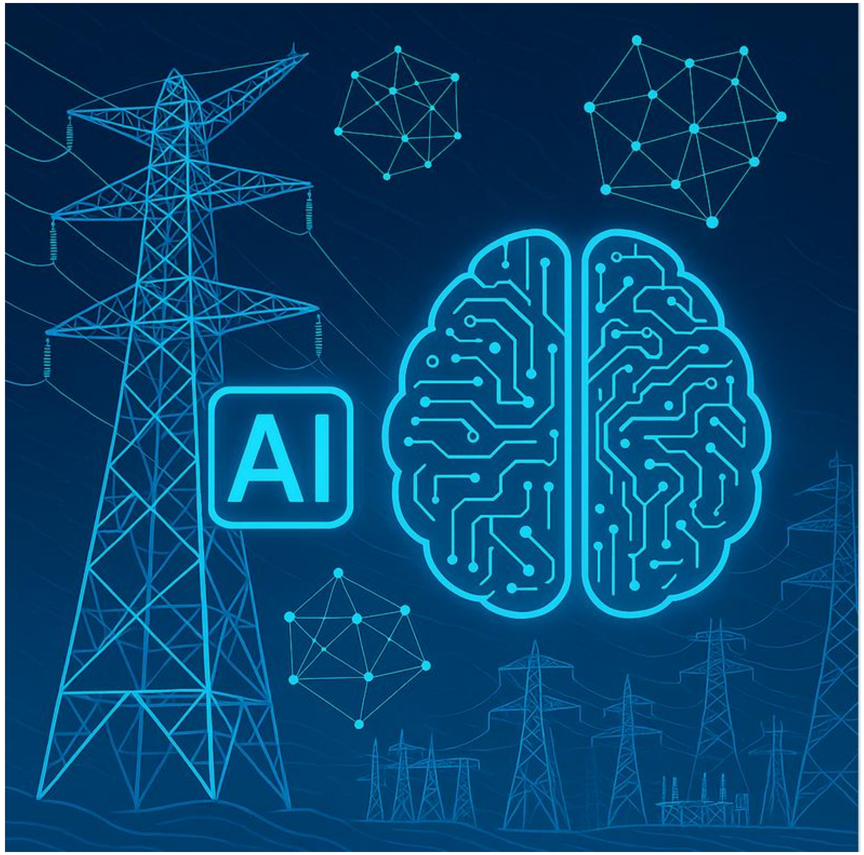
The AI/DC Energy Dilemma

Data Center Loads and Policy Issues for The Electric Utility Industry



Panel 3a- Fairhaven Room

Massoud Jourabchi

GREENWAY RESEARCH

Some Definitions

Types of Data Centers* (*excluding crypto-mining*)

- Embedded Data Centers: server closet, server room
- *Colocation Data Centers: Real Estate sites with data center services, typically in urban setting*
- *Hyperscale Data Centers: Google, AWS, Apple…, typically in rural setting*
- Edge Computing: Distributed data processing close to point of use

- **IP traffic: the flow of data across the internet. IP Traffic is also commonly referred to as web traffic.**
- **Moore Law: observation that the number of transistors on computer chips doubles approximately every two years (hardware improvements).**
- **Koomey Law: observation that the number of computations per joule of energy doubled about every 1.57 years. (software improvements)**
- **PUE : Ratio of electricity use for data services to total facility electricity consumption**

- EB: Exabyte is equal to 10^18 bytes or one billion gigabytes (GB)
- ZB: Zettabyte is equal 10^21 bytes

More definitions: Types of AI

There are a few types of AI as defined by Arend Hintze, researcher and professor of integrative biology at Michigan State University [1].

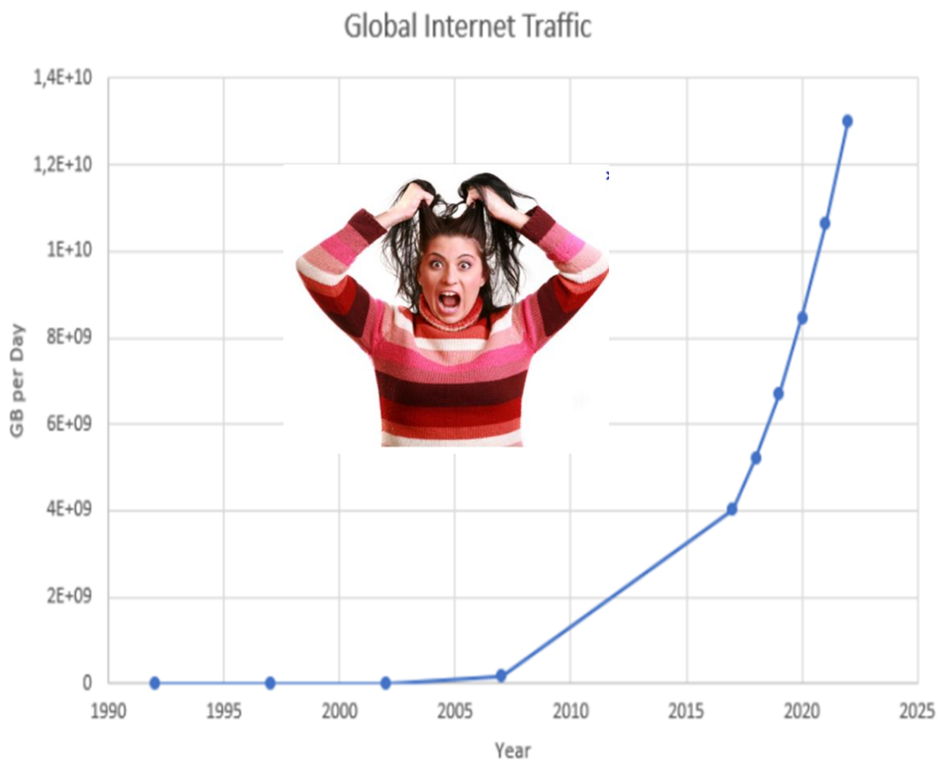
Reactive machines: AI systems that have no memory and are task-specific, meaning that an input always delivers the same output. Machine learning models tend to be reactive machines because they take customer data, such as purchase or search history, and use it to deliver recommendations to the same customers. Example: **Netflix recommendations**

Limited memory machines: This algorithm imitates the way our brains' neurons work together, meaning that it gets smarter as it receives more data to train on. Deep learning algorithms improve natural language processing (NLP), image recognition, and other types of reinforcement learning. Example: **Self-driving cars**

- **Self-aware:** The grand finale for the evolution of AI would be to design systems that have a sense of self, a conscious understanding of their existence. This type of AI does not exist yet.

This was Estimates of a Day in life of Data as of (2019)*

- 500 million tweets are sent
- 294 billion emails are sent
- 4 petabytes of data are created on Facebook
- 4 terabytes of data are created from each connected car
- 65 billion messages are sent on WhatsApp
- 5 billion searches are made
- 12,000,000,000 GB of Internet traffic Per Day



2: Increase in Global Internet Traffic in GB per Day from 1997 to 2022 (data from [Cisco, 2018])

- By 2025, it's estimated that 463 exabytes of data will be created each day globally – that's the equivalent of 212,765,957 DVDs per day

* Source: 2019 Visual Capitalist - https://www.visualcapitalist.com/how-much-data-is-generated-each-day

2024 updates show even greater Demand

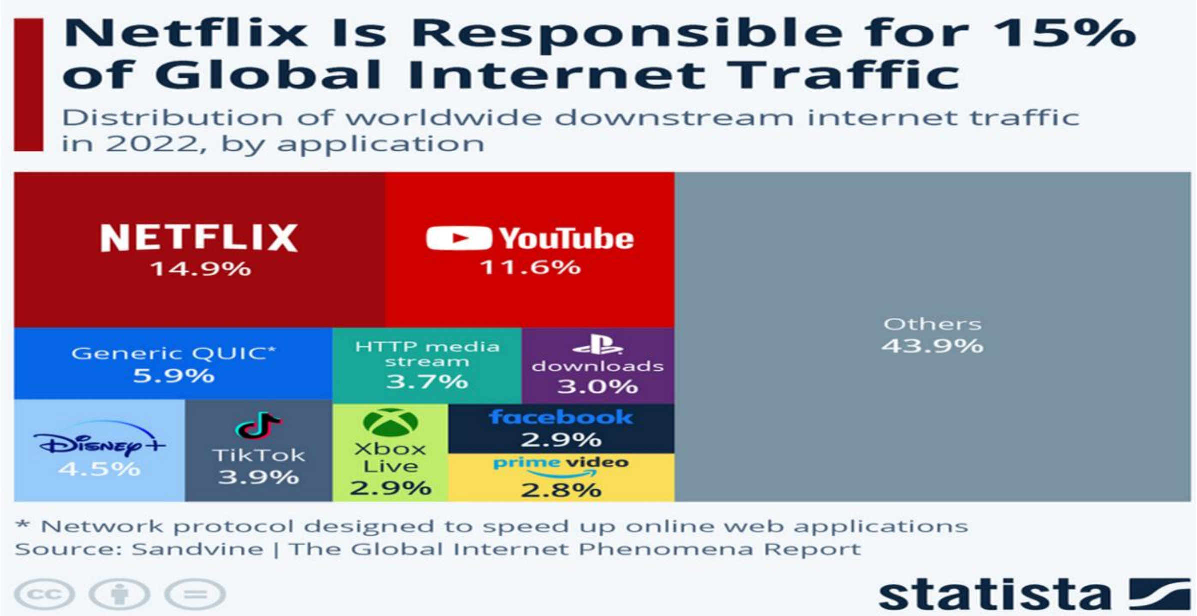
- According to the latest estimates from Statista, throughout 2024,

Approximately **402.74 million terabytes** of data are generated daily (equal to, encompassing newly created, captured, copied, and consumed information.

How much internet traffic is there globally?

According to radar.cloudflare.com in 2024

- Global data traffic has more than doubled in the five years since 2020.
- In year 2020, IP traffic was **32 exabytes**
- By 2024 it reached 68 exabytes. equates to the amount of data that it would take to stream in high definition a football game that lasts for 2 million years. Jan 22, 2025
- **41%** of traffic is from mobile devices
- USA Internet Traffic grow by 20%



GREENWAY RESEARCH

Power Consumption of Large Data Centers has been going up

Market Geography	Sum of Capacity (MW)
Northern Virginia	2552
Dallas	654
Phoenix	615
Silicon Valley	615
Chicago	555
New York City - Northern New Jersey	392
Portland	382
Atlanta	360
Los Angeles	206
Salt Lake City	203
Las Vegas	173
Seattle	105
Boston	95
Denver	78
Columbus	41
Grand Total	7026

Partial list.

Nationally power consumption of data centers is estimated at 9% of total electrical consumption in 2023.

Although share of electricity demand from large data centers has grown from 1.5% to 9% nationally, with an annual growth rate of 11%.

Growth in Total demand for electricity has been flat, indicating gains in efficiency in other sectors.

	2006	2021	2023	AAGR
Total US Electrical Consumption (TWH)	3660*	3800*	3680*	0.03%
Hyperscale Data Centers (TWH) National level	61	175	375	11%
Large DC electricity demand as % of national electricity demand	1.5%	5%	9%	

*- LBL National Assessment of Large data centers

Northwest was an earlier destination for Data Centers

In part because of:

- Ample power, (in part due to departure of large industrial customers)
- Low power and land cost,
- Availability of high-speed communication, global network of underwater cables
- Widespread internet bandwidth
- Reliability of power,
- Good reliable transmission system,
- Generous tax policies,
- Educated workforce,
- Wonderful weather

Between 2013 and 2022
Internet traffic grow at 40% per year (CGAR).
In the NW:
Data Center capacity grow at 20% (CAGR).
Revenue of data center entities grow at 23% (CAGR)

NW Regional	2014	2022
Colocation Count	70	104
Colocation aMW	226	525
Hyperscale DC Count	8	8
Hyperscale DC aMW	249	500

Average System loads in the Northwest

2014-2024 Balancing Area	Average Load	Market Share	Annual Growth
Bonneville	6,337	28%	0.8%
PUD No. 1 of Douglas County	205	1%	4.4%
Idaho Power	2,030	10%	1.0%
Pacificorp West	2,404	11%	0.1%
Chelan County PUD	275	1%	-6.7%
Portland General Electric Co.	2,402	11%	0.9%
Grant County PUD	588	2%	4.4%
Pudget Sound Energy	2,795	13%	0.2%
Seattle City Light	1,106	5%	-0.5%
Tacoma Power	558	3%	-0.8%
Northwestern	1,460	6%	-0.1%
Avista corp	1,300	6%	1.0%
PacifiCorp East (Idaho)	432	2%	-0.1%

Northwest load growth has been modest (about 1% per year).
Utilities with higher load growth typically have large Data Center loads

GREENWAY RESEARCH

Uncertainty in Utility Load Forecasts Has Increased due to

- Climate change has increased uncertainty on the long-term power planning process (Peak load fluctuations)
- Electrification of end-uses (timing and scale is uncertain)
- Transportation electrification (batteries , timing of charge is uncertain)
- Solar and wind generation availability
- Reshoring of key industrial sectors (chips)
- Data Centers Interconnection requests

Another reason for high load forecast is **uncertainty about energy consumption of AI**

- **How much energy does ChatGPT use?**

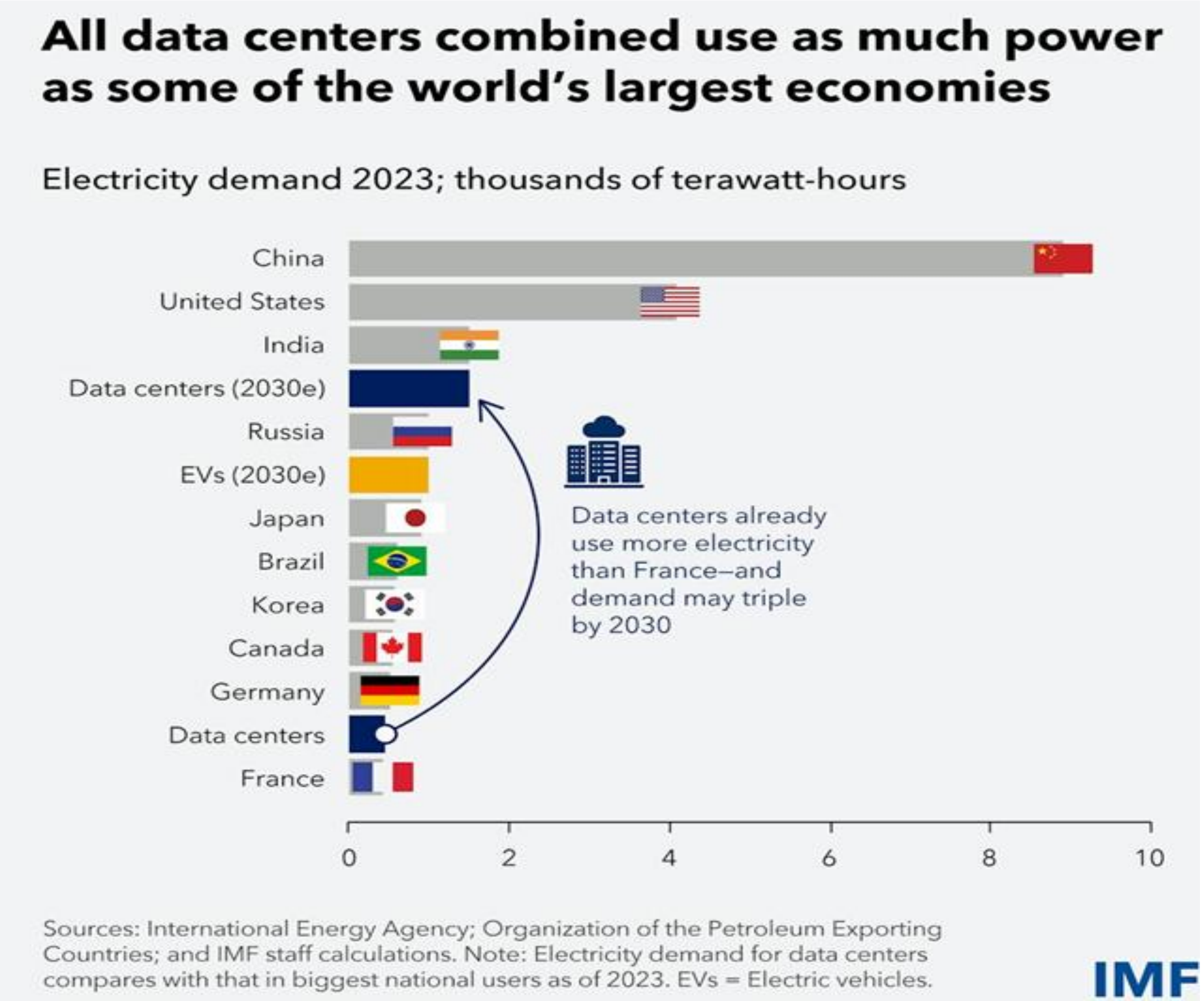
A commonly-cited claim is that powering an individual ChatGPT query requires around 3 watt-hours of electricity, or 10 times as much as a Google search. Alex de Vries (2023)
https://www.sciencedirect.com/science/article/pii/S2542435123003653?dgcid=author

However, According to Epoch AI (Feb 2025)

> "using up-to-date facts and clearer assumptions. We find that typical ChatGPT queries using GPT-4o likely consume roughly 0.3 watt-hours, which is *ten times* less than the older estimate. This difference comes from more efficient models and hardware compared to early 2023, and an overly pessimistic estimate of token counts in the original estimate."

This uncertainty leads to high and wide range of DC load forecasts

According to IMF report



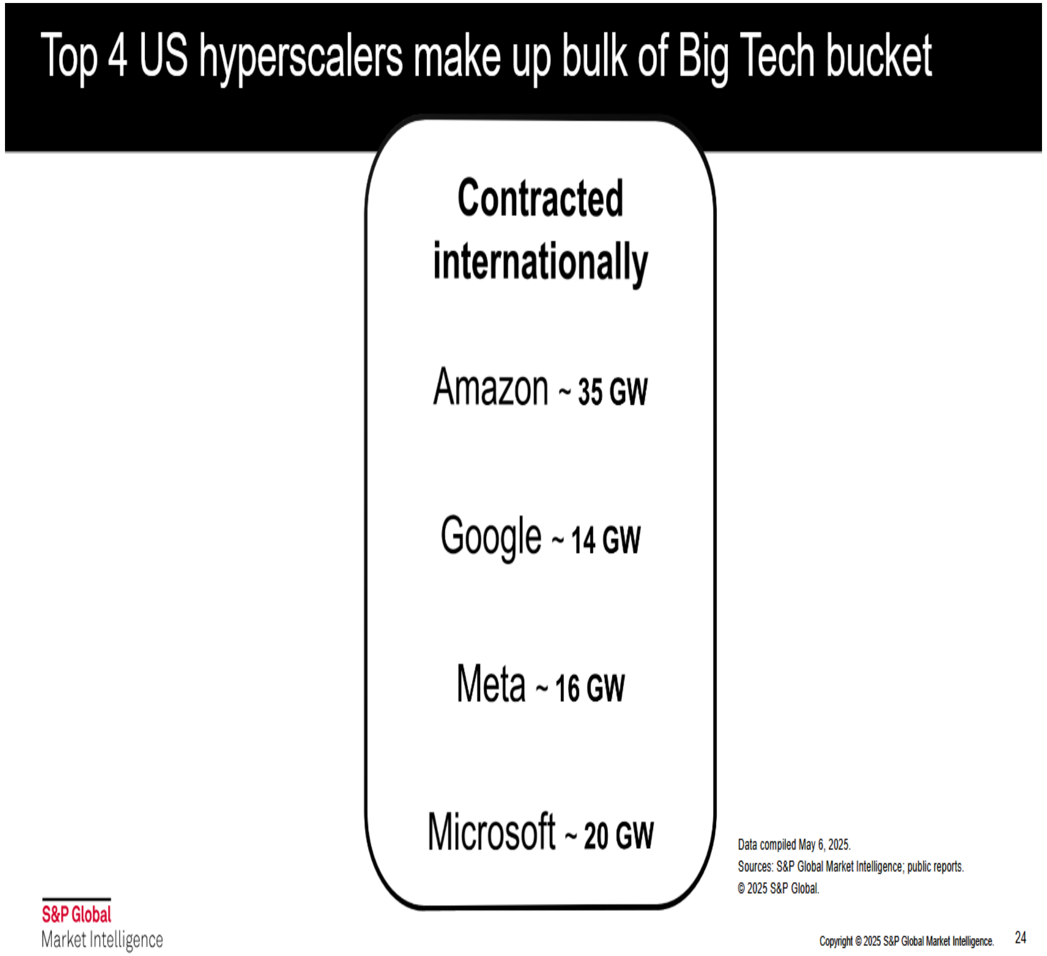
The world's data centers consumed as much as 500 terawatt-hours of electricity in 2023, according to the most recent full-year estimate by the Organization of the Petroleum Exporting Countries.

That total, which was more than double the annual levels from 2015-19, could triple to 1,500 terawatt-hours by 2030, OPEC projects.

Globally DC Construction is on hyperdrive

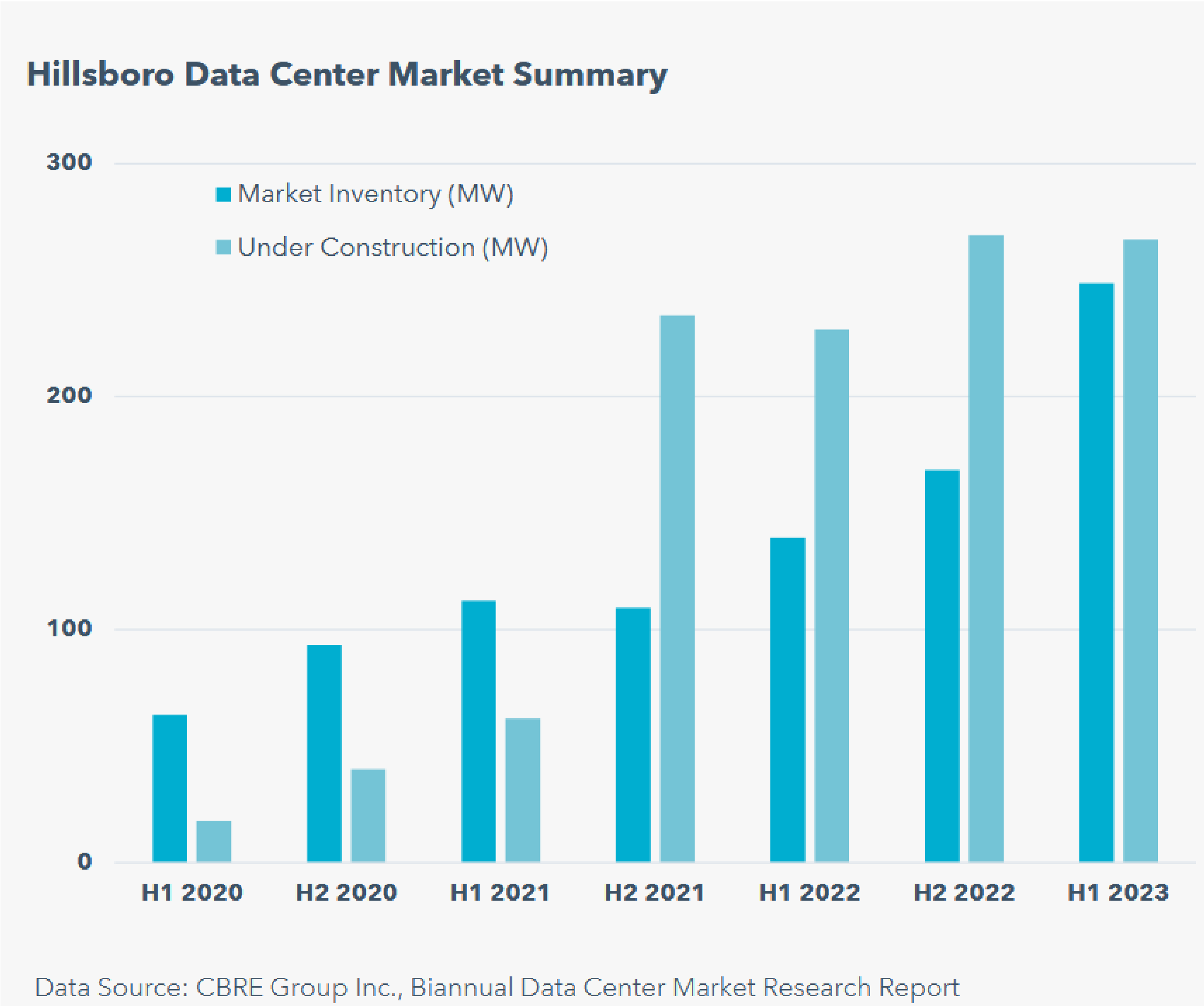
According to S&P market Intelligence there are over 85,000 MW of data center contracts are being constructed, world-wide

Various organizations have predicted the power demand of data centers would be greater than total power consumption of some countries.



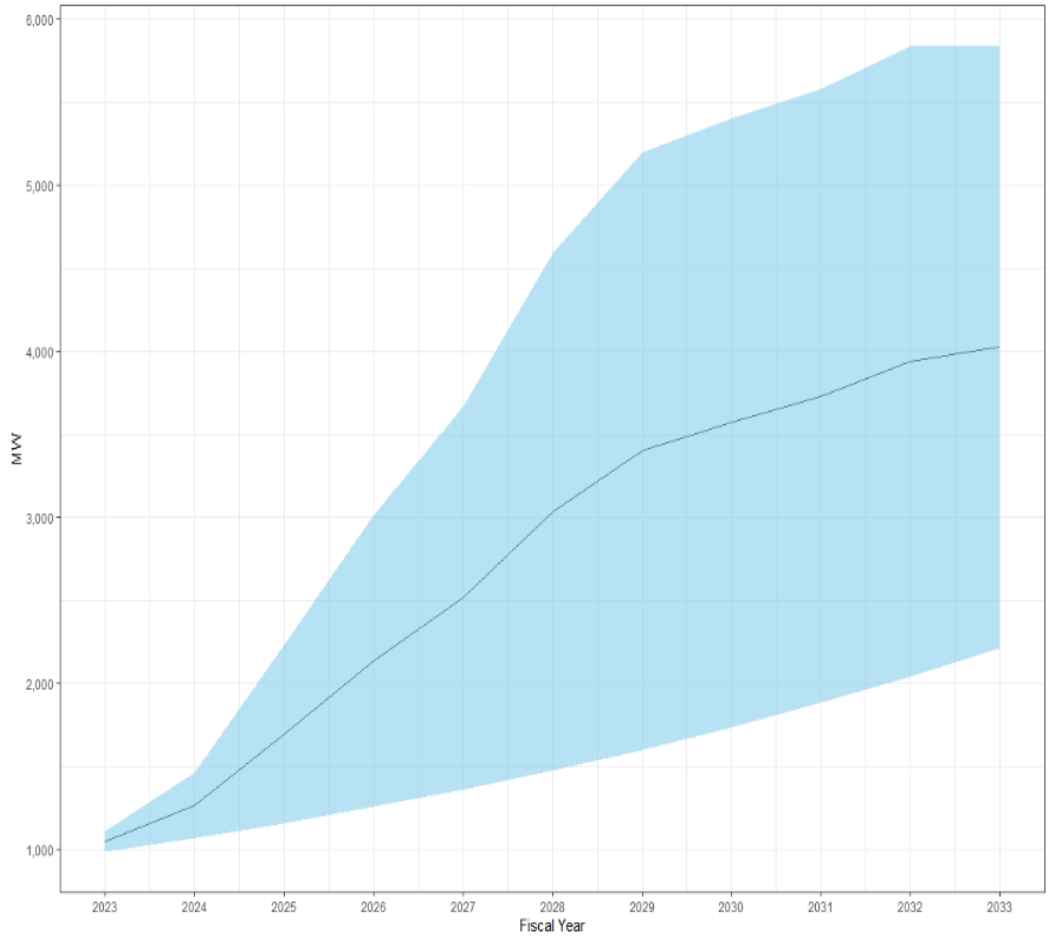
Locally, some utilities show significant increase in DC load

PGE reports Hillsboro DC market expect a doubling of capacity when new additions are online.



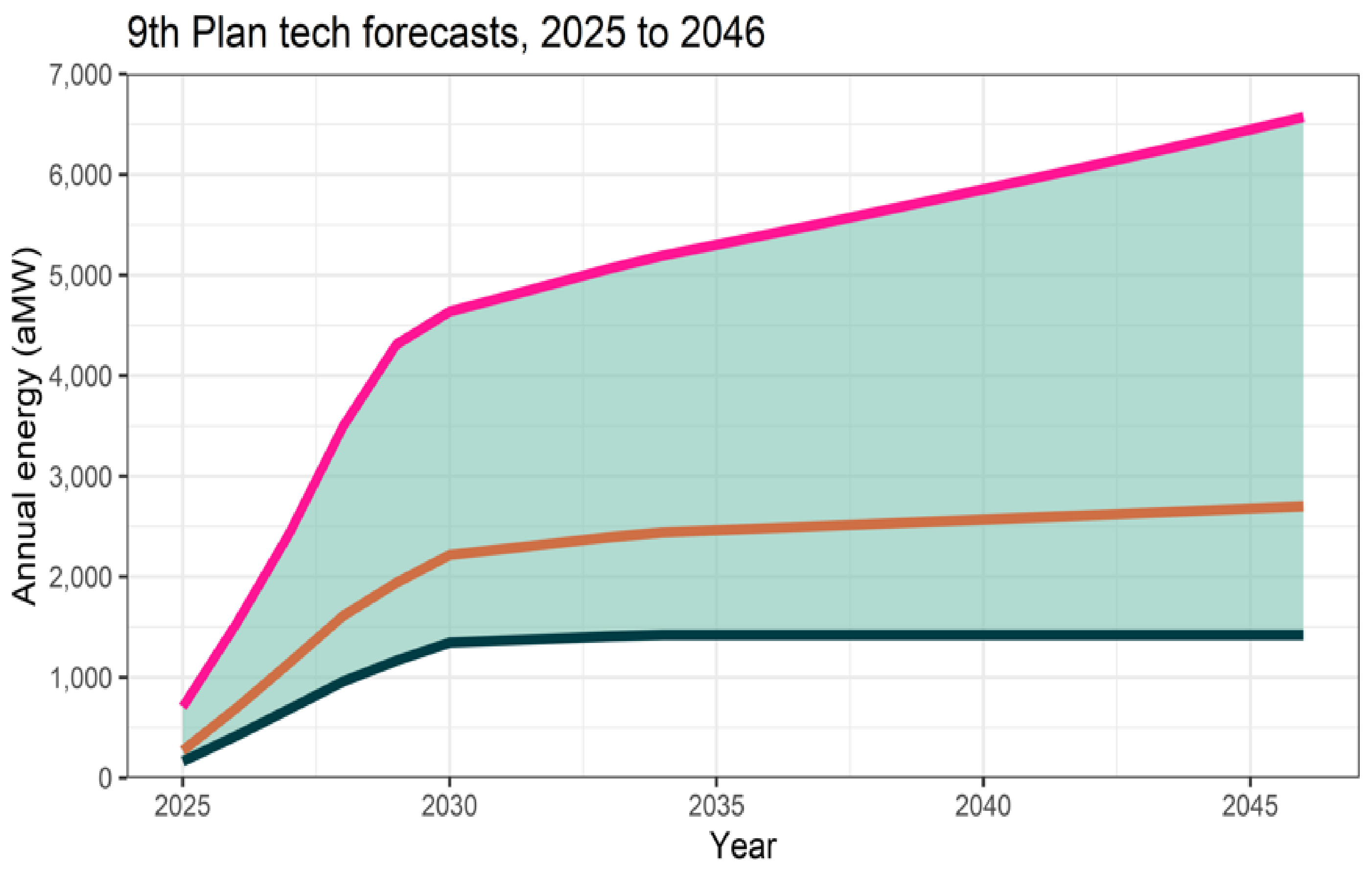
Bonneville Power Forecast (2025 draft)

- Shaded area indicate the range of potential data center load for the next 10 years
- Upper-bound is determined by an estimate of demand for FY2024 that includes the bulk of projects that have been reported to Transmission Planning, but may not meet the required confidence threshold for inclusion in the Agency forecast



- The solid line shows the Agency data center load forecast

Northwest Power and Conservation Council Forecast of incremental additions to data center plus chip fabrication



The **high forecast** through 2030 reflects utility and BPA growth expectations

The **mid forecast** through 2030 is a continuation of recent trends

The **low forecast** through 2030 has a slowing of recent trends

Post 2030 growth at a fixed rate depending on forecast

Testing two Methodologies to Forecast Long-term Demand for Large Data Center Power

Approach 1 (Bottom up- DC Model):

An engineering approach, Stock adjustment model that uses:

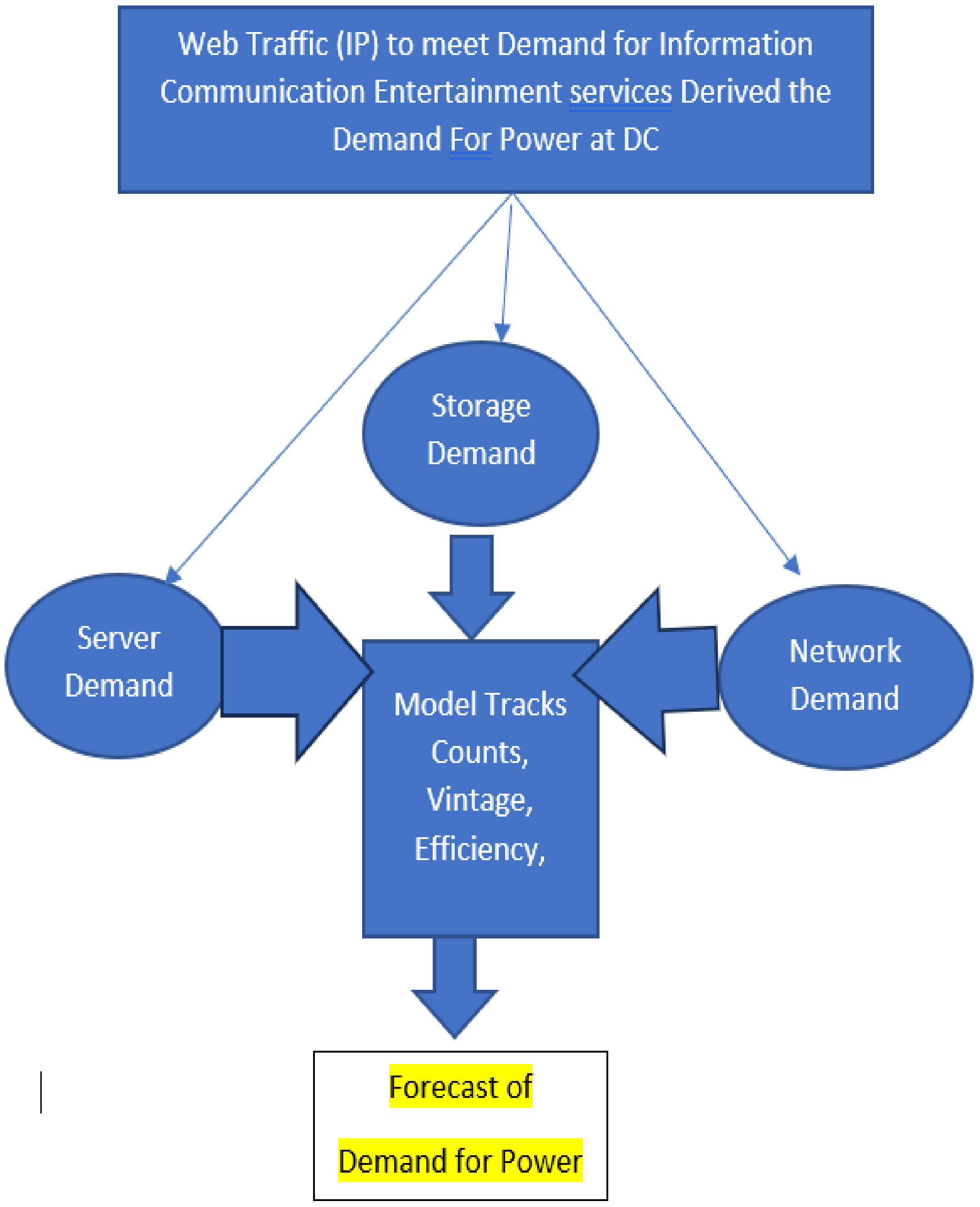
Projections of web traffic, counts, vintage, efficiency of DC components to estimates demand for power.

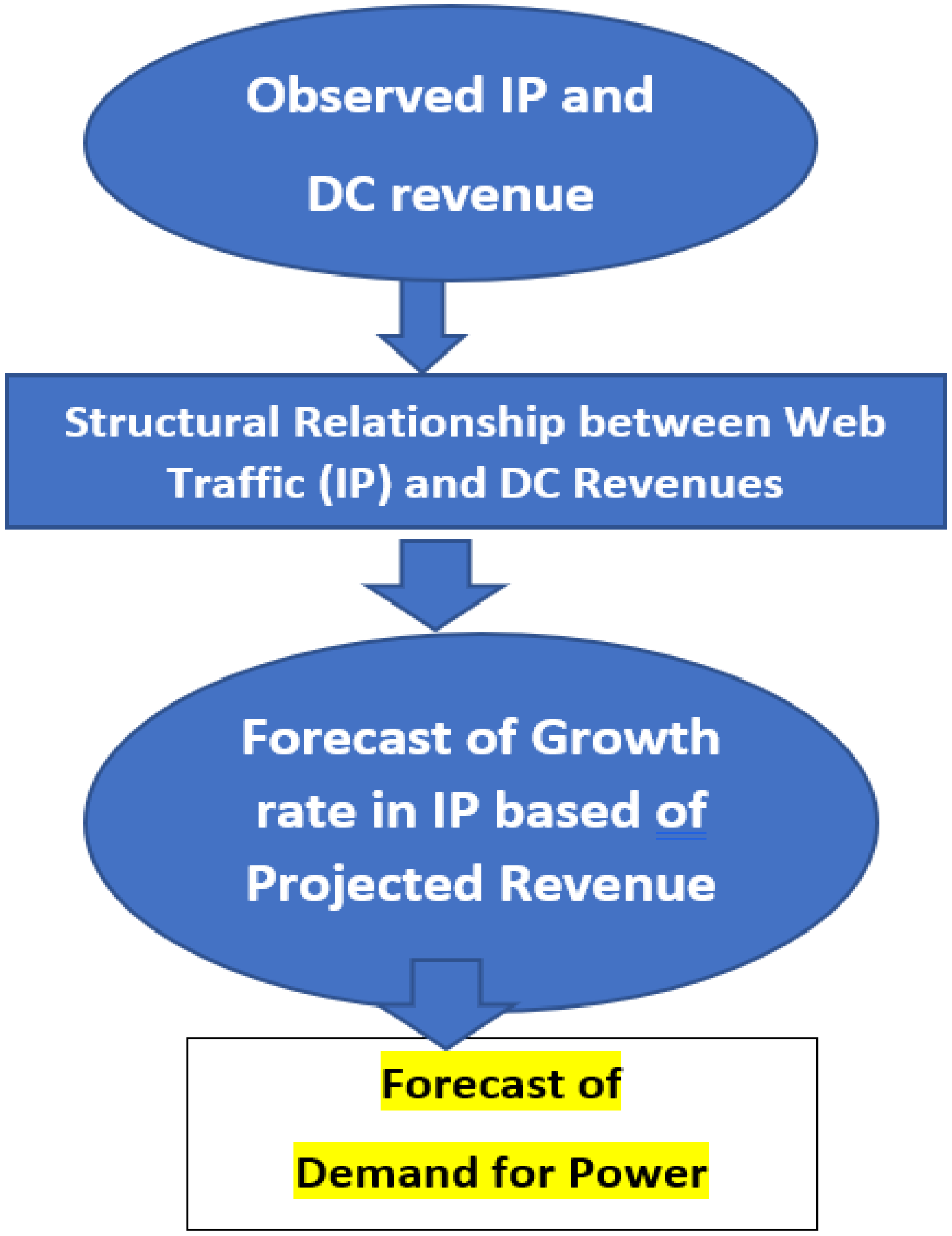
- Uses Moore's Law to project efficiency of servers.
- Uses Koomy's law to project power demand per calculation.

This approach requires more detail data on baseline efficiencies as well as projections on efficiency gains for servers, storage, and network. updating this information would require significant time and resources.

Approach 2 (top-down Sales Model):

- An econometric approach. It uses economic relationships between DC Sales/revenues, IP traffic and DC Capacity
- DC Sales and IP traffic are correlated strongly.
- Sales forecast was used as a rough proxy for IP forecast
- Forecast of DC capacity was estimated using Sales traffic.
- Forecasted growth rate in Sales was applied to existing DC capacity to project future DC capacity.
- This approach was simpler and more consistent with the economic growth trajectory for the region.


Web Traffic (IP) to meet Demand for Information Communication Entertainment services Derived the Demand For Power at DC
Storage Demand
Server Demand
Network Demand
Model Tracks Counts, Vintage, Efficiency,
Forecast of Demand for Power

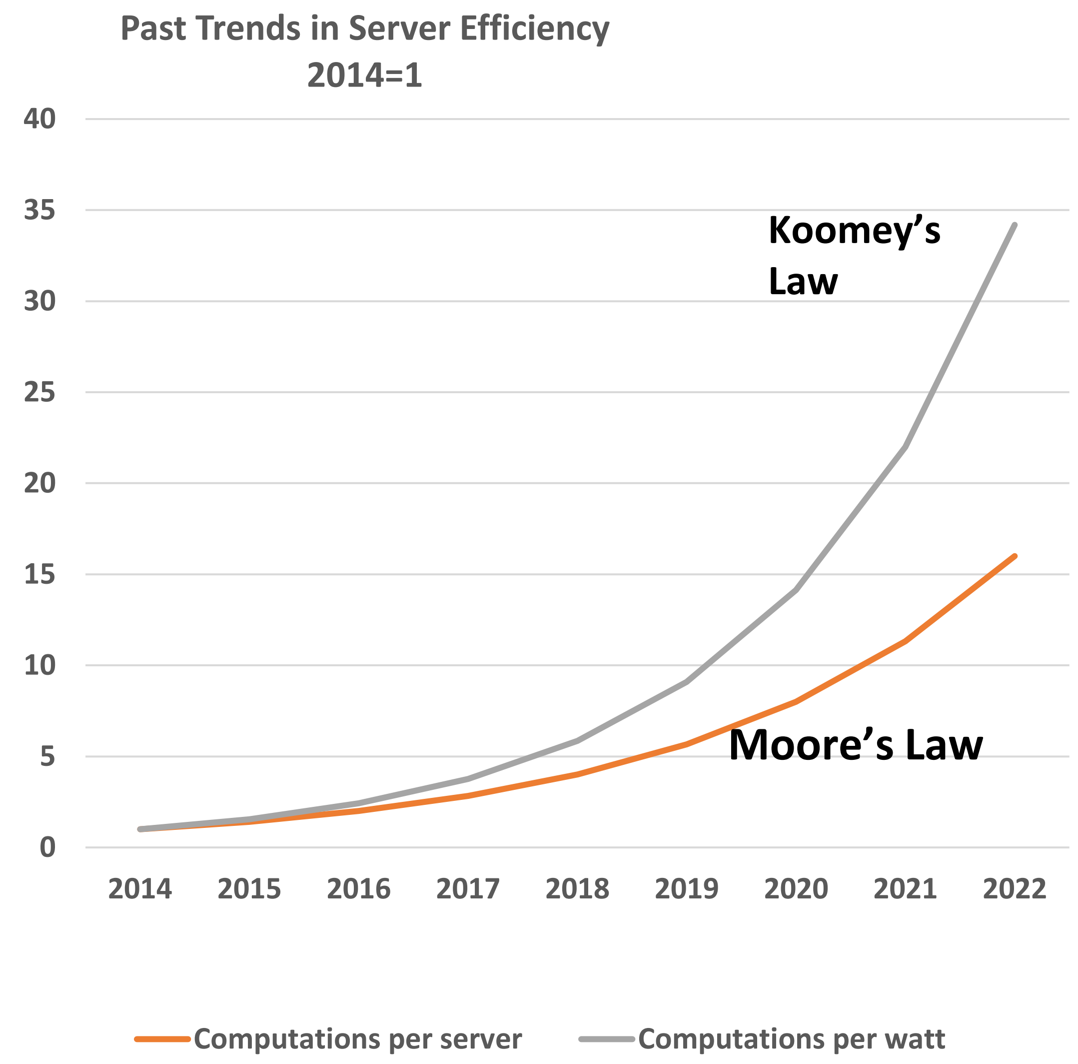

Observed IP and DC revenue
Structural Relationship between Web Traffic (IP) and DC Revenues
Forecast of Growth rate in IP based of Projected Revenue
Forecast of Demand for Power

Tracking revenue/sales of hyperscale DC as well as Growth in IP (Internet traffic) has grown at compound annual rate of 40%, Regionally Data Center Capacity grow at half that rate

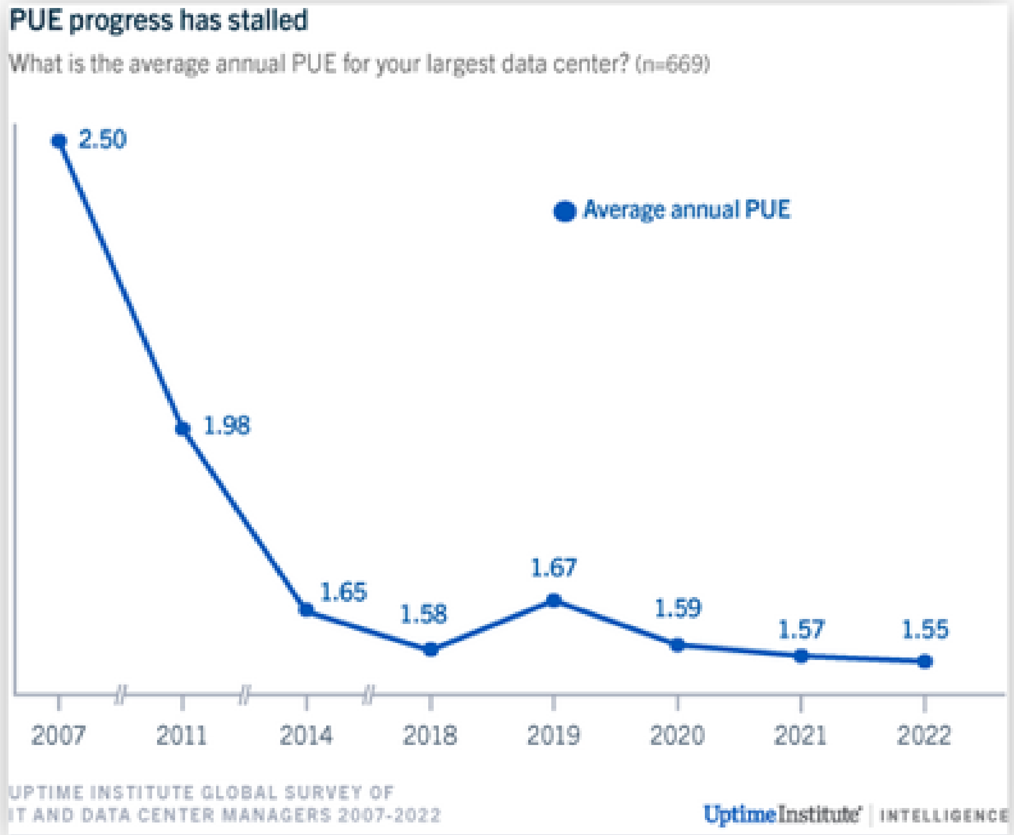
year	Sales ($B)	IP(EB)	estimated DC capacity (Colocator and Hyper-scale)
2013	13,371	1,926	
2014	15,597	2,774	474
2015	17,275	4,793	548
2016	20,615	6,819	641
2017	25,089	9,087	746
2018	31,706	11,557	863
2019	37,329	14,124	1,010
2020	44,424	17,116	1,172
2021	53,947	20,555	1,365
CGAR	23%	40%	19%

GREENWAY RESEARCH

Although demand for DC services has grown exponentially, power demand in DCs has not, in part due to Efficiency Gains



Typically, 43% of Electricity used in a data center is for servers, 11% for storage drivers, 3% for networks. Cooling and power provision systems use 43%. * 2016

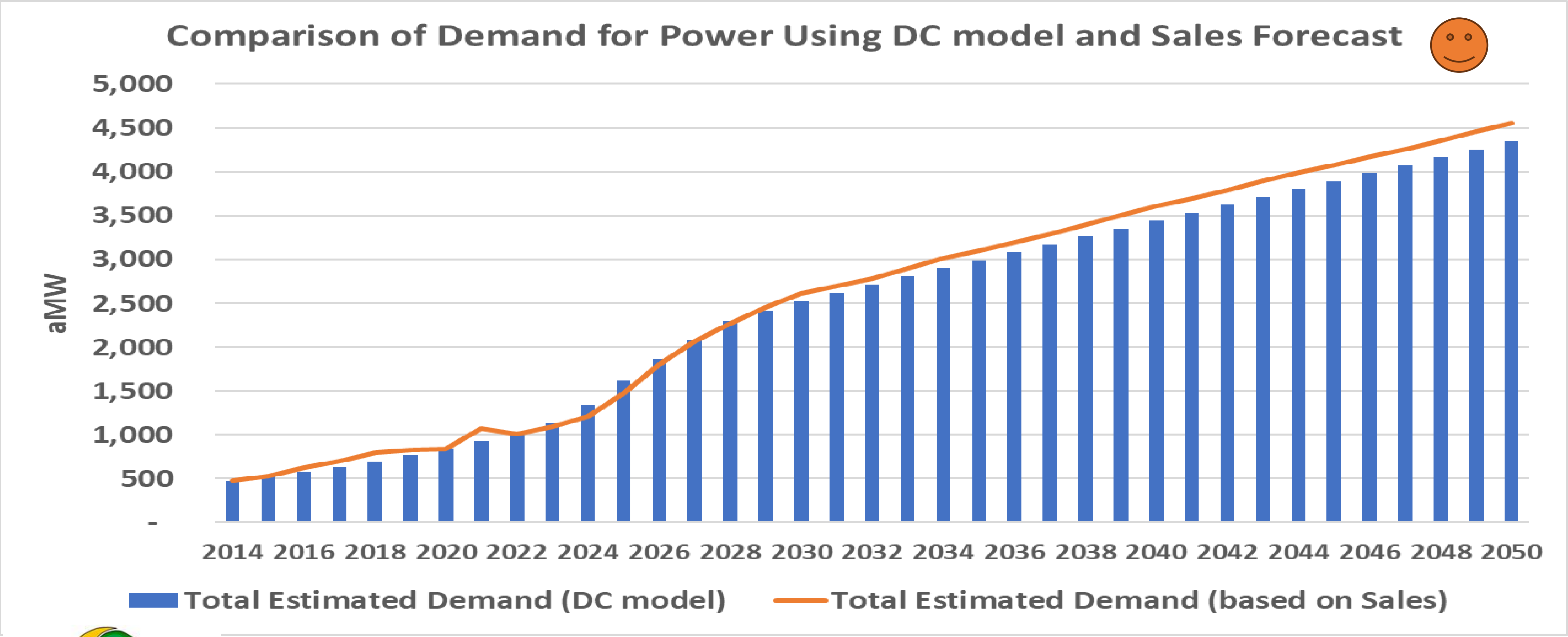


Source: CISCO Global Internet reports

Long-term Power Demand from the Large Data Centers

We used two methods for forecasting power needs

The Two approaches Produced Reasonably Close Forecasts



Would All the Data Centers be built?

- Recent EPRI survey of utility executives suggests- **Most likely not**.

The U.S. grid is flooded with data center proposals that will never get built. That's making it much more difficult for utilities and grid operators to plan for the future.

- *"Conservatively, you're seeing five to 10 times more interconnection requests than data centers actually being built," said Astrid Atkinson, a former Google senior director of software engineering and now co-founder and CEO of grid optimization software provider Camus Energy.*

- *Phantom" projects — or "vaporwatts," as they've been dubbed by Tim Hughes, chief development officer at Stack Infrastructure— take up utilities' resources and time, creating delays for all projects seeking connection, and inflating the country's load growth predictions.*

- *According to survey of 25 large utilities (EPRI survey in September 2024) , 48% expected data centers to account for at least 10% of peak load by 2030. Twenty-six percent expected double that share.*

Past trends in Global Data Center Energy Use shows that significant efficiency gains in *Servers, Storage, network and Average Power Usage effectiveness is possible*

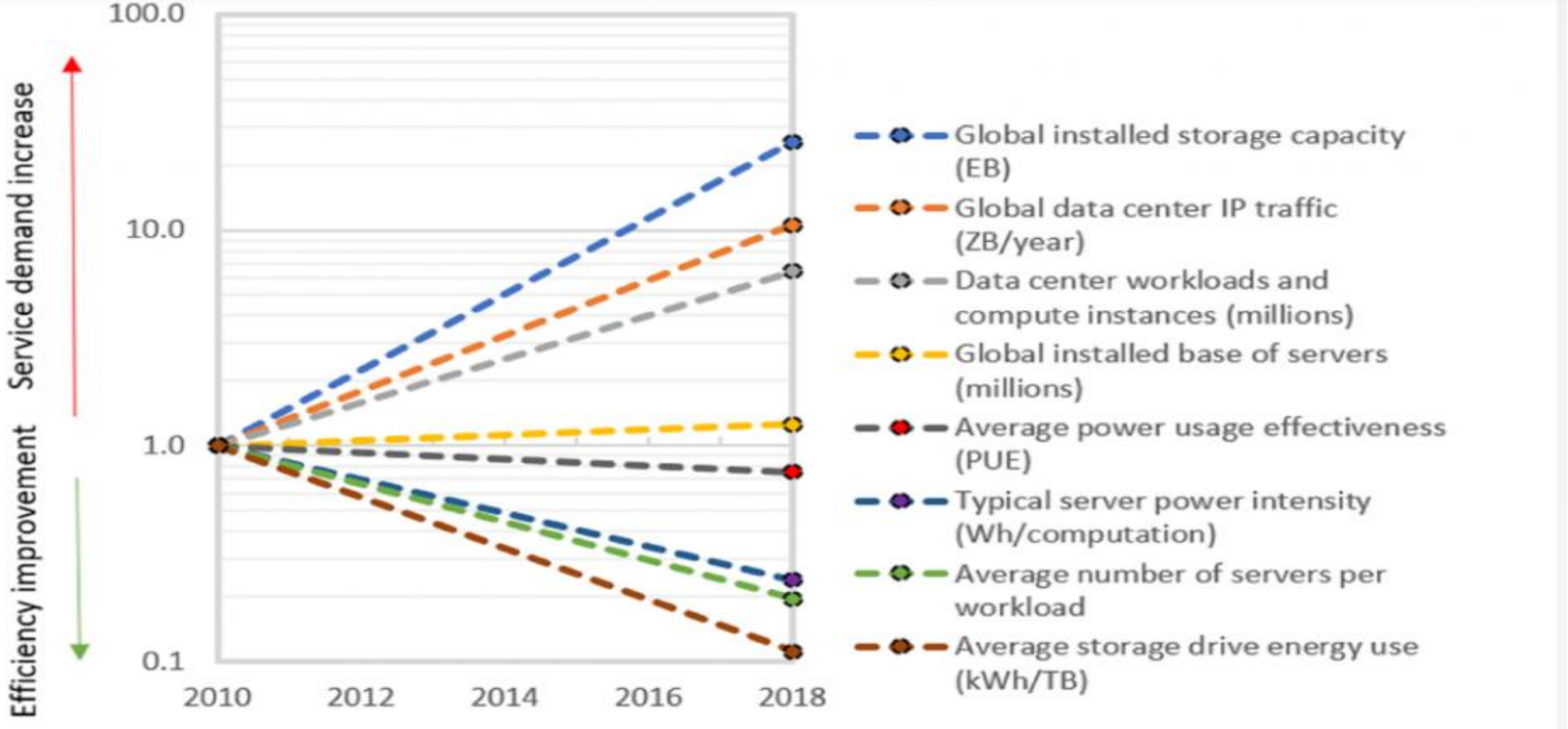
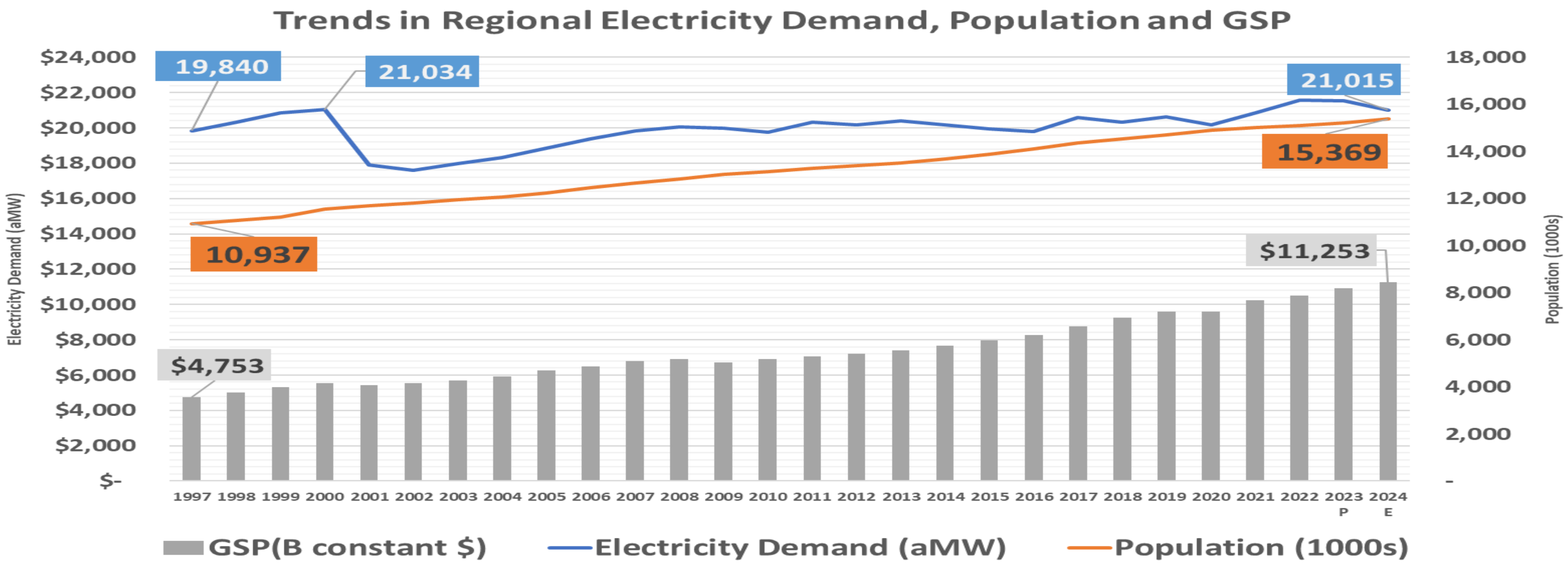


Figure 3. Relative change in global data center energy use drivers (2010=1). Source: Masanet et al. 2020.

Although regional population and economy have grown in the past 2 decades, Demand for power (including DC demand*) has not.

	IDAHO	MONTANA	OREGON	WASHINGTON	Growth since 2000
2024 Population (millions)	2	1.1	4.3	8	34%
2024 GDP (billions Real dollars)	100	58	265	702	203%
Electricity Demand 2024 P	2,900	1,800	6,000	10,000	-0.09%



GREENWAY RESEARCH

* Some large DC loads is met by non-utility providers

Future Trends and Constrains

Growth in Demand

- Artificial Intelligence (AI)
- Internet Of Things (IOT)
- Edge Computing
- Quantum Computing applications
- Future of Work-from-Home

Trends in supply

- Energy Efficiency (Koomy and Moore laws)*
- Data Center Cooling technologies
- Renewable Purchase Power Agreements
- Behind-the-meter renewable generation
- Automation using AI tools to predict and make suggestions for improving efficiency.
- Designing new buildings to meet LEED certification standards.
- Containerized/modular data centers

Constraints and Opportunities

Problem with powering Data Centers is not having enough power, but having right type of power at the right place will constrain pace of development of large data centers. Environmental concerns, high water usage, NIMBY would add another layers of opposition to rapid development of DC.

Innovations in AI systems including software, algorithms, and training methods could lead to substantial efficiency gains that reduce future electricity demand associated with AI technology compared to previous expectations

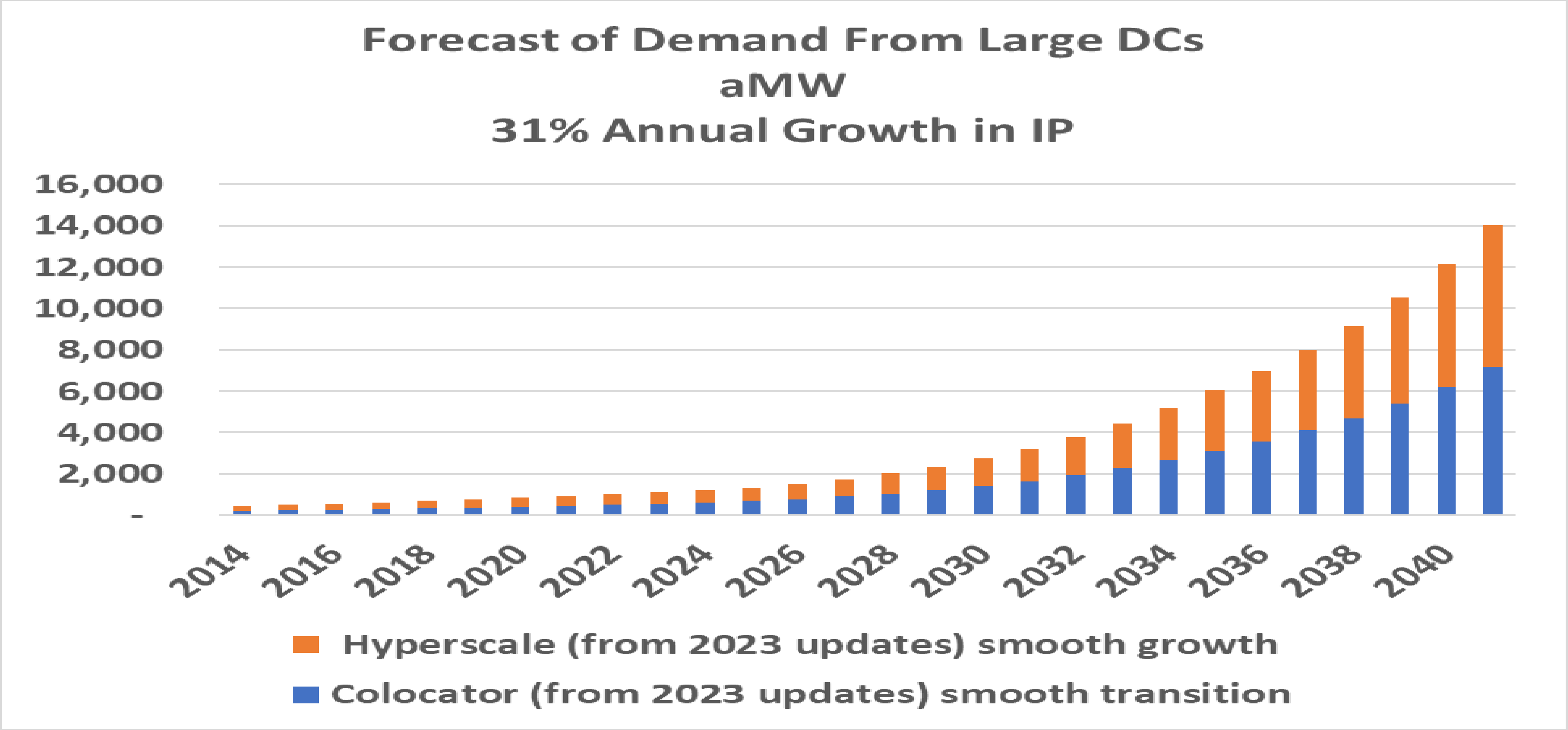
In summary

- There has been an amazing growth in demand for ICE (Information, Communication and Entertainment).
- Data Centers and internet traffic also have had phenomenal growth but at half the rate of growth in demand.
- Although Northwest and US have been on front-line of DC development, the global demand from is currently growing at a faster rate.
- Many forecasters, suggest large increases in power demand from DCs. These forecast are subject to wide range of uncertainty.
- Opportunities and Restriction on the demand and supply side could slowdown need for additional data centers.
- The AI/DC Energy Dilemma stems from interaction of two industries with different operating speeds
- DC players want fast response and implementation to monetize demand for ICE. Not knowing the speed of increase in demand, and uncertainty about availability of power supply, DC players typically overestimate their needs. DCs expect fast response from utility sector.
- The power sector is regulated in their operation and expansion. Utilities want to minimize costs of operating and investing in new resources. State regulatory bodies and local constraints will add the needed investments in generation, transmission and distribution more slowly.

Additional Slides

As a sensitivity we tested a “What if” analysis

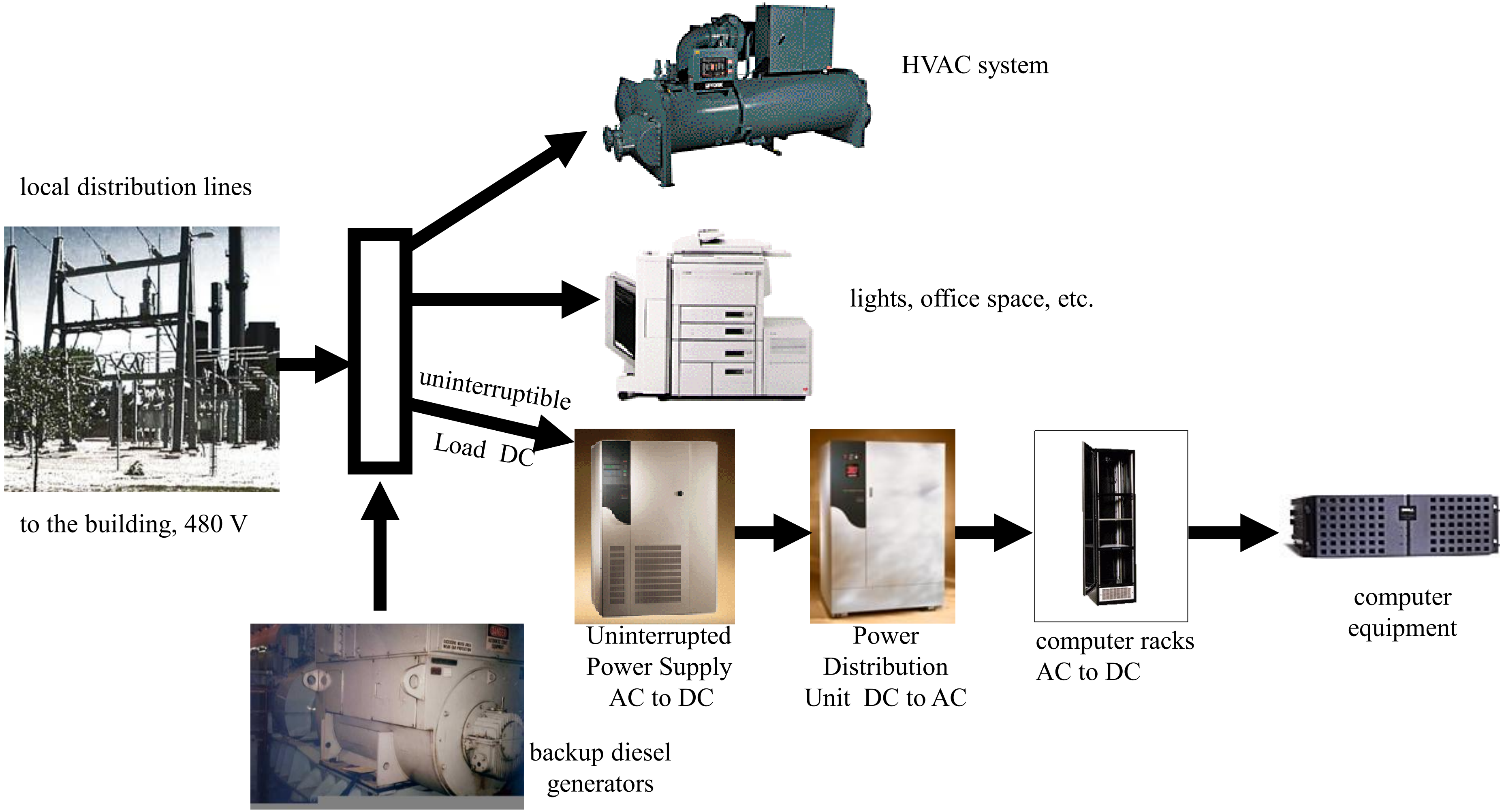
If server workloads continue to grow at the recent rates, ~ 31%
The demand for power would be tremendous – possibly representing over 50% regional Load



There are a variety of Data Center Sizes and Business Models

	Example	Approximate Energy Consumption	% of Data Centers in the US	% of Servers in the US	Typical Location	Some of Barriers to typical utility Energy Efficiency Programs	Opportunity for Energy Efficiency
Enterprise-class/hyper Data Centers	Google, Facebook, Amazon	10-100+ MW	0.3%	28%	non-metro area	secrecy, rapid market change, split incentives, identifying key player, baseline	comprehensive customized offerings/ requires long-term relationship, market movers
Mid-Tier Data Center	Colocators, EasyStreet	10 MW or less	0.4%	15%	Metro area	less secrecy, capital constrained, split incentives, baseline and incentive	comprehensive and customized/ requires long-term relationship
Localized Data Center	Hospitals, financial institutions, Government	10-500 KW	2.5%	16%	Metro area	Harder to locate, split incentives	Customized/Prescriptive, Training and information on energy efficiency options, long-term relationship
Server closets/Rooms	Small to Mid-size Company	5-10 KW	96%	~40%	business dependent	hard to locate, Small IT resources doing many tasks, IT not core business	Perscriptive program offering

Electricity Flows in Data Centers

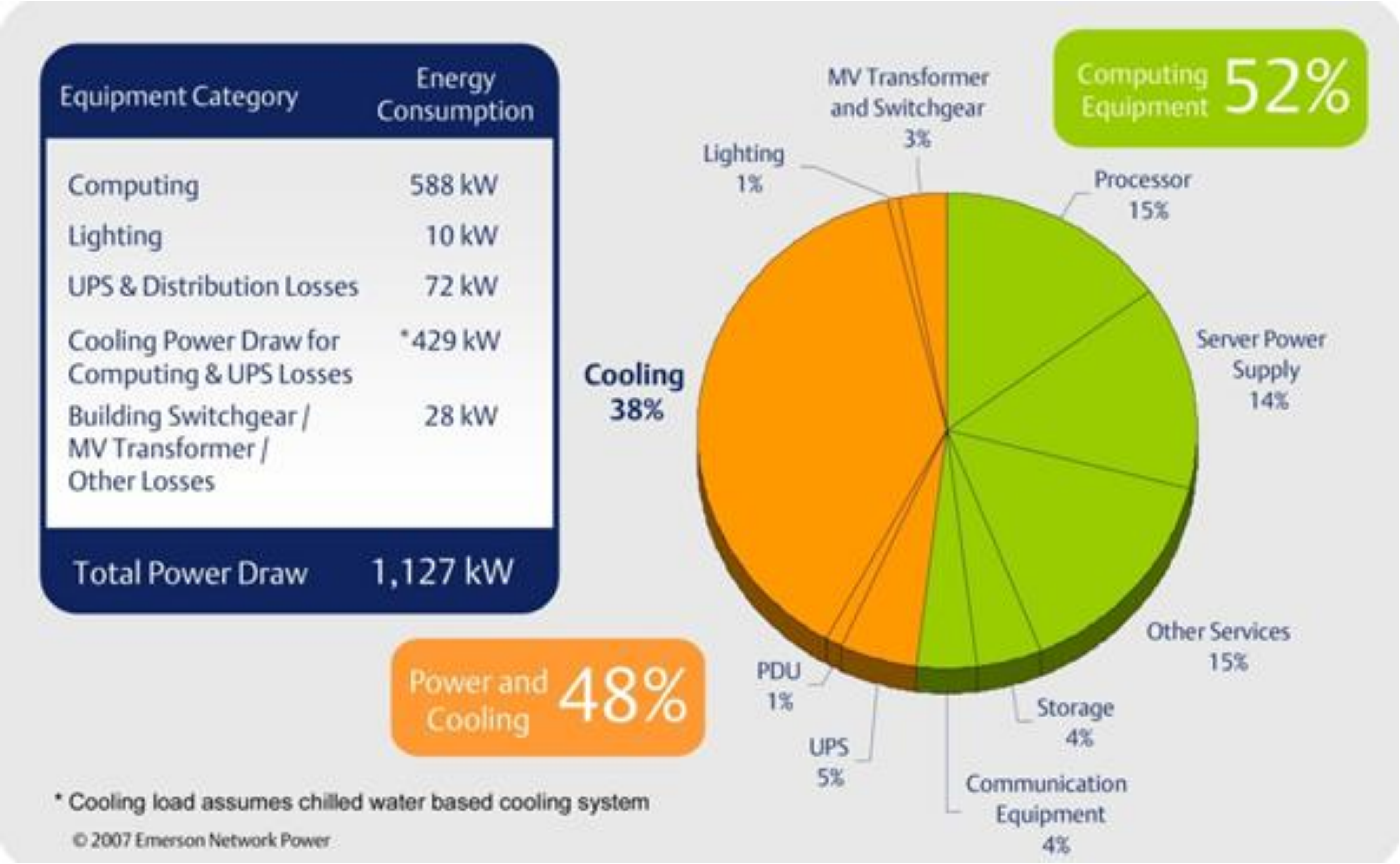


Jonathan Koomey, Ph.D.
Lawrence Berkeley National Laboratory

Northwest Power and Conservation Council

Power Distribution in Data Centers

newer data centers do much better



Extract from Data Center Market Assessment, Conducted by PECI, for Energy Trust of Oregon -2011