Bending the Energy Curve: Decoupling Digitalization Trends from Data Center Energy Growth

White Paper 212
Version 2

Data Center Research & Strategy

by Jim Simonelli
Victor Avelar
Wendy Torell

Executive summary

From 2010 to 2020, data center electricity consumption remained flat despite exponential data growth. However, since 2020, most industry forecast models now project significant growth in data center electricity consumption through 2030. This paper investigates the reasons behind this shift. We use a method that correlates annual information and compute technology (ICT) energy consumption with annual data production to frame the conversation between infrastructure and ICT technology providers. This correlation suggests a disruption in compute efficiency starting around 2020, with recovery projected to begin in 2026 and impact visible by 2030. The paper demonstrates that small improvements over currently planned projections in power usage effectiveness (PUE) and compute efficiency from 2026 can potentially "bend" the energy growth curve by 17%, decoupling data growth from compute energy consumption. This industry-led approach allows the world to prioritize the supply of clean energy. This focus fuels the economic and societal progress that comes with digitalization, rather than planning for seemingly unbounded data center energy growth.

RATE THIS PAPER ★★★★★

Key takeaways

1. **An industry-**led approach is necessary to "bend the curve"**, potentially reducing energy consumption by an estimated 17%.** Small annual gains in PUE and information and compute technology (ICT) efficiency, exceeding current projections from 2026 to 2030, can greatly alter the energy growth trajectory. This decoupling of data growth from energy consumption enables the industry to focus on the supply of sustainable energy rather than planning for seemingly unbounded data center energy growth.
2. **Global data center energy consumption remained flat from 2010 to 2020.** Despite a significant increase in data production during this period, from 2 to 64.2 zettabyte (ZB), data center electricity consumption remained relatively stable. This indicates that improvements in efficiency, including ICT efficiencies and PUE improvements, were effectively offsetting the demand for energy.
3. **Starting in 2020, energy became more closely coupled to the data growth.** Beginning in 2020, we observed a marked shift in data center energy consumption that aligns more closely with the exponential growth in data production. This is largely driven by energy-intensive workloads such as artificial intelligence (AI) and cloud computing.
4. **We recommend establishing an industry correlation metric, the "ICT performance factor" to explain and evaluate how performance improvements affect energy usage over time.** We propose an indicator metric of annual ICT TWh divided by data generated that year (zettabytes) to help us understand how industry performance impacts energy, and to demonstrate how improvements to that performance can affect the overall energy curve over time.
5. **Further PUE improvements will help reduce data center energy by 3.6% - or 38 TWh.** Average PUE has continued to trend down over the years, decreasing from 2.0 in 2010 to 1.51 in 2024. Following this trendline, the average data center would achieve 1.43 by 2030. Technologies enable *new* data centers to achieve better PUEs. We need an industry-wide effort for all *new* builds to achieve a PUE of 1.2, and for 10% of existing data centers to achieve a PUE of 1.3. If we do both by 2030, we will slow the curve by 3.6%.
6. **Further ICT performance factor improvements will reduce data center energy by 14.4% - or 153 TWh.** ICT performance factor is a more significant lever to the overall data center energy curve than PUE. Using industry-published data growth and energy forecasts, we show that the ICT performance factor reduces (improves) from 0.92 TWh/ZB in 2020 to 0.45 TWh/ZB by 2030. With an industry-wide effort, we can reduce this further and achieve 0.34 TWh/ZB by 2030. This is possible through full adoption of accelerated computing platforms for AI, software and dispatching algorithms to fully utilize compute, and continued acceleration of central processing units (CPU), graphics processing units (GPU), tensor processing units (TPU) and memory performance.
7. "Behind the meter" energy resources can further bend the utility**-demand energy curve.** Data centers deploying "behind the meter" power generation reduce the strain on the grid. Every megawatt of generation put in place saves 8.76 TWh. Sustainable options offering full-time grid relief are available today, with more on the way. These include solid oxide fuel cells with natural gas (with carbon sequestration), solid oxide fuel cells with green hydrogen, and gas turbines with carbon sequestration. In addition, small modular reactors (SMRs) will become a realistic alternative in the future.

Introduction

The rapid growth of data center energy consumption raises alarms about the ability of the electrical grid to keep pace. As demand for artificial intelligence (AI) and cloud computing drive data center expansion, electricity usage is projected to surge dramatically. This escalating demand poses regional challenges for power generation and grid capacity.

Also, developers are struggling with energy access and reliability, as connection lead times can be twelve to twenty-four months or more, exacerbating local and regional supply challenges. Data center developers often face additional local scrutiny, with pushback from authorities and the public at large over generator noise/pollution, land, and water use.

The concern over energy scarcity has led to various public forecasts of global data center energy consumption. These forecasts highlight the significant upward energy consumption trend through 2030, depicted in **Figure 1**. This alarming exponential growth pattern is gaining wide-spread attention. Numerous technical papers and articles on the subject look to explain the trends and the various levers impacting energy projections. A few also attempt to learn and extrapolate from historic trends.[1]

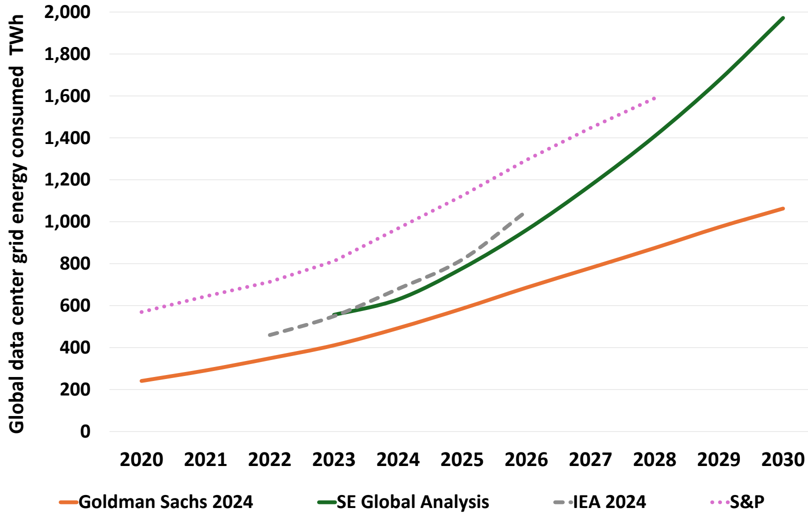


Figure 1

Forecast from various sources all indicating trends of continued TWh energy growth through 2030 [1]

We find that the rationale and methodologies for these projections are generally based on solid principles but with varying emphasis on the number of levers, their granularity, and their degree of impact on energy consumption. Through our exploration and analysis, we found a consistent overarching narrative to frame a new conversation, describing how information and compute technology (ICT) and infrastructure industries can collaborate (and have historically) to impact energy growth.

We believe there is a path forward for the industry.

In this document, we deliver three important elements to reframe the conversation. With the first, we quantify the energy growth and its drivers. In the second, we introduce a new industry correlation metric to explain the current situation. And for the final element, we model scenarios with varying efficiency improvements to quantify their impacts on the energy curve. We show how even minor energy improvements in ICT and infrastructure efficiency, beyond those already projected in studies, can

[1] Three examples include McKinsey & Company, How data centers and the energy sector can sate AI's hunger for power, S&P Global, Data Centers: Rapid Growth Creates Opportunities And Issues, and Deloitte, Powering artificial intelligence: A study of AI's environmental footprint—today and tomorrow

"bend the curve" between 2026 and 2030, allowing data center energy consumption to decouple from the planet's digitalization demands. In our analysis, we use the Goldman Sachs forecast, along with data from IEA, to establish a baseline curve. Combined, this provided a data set that spanned historic and future projections.

Inspiration and core methodology

The inspiration for this work comes from a paper by Schneider Electric's Sustainability Research Institute (SRI) entitled *Artificial Intelligence and Electricity – A System Dynamics Approach*. This SRI paper uses system dynamics to analyze foundational attributes and discuss potential futures for AI and its associated energy consumption. It offers insightful perspectives on how various internal (primarily technology-related) and external (mainly economic and societal) factors contribute to four scenarios specific to AI. The four scenarios and their trends are shown in **Figure 2**.

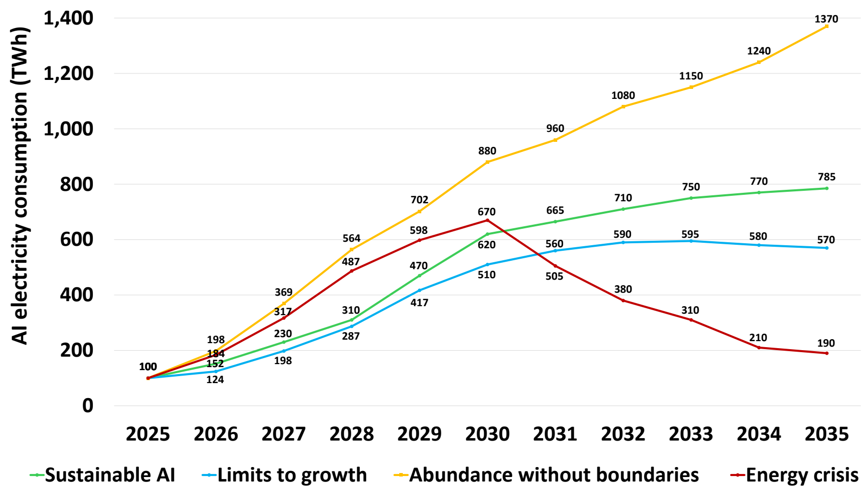


Figure 2

Results from the SRI paper, Artificial Intelligence and Electricity – A System Dynamics Approach, showing the TWh impact of four AI-based scenarios

As we reflected on these scenarios with a focus on the years leading up to 2030, we posed two additional questions. First, how do these trends relate to overall data center energy consumption, beyond just AI-related energy? Second, given that the growth pattern aligns with the trend shown in **Figure 1** and is based on similar efficiency assumptions, what would be the outcome if we could isolate the drivers and accelerate their energy impact? To accomplish this, we take a deeper look at historic trends starting in 2010 and focus on insights that emerge. With this broader view, a clear trend break appears around 2020 as illustrated *by the orange energy curve* in **Figure 3**.

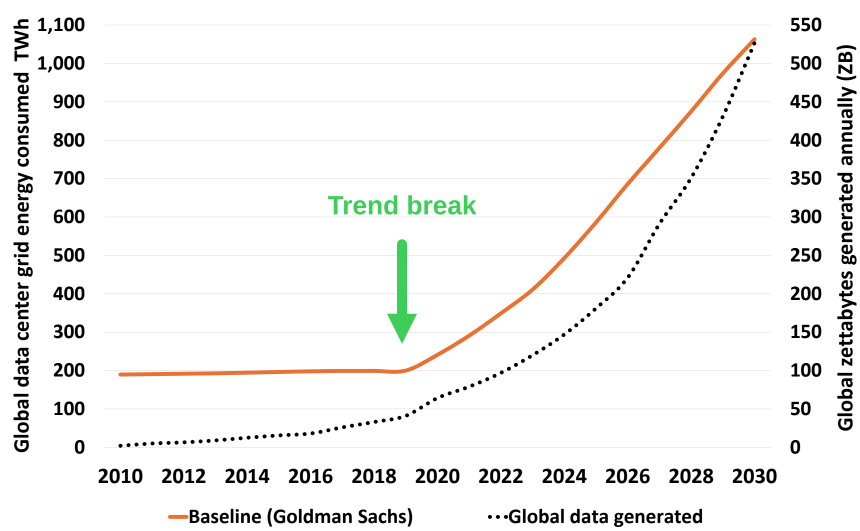


Figure 3

Graph showing total data center energy consumption. A clear trend break is seen in 2020 when the energy curve shifts from flat to increasing. We overlay global annual data production as a reference to the historic and future need for ICT.

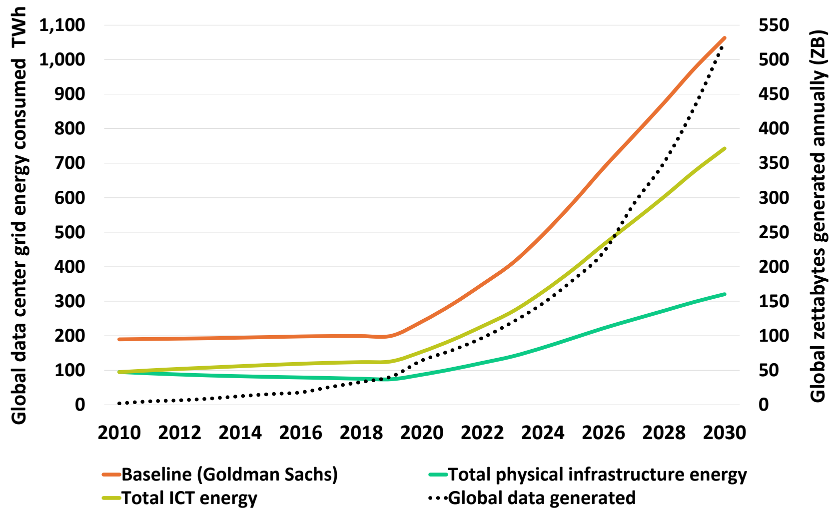
For this analysis to be meaningful, the energy forecast used must:

- reflect an internally consistent definition of what constitutes a "data center".
- include historical data and future projections of TWh.
- clearly outline assumptions impacting future growth (knowing these “baked-in” assumptions prevent double-counting our “bend the curve” improvements).

Our narrative leverages our core methodology, starting with a global data center energy forecast and a PUE forecast for years 2010 through 2030. Using the PUE data, we derive the ICT and physical infrastructure energy breakdown (**Figure 4**).

Figure 4

Plot showing breakdown of ICT and infrastructure energy for the baseline scenario



This becomes our **baseline scenario**. We then demonstrate how to “bend the curve” using separate PUE and ICT improvement factors for years 2026 through 2030. This results in three new scenarios: PUE improvements, ICT performance improvements, and both PUE and ICT performance improvements. A more detailed methodology is described in the **Appendix**.

In the next section, we describe an important aspect of this analysis: we recommend a new industry correlation metric, to provide insights on trends, guiding industry action and helping bend the energy curve.

Establishing the “ICT performance factor” for insights

It’s important we go beyond the forecast and understand the digital value gained from this energy. To be effective, we believe a correlation metric is useful to estimate the energy consumption per unit of compute. But we need to identify the proxy for information and communication technology (ICT) growth.

There are many approaches to understand the growth engines for ICT (servers, storage, networking) such as using estimates and projections for the following:

- Global Internet traffic (zettabytes)
- Global internet users (count)
- Workload growth (i.e. count or compute instances)
- Global data growth (zettabytes)

In Goldman Sachs’ report, *AI, data centers and the coming US power demand surge*, they use compute instances as a benchmark. All the above approaches provide similar trends and have some correlation to ICT energy needs. This is because they all exhibit three key attributes: (1) They are societal digitalization drivers for ICT, (2) they have a historic and future view, and (3) they have a generally consistent trend shape making them useful when studying correlations and sensitivities.[2] Any of these growth indicators could have been selected and lead to consistent findings. We chose data growth as the proxy for the following reasons: when you generate data, you rely not only on compute energy, but also storage, and networking energy. Furthermore, statistics and forecasts on data growth are readily available. In **Figure 3**, we overlay this data growth against TWh growth.

Using data growth, we propose a metric that we will call the **ICT performance factor** to correlate data production to new ICT energy usage as shown in **Equation 1**.

Equation 1

ICT performance factor

$$\text{ICT performance factor} = \frac{\text{ICT added per year (TWh)}}{\text{Data generated per year (Zettabytes)}}$$

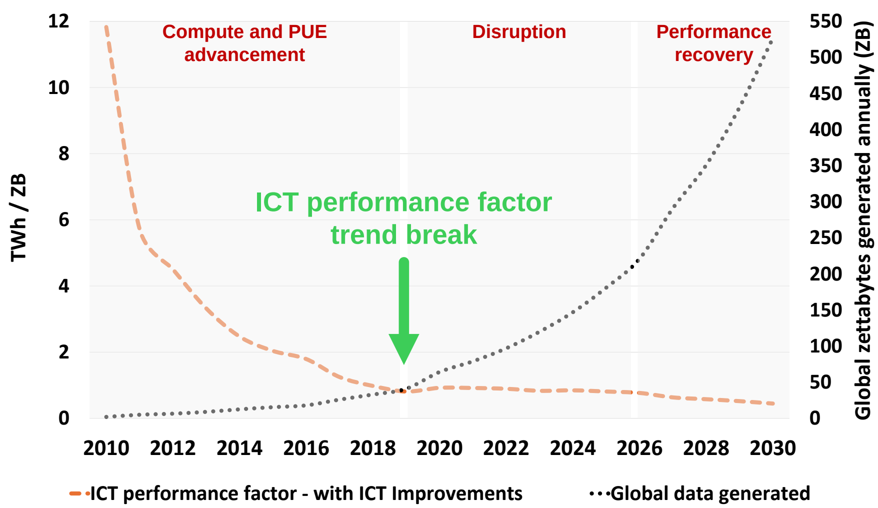
While this factor is imperfect and non-causal, it is useful in framing insights. It is based on industry-accepted data production trends that track well over historic and future years, thus helping to analyze trend breaks. It also enables us to build scenarios to demonstrate TWh changes.

Figure 5 shows the growth of data generation vs. ICT performance factor over time. A clear 3-phase trend is evident, despite a continuous and substantial rise in data production for 2010 to 2030.

- Phase 1 - the metric continually improves due to **data center advancements**.
- Phase 2 - a **disruption** occurs, and the metric remains relatively flat.
- Phase 3 - a **performance recovery** occurs (although at a slower rate than phase one).

Figure 5

Plot showing three phases of energy impact vs. ICT performance factor over a 20-year period



There are many hypotheses for explaining the trend break. These include reaching the limits of the percentage of workloads able to shift to more efficient cloud computing, diminishing CPU performance gains, Covid related delays in new silicon releases, acceleration of "new" AI workloads that were not yet fully optimized, and many others. In the Goldman Sachs' report, they use a kWh/workload metric and state "Data Center efficiency gains and the shift to the cloud/hyperscale have been critical drivers of the moderate increase in data center power demand, but decelerating efficiency gains helped to drive a pickup in power demand in recent years."

[2] Societal digitalization refers to the process of integrating digital technologies into various aspects of society, transforming how individuals interact, communicate, and engage with social institutions.

We call attention to the performance recovery phase beginning in 2026, following the initial trend break. This recovery is driven by continuous innovation from both infrastructure and ICT industries. Some of the continuous drivers for efficiency for both infrastructure and ICT are shown in **Table 1** with many of the ICT drivers also referenced by Goldman Sachs.

Modeling improvement scenarios

With the baseline established, we investigated how further improvements impact energy growth from 2026 to 2030. **Figure 6** illustrates the PUE trend and the ICT performance factor trend. We superimpose accelerated improvements (red dashed lines). These metric improvements become inputs into the three new scenarios.

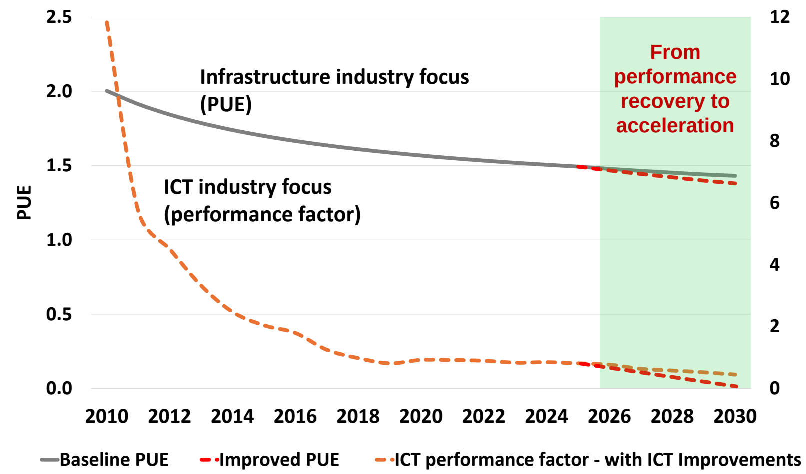


Figure 6

Plot showing existing trends in efficiency improvements. The dashed "red" lines hypothesize what could be the impact of reasonable acceleration of these trends.

Across the data center industry, ICT vendors are publishing data on how their technologies improve performance and/or efficiency. For instance, "the new family of NVIDIA Blackwell GPUs…will enable organizations to train AI models four times faster, improve AI inferencing performance by 30 times and do it with up to 25 times better energy efficiency than NVIDIA's previous generation Hopper™ architecture."[3] There are also academic papers[4] analyzing improvements at the component level.

We believe it is important to go a step further and demonstrate how both performance (ICT) and efficiency (infrastructure) impact the overall energy forecast.

Our scenarios (see **sidebar**) aim to accelerate efficiency trends in both infrastructure and ICT industries, as shown in **Table 1**. They do not rely on breakthrough technologies or disruptions. By taking this more realistic and conservative approach, we hope to foster more actionable collaboration between industries to enable this acceleration. A good example of potential outcomes from collaboration is aligning on liquid cooling supply temperature, flow rates and quality. With a clear, uniform roadmap aligned with the server industry, the infrastructure industry can better tune and time the associated cooling offers.

The four energy curve scenarios

- Baseline forecast scenario
- Scenario 1 - PUE improvements
- Scenario 2 - ICT performance improvements
- Scenario 3 - Both PUE and ICT performance improvements

[3] NVIDIA, Data Center Chips in 2024: Top Trends and Releases

[4] MDPI, Architectural and Technological Approaches for Efficient Energy Management in Multicore Processors

Table 1

Examples of continued efforts by both infrastructure and ICT industries to improve efficiencies

Infrastructure

- **Liquid-cooled servers to save fan energy**

  As Dell discusses, combining liquid cooling with other cooling technologies can significantly reduce the fan power required to cool IT equipment. Supermicro discusses how liquid cooling reduces the need for high-energy fans, leading to lower operational expenses.

- **Higher-supply-temperature cooling equipment to match higher-temperature-tolerant servers**

  Uptime Institute discusses how direct liquid cooling (DLC) systems can operate at higher supply and return water temperatures, which is beneficial for both free cooling and waste heat recovery.

- **Advanced high-temperature chillers with economization**

  Optimizing the chilled water temperature set point can lead to up to 40% energy savings without affecting cooling capacity or indoor climate. Additionally, advanced controls can reduce energy usage (White Paper 225).

- **Higher-density racks, row-based cooling, and containment to capture heat with shortest path**

  White Paper 130 describes how row and rack-based cooling architectures are designed to capture and reject heat more efficiently by minimizing air mixing and directing airflow along the shortest paths. This approach improves hot air capture and ensures that the cooling system operates more effectively (White Paper 208).

- **Continued improvement on UPS and power conversion efficiencies**

  According to White Paper 214, UPS efficiency has progressed from 85-92% years ago to 94-99% today. This has been driven by several technological advancements including eco-mode (White Paper 157).

- **Push to higher final distribution voltages of 415Vac to 480Vac and in the future to 600Vac or 800Vdc**

  A wire distributing 1 kW of power at 415/240Vac has less losses than the same wire at 208/120Vac with double the current. White Paper 128 states that data center efficiency can be improved by increasing final distribution voltages to 415V and higher.

- **Architectures optimized for low temperature climates**

  Data center architectures optimized for low-temperature climates can offer significant advantages over those in higher temperature climates.

ICT

- **Continued acceleration of CPU, GPU, TPU and memory performance**

  GPU performance "has increased roughly 7,000 times" since 2003. CPUs like NVIDIA Grace™ CPU Superchip are delivering breakthrough energy efficiency and performance. The latest TPU, Trillium, offers more than 4.7 times improvement in compute performance per chip compared to the previous generation. NVIDIA's Rubin architecture will utilize advanced memory technologies like HBM4, which will significantly enhance performance and efficiency compared to previous generations.

- **Adoption of accelerated computing platforms tuned for AI and non-AI workloads**

  Supermicro discusses how industries such as healthcare, finance, and automotive are increasingly adopting accelerated computing solutions to tackle complex challenges. IBM discusses how accelerated computing solutions are in high demand across many industries because they can perform calculations faster and more efficiently than traditional central processing units (CPUs). Advanced Processing Units (APUs) combine CPUs and GPUs onto a single die. This integration aims to enhance compute performance for specialized workloads that involve both sequential and parallel processing tasks. With this integration, APUs improve efficiency and speed to better meet the demands of complex applications.

- **Task-specific hardware accelerators (e.g., cryptography, data compression) using the most efficient algorithms**

  MIT discusses how domain-specific hardware accelerators, including those for data compression, can offer significant performance improvements by leveraging specialized operations & parallelism.

- **Chips developed with power-efficient architectures (e.g., 3D stacking & chiplet designs)**

  3D stacking technologies are becoming essential for next-generation high-performance energy-efficient designs, improving on-chip memory capacity and bandwidth. Chiplet technology is emerging as a groundbreaking approach that addresses many of the challenges faced by traditional monolithic System-on-Chip (SoC) designs.

- **Adjustable processor clock speed & voltage based on workload demands**

  Techniques that adjust processor clock speed and voltage based on workload demands, such as Dynamic Voltage and Frequency Scaling (DVFS), improve power efficiency.

- **Intelligent scheduling algorithms to distribute workloads**

  Intelligent load balancing and scheduling algorithms that distribute workloads across different hardware resources can improve compute and energy efficiency.

Scenario 1 - PUE improvements

As a baseline, we used the Uptime Institute survey data for yearly average PUEs to establish an extended PUE forecast from 2010 to 2030 (which closely matched IEA 2014 and 2020 snapshots). This trend line results in an average PUE of 1.49 in 2025.

This scenario models wider adoption of the infrastructure improvements (from **Table 1**) beginning in 2026. This model assumes we achieve PUEs of 1.2 for new data center builds and PUEs of 1.3 for a portion (10%) of existing data centers, by 2030. **This approach leads to a weighted average PUE of 1.38 vs. the baseline of 1.43, by 2030.** This improved trendline is shown in **Figure 7**.

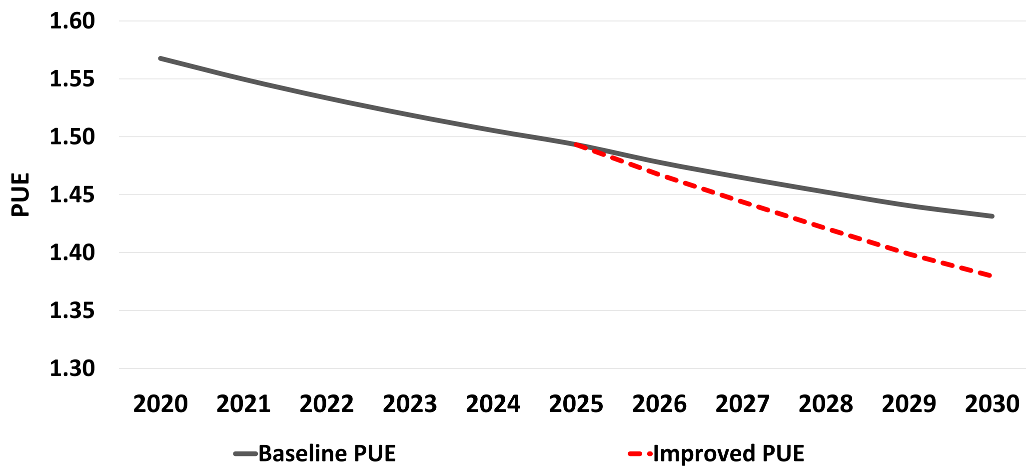


Figure 7

Plot showing PUE baseline trend and the accelerated PUE scenario

Scenario 2 - ICT performance improvements

From the baseline ICT performance factor trend, we modeled the impact of improving this performance factor by 5% in years 2026-2027, 15% in years 2028-2029, and 25% in 2030, This results in a 2030 value of 0.34 TWh/ZB vs. the baseline of 0.45 TWh/ZB through continued ICT actions in **Table 1. Figure 8** displays this improved trendline.

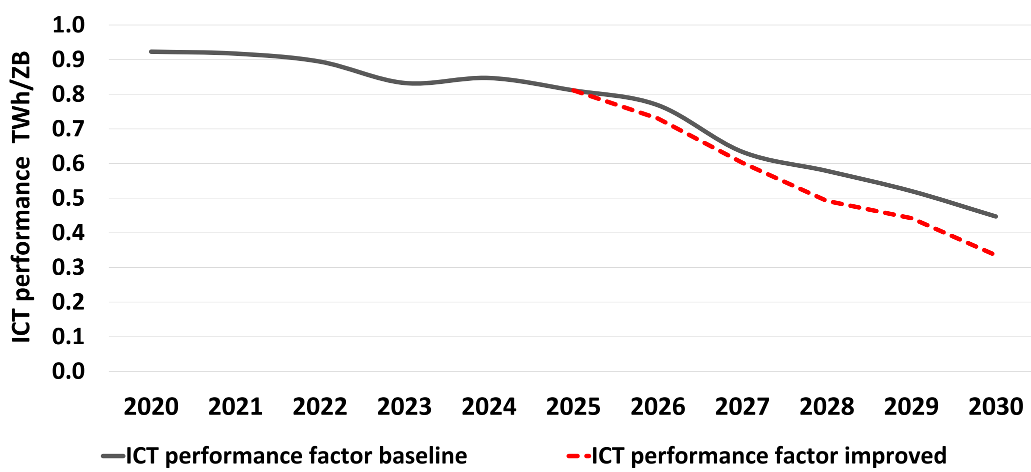


Figure 8

Plot showing existing ICT improvement baseline and the accelerated ICT improvement scenario

Accompanying the continuous ICT improvements from **Table 1**, the ICT performance factor is influenced by the growing portion of AI workloads to support AI inferencing. This workload has a usage profile different than AI training and is a new pathway for further ICT optimization yet to be fully realized. Two examples of companies focused on this specialization are Groq and SambaNova.

Groq specializes in high-performance, low-latency computing with an architecture enabling real-time processing with minimal latency by allowing data to be processed in smaller batches. This approach reduces energy consumption by minimizing data movement. It demonstrates significant performance improvements, such as 18 times faster inference for large language models compared to traditional cloud providers.

SambaNova focuses on integrating purpose-built hardware with advanced software to optimize AI workloads. Their unique architecture maximizes resource utilization and minimizes energy use, enabling higher throughput and lower latency for complex computations. This results in a more sustainable approach to data processing.

Scenario 3 - PUE and ICT performance improvement

This scenario models the total impact on energy if both scenario 1 and scenario 2 improvements are accomplished.

Results

In **Figure 9**, the energy curve is illustrated for the baseline scenario as well as the three improvement scenarios. This view shows **a potential of a total TWh reduction of 17% or 183 TWh** when the combined (modest) PUE and ICT efficiency improvement goals are achieved.

Looking closer, it becomes apparent that while PUE plays a role in bending the curve with a 3.6% or 38 TWh energy savings in 2030, ICT is the major lever, saving 14.4% or 153 TWh.

This is because the physical infrastructure efficiency gains are hitting a point of diminishing returns as PUEs approach 1. As we move beyond 2030, the focus on ICT improvements must continue as an industry. These improvements offer the potential to further flatten or bend the consumption curve. **Figure 10** provides a zoomed-in view of energy savings for the scenarios.

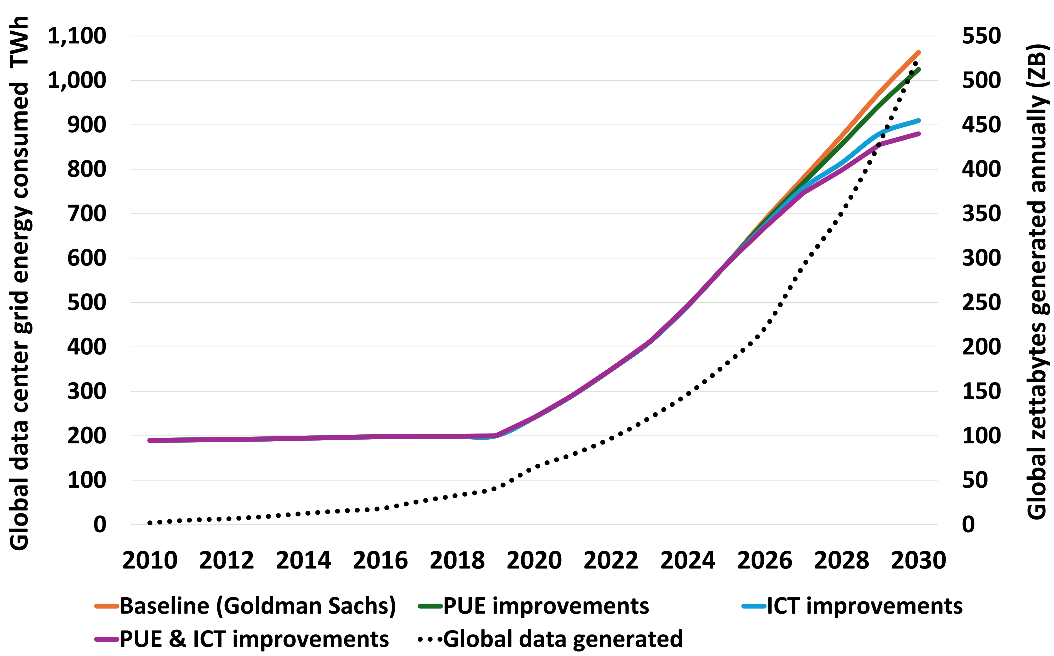


Figure 9

Results of modest acceleration to infrastructure and ICT improvements shows the impact on "bending" the growth curve and showing that, once again, it is possible to decouple data center energy growth from digitalization drivers

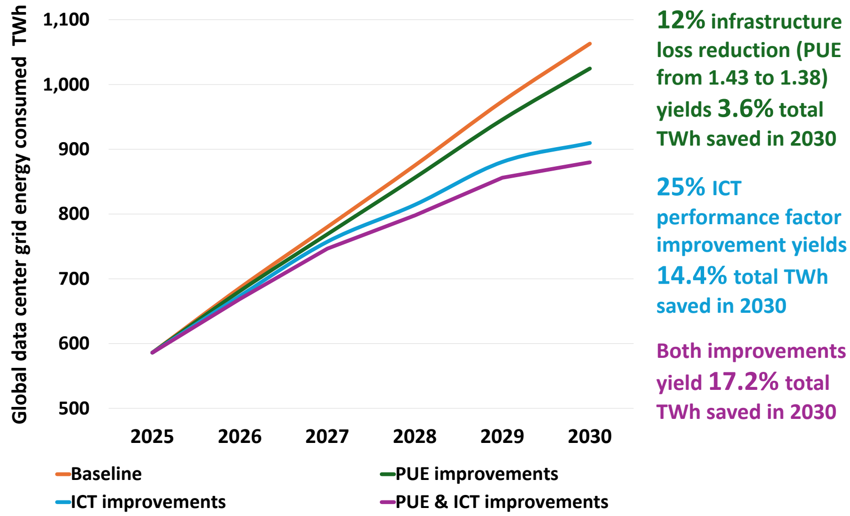


Figure 10

Zoomed in view of the energy saving scenarios

While these results are based on Goldman Sachs' energy forecast, **Table 2** demonstrates that the impact is relatively insensitive to the baseline forecast used. This holds true as long as the baseline data has a similar trend pattern.

Table 2

Comparison of results showing the similarity of global TWh reduction independent of base data source

	Baseline	PUE improvements	ICT improvements	PUE & ICT improvements	Forecast	Total savings
Energy saved from baseline (%)	0%	3.6%	14.4%	17.2%	Goldman Sachs	17.2%
Energy saved from baseline (TWh)	0	38	153	183		
Energy saved from baseline (%)	0%	3.9%	15.6%	18.7%	Schneider global analysis	18.7%
Energy saved from baseline (TWh)	0	77	309	368		
Energy saved from baseline (%)	0%	3.5%	14.0%	16.8%	S&P	16.8%
Energy saved from baseline (TWh)	0	66	261	312		

Because of this independence, we believe this fosters a more direct conversation focused on "bending the curve" driving industry specific actions for meaningful results.

The important role of behind-the-meter power

Although not a reduction in overall energy consumption, further "bending" can be achieved from the perspective of utility grid power. When data centers utilize "behind the meter" power generation, they alleviate pressure on the electrical grid. Each megawatt of generation implemented saves 8.76 TWh. There are now more sustainable options available for continuous grid relief. These include solid oxide fuel cells using natural gas with carbon sequestration, solid oxide fuel cells powered by green hydrogen, and gas turbines equipped with carbon capture technology. Additionally, SMRs are expected to become a viable alternative in the future.

Figure 11 illustrates a "hypothetical" further bend of the energy curve, where behind-the-meter power generation grows from 1% in 2026 to 10% in 2030. This demonstrates a utility viewpoint. Building upon the savings of Scenario 3, adding this energy generation behind the meter, the utility grid is forecasted to see an energy savings of 25.5% by 2030.

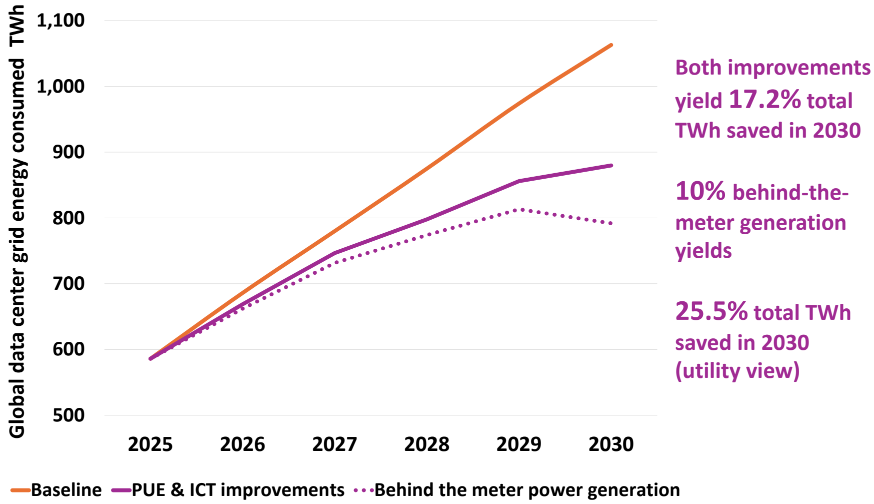


Figure 11

Utility grid view of the energy saving with behind-the-meter energy generation

Next steps

The rising energy consumption of global data centers has captured significant attention, highlighting the urgent need for effective solutions within the industry. However, there is a promising path forward. By understanding the factors driving energy growth and identifying strategies to "bend the curve," proactive measures can be implemented. Addressing these challenges directly will pave the way for a more sustainable future in data center operations. To this end, businesses should consider four proactive steps to enhance efficiency and reduce energy demand.

1. **Set ambitious PUE targets**. Establish ambitious goals for Power Usage Effectiveness (PUE) across all new data center builds, aiming for an average PUE of 1.2 or less, while also targeting improvements in your existing data center fleet to bring the average PUE down. This industry-wide initiative can significantly slow energy growth.
2. **Look at accelerating more efficient ways of computing.** To find efficient computing solutions, organizations can focus on several key strategies.
   - Full adoption of accelerated computing platforms for AI and the use of specialized hardware accelerators can enhance performance while reducing energy consumption.
   - Implementing intelligent scheduling algorithms and optimizing workload configurations will ensure that resources are utilized effectively.
   - Employing chips with power-efficient architectures, adjusting processor clock speeds based on workload demands, and utilizing the most efficient algorithms for specific tasks will further improve overall efficiency in computing operations.
3. **Invest in sustainable** "behind the meter" **energy solutions.** Prioritize investments in "behind the meter" energy generation options, such as solid oxide fuel cells and gas turbines with carbon sequestration, to reduce reliance on the grid and lower overall energy consumption. Also learn about the potential for SMRs for the future.
4. **Foster industry collaboration.** Engage in industry-led initiatives aimed at decoupling data growth from energy consumption, promoting collective efforts to achieve efficiency improvements that can bend the energy growth curve by at least 17% by 2030. This collaborative approach will support both sustainability goals and economic growth associated with digitalization.

About the authors

Jim Simonelli is Senior Vice President & Chief Technology Officer for Schneider Electric's Data Center and Networks division and segment. In this role, he is responsible for deployment of the R&D strategy for hardware, software and service offers and for forward-looking technology exploration activities. Jim is a long-time veteran with Schneider Electric holding many leadership positions relating to technology and innovation in the Data Center critical power and cooling business arena. Jim holds a BS and MS in Power Electronics from Worcester Polytechnic Institute.

Wendy Torell is a Senior Research Analyst in Schneider Electric's Data Center Research & Strategy group bringing 30 years of data center experience. Her focus is analyzing and measuring the value of emerging technologies and trends: providing practical, best practice guidance in data center design and operation. Beyond traditional thought leadership, she championed and leads development of interactive, web-based TradeOff Tools. These calculators help clients quantify business decisions, while optimizing their availability, sustainability, and cost of their data center environments. Her deep background in availability science approaches and design practices helps clients meet their current and future data center performance objectives. She brings a wealth of experience across Schneider Electric's broad portfolio and with the market at large. She holds a BS in Mechanical Engineering from Union College and an MBA from University of Rhode Island. Wendy is an ASQ Certified Reliability Engineer.

Victor Avelar is a seasoned expert in data center energy efficiency and design, serving as the Chief Research Analyst at Schneider Electric's Data Center Research & Strategy group. With over 25 years of experience, Victor leads cutting-edge research and best practice development for sustainability, risk management, and next-generation data center technologies. He's a trusted advisor to clients globally, providing actionable insights on enhancing infrastructure performance through innovative solutions such as liquid cooling and energy modeling. Known for his clear, practical guidance, Victor helps organizations tackle the evolving challenges of sustainable and efficient data center operations. He is central to the development of technology adoption forecasts for data centers. He also leads the peer review process for all EMRC content. Victor holds a bachelor's degree in mechanical engineering from Rensselaer Polytechnic Institute and an MBA from Babson College. He is a member of AFCOM and a sought-after speaker on AI infrastructure.

RATE THIS PAPER ★★★★★

Resources

[Artificial Intelligence and Electricity: A System Dynamics Approach](#)
Report

[High-Efficiency AC Power Distribution for Data Centers](#)
White Paper 128

[Choosing Between Room, Row, and Rack-based Cooling for Data Centers](#)
White Paper 130

[Eco-mode: Benefits and Risks of Energy-saving Modes of UPS Operation](#)
White Paper 157

[Small Modular Nuclear Reactors Suitability for Data Centers](#)
White Paper 186

[How Row-based Data Center Cooling Works](#)
White Paper 208

[Guidance on What to Do with an Older UPS](#)
White Paper 214

[Optimize Data Center Cooling with Effective Control Systems](#)
White Paper 225

Browse all white papers
whitepapers.apc.com

Browse all TradeOff Tools™
tools.apc.com

Note: Internet links can become obsolete over time. The referenced links were available at the time this paper was written but may no longer be available now.

Contact us

For feedback and comments about the content of this white paper:

Schneider Electric Data Center Research & Strategy
dcsc@schneider-electric.com

If you are a customer and have questions specific to your data center project:

Contact your Schneider Electric representative at
www.apc.com/support/contact/index.cfm

Appendix: Analysis methodology

In this section we provide a more detailed methodology including data sources and assumptions used to create the baseline model and the three improvement scenarios.

Baseline

The baseline energy curve illustrates the historic trend of energy consumption trend, highlighting a marked increase in energy use into the future, through to 2030. The curve was derived from Goldman Sachs (future) and IEA (historic) data, using linear extrapolation for gap years. Since most available forecasts don't provide data for this complete time range, we merged the two data sets. The IEA data set also served as the historic trendline for other baseline forecasts (see **Figure 10**).

The energy curve also allows us to break down energy consumption into new and existing data centers. We assumed that all of the incremental year-over-year energy consumption was attributed to new data centers, while the remaining consumption came from existing data centers. For example, 241 TWh were consumed in 2020, and 200 TWh were consumed in 2019. This means that in 2020, 41 TWh of energy was consumed by new data centers and 200 TWh by existing data centers.

To model the energy savings from PUE and ICT improvements, we first need to break down the total energy baseline into physical infrastructure and ICT energy. Having a baseline PUE forecast allows us to do this with the formula: Total energy / PUE = ICT energy. Annual PUE data was derived from a rational function curve fit of Uptime Institute survey data. We assumed that new and existing data centers used the same baseline PUE trend up to and including 2025. After 2025, new data centers follow a baseline PUE trajectory of 1.3 by 2030, while existing data centers use baseline PUEs that target 1.44 by 2030.

We are now able to break down the physical infrastructure and ICT energy into four categories for each year: new ICT & physical infrastructure (current year) and existing ICT & physical infrastructure (carried over from the previous year). Like all data centers, we accounted for ICT refresh cycles. We assumed that ICT equipment is replaced every four years. Therefore, every year ¼ of the previous year's ICT energy is categorized as refreshed ICT. This means that, for years 2026 to 2030, the ICT improvement factor used in the "ICT performance improvements" scenario is applied to incremental new ICT as well as the refreshed ICT.

PUE improvements

Using an improved target PUE of 1.2 for new data centers, we created a revised PUE trendline from 2026 to 2030, using linear interpolation. Likewise, we created a revised PUE trendline from 2026 to 2030 for existing data centers. We apply these improved PUEs to 10% of existing data centers. See **Table A1** below. These PUEs where used to calculate the lower physical infrastructure energy and therefore lower data center energy for years 2026 to 2030. Note, years prior to 2026 use baseline values.

Table A1

PUE comparison

	2026	2027	2028	2029	2030
Baseline PUE for new data centers	1.45	1.42	1.38	1.34	1.30
Improved PUE for new data centers	1.43	1.38	1.32	1.26	1.20
Baseline PUE for existing data centers	1.48	1.47	1.46	1.45	1.44
Improved PUE for existing data centers	1.45	1.42	1.38	1.34	1.30
Baseline weighted average PUE	**1.48**	**1.46**	**1.45**	**1.44**	**1.43**
Improved weighted average PUE	**1.47**	**1.44**	**1.42**	**1.40**	**1.38**
Percent improvement from baseline	**0.7%**	**1.4%**	**2.2%**	**3.0%**	**3.7%**

ICT performance factor improvements

As discussed in the body of this paper, the ICT performance factor provides a means of quantifying the ICT energy consumption per unit of data generated (TWh/ZB). We calculated it every year by dividing new ICT energy (including refreshed ICT) by that year's data growth. It enables us to clearly see inflection points over time. The annual data growth (the denominator) was derived from Statistica data using a Gaussian curve fit.

From the baseline ICT performance factor trend, we modeled the impact of improving this performance factor, by 5% in years 2026-2027, 15% in years 2028-2029, and 25% in 2030. This results in a 2030 value of 0.34 TWh/ZB vs. the baseline of 0.45 TWh/ZB. **Figure 7** illustrates this improved trendline.

Reduced ICT energy is calculated by multiplying the data generated (ZB) in that year by the improved performance factor (TWh/ZB). A beneficial side-effect of lower IT energy is a decrease in the physical infrastructure energy to support the IT. Note, years prior to 2026 used baseline values.