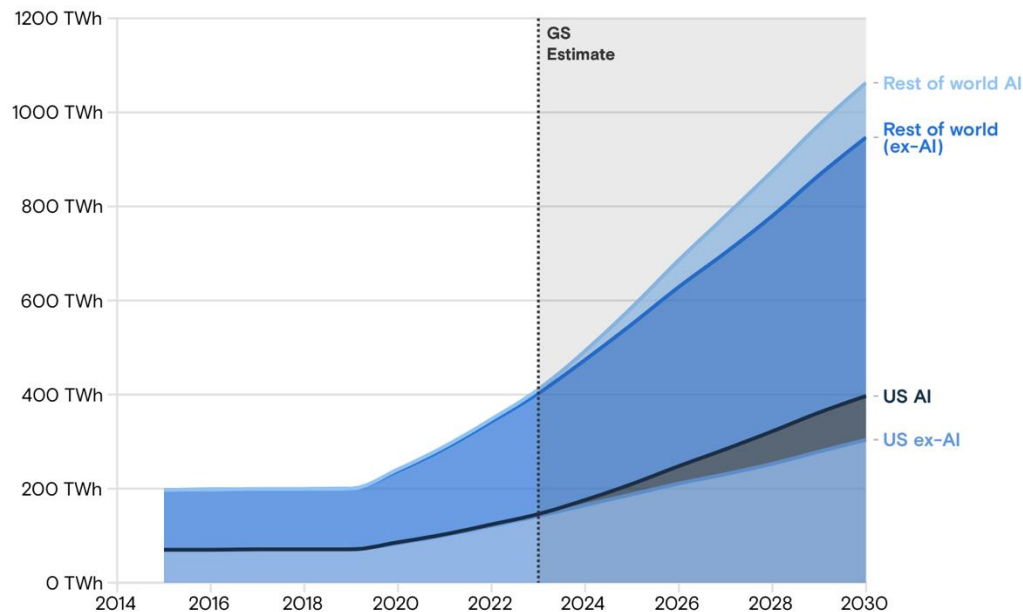
2. Consider which market (not HPC) drives these changes.

G. A. Moore. 1991. "Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers". HarperCollins Publishers, New York, NY.

And don't forget about Cloud...

- ~20% CAGR thru 2030!
- Has crossed chasm. (AWS, DC growth)

Data center power demand



Source: Masanet et al. (2020), Cisco, IEA, Goldman Sachs Research

Goldman
Sachs

Remember: Vendor tech will be driven by large markets! And the market is NOT HPC.

The next 20 years...



“The most important technologies of the next 20 years have not been invented yet.”

Kevin Kelly, Wired

Thought Experiment Takeaways

- Perf efficiency in SC gets worse with scale
- \$700k/MW/year IS disruptive to HPC
- ~~Laptops outselling PC's~~
 - ~~Laptop features will migrate to servers~~
 - ~~Power-aware computing will increase HPC efficiency~~
- High-performance, power-aware computing (relevant but siloed)

Thought Experiment Takeaways

- Existing tech are (again) not sustainable
 - Not all bad; more science now possible
- Markets (not HPC) drive new technologies
 - Except crisis: moon shot (\$250B / 12 years)
- Cost is ALWAYS a driver
 - Environmental impact seldom drives spend
 - Bill Gates (SC05)
- AI is going to eat our lunch
 - Upside: help the science
 - Upside: help the system that does the science
 - Downsides: training, cost, **understanding**
- The **next disruptive tech** is not known and probably has not been invented yet

Disruptive Tech

IF it requires refactoring, two things needed:

1) significant improvements that won't be out paced by advances in SOA

2) wide-spread adoption beyond HPC

→ aka economy of scale

Moore? Amdahl? Nope. Von Neumann (Data).

- Quantum: Good idea, Market bubble
- Neuromorphic: Good idea, Shaky market too
- Other HW: Good ideas, See above

What should we be doing?

- Adoption in HPC takes way too long
 - We (HPC) are the opposite of Agile; can we change?
- AI is doing pretty well on its own
 - The chasm is real: AI-company funerals TBD
 - What if the AI is wrong? Siri anyone? ChatGPT?
 - AI is coming, how can we best use it?
 - Science? Apps? Systems? Elsewhere?
- Holistic power measurement and management (Help from AI?)

Understanding

<shameless plug>

- Cost models for efficiency is critical to UNDERSTANDING what systems do
 - Power-aware Speedup
 - Is-energy efficiency
 - Compute-Overlap-Stall Model
- System-level power measurement critical
 - PowerPack; lots of others
- Role of AI is...complicated (runtime vs decision speed)
- Benchmarks need sustained funding!

Parting thoughts

- Short term game: efficiency and cost
- Long term game: environment
 - Optics of DOE vs environment...yikes
 - Demand side? Cradle to grave?
- I don't like the sound of "lots of accelerators"
 - There has to be a better way, doesn't there?
- We need to embrace cloud more (CAGR)
- From 15-20 years ago, incumbents?
 - Almost the entire stack and most apps
 - Ask why?
 - Need rewards for embracing change, being agile

Thank You!



Kirk W. Cameron, Ph.D.
Computer Science
Virginia Tech
cameron@vt.edu