WHITE PAPER

# REDEFINING PERFORMANCE FOR WORKLOADS WITH OPTIMIZED SYSTEMS USING BUILT-IN ACCELERATORS

*New Supermicro X13 Servers Are Designed to Meet Modern Workload Demands Utilizing the 4th Gen Intel® Xeon® Scalable Processors*

## Introduction

As digital transformation continues, new and innovative technologies are needed to meet the increasing demand for a wide range of workloads. The most performant, flexible, and secure servers with new CPUs are needed whether systems are deployed on-premises or in a shared cloud environment. Supermicro's wide range of servers and storage systems continue to satisfy the demands of leading organizations and are now available with the 4th Gen Intel Xeon Scalable processors.

The continuing increase in the performance per watt of the latest generations of Intel CPUs allows for more services and a reduction in execution time for a wide range of use cases and applications. With an increased reliance on many applications that did not exist a few years ago, enterprises continue to require new servers that can meet the Service Level Agreements (SLAs) that their employees and customers demand. In addition, data analytics and high-performance computing applications can continue to deliver results in decreased time. The business benefits of new generations of servers based on the latest CPUs cannot be underestimated. Faster and more secure computing can result in lower OPEX and increased customer satisfaction and employee productivity.

As the amount of data grows, much of the new data will be generated at the "edge" of the network, which will require new form factors for servers as well as new software to enable intelligent decisions to be made close to where the data is generated. Using AI technologies at the edge may require servers that can operate in environments that are not climate controlled. Servers

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

containing the latest CPUs that live at the edge require careful design to allow heat dissipation in non-data center environments. The continuum of computing from the edge to the data center requires a range of systems that can securely manage data, perform analytics and determine what data needs to be transmitted back to more extensive facilities.

Core counts and gigahertz per core continue to increase, generation to generation. Below is a table of the increase in core counts with the corresponding base clock speed, generation over generation.

| | 3rd Gen Intel Xeon Scalable Processors | 4th Gen Intel Xeon Scalable Processors | % Increase |
|---|---|---|---|
| Max Cores | 40 (8380) | 60 (8490H) | + 50% |
| Max GHz (at Max cores) | 2.3 | 1.9 | |
| Total GHz at Max Cores | 92 GHz | 114 GHz | + 24% |

*Table 1 - Comparison of Max GHz and Cores Between Generations*

However, a system's overall performance depends on several factors, including communication speeds from the CPU to the peripherals, the performance of the memory system to the CPU, specialized configurations of memory, and innovative software and hardware to enable large memory applications. These technologies are designed into the offerings from Supermicro and are discussed below.

## 4th Gen Intel Xeon Scalable processor Generation over Generation Benchmarks

The 4th Gen Intel Xeon Scalable processor in a 2U dual socket Supermicro Server (SYS-621H-TN12R LINK) has shown a remarkable increase in performance over a 2U dual socket Supermicro server based on the 3rd Gen Intel Xeon Scalable processor (SYS-220U-TNR)*. Specifically:

- 1.5X to 2.0X performance in compute-intensive floating point benchmarks (HPL, HPCG, Stream, GROMACS, and LAMMPs).

- Intel Accelerators:

  Intel® AMX for AI applications – ResNet50 shows up to a 3.5X performance increase in throughput and a 3.5X reduction in latencies over the previous generation of Intel CPUs. In addition, ResNet50 V1.5, using the data type Bfloat16, showed an amazing 10.3X reduction in latencies and throughput compared to the previous generation of Intel CPUs.

  Intel® IAA for In-Memory Analytics – Using RocksDB, the performance gain was over 2.5X compared to the previous generation of Intel CPUs.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

## Supermicro Advantages for Rack-Level Performance and Efficiency

Supermicro can bring a wide range of products to market faster because of the Supermicro Building Block Solutions® approach. In addition, the resulting server product families can be designed for specific workloads, which lowers the TCO for customers.

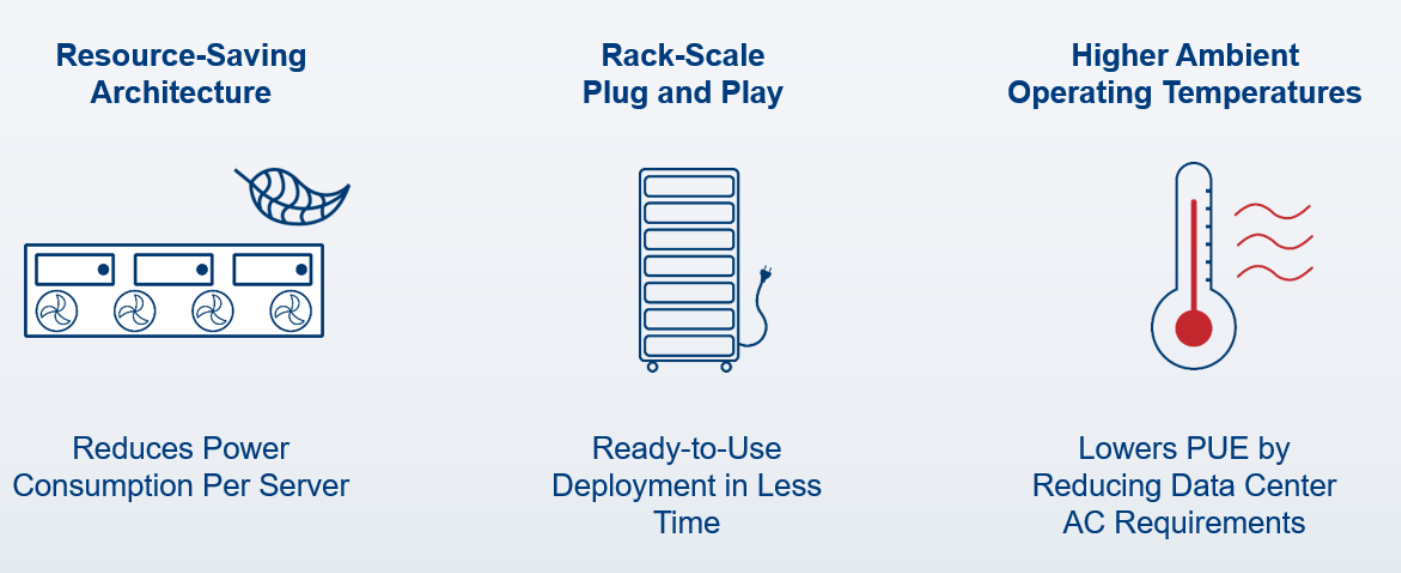


*Figure 1 - Supermicro Advantages*

The new Supermicro X13 servers are designed to reduce power usage while delivering maximum performance. In addition, Supermicro provides complete and ready-to-use racks of systems, resulting in a faster time for a customer to be productive.

- **Resource-Saving Architecture:** Supermicro servers, specifically the Supermicro BigTwin®, the Supermicro GrandTwin™, and the Supermicro SuperBlade®, are designed to share components, specifically the power supplies and the fans. Sharing these components by multiple servers reduces the power needed for the enclosure while maintaining the required performance.

- **Rack-Scale Plug and Play:** Supermicro factories deliver fully functioning racks to customers with L12 integration. By testing both hardware and software at the cluster level, end users can plug into power, connect to the network, and are ready to go. Simply put, Plug and Play.

- **Flexible Cooling Options:** Supermicro X13 servers can be cooled in various ways. Whether traditional (CRAC), free-air, or liquid cooling, many Supermicro servers are designed to run with high inlet temperatures, up to 40° C. This design feature can reduce the need for CRAC systems, reducing the data center PUE and thus reducing OPEX.

Supermicro designs and manufactures a wide range of servers and storage systems deployed from the Edge to hyperscale data centers. Different form factors with varying amounts of CPUs, memory capacity, storage types and capacity, and environmental considerations are engineered and delivered by Supermicro. The key to offering many different systems is an advanced engineering design process and teaming up with leading-edge CPU manufacturers, such as Intel.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

Using a Building Block design methodology, a wide range of systems can be quickly engineered to deliver optimized products for different workloads. These building blocks are then combined or slightly modified to be used in different form factor designs, higher airflow mechanical enclosures, or for more I/O capability, for example.

As CPUs run faster, with more cores, more heat is generated. Supermicro designs systems that efficiently remove this heat, lowering cooling costs and allowing CPUs to run up to their maximum thermal design power (TDP).

In this white paper, we look in-depth at Supermicro's latest X13 portfolio of servers and storage systems and how these systems help organizations thrive in today's digital landscape.

## Wide Range of Products for Varying Workloads

Supermicro's customers span many industries, with some common objectives:

- Ability to meet Service Level Agreements (SLAs) – Whether servicing employees or end-user customers, the CPU and I/O systems' responses are expected to fall within a specific time range.
- Provide new services to customers – As customers demand new services, which may run partially on edge devices as "apps," organizations must set up the back-end infrastructure to handle and respond to more data and processing than ever before.
- Reduce costs with more powerful systems – Some workloads do not increase at the same rate as new processors' computational and I/O power. Therefore, new CPUs allow them to reduce costs by assigning more work to lesser systems for these organizations.
- Enable new insights – By taking advantage of the latest CPU designs, scientists, engineers, and data analytics professionals can gain new insights and simulate physical systems more accurately.

Various workloads are all addressed by the Supermicro X13 servers and storage systems. These include:

- **Cloud** – Designing and implementing a cloud solution requires a wide range of optimized products for different workloads, not just for environments where the price-performance of the compute aspect is most important. Storage and networking are also critical for a productive and cost-effective cloud data center.
- **5G/Telco** – The rapid development and installation of 5G networks drive demand for fast CPUs resilient to the environment. Systems need to be efficiently cooled while performing full analytics. The new X13 lineup provides significantly more computing power per watt with reduced cooling requirements.
- **Artificial Intelligence (AI)** – Systems with fast CPUs and associated GPU sub-systems are required for the growing AI use cases. Supermicro X13 servers can house up to 10 GPUs in a 4U rack height and excel at AI applications, enabling faster training and inference applications. In addition, Supermicro designs servers specifically to accommodate many GPUs for maximum AI application performance.
- **High-Performance Computing (HPC)** – HPC systems are used by more than just university and national lab researchers. Enterprises integrate HPC systems into everyday workflows to bring products to market faster or discover new vaccines and drugs. HPC systems require fast cores, large amounts of memory, and fast networking between systems.
- **Big-Data Analysis** – As the volume of data generated everywhere explodes, the systems must access, analyze, and present structured and unstructured data to the user. These tasks require holding an increasing amount of data in memory, fast computation, and quick data communication to GPUs if needed.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

- **Streaming and Content Delivery** – New services deliver video to end users, both within corporate environments and from data centers to the Edge, in real-time. The X13 systems, with the new fast CPUs and communication channels from storage devices, are suited very well to this task.
- **Virtualization** – With many enterprises utilizing virtualization technologies to get higher utilization from existing servers, the new X13 servers, with the 4th Gen Intel Xeon Scalable processors, allow for higher-powered virtualization machines, as there are more cores available and faster CPUs.
- **Enterprise** – Typical enterprise workloads will benefit from the new X13 systems with increased performance and reduced costs. In addition, existing workloads will execute faster, using less power than previous generations of Supermicro servers.

## How 4th Gen Intel Xeon Scalable Processor Enhances Workloads and Highlights

### 4TH GEN INTEL XEON SCALABLE PROCESSOR DESCRIPTION

The 4th Gen Intel® Xeon® Scalable processors designed to accelerate performance across the fastest-growing workloads that businesses depend on today. Built-in accelerators improve performance across AI, data analytics, networking, storage, and HPC. By making the best use of CPU core resources, built-in accelerators can result in more efficient utilization and power efficiency advantages, helping businesses achieve their sustainability goals. 4th Gen Intel Xeon Scalable processors have advanced, hardware-enabled security technologies to help protect data while unlocking new opportunities for business collaboration and insights. No matter the deployment path, these processors enable solutions that help businesses scale infrastructure and achieve value fast.

While increasing the performance of computing systems continues over time with Intel's innovations, different workloads require this new performance, while other workloads benefit from the lower cost per unit of work. For example, while the performance of CPUs increases, typical Enterprise workloads (HR, ERP, Inventory Control, etc.) mainly do not require the performance gains from generation to generation but rather benefit from assigning more work to a given CPU. New Enterprise workloads, such as analytics, video conferencing, and application delivery, require performance improvements to take advantage of the new 4th Gen Intel Xeon Scalable processors' new performance levels. HPC and AI require both increased core numbers, increased GHz, and parallelization and networking outside of the system itself.


© 2023 Copyright Super Micro Computer, Inc. All rights reserved

In addition to the new technologies that the 3rd Gen Intel Xeon Scalable processors introduced, the 4th Gen Intel Xeon Scalable processors include:

- **50% More cores:** Maximum of 60 cores compared to 40 cores in the 3rd Gen Intel Xeon Scalable processors
- **Faster communication:** PCIe 5.0 is 2X faster and uses more lanes than the maximum PCIe available in previous Intel CPU generations
- **Memory performance:** The new Intel processors can utilize DDR5-4800MHz memory, 50% faster than previous generations.
- **Faster communication between CPUs:** More and faster Ultra Path Interconnects are available with the 4th Gen Intel Xeon Scalable processor than the 3rd Gen Intel Xeon Scalable processors
- **Accelerators built into the CPU that speed up specific workloads:**
    - Intel® Advanced Matrix Extensions (Intel® AMX) – Up to 8X performance increase in matrix multiply microbenchmarks
    - Intel® Data Streaming Accelerator (Intel® DSA)– up to 2.5X improvement in data movement performance
    - Intel® QuickAssist Technology (Intel® QAT) – up to 98% core workload capacity for encryption and related workloads
    - Intel® vRAN Boost – Reduces vRAN power consumption by up to 20%
    - Intel® Dynamic Load Balancer (Intel® DLB) – Improve the system performance related to handling network data on multi-core Intel® Xeon® Scalable processors.
    - Intel In-Memory Analytics Accelerator (Intel® IAA) - Faster queries and resource utilization for more efficient business decision making

There are several advantages to using the 4th Gen Intel Xeon Scalable processors for different workloads with different models for various workloads. Below is a table that describes the Models and the Advantages of the optimized CPUs.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

| Models | Advantages |
|---|---|
| Performance Family | 2 socket, 4th Gen Intel Xeon Scalable processors designed for the highest performing Cloud, Enterprise, and HPC workload requirements in terms of cores and clock rate |
| Mainline Family | 2 sockets, 4th Gen Intel Xeon Scalable processors for Enterprise workloads |
| Liquid Cooled (-Q) | Processors specifically designed for liquid cooling environments |
| Single socket (-U) | General purpose CPU core counts, optimized for single-socket servers |
| Enterprise Applications (-H) | 1S or 2S for in-memory databases, analytics, and virtualization |
| 5G and Networking Optimized (-N) | Specific optimizations for 5G and Networking applications |
| Cloud Optimized IaaS (-P) / SaaS (-V) / Media (-M) | Optimized for cloud environments |
| Storage & Hyperconverged Infrastructure (HCI) Optimized (-S) | CPU optimized for storage and HCI infrastructure |
| Long Life Use and Industrial Conditions (-T) | Optimized for NEBs and Industrial Environments |

*Table 2 - Optimized SKUs for Different Workloads*

## Supermicro X13 Server Product Family Overview

The Supermicro product family contains over 200 servers and storage systems designed for customer workloads. All take advantage of the new features and capabilities of the 4th Gen Intel Xeon Scalable processors.

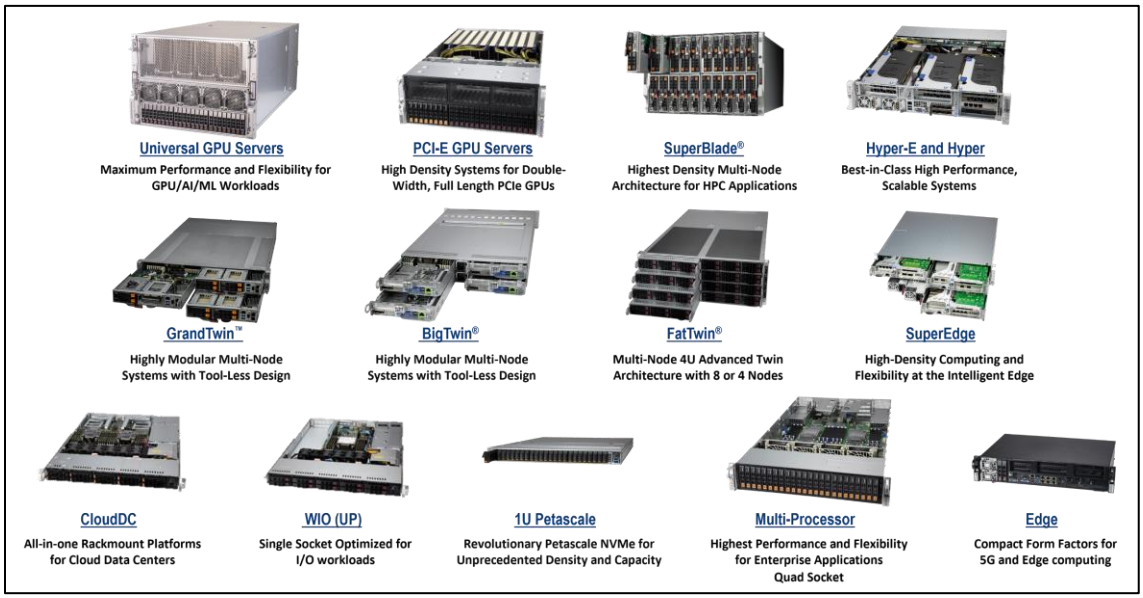


*Image 2 - Supermicro Lineup of X13 Servers*

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

The Supermicro product line can be segmented into the following product families.

**GPU Family** – The Supermicro GPU family of servers excels at HPC and AI applications. Systems have been designed to house multiple GPUs in a single server so applications can process data at tremendous rates. While many Supermicro server lines can accommodate one or two GPUs, the GPU family extends the quantity of GPUs in a single server up to 10 in a 4U form factor. The GPU family of servers can house multiple GPUs and is designed so that GPUs can efficiently communicate with each other, allowing GPU systems to bypass internal communication paths for faster results. The GPU systems can also address the maximum memory that the 4th Gen Intel Xeon Scalable processors up to 6TB per socket. The 4th Gen Intel® Xeon® Scalable processors have up to 80 lanes of PCIe 5.0—ideal for fast networking, high-bandwidth accelerators, and high-performance storage devices. PCIe 5.0 doubles the I/O bandwidth from PCIe 4.0, maintains backward compatibility, and provides foundational slots for Compute Express Link (CXL).

a. **Universal GPU Servers with OAM GPUs** – With Supermicro's advanced architecture and thermal design, including liquid cooling and custom heatsinks, the Supermicro 4U or 5U GPU system featuring four of NVIDIA's latest HGX H100 GPU or other OAM baseboards can deliver up to 10x AI training performance (Intel estimates) and 10x inference workload capacity (Intel estimates) compared to the previous generation of CPUs, and highest density in a flexible 4U system. The latest Supermicro GPU system in an 8U form factor can accommodate 8 x NVIDIA H100 GPUs. In addition, the Supermicro GPU systems feature the latest technology stacks, such as 400G networking, NVIDIA NVLink and NVSwitch, 1:1 GPUDirect RDMA, GPUDirect Storage, and NVMe-oF on InfiniBand.

   Common workloads include:

   AI/ML • Deep Learning Training and Inference • High-performance Computing (HPC) • Building Block for Scalable AI Infrastructure

   For the Supermicro GPU systems, all of the new features of the 4th Gen Intel Xeon Scalable processors will help high-end applications perform better and return faster with the latest GPU systems from Supermicro. More and faster cores, higher bandwidth to the GPUs and other devices, and the ability to address vast amounts of memory are exactly what large HPC and AI applications demand.

*Image 2 - Supermicro X13 Universal GPU*

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

b. **GPU Servers with PCIe GPUs**– The GPU systems that attach the GPU accelerators via the PCIe bus are ideal for environments that require multiple GPUs that perform their work with direct commands from the CPU. HPC and AI/ML environments will benefit significantly from the 4th Gen Intel Xeon Scalable processors. Various platforms can accommodate from one to 10 GPUs.

Common workloads include:

AI/ML • Deep Learning Training and Inference • High-performance Computing (HPC) • Rendering Platform for High-end Professional Graphics • Best-in-Class VDI Infrastructure Platform

GPU Systems with PCIe will benefit significantly from the PCIe 5.0 communications bus, the increased number of cores, and up to 6TB of memory per CPU for large applications.

*Image 3 - Supermicro X13 GPU Server w/PCIe GPUs*

**Supermicro SuperBlade** – SuperBlade is an ideal system for the most demanding workloads, with the support of the highest performance CPUs and the fastest networking available today. SuperBlade supports one or two CPUs and up to four GPU accelerators or PCIe cards. Supermicro's high-performance, density-optimized, and energy-efficient X13 SuperBlade can significantly reduce many organizations' initial capital and operational expenses. It utilizes shared, redundant components, including cooling fans, switches or pass-through modules, power supplies, and integrated system management, to deliver the compute performance of an entire server rack in a much smaller physical footprint. These systems support GPU-enabled blades and are optimized for AI/ML/DL, HPC, Cloud, EDA, Virtualization, and Health and Financial Services. In addition, SuperBlade is available in two form factors.

- 8U SuperBlade can support up to 20 single-width nodes or 10 double-width nodes and delivers high performance with advanced networking up to 400Gb NDR InfiniBand or 400Gb Ethernet.
- 6U SuperBlade can support 10 single-width nodes or five double-width nodes and 32 DIMM slots per node of DDR5-4800 MHz memory in a high-performance, dense infrastructure. The single-socket blades can support up to four GPUs for workload acceleration.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

*Image 4 - SuperBlade X13 8U Enclosure*

*Image 5 - SuperBlade X13 6U Enclosure*

A significant advantage of the SuperBlade is that cooling fans and power supplies are shared among the server blades, which reduces power consumption compared to individual rackmount servers with their own non-sharable fans and power supplies. The SuperBlade also offers both air and liquid cooling solutions to support dual 4th Gen Intel Xeon Scalable processors up to 350W TDP. Many of the 4th Gen Intel Xeon Scalable processor's new features will benefit all users of the Superblade. For example, the increased core count, performance, and amount of directly addressable memory with increased speed (up by 50%) are extremely valuable for workloads running on the SuperBlade. Also, support for PCIe 5.0 allows for faster communication with GPUs and external networking that are installed on the blades.

Common workloads for Supermicro X13 SuperBlade systems include:

AI/ML/DL • HPC • Hybrid Cloud • EDA • Virtualization • Health • Financial Services

*Image 6 - Supermicro X13 6U 1 Socket SuperBlade*

*Image 7 - Supermicro X13 4U 2 Socket SuperBlade*

*Image 8 - Supermicro X13 6U 2 Socket SuperBlade*

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

**Twin Family** – The Supermicro Twin product line comprises innovative systems that put multiple independent servers within the same enclosure. This lowers operating expenses by allowing the use of shared resources, such as the 2U enclosure, heavy-duty fans, backplane, and N+1 power supplies. Within the Twin product line are three product families:

a. **GrandTwin™** - The Supermicro GrandTwin is an all-new architecture purpose-built for single-processor performance. The design maximizes computing memory and efficiency to deliver maximum density. Powered by 4th Gen Intel® Xeon® Scalable processors, GrandTwin's flexible modular design can be easily adapted for a wide range of applications, with the ability to add or remove components as required, reducing cost. For front configurations, all I/O and node trays are fully accessible from the cold aisle, simplifying installation and servicing in space-constrained environments. In addition, flexible storage and networking options are available via front AIOM modules, allowing countless custom configurations.

   Common workloads for the Supermicro X13 GrandTwin family include: Virtualization • Cloud Hosting & Content Delivery • Hyperscale / Hyperconverged

*Image 9 - Supermicro GrandTwin*

b. **BigTwin®** – The BigTwin is a 2U design containing either two or four server nodes, with each server accommodating up to 2 CPUs and up to 6TB of DRAM in each node. BigTwin systems are designed to accommodate a wide variety of storage options, and flexible networking options with AIOM, offering a superset of OCP 3.0 features and performance.

   Common workloads for the Supermicro X13 BigTwin family include: Diskless HPC • All-Flash HCI • Hybrid Cloud • All-Flash NVMe Storage • High-Performance File Systems • Software-Defined Storage

*Image 10 - Supermicro BigTwin w/2 Nodes*

*Image 11 - Supermicro BigTwin w/4 Nodes*

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

c. **FatTwin®** – The FatTwin product comes in either four or eight nodes, with up to 16 DIMM slots. Each node can support 1 CPU and has versatile networking and storage options. FatTwin nodes are front-accessible to accommodate cold-aisle serviceability environments and increased ease of use. Because of FatTwin's shared component design, energy savings are quickly realized, and each node can be configured with a range of hot-swappable storage devices. The Supermicro FatTwin® systems will benefit from the new 4th Gen Intel Xeon Scalable processor features such as increased memory addressability, PCIe 5.0 communication speeds and lanes, and increased cores. With the ability to take advantage of FatTwin's modularity to support Smart NICs and other networking options to take advantage of the increased PCI-Lanes available in the 4th Gen Intel Xeon Scalable processor architecture. In addition, FatTwin is a battle-tested platform deployed worldwide in various use cases from the Edge to HPC. The new FatTwin X13 provides a seamless upgrade for these deployments with little operational risk in the upgrade process.

   Common workloads for the Supermicro X13 FatTwin family include:

   Data Center Infrastructure • Hyperscale / Hyperconverged • Cloud optimized servers • Data Center Enterprise Application • Scale-out of Storage Expansion • Telcom Data Center & ETSI Certified • Virtualization Server • Electronic Design Automation

*Image 12 - Supermicro X13 FatTwin w/4 Nodes*

*Image 13 - Supermicro X13 FatTwin w/8 Nodes*

**Hyper-E Family** - The Hyper-E servers are designed for maximum rackmount flexibility with rear and front I/O for tomorrow's data center requirements. These systems can handle up to 350 watts for the CPUs and the maximum number of DIMMs to accelerate a wide range of workloads. The Hyper-E family is designed for 5G and Telco environments, where NEBS Level 3 compliance is a must-have. With this requirement, the "N" and "T" SKUs (for 4thGen Intel Xeon Scalable processors) are ideal for the Hyper-E systems.

X13 Hyper-E brings the performance and flexibility of Supermicro's flagship Hyper series to the Edge with short-depth form factors designed for Edge data center and telco deployments. Telco-optimized configurations are NEBS Level 3 certified and feature optional DC power supplies on selected models. In addition, all I/O and expansion slots are front-accessible for easy servicing in space-constrained environments. At the same time, maintenance-friendly design innovations eliminate the need for tools when servicing, simplifying rollout and installation.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

The Hyper-E is also designed as a "short-depth" server, which means it can fit in tighter environments and features front or rear IO accessibility. The Hyper-E is also available with AC and DC power options. The Hyper-E systems sport many PCIe slots (up to eight) for extreme flexibility, are toolless for fast and easy servicing, and come with various storage devices (NVMe/SAS/SATA). The Hyper systems can also support up to 2 AIOM/OCP 3.0 NICs. Hyper-E systems will benefit from the increased core count at similar pricing with the 4th Gen Intel Xeon Scalable processors. In addition, the faster PCIe 5.0 communications bus will give more rapid access to storage devices, resulting in better performance.

Common workloads for the Supermicro X13 Hyper-E family include:

5G Core and Edge • Telecom Micro Data Center • Enterprise Server • Cloud Computing • Big Data Analytics • Hyperconverged Storage • AI Inference and Machine Learning • Network Function Virtualization

Hyper-E systems will benefit from the increased core count at similar pricing with the 4th Gen Intel Xeon Scalable processors. In addition, the faster PCIe 5.0 communications bus will give more rapid access to storage devices, resulting in better performance.

*Image 14 - Supermicro X13 2U Hyper-E Server*

**Hyper Family –** The X13 Hyper series brings next-generation performance to Supermicro's flagship range of rackmount servers, built to take on the most demanding workloads along with the storage & I/O flexibility that provides a custom fit for a wide range of application needs. Supermicro Hyper systems are available in 1U or 2U versions, with up to 32 DIMM slots. With the cooling capacity to accommodate the highest performing CPUs, the Supermicro Hyper product family is optimized for maximum compute performance.

Optimized for:

5G Core and Edge • Telco Micro Data Center • Enterprise Server • Cloud Computing • Big Data Analytics • Hyperconverged Storage • AI Inference and Machine Learning • Network Function Virtualization

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

*Image 15 – Supermicro X13 Hyper System*

**Supermicro Edge Servers** – Optimized for telco edge workloads, Supermicro X13 Edge systems offer high-density processing power in compact form factors. Flexible power with both AC and DC configurations available and enhanced operating temperatures up to 55° C (131° F) with selected components make these systems ideal for Multi-Access Edge Computing, Open RAN, and outdoor edge deployments. Supermicro SuperEdge brings high-density compute and flexibility to the intelligent edge, with three hot-swappable single processor nodes and front I/O in a short-depth 2U form factor.

Typical workloads for the Supermicro X13 edge servers include:

Multi-Access Edge Computing • Flex-RAN/Open RAN • Edge AI • 5G

*Image 16 - Supermicro X13 Edge Server*

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

**Supermicro SuperEdge** – Supermicro's X13 SuperEdge is designed to handle increasing compute and I/O density requirements of modern edge applications. With three customizable single-processor nodes, SuperEdge delivers high-class performance in a 2U, short-depth form factor. Each node is hot-swappable and offers front access I/O, making the system ideal for remote IoT, edge, or telco deployments. With flexible Ethernet or Fiber connectivity options to the BMC, SuperEdge makes it easy for customers to choose remote management connections per their deployment environments.

Common workloads for the Supermicro X13 SuperEdge include:

5G Open RAN/Flex-RAN • C-RAN (vRAN) • Telecom/Networking Appliance • Multi-Access Edge Computing • Edge Data Center • Enterprise Edge Computing

*Image 17 - Supermicro X13 SuperEdge Server*

**CloudDC Family** – The CloudDC family is explicitly designed for cloud data centers where space is premium. The CloudDC product line is toolless, meaning servicing these servers is quick and easy. The I/O options vary, and the systems can accommodate up to two double-width GPUs. The CloudDC family comes with dual AIOM OCP 3.0 support, which gives the product family tremendous expandability and flexibility. The CloudDC family also supports up to 6 PCIe 5.0 slots. The PCIe slots are equally split between the CPUs, which results in additional flexibility. 12 NVMe storage devices are supported for maximum I/O performance and capacity. The toolless mechanical design makes serviceability easy and provides hot-swappable storage capability.

Common workloads for the Supermicro X13 CloudDC family include:

Cloud Computing • Web Servers • Hyper-converged Storage • Virtualization • File Servers • Head-node Computing • 5G Telco • AI Inferencing

The 4th Gen Intel Xeon Scalable processor features that would benefit these applications include the increased core count and PCIe 5.0.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

*Image 18 – Supermicro X13 CloudDC 1U Server*

**WIO Family** – The Supermicro WIO systems offer a wide range of I/O options to deliver truly optimized systems for specific requirements. Users can optimize the storage and networking alternatives to accelerate performance, increase efficiency and find the perfect fit for their applications. In addition to enabling customizable configurations and optimization for multiple application requirements, Supermicro WIO servers also provide attractive cost advantages and investment protection. Supermicro X13 WIO systems accommodate double-width GPUs for accelerated AI/ML workloads through newly-designed top loading expansion slots.

Typical workloads for the Supermicro X13 WIO family include:

Enterprise Applications • Networking Appliance • Firewall/Security Appliances • General Purpose Computing • Cloud Computing • Media Entertainment

*Image 19 - Supermicro X13 WIO 1U Server*

**Petascale Storage** – The X13 All-Flash NVMe systems offer industry-leading storage density and performance with EDSFF drives, allowing unprecedented capacity and performance in a single 1U chassis. The first in a coming lineup of X13 storage systems, this latest E1.S server supports 9.5mm and 15mm EDSFF media, now shipping from all the industry-leading flash vendors.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

Typical workloads for the Supermicro X13 Petascale storage family include the following:

Data Intensive HPC/AI • Private & Hybrid Cloud • Software-Defined Storage • NVMe Over Fabrics Solution • In-Memory Computing • Composable Infrastructure Platform

*Image 20 – Supermicro X13 Petascale Server*

**Multi-Processor Family** – The X13 MP servers bring maximum configurability and scalability in a 2U or a 6U design, incorporating 4 or 8 4th Gen Intel® Xeon® Scalable processors to provide maximum memory density for mission-critical, in-memory compute and database workloads. X13 MP servers will be available in mid-2023.

Typical workloads for the Supermicro X13 Multi-Process family include:

In-Memory Computing • Virtualization • HPC • ERP Database • Scientific Research • Hyper-Scaling

*Image 21 - Supermicro X13 MP Server with 4 CPUs*

*Image 22 - Supermicro X13 MP Server with 8 CPUs*

All of the above Supermicro servers are designed with the following new system level technologies.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

## Additional New System Level Technologies for Increased Performance Using the 4th Gen Intel Xeon Scalable processor

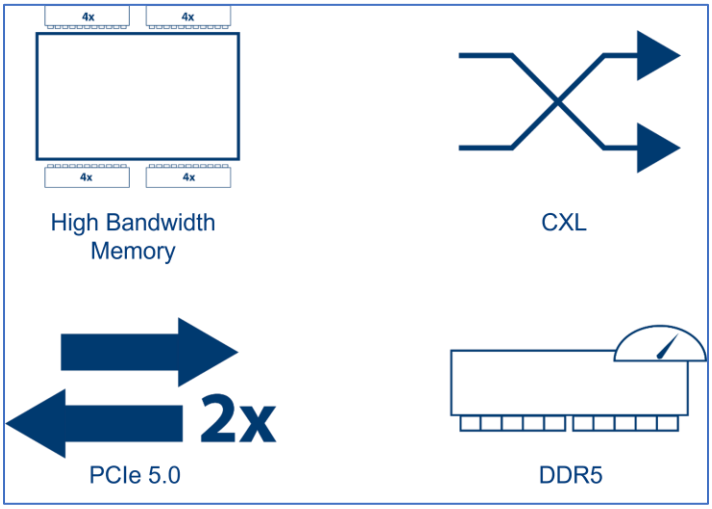


*Figure 2 – New System Level Technologies*

**High Bandwidth Memory (HBM)** - There are several applications and workloads that require high bandwidth access to memory. For these workloads, Intel has created the Intel® Max CPU Series for a limited number of CPU models. With HBM, specific processors can contain up to 64GB of HBM, which increases performance by up to 3.7X compared to a similar 3rd Gen Intel Xeon Scalable processor system.

Supermicro servers that are designed for the Intel® CPU Max Series include the following product lines:

| Supermicro Product Family | Form Factor | Number of Nodes / GPU Capability | Number of Sockets Per Server | Air/Liquid Cooled |
|---|---|---|---|---|
| Hyper | 2U | 1 | 1 or 2 | Air and Liquid |
| BigTwin | 2U | 2 | 1 | Air and Liquid |
| SuperBlade | 8U Chassis | 10 | 2 | Air |
| SuperBlade | 8U Chassis | 20 | 2 | Liquid |
| SuperBlade | 6U Chassis | 5 | 1 or 2 | Air |
| GPU Server | 4U or 5U | 1 node / 8 or 10 PCIe GPUs | 2 | Air and Liquid |
| GPU Server | 4U or 5U | 1 node / 4x H100 SXM GPUs | 2 | Air and Liquid |
| GPU Server | 8U | 1 node / 8x H100 SXM GPUs | 2 | Air and Liquid |

*Table 3 - Supermicro Serves with Intel MAX CPUs*

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

**PCIe 5.0** – With PCIe 5.0, the speed at which PCIe peripherals communicate with the CPU is double what was available with the previous generation of CPUs. This new technology will increase performance for a wide range of applications.

**CXL 1.1** – The Compute Express Link™ (CXL™) is an industry-supported Cache-Coherent Interconnect for Processors, Memory Expansion, and Accelerators. CXL technology maintains memory coherency between the CPU memory space and memory on attached devices, which allows resource sharing for higher performance, reduced software stack complexity, and lower overall system cost.

**DDR5 Memory** – Increased performance and capacities of DDR5 compared with DDR4 memory will increase the performance of applications while reducing power consumption.

## Supermicro Intelligent Management (SuperCloud Composer)

SuperCloud Composer is a composable cloud management platform that provides a unified dashboard to administer software-defined data centers. Supermicro's cloud infrastructure management software brings speed, agility, and simplicity to IT administration by integrating data center tasks into a single intelligent management solution.

Our robust composer engine can orchestrate cloud workloads through a streamlined industry-standard Redfish API. SuperCloud Composer monitors and manages the broad portfolio of multi-generation Supermicro servers and third-party systems through its data center lifecycle management feature set from a single unified console.

## Applications Benefits Summary - With the new 4th Gen Intel Xeon Scalable processors, applications will benefit from several innovations.

- More cores – performance will increase by default for applications that scale with the number of available cores.
- More extensive memory access – with more memory that can be accessed on the main memory bus, applications will perform better without waiting for data to be retrieved from storage devices.
- Faster memory access – with higher memory bandwidth, applications will execute faster, requiring less time to wait for critical data.
- Faster communication – with PCIe 5.0, applications can communicate with PCIe devices at twice the speed as before, resulting in overall application performance increases.
- Interconnect between sockets – for applications that require socket-to-socket communication, the faster UPI channels will reduce execution time.

### MEMORY CAPACITY

The 4th Gen Intel Xeon Scalable processor's memory capacity can be addressed directly per socket is 4 Terabytes.

Increased memory allows for more extensive applications to be run in less time. Data analytics, HPC, and more VMs can easily take advantage of this increased memory capacity to deliver results to users faster. Performance is improved by keeping more data in memory than on storage devices, and more extensive and complex simulations or analytics can be executed to gain more in-depth insight.

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

| | 3rd Gen Intel Xeon Scalable processors | 4th Gen Intel Xeon Scalable processors (1S-2S) | % Increase |
|---|---|---|---|
| Memory DIMMs (max/socket) | 16 | 16 | -- |
| Max Memory (DRAM)/socket | 4TB | 4TB | -- |

*Table 4 - Memory Capacity*

### CPU CORE COUNT AND PERFORMANCE

The 4th Gen Intel Xeon Scalable processor is built with the Intel 7 process. By reducing the transistor size, more transistors can be placed on a chip, leading to higher overall performance. The 3rd Gen Intel Xeon Scalable processors had a maximum of 40 cores per socket, while the Intel 4th Gen Scalable processors contain a maximum of 60 cores. For applications that scale well and can take advantage of the increased core count, a performance increase will be automatic as long as the application does not have an I/O bottleneck. At a minimum, the performance should increase linearly to the 40 core count of the previous generations of Intel CPUs (3rd Gen Intel Xeon Scalable processors). In addition, the 4th Generation Intel Xeon Scalable processors use the new Golden Cove core, a new microarchitecture containing new instructions, larger caches, and other enhancements available to all applications. Although some applications may need to be recompiled to take advantage of the new instructions, many CPU-dependent applications will see a 24% increase in performance due to the total GHz in the CPU.

| | 3rd Gen Intel Xeon Scalable processors | 4th Gen Intel Xeon Scalable processors (1S-2S) | % Increase |
|---|---|---|---|
| Max Cores | 40 | 60 | 50% |
| Base GHz at Max Cores | 2.3 | 1.9 | NA |

*Table 5 - Supermicro X13 Servers - Max Cores and GHz*

### FASTER INTERCONNECTS BETWEEN SOCKETS (ULTRA PATH INTERCONNECT)

The 4th Gen Intel Xeon Scalable processors have faster communication between sockets with up to four Intel Ultra Path Interconnects (Intel UPI) running at 16 GT/s. This feature is essential when applications use more than one socket, and the sockets must communicate. With faster communications for applications that run across sockets, performance will benefit and show a decrease in time to solve or deliver results.

| | 3rd Gen Intel Xeon Scalable processors | 4th Gen Intel Xeon Scalable processors (2S) | % Increase |
|---|---|---|---|
| Number of UPI links (max) | 3 | 4 | 33% |
| Performance | 11.2 GT/s | 16 GT/s | 43% |
| Total UPI Throughput | = 3 * 11.2 = 33.6 GT/s | = 4 * 16 = 64 GT/s | 90% |

*Table 6 - Interconnect Links and Speeds Between Sockets*

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

### MEMORY ACCESS PERFORMANCE

The speed at which the CPU can access memory dramatically affects the overall execution time of a task. The 4th Gen Intel Xeon Scalable processors have improved memory access bandwidth of up to 4800 Megatransfers per second (MT/s). The faster the MT/s rate, the faster the CPUs can retrieve data and act on it. The previous generation of Intel processors limit was 3200 MT/s, and six channels could deliver 8 x 3200 = 25,600 MT/s. The 4th Gen Intel Xeon Scalable processors use eight channels for memory access. Thus, the maximum performance per socket = 8 * 4800 MT/s = 38,400 MT/s, a 50% improvement.

| | 3rd Gen Intel Xeon Scalable processors | 4th Gen Intel Xeon Scalable processors | % Increase |
|---|---|---|---|
| Memory Performance | 3200 MHz | 4800 MHz | 50% |
| Number of Memory Channels | 8 | 8 | -- |
| Total Memory Bandwidth | = 8 * 3200 MT/s<br>= 25,600 MT/s | = 8 * 4800 MT/s<br>= 38,400 MT/s | 50% |

*Table 7 - Memory Performance Generation to Generation*

### FASTER CONNECTIONS TO ACCELERATORS

The 4th Gen Intel Xeon Scalable processor supports the PCIe 5.0 standard, which has a peak performance of twice that of the previous PCIe 4.0 standard. PCIe 5.0 delivers 32 GT/second per lane. The performance of a system for communicating with PCIe devices is computed as follows:

| | PCIe 4.0 | PCIe 5.0 | % Increase |
|---|---|---|---|
| Per Lane Performance | 16 GT/Second | 32 GT/Second | 100 % |

*Table 8 - PCIe Comparison - Generation to Generation*

## How Does Supermicro Do It?

Supermicro incorporates a Building Block® approach which allows us to design individual components with the latest technology and then engineer these different components together into various systems. Using this design process, Supermicro can create many variations, including additional CPUs, the number of memory slots, the number of PCIe lanes, and the number and type of storage devices. Depending on the form factor, cooling requirements, and memory requirements, application-optimized systems can quickly be developed. Innovative design allows for efficient cooling and the sharing of other mechanical components. Supermicro's servers can accommodate high-end CPUs in various form factors.

## Performance / Power over time. Why this is important to data centers.

Over time, with Intel's advancement of CPU technology, more computing power is available at a given price and a given amount of energy. As a result, Intel has increased the amount of work performed per unit of electricity by a factor of 9 over the past 13

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

years. This means that more work can be performed at a constant power draw, enabling organizations to offer more services and applications to their employees or the public.

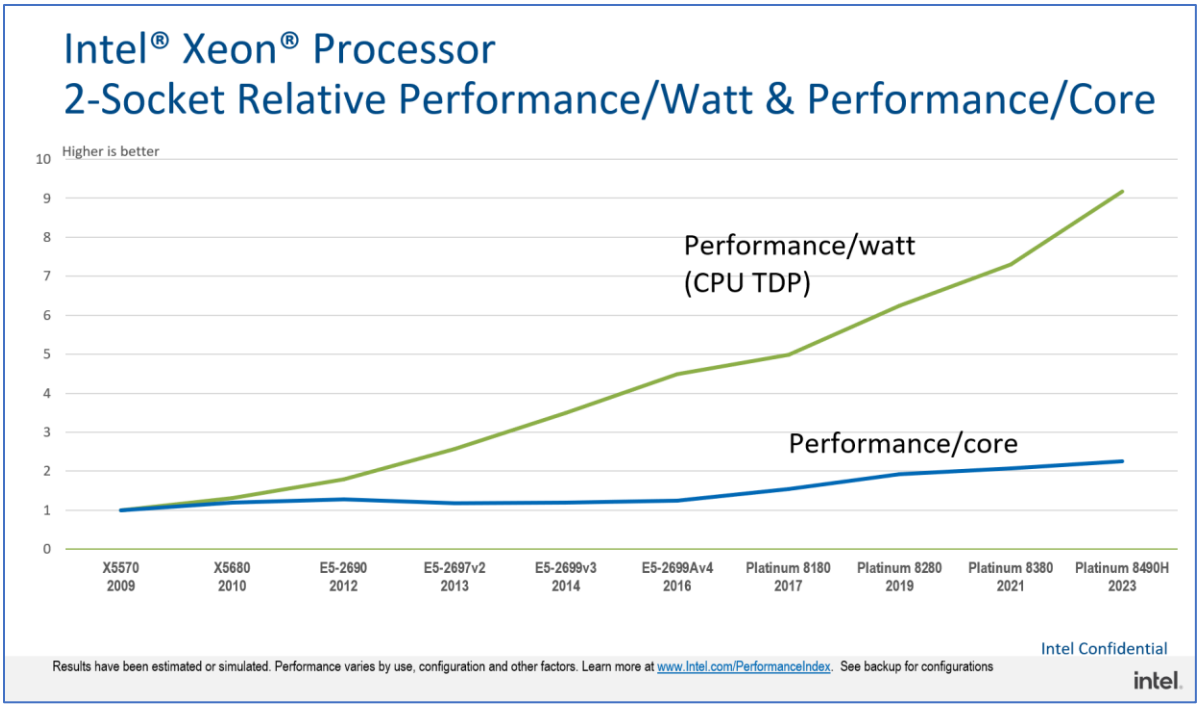


*Figure 3 - Performance Per Watt Over Time (Image Courtesy of Intel)*

## Summary

The new X13 product line from Supermicro enables all organizations to take full advantage of Intel's latest CPUs. Ranging from a single processor to the latest in blade technology and from 8 cores per socket to 60 cores per socket, Supermicro has a server designed for your workload. With the increase in the amount of memory that can be addressed and the performance of the memory sub-system, applications can access more data faster. The increase in core count numbers and clock rates results in a faster time-to-solution and more performance per watt. The Supermicro X13 product lineup is designed for workloads that range from the Edge to the Data Center.

## Resources

www.supermicro.com/x13

*Hardware Information: The SYS-621H-TN12R system (X13) has the following configuration HARDWARE DESCRIPTION QUANTITY SYSTEM SYS-621H-TN12R MOTHERBOARD X13DEM 1 FIRMWARE BIOS-Version=1.0, BMC-Version=1.00.(18), CPLD-Version=F2.41.24 N/A CPU Intel Xeon Platinum 8480+ 56C112T 2 MEMORY Samsung M321R8GA0BB0 64GB DDR5 4800MHz 16 DRIVE Intel DC P4600 1.6TB NVMe PCIe 3.0 3D TLC 2.5" 3DWPD 1 NIC INTEL I350 1GbE

*The SYS-220U-TNR system (X12) has the following configuration HARDWARE DESCRIPTION QUANTITY SYSTEM SYS-220U-TNR MOTHERBOARD X12DPU-6 1 FIRMWARE BIOS-Version=1.4, BMC-Version=1.01.(24), CPLD-Version=F1.00.D5 N/A CPU Intel Xeon

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

Platinum 8358P 32C64T 2 MEMORY SK Hynix HMAA8GR7AJR4N 64GB DDR4 3200MHz 16 DRIVE Intel DC P4600 1.6TB NVMe PCIe 3.0 3D TLC 2.5" 3DWPD 1 NIC NVIDIA ConnectX-6 Lx 25GbE SFP+ 1 3. Software Environment SOFTWARE VERSION OS Ubuntu 22.04 OPEN MPI 4.0.4

## Appendix A – Supermicro X13 Summary

| Family | SKU | Form Factor | # of CPUs | # of DIMMs | Drives | Key Applications/Notes |
|---|---|---|---|---|---|---|
| All-Flash EDSFF | SSG-121E-NES24R | 1U | 2 | 32 | 24x EDSFF (E1.S 15mm) NVMe SSD | Data Intensive HPC/AI, Private & Hybrid Cloud, Soft-ware-defined Storage, NVMe Over Fabrics Solution, In-memory Computing, Composable Infrastructure Platform |
| GrandTwin™ | SYS-211GT-HNTF | 2U4N | 1 (node) | 16 (node) | 4x U.2 NVMe/SATA (node) | HPC, Financial Services, Content , Delivery/Streaming, Electronic Design Automation |
| | SYS-211GT-HNC8F | | | | 4x U.2 NVMe/SAS/SATA (node) | |
| | SYS-211GT-HNTR | | | | 6x U.2 NVMe/SATA (node) | |
| | SYS-211GT-HNC8R | | | | 6x U.2 NVMe/SATA (node) | |
| Universal GPU | SYS-421GU-TNXR | 4U/5U | 2 | 32 | 10x 2.5" U.2 NVMe | AI/ML, HPC, Vertical Markets (thermal modeling and other parallel-processing intensive tasks), Big Data Analytics |
| | SYS-821GE-TNHR | 8U | | | 8x 2.5" U.2 NVMe via PCI-E switch + 8x 2.5" U.2 NVMe (optional) | Supports next-generation HGX H-100 8-GPU for AI training Front/Rear I/O options available |

| Family | SKU | Form Factor | # of CPUs | # of DIMMs | Drives | Key Applications/Notes |
|---|---|---|---|---|---|---|
| GPU-optimized | SYS-421GE-TNRT/TNRT2 | 4U | 2 | 32 | 16x 2.5" SAS/SATA + 8x 2.5" NVMe | AI/ML, DL Training and Inference, HPC, Rendering Platform for High-end Professional Graphics, Best-in-Class VDI Infrastructure Platform. Single/Dual root configurations available |
| | SYS-421GE-TNRT3 | | | | 8x 2.5" SATA + 8x 2.5" U.2 NVMe (direct to CPU) | AI Compute/Model Training/Deep Learning/HPC Direct Connect Configuration |
| | SYS-521GE-TNRT | 5U | | | 16x 2.5" SATA/SAS + 8x 2.5" U.2 NVMe | AI Compute/Model Training/Deep Learning/HPC High Performance Simulations of Complex 3D Environments |
| | SYS-741GE-TNRT | 4U Tower | | | 8x 3.5" SATA + 8x 2.5" U.2 NVMe direct to CPU | AI Compute/Model Training/Deep Learning/HPC Real-time High Quality Multi-GPU Ray Tracing High Performance Simulations of Complex 3D Environments |
| FatTwin® | SYS-F511E2-RT | 4U8N | 1 (node) | 16 (node) | 6x 2.5" NVMe/SAS/SATA (node) | Hyperscale/Hyperconverged, Cloud Optimized Servers, Data Center Enterprise Applications, Scale out of Storage Expansion, Telcom Data Center & ETSI Certified |
| | SYS-F521E3-RTB | 4U4N | | | 8x 3.5" NVMe/SAS/SATA (node) | |
| BigTwin® | SYS-621BT-H Series | 2U4N | 2 (node) | 16 (node) | 3x 3.5" NVMe/SATA (node) | HCI, HPC, CDN, Hybrid Cloud, CaaS, Cloud Computing, Big Data Analytics, Back-up and Recovery, Scale-out-Storage |
| | SYS-221BT-H Series | | | | 6x 2.5" NVMe/SATA (node) | |
| | SYS-621BT-D Series | 2U2N | | | 6x 3.5" NVMe/SATA (node) | |
| | SYS-221BT-D Series | | | | 12x 2.5" NVMe/SATA (node) | |

| Family | SKU | Form Factor | # of CPUs | # of DIMMs | Drives | Key Applications/Notes |
|---|---|---|---|---|---|---|
| Hyper | SYS-221H-TNR/TN24R | 2U | 2 | 32 | 24x 2.5" NVMe/SAS/SATA | Enterprise Server, Cloud Computing, Big Data Analytics, Hyperconverged Storage, AI Inference and ML, Network Function Virtualization |
| | SYS-621H-TN12R | | | | 12x 3.5" NVMe/SAS/SATA | |
| | SYS-121H-TNR | 1U | | | 12x 2.5" NVMe/SAS/SATA | |
| Hyper-E | SYS-221HE-FTNR | 2U | 2 | 32 | 6x 2.5" NVMe/SAS/SATA | 5G Core and Edge, Telco Micro Data Center |
| | SYS-221HE-FTNRD | | | | | |
| SuperEdge | SYS-211SE-31A | 2U3N | 1 | 8 | 2x internal M.2 NVMe 2280/22110 | 5G Open RAN/Flex-RAN, C-RAN (vRAN), Telecom/Networking Appliance, Multi-Access Edge Computing, Edge Data Center Enterprise Edge Computing. AC or DC Power Supplies |
| | SYS-211SE-31D | 2U3N | | | | |
| 5G/Edge | SYS-111E-FWTR | 1U Short-depth | 1 | 8 | 2x 2.5" internal SATA | Multi-Access Edge Computing, Outdoor DU for 5G Applications, Flex-RAN, Open RAN vBBU, Edge AI, ML AC or DC Power Supplies |
| | SYS-111E-FDWTR | | | | | |
| | SYS-211E-FRN2T | 2U Compact | 1 | 8 | 2x 2.5" hot-swap NVMe | Multi-Access Edge Computing (MEC), C-RAN (vRAN), uCPE, Edge AI, Deep Learning, Advanced Network Security AC or DC Power Supplies |
| | SYS-211E-FRDN2T | | | | | |

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

| Family | SKU | Form Factor | # of CPUs | # of DIMMs | Drives | Key Applications/Notes |
|---|---|---|---|---|---|---|
| CloudDC | SYS-621C-TN12R | 2U | 2 | 16 | 12x 2.5/3.5" NVMe/SAS/SATA | Cloud Computing, Web Servers, Hyper-converged Storage, Virtualization, File Servers, Head-node Computing, 5G Telco AI Inferencing |
| | SYS-521C-NR | | 1 | 16 | 12x 3.5" SAS/SATA | |
| | SYS-121C-TN2R | 1U | 2 | 16 | 8x 2.5" SAS/SATA (2x NVMe hybrid) | |
| | SYS-121C-TN10R | | | | 10x 2.5" NVMe/SAS/SATA | |
| | SYS-611C-TN4R | | | | 4x 3.5/2.5" NVMe/SAS/SATA | |
| | SYS-111C-NR | | 1 | 16 | 10x 2.5" NVMe/SAS/SATA | |
| UP WIO | SYS-511E-WR | 1U | 1 | 8 | 4x 3.5" SAS/SATA | Enterprise Applications, Networking Appliance, Firewall/Security Appliances, General Purpose Computing, Cloud Computing, Media Entertainment |
| | SYS-111E-WR | | | | 10x 2.5" NVMe/SAS/SATA | |
| | SYS-521E-WR | 2U | | | 8x 3.5" SAS/SATA | |
| Multi-processor | SYS-241H-TNRTTP | 2U | 4 | 64 | 8x NVMe hybrid | AI, Business Intelligence, ERP, CRM, Scientific Virtualization, In-memory Database, HCI, SAP HANA |
| | SYS-241E-TNRTTP | | | | 24x NVMe hybrid | |
| | SYS-681E-TR | 6U | 8 | 128 | | |

| Family | SKU | Enclosure | # of CPUs | # of DIMMs | Drives | Key Applications/Notes |
|---|---|---|---|---|---|---|
| SuperBlade | SBI-421E-1T3N | SBE-820J2-830/630 | 2 | 16 | 3x SATA (with 2x NVMe) | AI/ML/DL, HPC, Hybrid Cloud, EDA, Virtualization, Health, Financial Services |
| | SBI-421E-5T3N | | | | 3x SATA (with 2x NVMe) | |
| | SBI-621E-1C3N | SBE-610J2-830/630 | | 32 | 3x SAS (with 2x NVMe) | |
| | SBI-621E-1T3N/5T3N | | | | 3x NVMe/SATA | |
| | SBI-611E-1C2N | SBE-610J2-830/630 | 1 | 16 | 2x NVMe/SAS | |
| | SBI-611E-IT2N/5T2N | | | | 1x NVMe/SATA | |

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

# Appendix B – (courtesy of Intel)

## 4th Gen Intel® Xeon® Scalable Processors

| SKU | CORES | BASE | ALL CORE TURBO (GHz) | Max TURBO (GHz) | CACHE (MB) | TDP (Watts) | Maximum Scalability | DDR5 Memory Speed | UPI Links Enabled | Default DSA Devices | Default IAA Devices | Default QAT Devices | Default DLB Devices | Intel SGX Enclave Capacity (Per Processor) | Long Life Use | Intel® On Demand Capable |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **IMDB/ANALYTICS/VIRTUALIZATION OPTIMIZED (-H) – SOCKET SCALABLE** | | | | | | | | | | | | | | | | |
| 8490H | 60 | 1.9 | 2.9 | 3.5 | 112.5 | 350 | 8S | 4800 | 4 | 4 | 4 | 4 | 4 | 512GB | | |
| 8468H | 48 | 2.1 | 3.0 | 3.8 | 105 | 330 | 8S | 4800 | 4 | 4 | 4 | 4 | 4 | 512GB | | |
| 8460H | 40 | 2.2 | 3.1 | 3.8 | 105 | 330 | 8S | 4800 | 4 | 4 | 4 | 0 | 0 | 512GB | | |
| 8454H | 32 | 2.1 | 2.7 | 3.4 | 82.5 | 270 | 8S | 4800 | 4 | 4 | 4 | 4 | 4 | 512GB | | |
| 8450H | 28 | 2.0 | 2.6 | 3.5 | 75 | 250 | 8S | 4800 | 4 | 4 | 4 | 0 | 0 | 512GB | | |
| 8444H | 16 | 2.9 | 3.2 | 4.0 | 45 | 270 | 8S | 4800 | 4 | 4 | 4 | 0 | 0 | 512GB | | |
| 6448H | 32 | 2.4 | 3.2 | 4.1 | 60 | 250 | 4S | 4800 | 3 | 1 | 1 | 2 | 2 | 512GB | | |
| 6418H | 24 | 2.1 | 2.9 | 4.0 | 60 | 185 | 4S | 4800 | 3 | 1 | 1 | 0 | 0 | 512GB | ✓ | |
| 6416H | 18 | 2.2 | 2.9 | 4.2 | 45 | 165 | 4S | 4800 | 3 | 1 | 1 | 0 | 0 | 512GB | | |
| 6434H | 8 | 3.7 | 4.1 | 4.1 | 22.5 | 195 | 4S | 4800 | 3 | 1 | 1 | 0 | 0 | 512GB | | |
| **5G / NETWORKING OPTIMIZED (-N)** | | | | | | | | | | | | | | | | |
| 8470N | 52 | 1.7 | 2.7 | 3.6 | 97.5 | 300 | 2S | 4800 | 4 | 4 | 0 | 4 | 4 | 128GB | ✓ | ✓ |
| 8471N | 52 | 1.8 | 2.8 | 3.6 | 97.5 | 300 | 1S | 4800 | 0 | 4 | 0 | 4 | 4 | 128GB | ✓ | ✓ |
| 6438N | 32 | 2.0 | 2.7 | 3.6 | 60 | 205 | 2S | 4800 | 3 | 1 | 0 | 2 | 2 | 128GB | ✓ | ✓ |
| 6428N | 32 | 1.8 | 2.5 | 3.8 | 60 | 185 | 2S | 4000 | 3 | 1 | 0 | 2 | 2 | 128GB | ✓ | ✓ |
| 6421N | 32 | 1.8 | 2.6 | 3.6 | 60 | 185 | 1S | 4400 | 0 | 1 | 0 | 0 | 0 | 128GB | ✓ | ✓ |
| 5418N | 24 | 1.8 | 2.6 | 3.8 | 45 | 165 | 2S | 4000 | 3 | 1 | 0 | 2 | 2 | 128GB | ✓ | ✓ |
| 5411N | 24 | 1.9 | 2.8 | 3.9 | 45 | 165 | 1S | 4400 | 0 | 1 | 0 | 2 | 2 | 128GB | ✓ | ✓ |
| **CLOUD OPTIMIZED IaaS (-P) / SaaS(-V) / Media(-M)** | | | | | | | | | | | | | | | | |
| 8468V | 48 | 2.4 | 2.9 | 3.8 | 97.5 | 330 | 2S | 4800 | 3 | 1 | 1 | 1 | 1 | 128GB | | ✓ |
| 8458P | 44 | 2.7 | 3.2 | 3.8 | 82.5 | 350 | 2S | 4800 | 3 | 1 | 1 | 1 | 1 | 512GB | | ✓ |
| 8461V | 48 | 2.2 | 2.8 | 3.7 | 97.5 | 300 | 1S | 4800 | 0 | 1 | 1 | 1 | 1 | 128GB | | ✓ |
| 6438M | 32 | 2.2 | 2.8 | 3.9 | 60 | 205 | 2S | 4800 | 3 | 1 | 1 | 0 | 0 | 128GB | | ✓ |
| **STORAGE & HYPERCONVERGED INFRASTRUCTURE (HCI) OPTIMIZED (-S)** | | | | | | | | | | | | | | | | |
| 6454S | 32 | 2.2 | 2.8 | 3.4 | 60 | 270 | 2S | 4800 | 4 | 4 | 0 | 4 | 4 | 128GB | | ✓ |
| 5416S | 16 | 2.0 | 2.8 | 4.0 | 30 | 150 | 2S | 4400 | 3 | 1 | 0 | 2 | 2 | 128GB | | ✓ |

Y Supports Intel Speed Select Technology – Performance Profile 2.0 (Intel SST-PP)

Intel may make changes to specifications and product descriptions at any time, without notice. Please visit intel.com/xeon or contact your Intel representative to obtain the latest product specifications. Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. All processors support Intel Virtualization Technology (Intel VT-x).

© 2023 Copyright Super Micro Computer, Inc. All rights reserved

## 4th Gen Intel® Xeon® Scalable Processors

| SKU | CORES | BASE | ALL CORE TURBO (GHz) | Max TURBO (GHz) | CACHE (MB) | TDP (Watts) | Maximum Scalability | DDR5 Memory Speed | UPI Links Enabled | Default DSA Devices | Default IAA Devices | Default QAT Devices | Default DLB Devices | Intel SGX Enclave Capacity (Per Processor) | Long Life Use | Intel® On Demand Capable |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2S PERFORMANCE GENERAL PURPOSE** | | | | | | | | | | | | | | | | |
| 8480+ | 56 | 2.0 | 3.0 | 3.8 | 105 | 350 | 2S | 4800 | 4 | 1 | 1 | 1 | 1 | 512GB | | ✓ |
| 8470 | 52 | 2.0 | 3.0 | 3.8 | 105 | 350 | 2S | 4800 | 4 | 1 | 0 | 0 | 0 | 512GB | ✓ | ✓ |
| 8468 | 48 | 2.1 | 3.1 | 3.8 | 105 | 350 | 2S | 4800 | 4 | 1 | 0 | 0 | 0 | 512GB | | ✓ |
| 8460Y+ | 40 | 2.0 | 2.8 | 3.7 | 105 | 300 | 2S | 4800 | 4 | 1 | 1 | 1 | 1 | 128GB | | ✓ |
| 8462Y+ | 32 | 2.8 | 3.6 | 4.1 | 60 | 300 | 2S | 4800 | 3 | 1 | 1 | 1 | 1 | 128GB | ✓ | ✓ |
| 6448Y | 32 | 2.1 | 3.0 | 4.1 | 60 | 225 | 2S | 4800 | 3 | 1 | 0 | 0 | 0 | 128GB | ✓ | ✓ |
| 6442Y | 24 | 2.6 | 3.3 | 4.0 | 60 | 225 | 2S | 4800 | 3 | 1 | 0 | 0 | 0 | 128GB | | ✓ |
| 6444Y | 16 | 3.6 | 4.0 | 4.1 | 45 | 270 | 2S | 4800 | 3 | 1 | 0 | 0 | 0 | 128GB | | ✓ |
| 6426Y | 16 | 2.5 | 3.3 | 4.1 | 37.5 | 185 | 2S | 4800 | 3 | 1 | 0 | 0 | 0 | 128GB | ✓ | ✓ |
| 6434 | 8 | 3.7 | 4.1 | 4.1 | 22.5 | 195 | 2S | 4800 | 3 | 1 | 0 | 0 | 0 | 128GB | | ✓ |
| 5415+ | 8 | 2.9 | 3.6 | 4.1 | 22.5 | 150 | 2S | 4400 | 3 | 1 | 1 | 1 | 1 | 128GB | ✓ | ✓ |
| **2S MAINLINE GENERAL PURPOSE** | | | | | | | | | | | | | | | | |
| 8452Y | 36 | 2.0 | 2.8 | 3.2 | 67.5 | 300 | 2S | 4800 | 4 | 1 | 0 | 0 | 0 | 128GB | ✓ | ✓ |
| 6438Y+ | 32 | 2.0 | 2.8 | 4.0 | 60 | 205 | 2S | 4800 | 3 | 1 | 1 | 1 | 1 | 128GB | | ✓ |
| 6430 | 32 | 2.1 | 2.6 | 3.4 | 60 | 270 | 2S | 4400 | 3 | 1 | 0 | 0 | 0 | 128GB | ✓ | ✓ |
| 5420+ | 28 | 2.0 | 2.7 | 4.1 | 52.5 | 205 | 2S | 4400 | 3 | 1 | 1 | 1 | 1 | 128GB | ✓ | ✓ |
| 5418Y | 24 | 2.0 | 2.8 | 3.8 | 45 | 185 | 2S | 4400 | 3 | 1 | 0 | 0 | 0 | 128GB | ✓ | ✓ |
| 4416+ | 20 | 2.0 | 2.9 | 3.9 | 37.5 | 165 | 2S | 4000 | 2 | 1 | 1 | 1 | 1 | 64GB | ✓ | ✓ |
| 4410Y | 12 | 2.0 | 2.8 | 3.9 | 30 | 150 | 2S | 4000 | 2 | 1 | 0 | 0 | 0 | 64GB | ✓ | ✓ |
| **LIQUID COOLED GENERAL PURPOSE(-Q)** | | | | | | | | | | | | | | | | |
| 8470Q | 52 | 2.1 | 3.2 | 3.8 | 105 | 350 | 2S | 4800 | 4 | 1 | 0 | 0 | 0 | 512GB | | ✓ |
| 6458Q | 32 | 3.1 | 4.0 | 4.0 | 60 | 350 | 2S | 4800 | 3 | 1 | 0 | 0 | 0 | 128GB | | ✓ |
| **SINGLE SOCKET GENERAL PURPOSE(-U)** | | | | | | | | | | | | | | | | |
| 6414U | 32 | 2.0 | 2.6 | 3.4 | 60 | 350 | 1S | 4800 | 0 | 1 | 0 | 0 | 0 | 128GB | | ✓ |
| 5412U | 24 | 2.1 | 2.9 | 3.9 | 45 | 185 | 1S | 4400 | 0 | 1 | 0 | 0 | 0 | 128GB | | ✓ |
| 3408U | 8 | 1.8 | 1.9 | 1.9 | 22.5 | 125 | 1S | 4000 | 0 | 1 | 0 | 0 | 0 | 64GB | ✓ | |
| **LONG - LIFE USE (IOT) GENERAL PURPOSE (-T)** | | | | | | | | | | | | | | | | |
| 4410T | 10 | 2.7 | 3.4 | 4.0 | 26.25 | 150 | 2S | 4000 | 2 | 1 | 0 | 0 | 0 | 64GB | ✓ | ✓ |

Intel may make changes to specifications and product descriptions at any time, without notice. Please visit intel.com/xeon or contact your Intel representative to obtain the latest product specifications. Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. All processors support Intel Virtualization Technology (Intel VT-x).

© 2023 Copyright Super Micro Computer, Inc. All rights reserved