White Paper

Future-Ready Cooling Solutions and Easy Service Design for AI Training Systems

October 2022

Swappable Modular Design for Air Cooling and Liquid Cooling

Authors

Ash Liao
Assistant Technical Manager, Prospect Technology Research Division
Wiwynn

Jack Yeh
Assistant Technical Manager, Mechanical Department 5
Wiwynn

Eric CW Tseng
Sr. Engineer, Mechanical Department 1
Wiwynn

Table of Contents

Introduction 1
Liquid-Cooling Solution to Support High Power Accelerators 2
Swappable Modular Design for Air Cooling and Liquid Cooling 2
Rack-Level Facility of Liquid Cooling 3
Labor-Saving Mechanism 4
Wiwynn Integrated AI Training System - SV600G2 5
L10 On-Rack Serviceability 7
Two-Phase Cooling Implementation 7
Conclusion 9

Summary

Air cooling is the traditional way to remove heat generated from processing units by directing cool air through hot surfaces to dissipate heat. In recent years, the rapid expansion of AI and HPC is in demand of higher performance in both the processing units and in heat dissipation. Aimed to address the heat dissipation limit of air cooling, liquid cooling is introduced as an advanced cooling method to have coldplate array dissipate heat directly from the processing units. With significant research into AI training system and other data center applications, we came up with multiple swappable cooling solutions for cooling hot processing units (OCP accelerator module) within the same AI training system. This swappable cooling solution design can provide quick and easy upgrade of cooling in most scenarios.

Introduction

Thermal solution is now the main challenge for unleashing a processor's full potential. The required heat flux increases along with increased performance. Furthermore, cooling graphics processing units (GPU) and AI accelerators is now even more important and is what most new cooling technology are specifically targeting.

Gaming, 3D visualization, and AI training with big data are three of the major trending applications that rely heavily on quick and parallel processing of GPUs. As the trend pushes the expected performance further, the limitations of processing speed for add-on-card form factor GPU forced the adoption and the launch of mezzanine-type accelerator module. Moreover, data transfer between GPU cards relies on CPU processing. The interconnecting topology of mezzanine-type accelerator modules offers direct data transfer and thus provides low latency transfers between accelerators.

Accelerator modules, while having more bandwidth and power consumption, are increasingly playing a key role in AI training systems. The Open Accelerator Infrastructure (OAI) group, a subgroup of the Open Compute Project Foundation (OCP), also defined a unified and open source specification for a mezzanine-type accelerator (OCP Accelerator Module, OAM) and universal baseboard (UBB) in order to increase the innovative pace.

Air cooling is the most common solution for server products. When it comes to thermal solution upgrades, such as to support increased accelerator TDP, most users would prefer keeping existing server architecture with only minor changes or upgrades. Liquid cooling is one cost-effective way to achieve this. Upgrading to such thermal solution has high acceptability as it requires minimum change from air-cooling system.

Liquid cooling server products have surfaced in recent years. Compliance of such cooling facility is an important mechanical design consideration. Another point of consideration is the user-friendliness of the design that should decrease the time cost and smooth the maintenance procedure.

In this whitepaper, Wiwynn provides a mature example of integrated AI training system with both air-cooling and liquid-cooling solutions. The maximized modular and quick-swappable design and a OAM system with various power dissipation support, the example demonstrates that the system is ready to satisfy all AI application demands.

Liquid-Cooling Solution to Support High Power Accelerators

There are 2 approaches for system cooling: air-cooling (AC) and liquid-cooling (LC). The two systems can share the same architecture, and the modular design of the two cooling solutions allows interchangeability for extra flexibility.

The supported power consumption depends on the system cooling capability. For example, an air cooling solution can support Habana HL225H accelerator, which features up to 600 W TDP. Liquid cooling solution is used for accelerators exceeding 600 W power consumption.

Swappable Modular Design for Air Cooling and Liquid Cooling

To lower the cost and maximize the quantity of common parts, Wiwynn keeps the same architecture but goes into the modular design route for air-cooling and liquid-cooling systems. For the liquid-cooling systems, the two fan modules can be removed from the rear of the air-cooling system and replaced by quick connector (QC) brackets of liquid-cooling cold plate. The brackets support the cold plate terminal connectors on the liquid-cooling system. As shown in Figure 1, the air-cooling and liquid-cooling systems leverage most of the parts except the interchangeable modules between the two cooling solutions such as the quick connector brackets of cold plate, OAM heatsinks, cold plate array and fan modules. It is easy and efficient to replace one cooling solution with the other without the need to change the entire system.

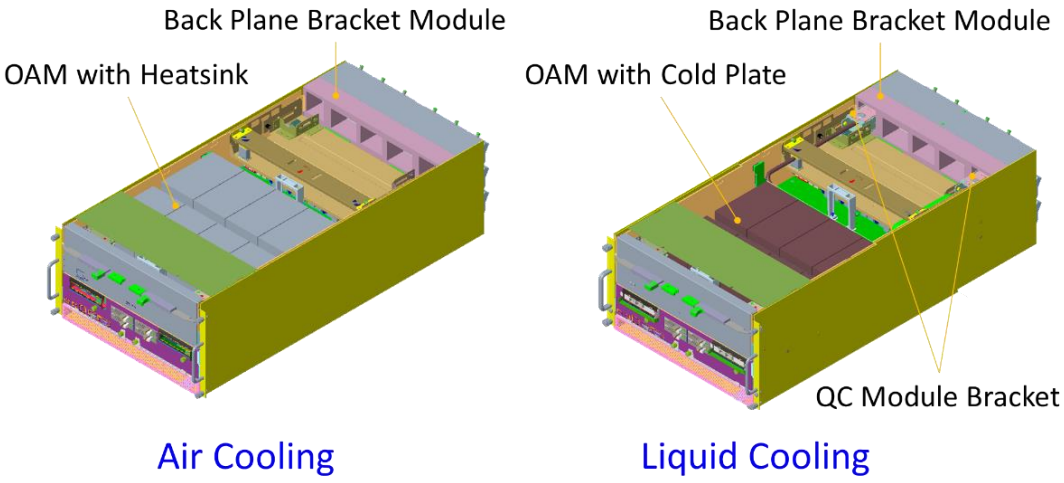


Figure 1 Modular Design for Air Cooling and Liquid Cooling

Swappable Module	Air-Cooling System	Liquid-Cooling System
OAM Cooling Module	OAM with air-cooling heatsink	OAM and cold plate array
Coolant I/O	2x fan module	QC bracket of cold plate

Table 1 Modular Design for Air Cooling and Liquid Cooling

Rack-Level Facility of Liquid Cooling

In addition to the thermal design parts, there are also facilities at the rack level that correspond to the entire cooling system. At the rear area of rack, Wiwynn has developed 2 types of liquid-cooling connections – blind-mate and manual-mate connections. As shown in Figure 2, the blind-mate connection is easier to connect between the cold plate terminals and the manifolds. However, the manifolds are placed right behind the L10 system that partially block the airflow, influencing the thermal efficiency. Additionally, blind-mate quick connectors can only withstand a smaller pressure range, making the applicable scenarios limited.

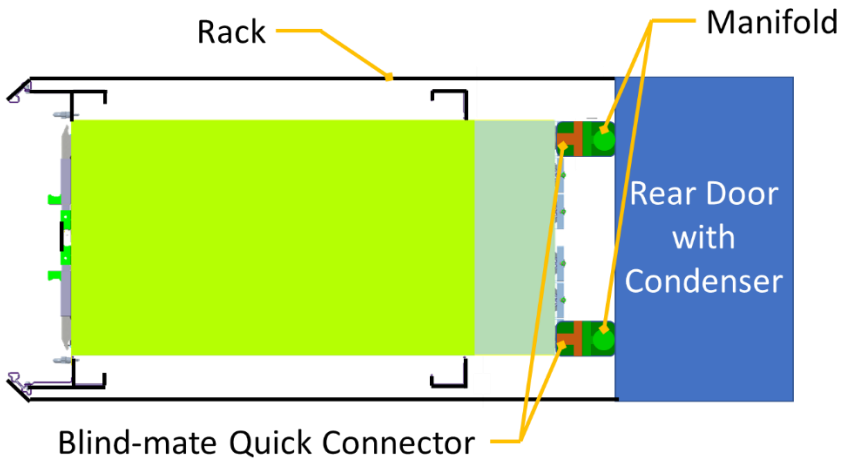


Figure 2 Blind-Mate Connection at the Rack Level

On the other hand, a manual-mate connector can support a higher range of liquid pressure within the cold plate array. The locations of the manifolds can also be adjusted.

Labor-Saving Mechanism

The system uses modularity and a commonly-used design, which allows easy interchangeability between air-cooling and liquid-cooling use cases. This design also takes into account the assembly tolerances and connector engagement.

Since the system requires high power supply and high-speed signal transmission, smaller (relatively) connectors within the scope of the design requirements are preferred to avoid affecting the air flow. There are many high density board-to-board connectors for signal connection and power delivery.

Due to the strong forces required when connecting / disconnecting board-to-board connectors, a mechanism to reduce such forces is required.

Wiwynn provides a patented labor-saving ejector to help such operation and maintenance for end users.

Before designing the force-saving ejector, we needed to analyze the insertion force and desertion force first. The following items should be considered within both force equilibrium and moment equilibrium calculations: board-to-board connector engagement force and the corresponding frictional force of the insertion and detachment operation. The spring force of blind-mate QC should also be considered if it is a blind mate liquid-cooling system. Since the system adopts manual-mate connection in liquid-cooling system, the QC spring force is not considered here.

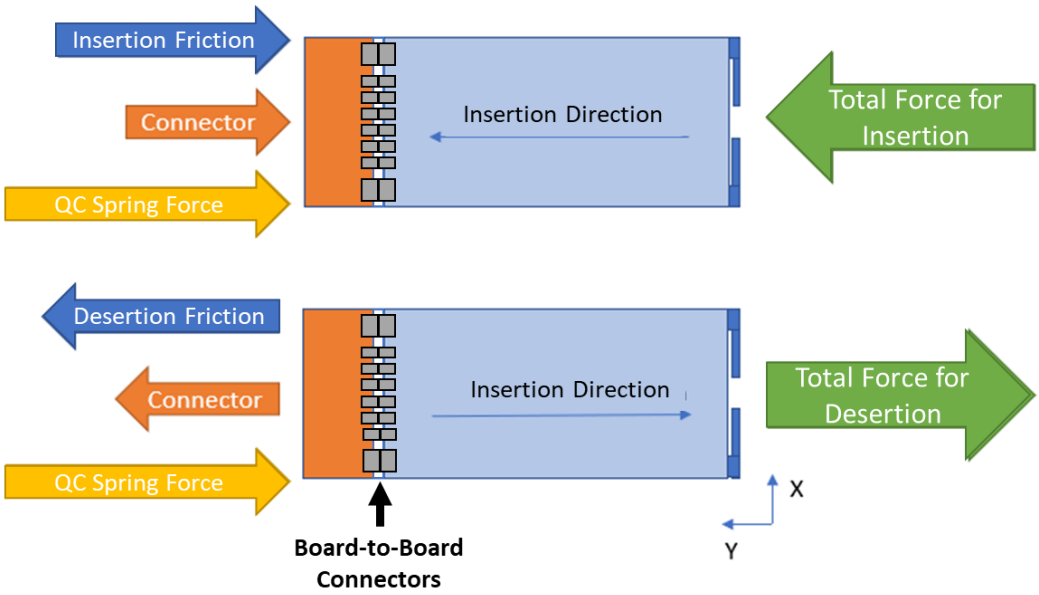
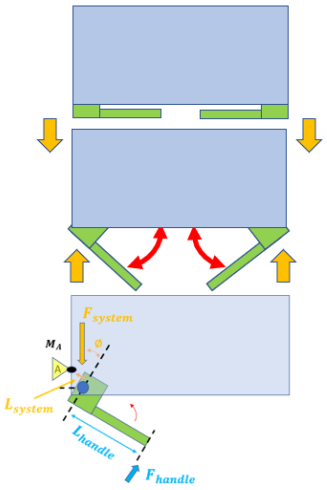


Figure 3 Operation Force Analysis

According to the calculation results in Fig. 4, the labor-saving ratio looks to be greater than 10 (The difference between pushing the system chassis directly by hand and pushing it with a labor-saving structure). This will significantly help ease the users when maintaining the system.



$$\sum M_A = 0$$

$$F_{system} = (F_{friction} + F_{QC} + F_{connector})/2, \text{per side}$$

$$F_{handle} \times L_{handle} - (F_{system} \times \cos\emptyset) \times L_{system} = 0$$

$$\rightarrow F_{handle} = \frac{F_{system} \times L_{system} \times \cos\emptyset}{L_{handle}}$$

- Handle force is related to the handle angle (∅) and system force at different stages.
- Operation Force Analysis, Labor-saving ratio ≒11.43

Figure 4 Analysis of Labor-Saving Mechanism

Moreover, this ejector takes into account the long-stroke requirements of the board-to-board connector engagement. The maximum supported engaging travel of this ejector design is 38.7 mm. For system designs requiring a longer connection and disconnection path between board-to-board connectors, this labor-saving ejector is still a suitable solution.

The labor-saving mechanism can be utilized and applied to air-cooling, manual-mate liquid cooling, and even blind-mate liquid cooling, which is almost the entirety of the cooling system design.

Wiwynn Integrated AI Training System - SV600G2

As a leading company in designing high performance, innovative and reliable servers, Wiwynn offers an AI training system, SV600G2, the best option for AI training.

As listed in Table 2, SV600G2 is a 6U server (447 mm (W) x 915 mm (D) x 264 mm (H)) compatible with EIA-19” rack. For accelerator, SV600G2 has OAI-UBB and Habana HL225 OAM. For CPU, SV600G2 leverages 3rd Gen Intel Xeon scalable 64-bit processor (codename: Ice Lake). It also supports 3+1 redundant PSU to ensure system reliability.

SV600G2 is ready for small-scale and large-scale computing. On-chip integration of 24 x 100 RDMA over Ethernet (RoCE) ports on the HL225H provides flexibility as a standalone training server or in scale-out clusters to meet customer system requirements.

The system is designed with the consideration of all-in-one integration, high power efficiency, and minimum system dimensions. OAM cooling is our most focused part of the whole system. Even though SV600G2 currently

contains up to 600W OAM with air cooling solution only, we offer a flexible, robust, yet interchangeable modular design between air-cooling and liquid-cooling systems in the event of an OAM upgrade.

System	6U AI Training System
System Dimensions	6U, EIA-compliant, 915 mm depth
Accelerator	8x Habana HL225H (OAM TDP is up to 600 W) OAM
CPU	2x Intel 3rd generation Xeon-SP (Ice lake), up to 270 W
Memory	32x DDR4 RDIMM
Storage	2x on-board PCIe x4 M.2 2280 or 22110
	4x SSD PCIe x4 U.2 (15/7 mm)
PCIe Card	Up to 2x PCIe FHHL slots with configuration: - 2x PCIe NIC x16
Cooling System	Air-cooling system: - 18x Hot-swappable 6056 Dual Rotor Fan
	Liquid-cooling system (reserved for future upgrade): - 1x 8-sets-in-series cold plate array - 16x Hot-swappable 6056 Dual Rotor Fan
Power Supply Unit	Redundant 3+1 4000 W CRPS
	Universal AC-in

Table 2 Server Feature Table

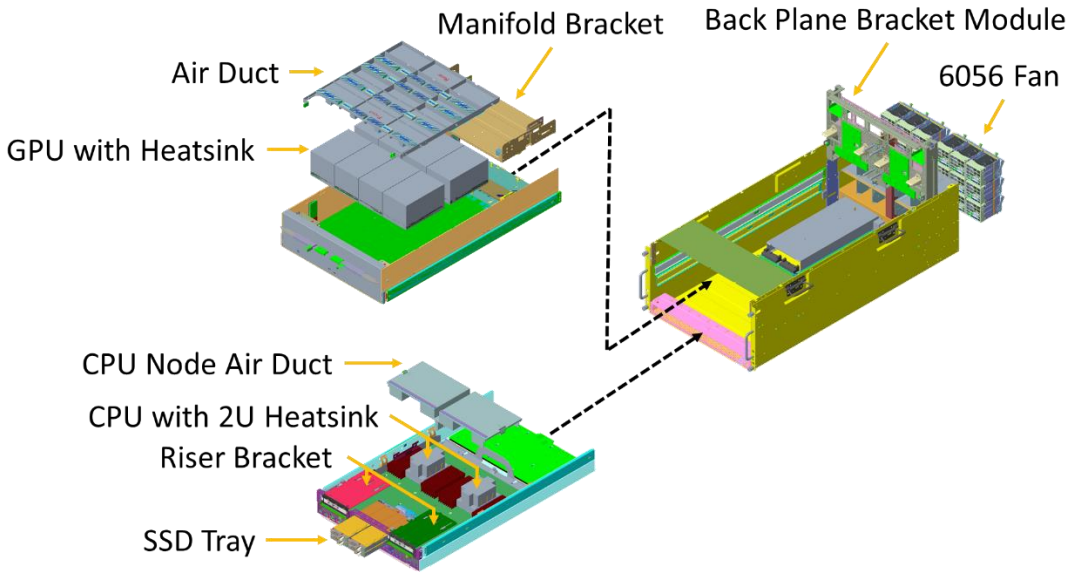


Figure 5 SV600G2 System Exploded Diagram (Air Cooling)

The mechanical design of SV600G2 not only fulfills thermal requirements, but also takes into account all aspects of the server system. To ensure the system output and board-to-board signal quality of both air-cooling and liquid-cooling systems, a simplified yet easy-to-operate robust mechanical design emerged to meet this complex architecture. Most components have toolless designs to reduce potential maintenance errors. All key commodities can be maintained on the front side to for easy and efficient operations.

L10 On-Rack Serviceability

Considering the heavy weight of the system, it is inconvenient to dismount the SV600G2 systems. To address this problem, SV600G2 is separated into 2 drawers with inner rails support. The main commodities with higher serviceability frequency such as OAM with air-cooling heatsink, CPU, NIC, DIMM and SSD can therefore be served directly on rack. Not only does this design decrease the risk of injury by avoiding the removal of the entire system from the rack, but also shortens service time. The less time used for servicing, the more benefits for operational servers.

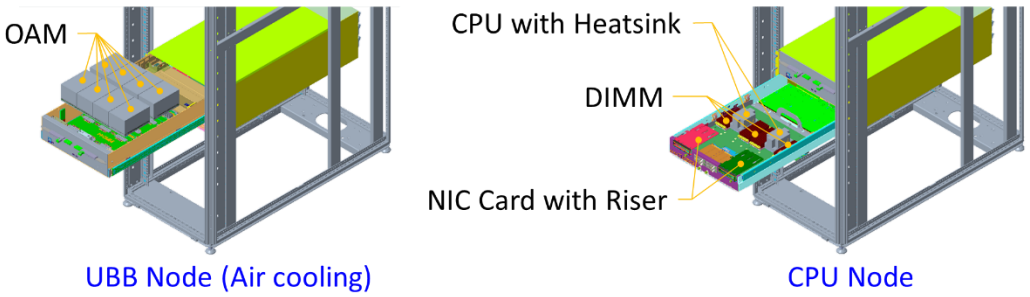


Figure 6 SV600G2 Key Commodity for Rack Service

Two-Phase Cooling Implementation

In SV600G2, Wiwynn adopts the manual-mate connection. The manifold is placed on two sides of the SV600G2 system for easy maintenance as shown in Fig. 7. In addition, this placement reduces the impedance of the airflow since the manifolds are not deployed directly behind the fan air outlet but, rather, on the sides of the L10 system.

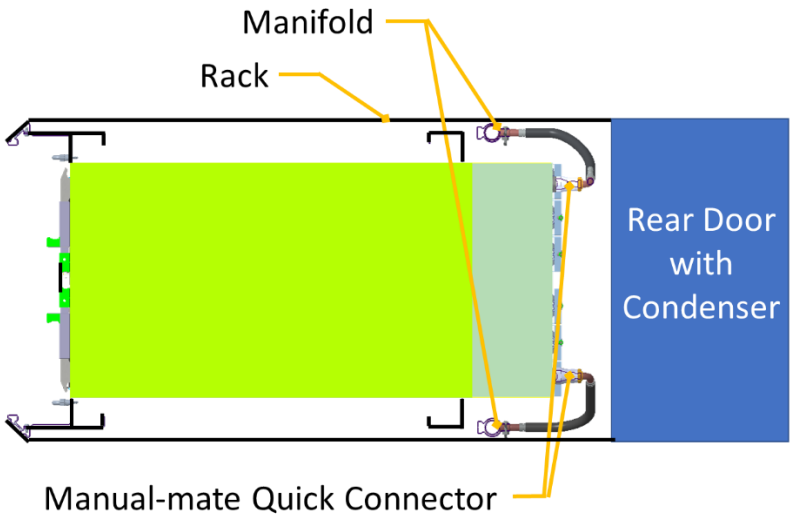


Figure 7 SV600G2 with Manual-Mate Connection at the Rack Level

SV600G2 has been designed with 2-phase liquid cooling alternative. This design utilizes the heat of vaporization to remove the heat that offers better thermal dissipation performance than 1-phase liquid cooling. Due to the uniform coolant temperature of liquid–vapor boundary state within the cold plate array, the coolant temperature at the final point of the accelerator series array (downstream) is the same as the starting point (upstream). This offers the unbiased cooling capacity for each accelerator. The dielectric fluid in SV600G2's cold plate, R1234YF, is an eco-friendly and readily available refrigerant. Comparing R1234YF with water, R1234YF is non-conductive and much safer when accidental leakage occurs. At 1 atm standard atmosphere, the boiling point temperature of -29℃ for R1234YF is in its vapor state. To maintain the cooling condition of the liquid-vapor boundary state, Cooling Distribution Unit (CDU) raises the pressure to 9.7 atm in the cooling circuit so the boiling point of R1234YF is at the designed working temperature of 48℃. To support this pressure, SV600G2 implements the manual-mate quick connection, where threading fastening is much reliable than the float-spring fastening used in blind-mate connection. Although the blind-mate design has easier installation / removal, the manual-mate connection can ensure that the cooling pipe is more securely fastened when servicing the cooling system under such high pressure.

Conclusion

According to the knowledge base of system architecture and consideration of all aspects of server design, Wiwynn provides an integrated AI training system benchmark and design practice. Modular interchangeability between air and liquid-cooling systems was feasible due to the carefully chosen cold plate terminal connector, allowing air cooling options for OAM power consumption equal or below 600 W, and liquid cooling options for over 600 W power consumption. The connector was minimized to the size of a small fan module but able to support a big pressure range. The ExaMAX high-speed signal connectors and XCEDE power connectors were considered when the labor-saving mechanism was designed. On-rack service of key commodities were also considered. Wiwynn cares about the maintenance time cost of all end customers.

To fully satisfy the cooling requirements to enter the high power OAM era, SV600G2 features the following key advantages:

1. Two thermal solutions for different OAM power consumptions – This thermal collaborative design can be applied for different OAM vendors and their corresponding UBB. The maximized modularized design shared between air and liquid cooling systems is more than ready to satisfy the power consumption requirements for present and future systems.
2. Reserved Manual-Mate with eco-friendly refrigerant – The use of eco-friendly and non-conductive refrigerant R1234YF in SV600G2 complies with the spirit of the ESG and is also safe in the long run. The 2-phase liquid cooling offers higher heat dissipation capability. Moreover, the manual-mate connection between cold plate terminals and manifolds can withstand a high-pressure range inside the cold plate.
3. Reliable mechanical design – The system passed strict shock and vibration tests both at the system and rack level.
4. Easy service - We take care of all aspects of system maintenance, including on-rack serviceability of key commodities and mechanical design that are easy to operate and labor-saving.

The holistic and complete solution can help users gain business and research advantages. Not only is SV600G2 a mature and future-ready AI training system, it is also a flexible system that introduces interchangeable cooling options for OAM.

About Wiwynn

Wiwynn® is an innovative cloud IT infrastructure provider of high quality computing and storage products, plus rack solutions for leading data centers. We aggressively invest in next generation technologies for workload optimization and best TCO (Total Cost of Ownership). As an OCP (Open Compute Project) solution provider and platinum member, Wiwynn actively participates in advanced computing and storage system designs while constantly implementing the benefits of OCP into traditional data centers.

For more information, please visit Wiwynn website or contact sales@wiwynn.com
Follow Wiwynn on Facebook and Linkedin for the latest news and market trends