



IMPLEMENTING A GREEN DATA CENTER

INNOVATIVE SERVER DESIGNS LOWER POWER USAGE AND REDUCE E-WASTE



TABLE OF CONTENTS

Executive Summary 1
Server Path Forward2
Supermicro Product Design 2
Supermicro Product – Leading With Innovation4
Supermicro Systems w/Resource Saving Architecture7
Lifetime Electricity Usage9
Customer Examples-Reducing PUE 12
Summary13

SUPERMICRO

Supermicro (Nasdaq: SMCI), the leading innovator in high-performance, highefficiency server and storage technology is a premier provider of advanced server Building Block Solutions® for Enterprise Data Center, Cloud Computing, Artificial Intelligence, and Edge Computing Systems worldwide. Supermicro is committed to protecting the environment through its "We Keep IT Green®" initiative and provides customers with the most energy-efficient, environmentally-friendly solutions available on the market.

SUPERMICRO PRODUCTS REDUCE ENVIRONMENTAL IMPACT

Innovative Server Designs Lower Power Usage and Reduce E-Waste



Executive Summary

Climate change is affecting billions of people, and the contribution to this phenomenon by greenhouse gasses is well understood. Using finite resources (oil, natural gas, and coal) for electricity generation produces CO2, which traps heat in the earth's atmosphere. Data centers, comprising thousands of servers, are huge electricity consumers. According to a Forbes report, data centers use about 1% of all electricity¹.

According to the International Energy Agency, data centers consume approximately 200 terawatt-hours (TWh) of electricity, or nearly 1% of global electricity demand, contributing to 0.3% of all global CO2 emissions. Estimates for the power used by data centers in the near future range from 353 TWh ² to 1,000 TWh in 2025³. Another study ⁴ estimates data center power requirements are growing at about 6.9 % per year.

Server Path Forward

Although the performance per watt improvement of the latest generation of CPUs continues to improve, there are several reasons why servers from Supermicro can help organizations reduce their electricity usage and, therefore, costs. In addition, depending on the type of power purchased (fossil fuels or renewable), a reduction in CO2 released into the atmosphere can also be accomplished. CO2 is classified as a greenhouse gas (GHG) known to trap heat in the earth's atmosphere.

Creating a computer server (or workstation) is a complex process involving worldwide supply chains. The environmental impact of a production process includes the GHG emissions in the production of all components, transportation to the final destination, the packaging of the shipping box, and retirement when the server has ended its useful life (E-waste). Over the life of a server, about 75% to 80% of the GHG emissions can be attributed to running the systems in a data center. Therefore, a plan to reduce the power consumption and PUE of servers and the data center is critical for any CIOs and CFOs.

PUE = (Power Usage Effectiveness) measures the effectiveness of overall power needed for a Data Center, compared to the power required for the IT gear, the power for servers, storage, networking, etc.

PUE is defined as the (Total Electricity into a data center) / (Electricity used by IT components)

Data centers typically have PUEs averaging 1.59⁵, with the most efficient deployments measuring in the 1.10 or lower range, with the upper range around 2.2. This means that at best, 10% of the power used by a data center is going toward air conditioning and other non-IT needs. At the upper end, more electricity is being used to cool the IT equipment than is being used by the IT equipment. Decreasing the PUE of large data centers is an ongoing effort and taking advantage of new technologies is a crucial component of this critical effort.

Supermicro Product Design

While an industry-standard server looks simple from the outside, the design for the internals is complex and involves knowledge of electronics, mechanical design, airflow simulation, thermal dynamics, and packaging. Supermicro uses advanced computational fluid dynamics (CFD) simulations, including complex algorithms that simulate the paths of air molecules overall heat-sensitive components (especially the most challenging components like the CPUs & GPUs), and through the chassis to optimize the performance and efficiency of our servers. This CFD simulation shows the warm air crossing over the electronics and detecting airflow obstructions. The result is airflow traces, as shown below. Note that these simulations can also be used to measure the airflow and cooling characteristics of an entire data center.









Figure 1 - Airflow with CFD Simulatoin of Supermicro Server

Servers are built with several sub-systems that must all

Figure 2 - CFD Simulation of airflow in a data center.

work together and should be designed to reduce power consumption yet deliver the required SLAs.

Power Supply Considerations

The choice of power supplies affects the power consumption of each server. A power supply is rated (Titanium, Platinum, Gold), which refers to the efficiency when converting AC power to DC power. Power supplies are typically most efficient at 50% loads. While a base power supply will only be 80% efficient at this load, a Titanium power supply will be greater than 96% efficient. All Supermicro servers are designed to utilize 80 Plus Platinum power supplies, which waste less electricity in the AC to DC conversion process, enabling a more efficient overall server. Supermicro optimizes the design to support the most efficient power supplies and gives customers a choice if their budget cannot accommodate the incremental cost.

Next generation technologies produce higher performance, and due to the laws of physics, these technologies consume more power. Supermicro's most demanding customers want these technologies in the same or similar density. The Big Twin, for example, will use 2x 3,000 Watt Titanium Power Supplies. This will be a 13% increase in power density while maintaining Titanium level efficiency.

Fan Considerations

The most common way to cool a server is to draw cooler air over the hot electronics. An efficient data center will have cold aisles, where the air temperature is cool and hot aisles, where the air temperature is hot, from the exhaust from the servers. Fans are located at the rear of the server, drawing the cooler air from the front across the electronics where the air warms up. The size of the fan and the rotating speed determine the Cubic Feet per Minute (CFM) of airflow over the CPUs, GPUs, Memory, Storage, and Networking hardware. The needed CFM can be calculated from the inlet temperature and the amount of heat produced.

$$Q = Airflow \times deltaT \times c_p \times \rho \qquad \qquad Q = \frac{Airflow \times deltaT \times 1.08}{3,412}$$

A more detailed discussion of the amount of airflow required over the hot electronics is found in this article ⁶.

Overall, the mechanical design of a server that includes the airflow, fan size and speed, and the power supplies significantly affect the electricity that a server requires for optimum performance. Therefore, Supermicro invests heavily in the design of servers, which gives maximum performance and reduces the electricity draw.



When possible, use larger fans because larger fans can provide the same CFM at slower RPMs and less power. With fans consuming as much as 15% of the system power, multi-node systems like the Supermicro Big Twin and Supermicro SuperBlades are the most efficient platforms of choice in the most energy air-cooled data centers.

Overall, the strategic design of a server that includes the airflow, fan size, fan speed, and the power supplies significantly affect the energy efficiency that a server requires for optimum performance. This is why Supermicro invests heavily in servers' in-house design and engineering that maximize performance while reducing energy consumption.

Supermicro Products - Leading with Innovative and purpose built design

Supermicro is known as having the first to market servers available containing the latest CPUs and GPUs from Intel, AMD, and NVIDIA. Supermicro designs the most advanced server and storage systems available today. By using the most advanced CPUs and GPUs available, customers are able to perform more work per watt than ever before. The industry-standard technologies that are used ensure compatibility, and the Supermicro design expertise gives customers the best performance per watt per dollar.

Modular designs allow for the sharing of components that contribute to electricity usage. This approach also enables specific and optimized configurations for every workload and environmental requirement. For example, since larger components (i.e., fans and power supplies) operate more efficiently than smaller components, it is also critical that fans and power supplies only deliver the power needed at a given time. This drives energy efficiency yet allows for scaling within a chassis.

While standard 1U or 2U servers remain popular, multi-node and blade servers offer energy-saving efficiencies not available in typical rack-level servers. A multi-node server contains up to 4 or 8 nodes, independent servers housed in a larger chassis. The larger chassis allows for larger and shared fans and power supplies, which are more efficient. For example, the Supermicro BigTwin system is shown below, with four servers installed. Supermicro offers a number of multi-node servers, containing from 2 to 8 servers and from 2U to 4U chassis.



Figure 3 - Supermicro BigTwin with 4 Nodes

Supermicro SuperBlades offer a similar solution as a multi-node system but with even higher densities. A Supermicro SuperBlade chassis can contain up to 20 blades and 40 CPUs in just a height of 8U. The SuperBlade also includes a network switch internal to the chassis, relying on the shared power supply and fans with shared power supplies and larger fans. This increases the power efficiency of the entire system.





Figure 4 - Supermicro SuperBlade - Front View - 20 Blades



Figure 5 - Supermicro SuperBlade - Rear View

In summary, server configuration optimized server designs drive

excellent energy efficiency, but couple this with the broadest multimode portfolio, and the results are superior energy efficiency at the right configuration for customers' demands

Besides the technological advances which lead to more cores, faster data speeds, new instructions, and higher communication rates, these new processors deliver higher performance per watt than previous generations. Thus, for the same amount of "work," less power is used by the CPU, reducing server electricity requirements. Likewise, more data processing can be delivered for a similar power draw. The chart below (courtesy of Intel) shows that the amount of work performed per watt increases over time, allowing for more innovative solutions to be created without the associated increase in power usage using Supermicro servers.

January, 2022





Figure 6 - Work Per Watt over time (image courtesy of Intel Corp.)

Supermicro Resource Saving Architecture

Supermicro designs a broad range of servers and storage systems that provide state-of-the-art services to many SMBs, enterprises, and research organizations. Supermicro servers support a wide range of workloads from the edge to the cloud. Specifically, the data center systems that consume the most electricity also have the potential to generate the most E-Waste. Therefore, Supermicro systems are designed with a resource-saving architecture that reduces E-waste.

Resource Saving Architecture, which Supermicro employs, reduces a significant amount of e-waste, as the entire system does not have to be replaced. Estimates are that a chassis can be useful for seven years. In addition, the Resource Saving Architecture of Supermicro systems allows for replacing components as needed. For instance, multiple generations of compute can be used without replacing the chassis or fans.

Supermicro Disaggregated Server Technology

A disaggregated server is designed for and allows different electronic components to be replaced through a defined interface. The figure below shows that the CPU and RAM module could be replaced with new and more energy-efficient components without replacing the I/O module.

Today's servers contain many components that are constantly being improved with more advanced technologies. As a result, completely replacing an entire server to upgrade a single component (i.e., CPU, Memory, I/O) is no longer necessary with Supermicro servers. Instead, these servers are designed so technicians can replace specific components individually without



affecting other sub-systems. This reduces significant amounts of e-waste (estimated to be 56 million tons in 2020), as the entire system does not have to be replaced.



Figure 7 - Disaggregated Server Example

Focusing on the power usage of just a CPU ignores the requirements of other components in high performance servers. Memory, storage, and networking components all require electricity and must be figured into the overall power budget. Over time, the power required by the memory sub-system decreases per GB. Additionally, the power needed by Solid State Disks (SSDs) is 1/2 to 1/3 of that HDD consumes while being significantly faster for storing or retrieving data. Supermicro gives customers the ability to choose from various storage vendors to maximize their unique workload requirements.

Supermicro Systems with Resource Saving Architectures

Supermicro designs and manufactures a broad range of servers and storage systems that address many workloads and environmental constraints. Supermicro servers are designed to incorporate the latest technologies from partners such as Intel, AMD, and NVIDIA.

a. Ultra Servers: The Supermicro Ultra servers are designed for a wide range of workloads in enterprise and cloud data centers. These systems are available with the latest 3rd Gen Intel Xeon Scalable processors or the 3rd Gen AMD EPYC processors. Ultra servers are available in either a 1U or 2U chassis height and can accommodate up to 8 TB of memory. In addition, these systems can incorporate up to 4 GPUs for HPC, AI, and ML applications.





Figure 8 - Supermicro Ultra Server

b. BigTwin® servers: The Supermicro line of BigTwin servers offers a resource-saving design, where shared components contribute to lower operating power usage. The BigTwin line of systems is optimized for demanding workloads such as Virtualization, Cloud, Software-defined Storage, Hosting & Content Delivery, HPC, and Hyperscale. The innovative design can accommodate up to 4 individual systems housed in a 1U or 2U enclosure. In addition, since the BigTwin shares power supplies and cooling fans across multiple servers, efficiencies are gained by using larger, more efficient fans and power supplies.



Figure 9 - Supermicro BigTwin Server

c. SuperBlade[®] servers: The highest density computing for demanding applications from Supermicro is the SuperBlade platform. Consisting of a varying number of blades depending on the size (and thus capabilities) and a chassis with an integrated InfiniBand switch, the SuperBlade reduces electricity requirements through the use of shared components. For example, the SuperBlade platform shares power supplies and fans among the individual blades, reducing electricity requirements. In addition, due to the inclusion of a networking switch in the chassis itself, there are no cables between the server and the switch, which can increase airflow and allow the fans to operate at a lower speed, reducing power consumption.







Figure 10 - Supermicro SuperBlade with 2 CPUs



Even with the same components, systems from various vendors will use different amounts of electricity when running the same workloads. However, as discussed earlier, Supermicro design expertise can reduce the electricity used compared to industry standard servers.

Comparing the relative energy efficiency of different systems can be complex. Benchmarks can be chosen that represent different workloads but may not exactly represent what an enterprise will run daily. For example, ServeTheHome⁷ has run a series of benchmarks that compare four industry standard servers to a Supermicro BigTwin 2U-4Node system. Under a compute-heavy load (75% to 100%), the Supermicro BigTwin system used 10% less power than the four individual servers combined, at 50% less space. Using less power directly affects the bottom line and reduces OPEX.

Lifetime Electricity Savings

Electricity savings based on Supermicro products can add up over the life of a server. For example, if a server is running at 100% use over three years, with the cost per kWh being \$ 0.11.70 (Across all sectors, USA, September 2021) https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_6_a then:

Standard server: 2 kW X 8760 hours = 17520 kwH. X \$ 0.12/kWh = \$ 2102.40 per year X 3 yrs = \$ 6307.20. If an entire rack (42U) is populated with these standard systems, the total cost for electricity for three years would be: \$ 6307.20 x 40 = \$ 252288. With a 30% reduction in the power used for the same workload for a Supermicro BigTwin, the savings per rack would be \$ 75,686 over three years.

For illustrative purposes, in the table below, we compare a standard 1U system with System Type 1 and System Type 2. System Type 1 would occupy 4U and have a lower power draw than Std and higher performance. System Type 2 is a larger system (8U) with more computing power and a lower power draw.





Std. (1U) 1	System Type 1	System Type 2
	Type 1	Type 2
1		, ,
—	4	8
100	130	150
2	1.7	1.6
8760	8760	8760
3	3	3
52560	44676	42048
0.12	0.12	0.12
6307.2	5361.12	5045.76
40	10	5
\$ 252,288	\$ 53 <i>,</i> 612	\$ 25,229
0	\$ 198,676	\$ 227,059
37.2	31.7	29.8
45.6	38.8	36.5
31,920	27,160	25,550
	2 8760 3 52560 0.12 6307.2 40 5252,288 0 37.2 45.6 31,920	2 1.7 8760 8760 3 3 52560 44676 0.12 0.12 6307.2 5361.12 40 10 5252,288 \$53,612 0 \$198,676 37.2 31.7 45.6 38.8

Table 1 - System Comparisons for Power used and Environmental Effects

The environmental effects of using industry-standard servers should be considered compared to the savings of using an industry standard server (Std.)

Cooling Required - How to minimize power usage at server/rack level

Supermicro servers are designed to function at the specified levels with air cooling. The combination of innovative mechanical design and placement of components and the most efficient fans allows the servers to be placed in standard data centers. However, in many cases, it was discovered ⁸ that data center operators are keeping their data centers cooler than required. Increasing the inlet temperature for servers reduces the energy consumption used for cooling, decreasing the PUE.

The most impactful method to reduce the PUE in a data center is liquid cooling. Liquids are more efficient than air at removing heat from CPUs and GPUs. A cool fluid is pumped over the CPU or GPU and heats up, drawing the heat away from the microprocessor. The liquid is then pumped to a cooling device external to the server. This external chiller can be located within the rack or external to the machine room. Below are some comparisons as to why water (or other fluid) is significantly better at removing heat than air is.





The benefits of using liquid cooling are:

- Reduction in the CRAC requirements CRACs can be operated at a lower cooling capacity, thus reducing energy consumption.
- Higher computing capacity When CPUs begin to overheat, the microprocessor will decrease the operating frequency to avoid shutting down. By keeping the CPU or GPU cooler, higher frequencies can be maintained, leading to faster processing and reduced time to complete a given task.

Many data centers in existence do not have the air cooling capacity for today's highest performing (and hottest) CPUs and GPUS. This may be because the airflow in the entire data center is insufficient, or the floor vents (for the cool air) are not large or deep enough, restricting air movement.

Below is a Supermicro SuperBlade and a BigTwin with liquid cooling designed into the server.



Figure 12 - Supermicro SuperBlade with Liquid Cooling Installed



Figure 13 - Supermicro BigTwin with Liquid Cooling Installed



Customer Examples That Reduce PUE

Supermicro Liquid Cooling Examples:

Osaka University - Osaka University is a world-renowned university known for leading-edge research in various disciplines. As a 90-year-old institution, Osaka University serves a wide range of researchers, students, and external experts, making available a supercomputer system that runs state-of-the-art software on the latest CPUs and GPUs. The Osaka University SQUID system uses Supermicro SuperBlades and liquid cooling to reduce the data center PUE and allow the CPUs and GPUs to run at maximum performance.



LLNL Ruby - Lawrence Livermore National Laboratory (LLNL) has deployed Ruby, a supercomputer dedicated to open science. The lab, which works under the National Nuclear Security Administration (NNSA) umbrella, says the new cluster will serve research tasks relating to nuclear security and COVID-19 research.

Ruby, provided by Supermicro, is already operational and consists of 1,512 nodes: 1,480 batch nodes equipped with Intel Xeon CLX-8276L CPUs, 24 debug nodes, and eight login nodes, all networked with Cornelis Networks Omni-Path and liquid-cooled using a direct-to-chip approach





Supermicro Air Cooled Examples that reduce PUE:

Intel - Intel has deployed over 30,000 Supermicro[®] MicroBlade[™] disaggregated Intel[®] Xeon[®] processor based servers at its Silicon Valley data center, one of the world's most energy efficient data centers with a PUE of 1.06, to support its growing computing needs ¹⁰.



Summary

Supermicro servers and storage systems are designed to reduce power consumption. With innovative designs compared to industry-standard servers, Supermicro servers use less electricity per unit of work. The shared component design of the BigTwin and SuperBlade product lines reduces the electricity requirements for a given workload. Although the savings in terms of electricity and thus costs are small for an individual server when multiplied by the thousands that a modern data center (cloud or on-prem) requires, the reduction in CAPEX is significant. And not just for the costs savings, but in terms of the amount of greenhouse gas emitted from fossil fuel-based power plants. As corporate social responsibility gains momentum that includes an enterprise's climate responsibility, the need for more efficient servers will grow, and Supermicro is at the forefront of designing and offering better systems for the environment.

Appendix

Resource Saving Architecture - Resource-saving architecture is where the systems are designed so that they can be independently updated compute modules without having to update shared resources like fans, PSUs, and management and networking modules. You can continue to use the enclosure and shared resources through multiple generations of compute technology.

Disaggregated Servers - It's within a blade module where you have multiple PCBs. So you can update compute board without having to update the storage controller board.

https://insidehpc.com/2021/12/composable-hpc-ai-at-scale-the-emergence-of-optical-i-o-chiplets/

Further Information:

www.supermicro.com

www.supermicro.com/liquidcooling

January, 2022

13

REFERENCES

<u>1 - (https://www.forbes.com/sites/forbestechcouncil/2021/05/03/renewable-energy-alone-cant-address-data-centers-adverse-environmental-impact/?sh=9bd39f15ddc9</u>).

2. https://www.sciencedirect.com/science/article/pii/S0306261921003019

3 - <u>https://www.datacenterdynamics.com/en/opinions/5-year-forecast-huawei-predicts-10-trends-shaping-data-center-facilities/</u>

4 - <u>https://www.grandviewresearch.com/press-release/global-data-center-power-</u> market#:~:text=The%20global%20data%20center%20power,by%20Grand%20View%20Research%2C%20Inc.

5. https://journal.uptimeinstitute.com/datacenter-energy-efficiency-by-region/

6 - <u>https://blog.se.com/datacenter/power-and-cooling/2017/09/08/airflow-cooling-information-technology-equipment-required/</u>

7 - <u>https://www.servethehome.com/supermicro-bigtwin-nvme-2u-4-node-nvme-chassis/</u>

8 - <u>https://www.supermicro.com/en/white-paper/datacenter-report</u>

9 https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator

10 - <u>https://www.intel.com/content/www/us/en/it-management/intel-it-best-practices/green-computing-at-scale-paper.html?wapkw=pue%201.06</u>

