White Paper

Intel NEX
Liquid Cooling Solution

Delivering a High-Performance Compute for Extreme Edge Environments

Evaluation of Nexalus liquid cooled solution, powered by a 4th Gen Intel® Xeon® processor to deliver an efficient and optimized compute for ruggedized edge deployments.

Authors:

Uzair Qureshi
Edge System Architecture, Intel

Marek Piotrowski
Edge System Architecture, Intel

Dr. Anthony Robinson
Chief Science Officer, Nexalus

Dr Sara Battaglioli
Project Manager, Nexalus

Table of Contents

Executive Summary

The rise of edge computing, fueled by technologies like private 5G networks, poses significant challenges in power consumption and thermal management. To address these challenges, Intel is collaborating with the broader ecosystem from the data center to the edge. In collaboration with Nexalus their Nexalus Endeavor DX-800 was used, which provides an innovative solution that was evaluated for environmentally-challenged edge deployments.

This paper demonstrates the effectiveness of cold plate jet impingement solution within an IP66 sealed enclosure through a retrofit case study of a 4th Gen Intel® Xeon® and Intel® Data Center GPU Flex Series for Edge AI based solution utilizing a Dell PowerEdge XR5610 server. As part of our evaluation, we compared the thermal performance of the standard air-cooled server against a similar server retrofitted with Nexalus Endeavor DX-800 solution.

Introduction

Edge computing marks a significant evolution in the way data processing and computing tasks are handled. It minimizes latency and conserves bandwidth - by bringing computation, AI, and data storage closer to the location where it is needed - leading to more efficient operations in a wide range of applications. This decentralized approach contrasts with traditional cloud computing, where data is processed in a central location far from the end-user. However, the shift towards edge computing brings forth unique challenges, particularly in the areas of thermal management, device durability, and system maintenance.

In edge computing, devices are often deployed in non-traditional, sometimes harsh environments. These can range from outdoor settings exposed to the elements to industrial environments with extreme temperatures or dust to hospital settings where the absence of contamination and low noise are paramount. Such conditions demand not only robust and resilient hardware but also innovative solutions to manage heat effectively and ensure consistent performance.

The main critical aspects intrinsic to edge computing are:

High Power Density and Heat Generation:
Higher compute capabilities result in greater power consumption, generating increased heat within the system. This elevated power density poses challenges as it can lead to localized hotspots and thermal throttling, which in turn lead to performance degradation and inconsistent latency which is a key value for edge if not effectively managed. The primary concern revolves around the efficient distribution and transfer of this heat flow throughout the system.

High Variety of Ambient Conditions and Rugged Environments:
Edge devices are frequently deployed in diverse environments where ambient conditions vary significantly. Extreme temperatures or dusty surroundings can adversely impact cooling efficiency and overall platform performance, which negatively impact the service-level agreements (SLAs). To mitigate these challenges, the solution should be robust and environmentally sealed, and ruggedized design is needed for such deployment. Such measures protect devices from external environmental factors without the need for substantial building infrastructure costs. Furthermore, adaptive cooling systems, capable of adjusting their operation based on ambient conditions, play a vital role in maintaining optimal temperatures.

Limited Space for Cooling Systems:
Compact and spatially constrained edge devices face challenges in accommodating advanced cooling solutions. With the increase in computing power, the concentration of generated heat becomes more pronounced, making effective heat dissipation a significant challenge. The key lies in designing cooling solutions that are not only compact but also highly efficient and sustainable, tailored specifically for edge devices. This approach eliminates the necessity for temperature and humidity-controlled buildings, addressing the limitations posed by restricted space.

These challenges require solutions that are robust, low-maintenance, and can fit into existing telco/edge infrastructures.

The following sections provide an overview of existing cooling technologies that are adopted for various deployment within the data center and edge:

Technology Overview

Air cooling, the predominant method for managing temperatures in today's servers, leverages air as a coolant due to its dielectric nature, non-toxicity, abundance, and lack of intrinsic global warming potential. This approach is particularly appealing due to its simplicity and accessibility. However, the limitations of air cooling are becoming increasingly apparent, especially as server power and heat flux levels rise.

Air's poor thermophysical properties make it challenging to manage the heat fluxes from powerful components like CPUs and GPUs while maintaining them within safe temperature limits. Pushing the limits of air cooling involves an escalation in airflow that results in significantly higher acoustic noise and fan power consumption, in addition to increased server height to accommodate larger heat sinks, which contradicts the essential need for compact and efficient design in edge computing.

Furthermore, in edge environments, where devices often operate in harsh conditions, the inefficiency of air cooling is exacerbated, as dust, humidity, and other particulates can clog air filters and cooling fans, leading to increased maintenance needs and a higher risk of system failures.

Considering the above, the imperative for more efficient cooling solutions becomes increasingly pronounced. Liquid cooling emerges as a forefront contender, distinguished by its efficiency and effectiveness in dissipating heat generated by high-powered servers. This cooling method can be broadly categorized into immersion cooling and direct liquid cooling.

Immersion cooling technologies, whether single-phase or two-phase, involve submerging the whole server in a tank with dielectric liquid. This method is particularly effective for managing high heat fluxes from components like CPUs and GPUs, thanks to the better thermophysical properties of the cooling liquids.

Unlike air cooling, immersion cooling is less affected by external factors such as ambient temperature fluctuations or dust, making it a potential choice for the diverse conditions often encountered in edge deployments. However, cooling using immersion tanks comes with their own set of challenges, especially deploying in remote or non-traditional edge locations, and there are also concerns on liquid containment. Numerous initiatives around the industry are actively working to address these challenges and immersion is a key emerging solution for future edge deployments.

Direct liquid cooling technology is an alternative, but another efficient method of removing heat from high-performance computing systems and servers. This technology employs cold plate heat exchangers, typically utilizing microchannel designs, to cool electronic components directly. The cooling process involves the coolant circulating through the cold plates attached directly to heat-generating components like CPUs and GPUs.

The key engineering motives for direct liquid cooling are:

- Cold plates have undergone extensive research and optimization and offer such low thermal resistance that they significantly reduce the temperature difference between the processors and the coolant, to the extent that elevated coolant temperatures are feasible.
- This extra headroom in the thermal budget means that above-ambient coolant temperatures can be used, which eliminates the need to chill the coolant; the coolant can dissipate heat directly to the ambient environment or repurpose it for other heating applications, contributing to energy efficiency.

Direct liquid cooling is usually used in conjunction with traditional air-cooling methods in hybrid cooling systems, where cold plates are used for high-powered electronics (like CPUs, GPUs, and other accelerators) and conventional air cooling for other lower power peripheral components (like memory, network controllers, storage media, power supply units, and voltage regulators, etc.).

Nexalus Solution for the Edge

In the realm of hybrid cooling solutions, Nexalus technology presents an innovative solution to the thermal challenges faced by edge devices. The introduction of the Nexalus Endeavor DX-800 (Figure 1) represents a paradigm shift in thermal management, addressing the limitations posed by space constraints and varying ambient conditions:

- **Compact and Efficient Thermal Management:** One of the core innovations of Nexalus technology lies in its compact and efficient thermal management system, utilizing Nexalus patented liquid jet impingement cooling cold plates for superior heat dissipation, ensuring optimum temperatures even within the tight confines of edge devices. In addition, by incorporating advanced components such as the Nexalus NXQE Heat Rejection Unit, NXP20 High Flow Pumps, the Nexalus Endeavor DX-800 retrofit achieves unparalleled efficiency.
- **Adaptability to Ambient Conditions:** Edge devices operate in diverse environmental conditions, ranging from extreme cold to scorching heat. The Nexalus technology addresses this challenge by dynamically adjusting its cooling mechanisms, such as coolant flow rates and heat rejection unit fan's speed. This adaptability ensures consistent performance even in temperatures ranging from -20°C (-4°F) to 60°C (140°F), surpassing the capabilities of conventional air-cooled systems. In addition, recognizing the diverse and harsh environments where edge devices operate, Nexalus technology incorporates a robust IP66-rated enclosure. This provides a safeguard against moisture, dust, and extreme temperatures, preserving the integrity of the server and guaranteeing uninterrupted performance.
- **Plug and Play Deployment:** The Nexalus Endeavor DX-800 boasts a user-friendly and seamless deployment process, allowing for quick and straightforward setup, ensuring swift and efficient integration into various environments.
- **Portable Size, Lightweight Module:** The compact and portable design of the Nexalus solution, comparable to the size of a suitcase, enhances its mobility and adaptability. This feature, along with the lightweight design, facilitates easy transportation and mounting in diverse locations, making it an efficient alternative solution for edge deployments.
- **No External Services or Specialist Skills Required:** The Nexalus Endeavor DX-800 eliminates the need for external services or specialized skills during deployment and operation. Its design emphasizes user-friendliness, allowing organizations to leverage the benefits of edge computing without requiring extensive expertise or additional services.
- **Full System Integration Before On-site Deployment:** The Nexalus Endeavor DX-800 supports full system integration before on-site deployment, streamlining the setup process. This feature enhances efficiency by allowing comprehensive testing and configuration adjustments before installation.
- **Swap-out or Upgrade Easily Completed by Removal of a Single Panel:** The modular design of the Nexalus Endeavor DX-800 facilitates easy maintenance and upgrades. Any necessary component swap-out or upgrades can be conveniently accomplished by removing a single panel, minimizing downtime and operational disruptions.
- **Quiet Operation for Easy Location:** With a focus on quiet operation, the Nexalus Endeavor DX-800 can be easily located without causing disturbances. This feature enhances its adaptability for deployment in various settings, including those with noise-sensitive considerations.

In conclusion, Nexalus technology offers a unique solution that is IP66 certified with extended temperature range enabling thermally-challenged deployments. In this context, a tailored experiment procedure and precise test approach were executed to validate Nexalus technology impact on edge devices. As an example, the following sections focus on a detailed comparison between an air-cooled Dell PowerEdge XR5610 server and its Nexalus Endeavor DX-800 retrofit counterpart. The fundamental characteristics of both configurations are described, and the key components of the Nexalus system are detailed. Through comparisons of test results, our analysis demonstrates the significant advancements in thermal management, energy efficiency, and overall performance introduced by Nexalus technology in the context of edge devices.

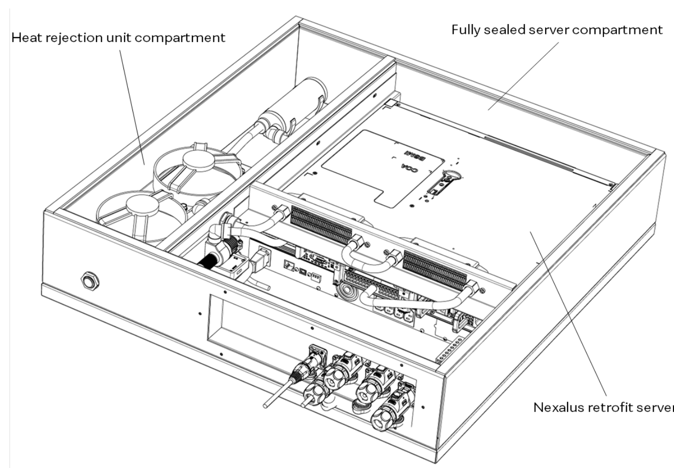


Figure 1. Representation of a server embedded in a Nexalus Endeavor DX-800

Experiment Procedure and Testing Approach

Overview

The initial phase of the experiment involved thermal testing of an air-cooled Dell PowerEdge XR5610 server. These tests were conducted by Intel under controlled conditions at 30°C (86°F) and 50°C (122°F) ambient temperature with the objective to establish baseline metrics for the server operating under standard conditions.

Parameters such as processing power, CPU and GPU temperatures, and energy consumption were measured and recorded. Following the testing of the air-cooled server, the Nexalus Endeavor DX-800 with a Dell PowerEdge XR5610 server underwent evaluation within a controlled environmental chamber. This phase of testing allowed simulation of a wide range of ambient temperatures (30°C (86°F) to 60°C (140°F)).

Temperature measurements were taken using ten thermocouples strategically placed to monitor various liquid and air temperatures at different locations within the server. Data collected provided valuable insights into the server's thermal behavior under varying workloads and ambient temperatures. In the series of tests listed above, the assessment of server performance involved the utilization of specific software tools to induce stress on the CPU and GPU and monitoring of the server's temperature levels. The collected temperature and performance data from both the air-cooled and Nexalus Endeavor DX-800 retrofit configurations were analyzed comprehensively. This detailed analysis, presented in the subsequent sections, provides a clear understanding of the thermal dynamics, workload handling, and power efficiency of the Nexalus Endeavor DX-800 server in comparison to its air-cooled counterpart.

Air-cooled Solution for the Edge based on Dell PowerEdge XR5610

System Characteristics

The air-cooled Dell PowerEdge XR5610 server (Figure 2), in its base air-cooled configuration, represents a typical example of conventional thermal management systems. Its essential characteristics provide a foundation for understanding the baseline performance before the implementation of Nexalus technology. The key specifications of the server are:

- **Case Dimensions:** The server is housed in a case measuring 483 mm x 477 mm x 43 mm (19" x 18.8" x 1.7").
- **Processor:** Intel® Xeon® Gold 6421N processor with a Thermal Design Power (TDP) of 185 W.
- **Graphics Card:** Intel® Data Center GPU Flex 140 with a TDP of 75 W.
- **Memory:** 8 × 16 GB DDR5 RAM.
- **Storage:** 2 × 1.6 TB Solid State Drives (SSD).

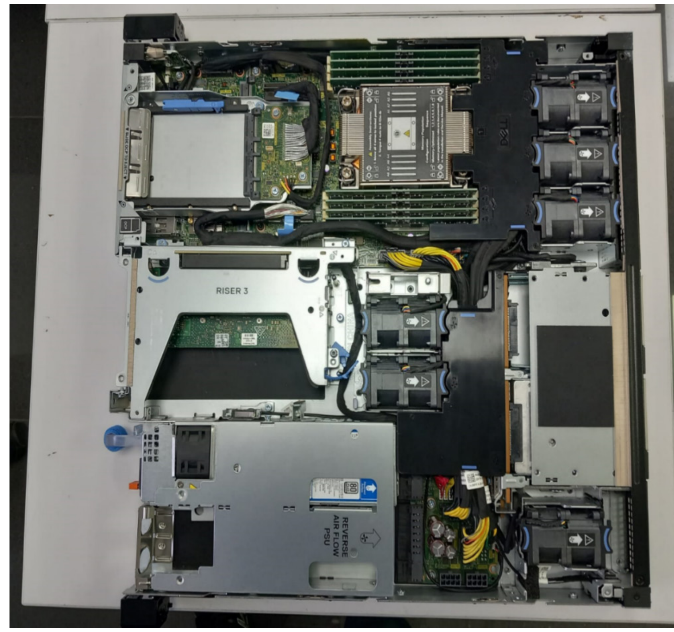


Figure 2. Air-cooled Dell PowerEdge XR5610 server

Thermal Management System

The server cooling system relies on 6 × 40 mm (1.6") fans strategically placed within the case to draw ambient air into the server through rear panel vents. This air is then forced over the internal electronics, including the CPU and GPU, providing essential cooling to maintain optimal operating temperatures.

Critical components like the CPU and GPU are located in close proximity to the fans and are equipped with finned heat sinks to ensure effective heat dissipation and prevent overheating.

Thermal Performance Evaluation

Intel conducted thermal testing of the air-cooled Dell PowerEdge XR5610 server under a controlled environment (utilizing thermal chambers) at various ambient temperatures of 30°C (86°F) and 50°C (122°F). Three different system configurations were used and workload performance was observed at various temperature, those configurations are listed as follows:

- 100% stress on CPU and no stress on GPU
- 100% stress on GPU and no stress on CPU
- 100% stress on CPU and 100% stress on GPU

Nexalus Endeavor DX-800 Cold Plate Solution

System Characteristics

The Nexalus Endeavor DX-800 retrofit for the Dell PowerEdge XR5610 server (Figure 3) comprises several key components, each designed to enhance cooling efficiency and ensure reliable performance in edge deployment scenarios.

The main components of Nexalus retrofit solution are:

- **Nexalus Enflux CPU Cold Plate:** The Enflux is specifically modified for 4th/5th Gen Intel® Xeon® product family. Contact with the CPU is further enhanced with Nexalus Hydronex Thermal Interface Material (TIM), which significantly reduce contact resistance between CPU & the cold plate.
- **Nexalus GPU Flex 140 Cold Plate:** Similarly, the GPU Flex 140 cold plate was mounted to Intel® Data Center GPU Flex 140, to allow effective heat dissipation contributing to the overall thermal efficiency of the system.
- **Nexalus NXP20 High Flow Pumps:** The solution is equipped with two NXP20 high flow pumps to allow efficient circulation of the liquid coolant with redundancy while maintaining steady flow rate to critical components i.e. CPU and GPU.
- **Nexalus NXQ 1U Heat Exchangers:** The Nexalus NXQ air-liquid heat exchangers are strategically positioned within the server. It facilitates the transfer of heat from the internal (non-processor) system components to the liquid coolant, ensuring adequate cooling.
- **Nexalus FluidX Coolant:** FluidX, formulated with a 5% dilution, is a water-based liquid coolant combined with organic and corrosion inhibitors package that offers all-round protection.
- **Nexalus Reservoir:** The Nexalus reservoir provides a stable and continuous supply of coolant to the system. It ensures consistent thermal performance by maintaining the proper coolant levels and flow rates within the server.
- **Nexalus Heat Rejection Unit:** At the core of the Nexalus solution is the heat rejection unit, featuring a miniature dry cooler fitted with two miniature outdoor-grade fans that is used to dissipate heat generated by the server.
- **Power Supply and PTC Heater:** The Nexalus system includes three 12 V / 5A power supplies for heat rejection unit fans and pumps, ensuring a reliable power source. Additionally, a Positive Temperature Coefficient (PTC) heater with a dedicated power supply is integrated to prevent condensation within the enclosure at low temperatures. This feature maintains a controlled internal environment, safeguarding the server components from moisture-related issues.
- **Nexalus Enclosure:** The Nexalus enclosure provides an IP66 fully sealed server compartment, offering protection against environmental factors such as precipitation, humidity, temperature fluctuations, salt, and dust. This sealed design ensures the server's integrity even in harsh outdoor environments, making it ideal for edge deployment.

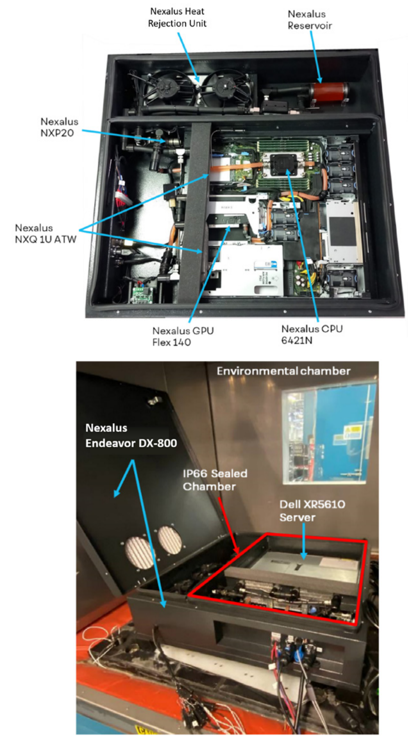


Figure 3. (top) Nexalus Endeavor DX-800 (bottom) Nexalus Endeavor DX-800 in the environmental test chamber (cover open for clarity)

Thermal Management System

Coolant Circulation and Heat Dissipation (Figure 4): The Nexalus NXP20 pumps circulate the liquid coolant into the IP66 sealed compartment of the enclosure, which houses the server. The coolant is then routed to Nexalus Enflux cold plate mounted on Intel® Xeon® Gold 6421N. Hydraulically in series, the liquid coolant is then passed through a Nexalus cold plate attached to an Intel® Flex 140 graphics card.

The working fluid then passes through the two internal Nexalus NXQ air-liquid heat exchangers. To facilitate the cooling of all supporting electronics of the server, the air within the sealed compartment circulates throughout this internal enclosure and across the Nexalus NXQ 1U ATW heat exchangers, where the collected heat is transferred to the liquid coolant.

For air circulation Nexalus utilized the current system fans to cool the server components. In this configuration all the heat generated by the server transferred into the liquid coolant stream is finally directed out of the server compartment to the ambient-exposed heat rejection unit compartment. Here, it circulates through the Nexalus dry cooler, which dissipates all the heat generated within the sealed server compartment to the ambient surrounding air through vents in the enclosure lid.

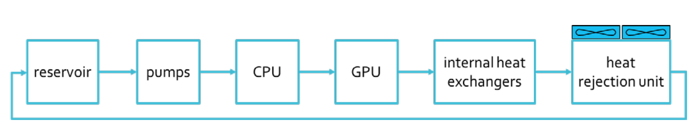


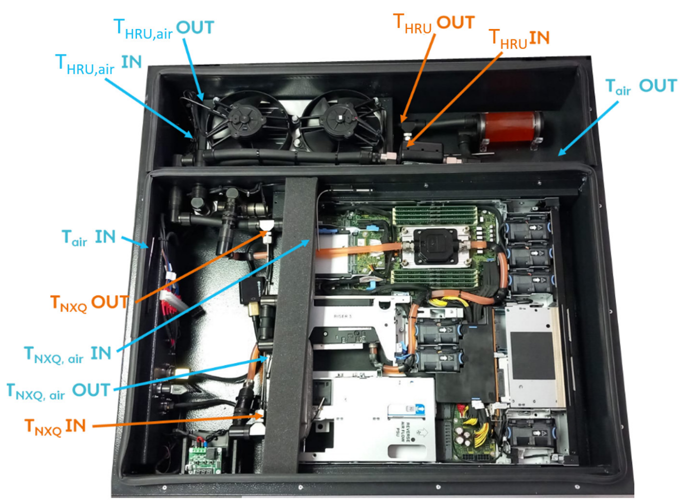
Figure 4. Nexalus Endeavor DX-800 Cooling Loop

Thermal Performance Evaluation

The Dell PowerEdge XR5610 retrofitted with the Nexalus Endeavor DX-800 was tested using an environmental chamber that provides a temperature range from -100°C (-148°F) to 177°C (350.6°F), with a temperature control accuracy of ±1.1°C (±1.98°F). A wide series of test scenarios was considered, by controlling the following variables:

- **Workload:**
  - 100% stress on CPU and no stress on GPU
  - 100% stress on GPU and no stress on CPU
  - 100% stress on CPU and 100% stress on GPU
- **Power setting (BIOS):**
  - "The Nexalus system was tested in **"Power Management Mode"** to showcase the power efficiency of Nexalus solution, and also same was done with BIOS updated configured to **"Cooling Boost Mode"** to demonstrate the maximum cooling potential at higher temperatures.
- **Environmental test chamber temperature:** 30°C, 50°C and 60°C (86°F to 140°F)

Ten thermocouples were strategically placed to monitor liquid and air temperatures at different locations during testing, as depicted in Figure 5.



Thermocouple Name	Location
T_{air} IN	Air inside sealed enclosure
T_{air} OUT	Air inside exhaust enclosure
$T_{NXQ,air}$ IN	NXQ heat exchanger inlet air
$T_{NXQ,air}$ OUT	NXQ heat exchanger outlet air
$T_{HRU,air}$ IN	Dry cooler inlet air
$T_{HRU,air}$ OUT	Dry cooler outlet air
T_{NXQ} IN	NXQ heat exchanger inlet liquid
T_{NXQ} OUT	NXQ heat exchanger outlet liquid
T_{HRU} IN	Dry cooler inlet liquid
T_{HRU} OUT	Dry cooler outlet liquid

Figure 5. Location of liquid (orange) and air (blue) thermocouples

Results

Table 1. Results of performance test of air-cooled Dell XR5610 server.

Stress Test	Ambient T °C	CPU power W	GPU power W	Total power W	CPU T °C	GPU T °C
GPU	30	111.5	42.2*	259.4	53.0	78.0
	50	114.5	42.0*	450.6	65.1	79.9
CPU	30	186.5	NA	319.8	80.7	NA
	50	185.1	NA	378.5	82.7	NA
CPU+GPU	30	186.0	40.0*	343.4	76.3	77.3
	50	185.6	40.0*	558.3	76.1	76.2

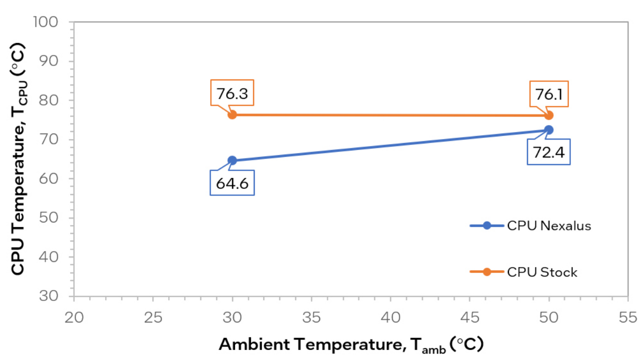
*Power measured from GPU die only, the actual listed GPU total power is 75W

Table 2. Results of performance test of Nexalus Endeavor DX-800.

Stress Test	Ambient T °C	CPU power W	GPU power W	Total power W	CPU T °C	GPU T °C
GPU	30	101.3	42.4*	240.6**	53.8	55.0
	50	102.5	41.9*	260.6**	61.3	61.4
	60	103.7	41.1*	590.0***	71.5	71.0
CPU	30	184.6	NA	318.3**	63.9	NA
	50	184.7	NA	338.1**	70.3	NA
	60	184.4	NA	664.8***	80.6	NA
CPU+GPU	30	185.1	37.3*	324.6**	64.6	56.6
	50	185.1	37.7*	362.8**	72.4	62.3
	60	183.0	38.5*	687.2***	80.9	71.6

*Power measured from GPU die only, the actual listed GPU total power is 75W
** Platform BIOS settings set to "Power Management Mode"
*** Platform BIOS setting set to "Cooling Boost Mode"

(a) 100% CPU+GPU Workload



(b) 100% CPU+GPU Workload

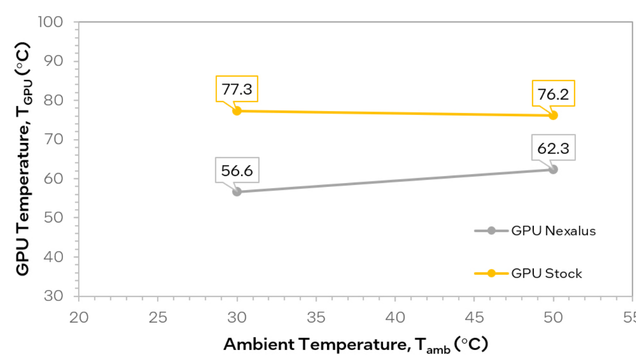
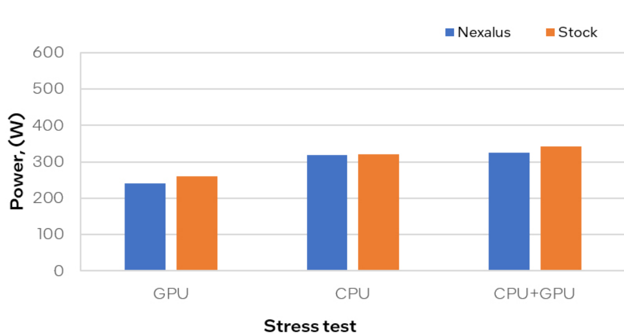


Figure 6: CPU (a) and GPU (b) core temperatures over test temperature range for air-cooled server vs. Nexalus Endeavor DX-800 in power management mode.

(a) Power consumption at 30°C



(b) Power consumption at 50°C

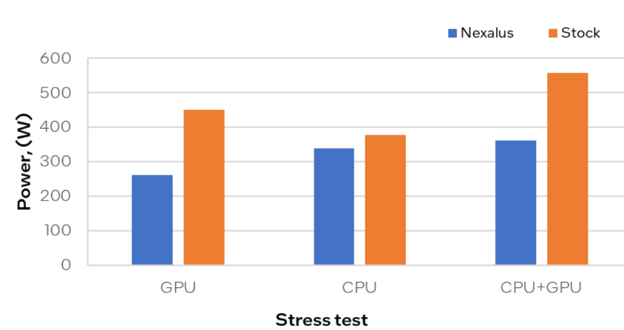


Figure 7: Comparison of total power consumption of air-cooled server vs. Nexalus Endeavor DX-800 in power management mode.

Conclusion

Based on the results presented above, it is evident that the Nexalus Endeavor DX-800 offers a range of significant advantages over the traditional air-cooled server configuration for edge computing:

- **Enhanced Cooling Performance:**
  As shown from the data presented above, Nexalus Endeavor DX-800 system consistently maintains lower temperatures for both CPU and GPU components and delivers a highly efficient solution for ruggedized edge.
- **Extended Operational Range:**
  Higher temperature of 60°C without CPU & GPU throttling were validated, compared to throttling observed with air cooling.
- **Noise reduction:**
  Noise measurements conducted in the Nexalus lab revealed a significant 23 dB reduction in noise levels between the Nexalus server (45.5 dB) and the air-cooled server (68.5 dB), demonstrating that the Nexalus solution is significantly quieter operation owing to the to its sealed architecture and optimized cooling system.
- **Power Consumption Dynamics:**
  The system was tested by configuring the BIOS in "Power Management Mode" mode which substantially reduces the fan speed and system power. Taking into consideration the harsh environment, power savings across all observed tests were roughly 16%.

Overall, the Nexalus Endeavor DX-800 system is an efficient solution for edge and telco customers that are looking to deploy scalable solution with 4th/5th Gen Intel® Xeon® processor family and Intel® Data Center GPU Flex Series that requires extended operating range with low latency to meet customer SLAs.