Appears in *The 3rd Workshop on Hot Topics in System Infrastructure (HotInfra'26)*

Power Breakdowns and PUE Implications of Air-Cooled and Direct-Liquid-Cooled Servers

August Ning
EcoCloud, EPFL
Lausanne, Switzerland
august.ning@epfl.ch

Xavier Ouvrard
EcoCloud, EPFL
Lausanne, Switzerland
xavier.ouvrard@epfl.ch

Babak Falsafi
EcoCloud, EPFL
Lausanne, Switzerland
babak.falsafi@epfl.ch

Abstract

Power usage *effectiveness* (PUE) has guided years of datacenter energy efficiency improvements, but only differentiating between facility and IT power obscures insights into how IT power is used within the servers and may misdirect optimizations. In this work, we measure air-cooled and direct liquid-cooled (DLC) servers and break down their power into power supply (PSU) loss, fan, idle, and compute power via power distribution unit (PDU)-level measurements. We apply our insights to show how high ambient temperature operation affects server power profiles and distorts PUE: increasing fan power improves PUE even though energy efficiency may decrease, while transitioning to DLC can worsen PUE despite reducing total datacenter energy consumption.

1 Introduction and Motivation

Datacenters have rapidly become one of the largest and fastest-growing consumers of electricity worldwide. Datacenter power consumption growth has outpaced power grid capacity growth, leading to delayed build-out timelines, increased electricity costs for operators and household consumers alike, and lasting implications on society as a whole. The International Energy Agency projects global datacenter electricity consumption to reach nearly 1 petawatt-hour by 2030 [10], which is roughly the current annual energy consumption of all European and African households combined [3, 5, 6, 9]. Datacenter operators are building gigawatt-scale campuses and must negotiate with utility companies and/or construct their own power generation to bring them online [18].

This pressure has made power efficiency a first-class datacenter design principle, and PUE [20], defined as the ratio of total facility energy to IT equipment energy, is the industry standard for measuring datacenter power efficiency.

$$\mathrm{PUE} = \frac{\mathrm{E_{Total}}}{\mathrm{E_{IT}}} \in [1, \infty)$$

The best possible PUE is 1, where all facility power is consumed by IT equipment. Figure 1 shows a high-level datacenter power breakdown. In practice, hyperscalers report fleet-wide PUEs approaching 1.2 and lower [7], and the metric has driven sustained investments in reducing non-IT power consumption. Despite PUE's wide adoption, it does not further break down IT power, limiting its insights for server designers. Furthermore, operators can improve PUE in misleading ways: higher ambient temperature decreases cooling costs by increasing servers' fan power (considered IT power), improving energy efficiency and PUE. However, IT operators can continue to lower PUE by increasing fan speeds even if compute performance and facility cooling savings have already saturated, leading to decreased overall energy efficiency.

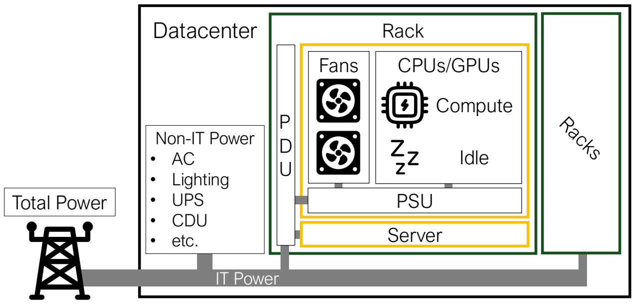

Figure 1. Datacenter power breakdown.

The Open Compute Project's (OCP) Datacenter and Computing Efficiency Metrics (DCEM) workstream [14] introduced infrastructure utilization *efficiency* (IUE) [13], a metric designed to scale across server-, rack-, and datacenter-level energy efficiency evaluation. IUE breaks down server power into PSU loss, fan, idle, and compute power, and is defined as the fraction of compute power relative to total power:

$$\mathrm{IUE} = \frac{\mathrm{E_{Compute}}}{\mathrm{E_{PSU\ Loss}} + \mathrm{E_{Fans}} + \mathrm{E_{Idle}} + \mathrm{E_{Compute}}} \in [0, 1]$$

A higher IUE means more server power is used for compute and provides a relevant metric for system and computer architects to guide their designs. IUE extends Patterson et al.'s IT-power usage effectiveness (ITUE) [16], which decomposes IT power into compute and overheads, by isolating fan power and PSU loss and further breaking down ITUE's compute power into idle and compute components.

In this work, we measure air-cooled and DLC servers and characterize their power draw according to the four

The 3rd Workshop on Hot Topics in System Infrastructure (HotInfra'26), co-located with ISCA'26, June 28, 2026, Raleigh, NC, USA

Table 1. Server configurations

Server	HPE ProLiant DL380 Gen11 [8]
CPUs	Dual-socket Intel Xeon Gold 6448Y
RAM	256 GB
Fans*	HPE High Performance Fan Kit (6× fans)
PSU†	2× 1600 W HPE Artesyn (80 Plus Platinum)
PDU	RNX RN3517
CDU‡	CoolIT CHx200
Coolant	25% propylene glycol

* SUNON VG60561BX-Q155-Q9T
† Our experiments only use one PSU to isolate PSU loss.
‡ CDU only required for the DLC server.

components of IUE. Rather than requiring dedicated instrumentation on each component, we propose a methodology that calculates expected power draw based on PDU-level measurements via controlled benchmarks to isolate fan, idle, and compute power. Furthermore, we independently verify PSU efficiency curves against 80 PLUS certification [4].

We use our characterization to project implications for datacenter operators and server designers. First, we show that increasing ambient temperatures (in order to reduce facility cooling energy) requires increasing server fan speeds, resulting in fan power accounting for 10% of server power in high-ambient-temperature co-location datacenters. Second, we find that transitioning from air cooling operations to DLC can actually *increase* reported PUE because of lower server fan power (which PUE counts as IT energy) and additional coolant distribution unit (CDU) pump power (which PUE counts as non-IT overhead), even though estimated overall datacenter power consumption is lower. Third, we argue that server fan power is fundamentally a cooling cost and propose a fan-adjusted PUE that reclassifies it as non-IT power. With this metric, datacenters that use less power on facility and server cooling achieve a lower PUE which better aligns with good datacenter design practice.

2 Server Power Breakdown

We characterize two identical production-grade CPU servers with one server's CPUs configured with cold plates for direct liquid cooling (other components such as RAM and PSUs remain air-cooled). Table 1 shows our servers' configuration.

PSU Loss: PSUs cannot perfectly convert the PDUs' AC power to DC. Although programs such as 80 PLUS [4] certify PSUs that achieve at least 80% power conversion yield (ratio of the PSU's output power to its input power) at varying load points, the actual efficiency varies based on load. We characterize PSU loss in an isolated setup with one PDU and one PSU (removed from the server), sweep the PSU's output load in 10% increments of the total rated power, and measure the PDU's output AC power, the PSU's input AC power, and the PSU's output DC power. Figure 2shows the PSU's yield against PSU output load. At 10% load (~160 W), the PSU's yield is 0.886, and yield from 20% and 100% load ranges between 0.913-0.941. We estimate a 0.2% yield loss within the cable between the PDU and PSU. While narrow, this range still matters at datacenter scale where small yield variations compound into substantial energy waste.

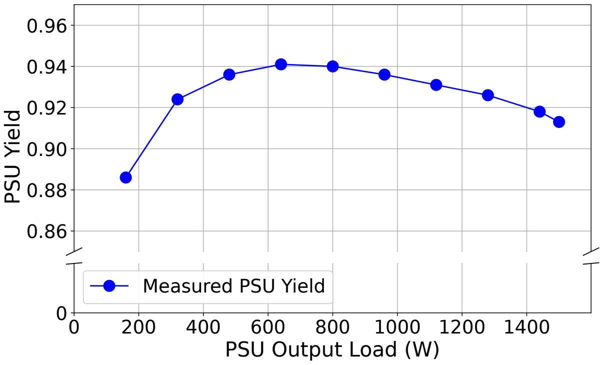

Figure 2. PSU yield against PSU output load.

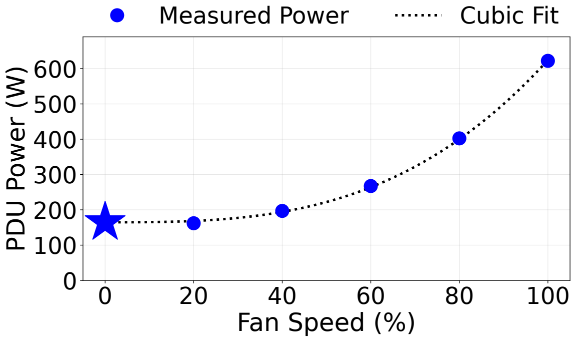

Figure 3. Air-cooled server's PDU's power against fan speed when the processors are not running workloads. The blue star indicates the extrapolated idle power.

Fan Power: Server fans cool internal components by blowing ambient air across the server chassis, and their power consumption scales cubically with fan speed [2, 15]. Although the intelligent platform management interface (IPMI) cannot directly measure fan power, we can measure the fan speed and use a cubic fit to estimate fan power. Each server is equipped with six high-performance fans which are required for our configuration (CPU TDP > 205 W or RAM capacity > 128 GB [8]) and each fan draws 72 W at 100% speed; server fan power can reach up to 432 W, which almost matches the combined TDP of the CPUs. However, at 24°C ambient operating conditions, server fans run at ~20% speed [17] and only consume 3 W total.

Idle Power: We define idle power as when the processors are powered on and not running any user workloads and when the fans are not spinning. This state also captures the power

Table 2. Case study assumptions

Metric	Hyperscaler	Co-location
PUE (@ 24°C)	1.1	1.3
Server Utilization	60%	15%
DLC Fan Speed	20%	20%
DLC CDU Power	20 W	20 W

overheads of RAM and other compute peripherals. Isolating fan power is complicated by the fact that fans run even when the server idles. Therefore we estimate idle power by idling the processors, independently sweeping minimum fan speed from the IPMI, and measuring the server's PDU power. We then apply a cubic fit to measurements and extrapolate the server's power when fans are off; our results are shown in Figure 3. From our fan-off PDU measurement, we account for the PSU loss to find idle power: the idle power for the air-cooled and DLC servers are 155 W and 165 W, respectively. **Compute Power:** Compute power is the portion of server power that directly performs useful work and, by definition, is the balance of the server power once PSU loss, fan power, and idle power are discounted. We define server utilization as the percentage of physical cores that are actively running workloads. In our servers, we are able to decouple fan and compute power because even at 100% server utilization, fan speed does not increase above 20%. In our case studies, our servers run Ubuntu 24.04.4 LTS and we use stress-ng's [11] `matrix` workload to load processor cores and increase server utilization by replicating the benchmark on additional cores. For each measurement, we run the server for 10 minutes and measure the average of the PDU power.

3 Case Studies

Assumptions: In this section, we use our breakdown methodology to estimate server power profiles and PUE implications for hyperscaler and co-location datacenters. Table 2 summarizes our assumptions. Hyperscalers own and operate the entire datacenter, allowing them to co-design servers, facilities, and workload scheduling and resulting in high server utilization and efficient PUE. Co-location datacenters lease rack space and power to tenants who supply their own servers. The operator controls only the facility (cooling, power distribution) while tenants control workloads, resulting in lower average server utilization and higher PUE.

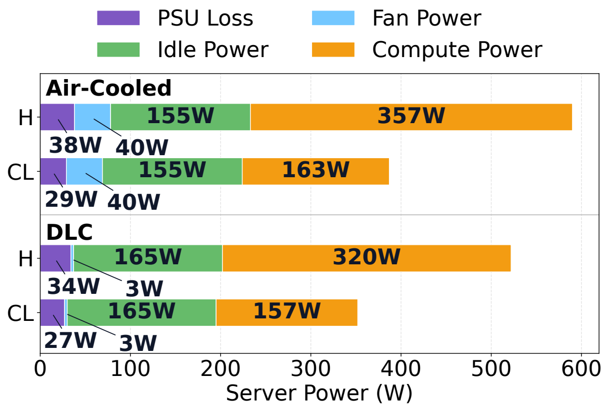

Figure 4. Projected server power breakdown for hyperscaler and co-location datacenters operating at 32°C.

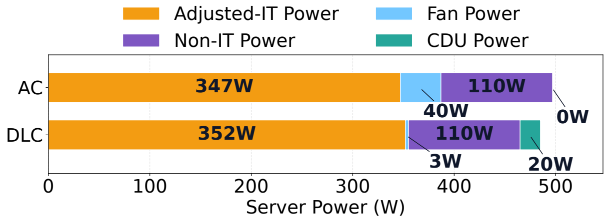

Figure 5. Projected per-server power breakdown for co-location datacenters operating at 32°C. IT power = adjusted-IT power + fan power.

Assuming a datacenter with identical racks and cooling solutions and negligible PDU loss, we use these PUE values to estimate per-server non-IT power. Servers that liquid cool their CPUs only require fans to cool peripherals and can sustain low fan speeds even at elevated ambient temperatures but require CDUs to pump coolant and add non-IT power overheads. We conservatively assume our CDU requires 200 W of pump power for a rack with ten servers, resulting in 20 W of CDU power per server.

High Ambient Temperatures: Traditionally, datacenter ambient temperatures range between 18°C-27°C [1], but operators have been elevating ambient temperatures (up to 32°C [12]) to reduce facility cooling costs. To model high ambient temperature operation, we project air-cooled server fan speed based on vendor data [17]. When operating air-cooled servers at 32°C, fan speed increases to 45% to adequately cool components [17]. From our power breakdowns, we expect fan power to increase to 40 W, a >10× increase over the 24°C baseline. Figure 4 shows the hyperscaler and co-location power breakdowns. For hyperscaler servers, fan power accounts for 6.7% of server power, but for co-location servers running at lower utilization, fan power grows to 10.2%. Transitioning to DLC reduces IT power by 11.1% and 8.3% for hyperscaler and co-location operations, respectively.

PUE Implications: Figure 5 shows the per-server power breakdown of the co-location datacenter operating at 32°C. Air-cooled operation sees PUE *improve* to 1.28 because the additional fan power increases IT power while DLC operation reaches PUE = 1.37 because IT power is lower and CDU power is accounted for as non-IT power. The DLC datacenter therefore reports a worse PUE despite consuming less total power (485 W vs. 497 W per server). This scenario exposes a fundamental flaw in how PUE accounts for fan power as IT power rather than cooling power. By simply increasing fan power, operators can improve PUE even though energy efficiency may decrease.

To correct this, we propose a fan-adjusted PUE that reclassifies server fan power as non-IT:

$$\mathrm{PUE_{FA}} = \frac{\mathrm{E_{Total}}}{\mathrm{E_{IT}} - \mathrm{E_{Fans}}} \in [1, \infty)$$

Under this metric, the air-cooled datacenter's $\mathrm{PUE_{FA}}$ rises to 1.43 compared to the DLC datacenter's 1.38. $\mathrm{PUE_{FA}}$ and IUE are not meant to be ironclad sustainability metrics to replace PUE, but rather illustrate how small accounting differences can significantly distort PUE and why efficient datacenter design needs to further break down IT power.

4 Conclusion and Future Work

In conclusion, we characterize air-cooled and DLC servers to understand their power profiles and to show the PUE implications of high ambient temperature operation. Breaking down server power into PSU loss, fan, idle, and compute power provides more relevant insights to system and computer architects in our pursuit of energy-efficient server design. Servers keep redundant PSUs online to provide automatic failover, so understanding their yield curves can drive load-based power distribution that minimizes PSU loss. From Figure 4, idle power ranges between 26.3%-46.5% of total server power. Reducing idle power is a full-system challenge demanding coordination among the OS, processors, memory, and peripherals.

For future work, we plan to control ambient temperature to verify expected fan behavior and characterize the tradeoffs between server fan power and facility cooling power; characterize GPU servers and their impact on fan power, as GPU servers require higher fan speeds [19] and are driving the transition towards liquid cooling; extend our comparison to immersion cooling servers, which eliminate fan power at the cost of additional cooling power and capital expenditures; and characterize representative datacenter workloads on x86 and ARM platforms.

Acknowledgments

We would like to thank Baihan Liu and Prof. Drazen Dujic of the EPFL Power Electronics Laboratory for their help characterizing PSU loss and our anonymous reviewers for their comments and feedback. The experiments in this work were conducted at the EcoCloud experimental facility; EcoCloud is an EPFL research center focusing on energy efficient and sustainable IT infrastructure. The authors are OCP DCEM workstream members and portions of this work were previously presented at the 2026 OCP EMEA Summit.

References

[1] ASHRAE. 2021. 2021 Equipment Thermal Guidelines for Data Processing Environments. Fifth Edition, Revised and Expanded.

[2] Richard Aynsley. 2002. Fan size and energy efficiency. *International Journal of Ventilation* 1, 1 (2002), 33–38. https://doi.org/10.1080/14733315.2002.11683619

[3] BloombergNEF. 2024. Africa Power Transition Factbook 2024. Accessed: May 27, 2026.

[4] CLEAResult. 2026. 80 PLUS Power Supply Certification Program. https://www.clearesult.com/80plus/ Accessed: May 27, 2026.

[5] Eurostat. 2025. Energy Use in EU Households Down Second Year in a Row. https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20250625-2. Accessed: May 27, 2026.

[6] Eurostat. 2026. Energy Consumption in Households. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_consumption_in_households. Accessed: May 27, 2026.

[7] Babak Falsafi. 2025. To Measure is To Know: Breaking Down Datacenter Power Consumption. Computer Architecture Today (ACM SIGARCH Blog). https://www.sigarch.org/to-measure-is-to-know-breaking-down-datacenter-power-consumption/ Accessed: May 27, 2026.

[8] Hewlett Packard Enterprise. [n. d.]. HPE ProLiant DL380 Gen11 QuickSpecs. https://www.hpe.com/psnow/doc/a50004307enw. Accessed: May 27, 2026.

[9] International Energy Agency. 2026. Africa – Countries & Regions: Electricity. https://www.iea.org/regions/africa/electricity. IEA Regional Analysis. Accessed: May 27, 2026.

[10] International Energy Agency. 2026. *Key Questions on Energy and AI*. Technical Report. IEA. https://www.iea.org/reports/key-questions-on-energy-and-ai Accessed: May 27, 2026.

[11] Colin Ian King. 2026. stress-ng. https://github.com/ColinIanKing/stress-ng/. Version 0.17.06 (gcc 13.2.0, x86_64 Linux 6.8.0-111-generic).

[12] Sebastian Moss. 2022. Meta ups server room temperatures to 90 degrees F, in effort to reduce water usage. https://www.datacenterdynamics.com/en/news/meta-ups-server-room-temperatures-to-90-degrees-f-in-effort-to-reduce-water-usage/. Accessed: May 27, 2026.

[13] Murugasamy Nachimuthu. 2024. Sustainability Metrics - Infrastructure Utilization Efficiency (IUE) Considerations. OCP Regional Summit. Open Compute Project, Data Center Efficiency Metrics (DCEM) Workstream.

[14] Open Compute Project. 2026. OCP Sustainability Project. https://www.opencompute.org/w/index.php?title=OCP_Sustainability_Project. Accessed: May 27, 2026.

[15] Michael K. Patterson. 2008. The effect of data center temperature on energy efficiency. In *2008 11th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*. 1167–1174. https://doi.org/10.1109/ITHERM.2008.4544393

[16] Michael K. Patterson, Stephen W. Poole, Chung-Hsing Hsu, Don Maxwell, William Tschudi, Henry Coles, David J. Martinez, and Natalie Bates. 2013. TUE, a New Energy-Efficiency Metric Applied at ORNL's Jaguar. In *Supercomputing*, Julian Martin Kunkel, Thomas Ludwig, and Hans Werner Meuer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 372–382.

[17] Hasnain Shabbir, Carlos Henry, Danny Whittington, and Jon Brown. 2019. *Custom Cooling Fan Options for Dell EMC PowerEdge Servers*. Technical White Paper. Dell Inc. https://dl.dell.com/manuals/common/customcooling_poweredge_idrac9.pdf

[18] Matthew S. Smith. 2026. The 5-GW Data Center. *IEEE Spectrum* (March 2026). https://spectrum.ieee.org/5gw-data-center Accessed: May 27, 2026.

[19] Super Micro Computer, Inc. 2025. *Comparison of Air-Cooled versus Liquid-Cooled NVIDIA GPU Systems*. White Paper. Supermicro. https://www.supermicro.com/white_paper/white_paper_SMCI_

Air_VS_Liquid-Cooled_NVIDIA_GPU_Systems.pdf

[20] The Green Grid Technical Committee. 2007. *Green Grid Metrics: Describing Data Center Power Efficiency*. White Paper WP#1. The Green Grid. https://www.thegreengrid.org/resources/library-and-tools/green-grid-metrics-describing-data-center-power-efficiency