

Open Systems for AI: Blueprint for Scalable Infrastructure

Version 1.0.0

Editors (s) :

Madhusudan Iyengar, Google

John Leung, Intel

Greg Sellman, AMD

John Stuewe, Dell

Contributors

Nishi Ahuja, Intel
Boon Chong Ang, Intel
Brian Aspnes, Intel
Allan Cantle, Nallasway/Untether Ai
Rob Coyle, OCP
Olivier Franza, Intel
Jessica Gullbrand, Intel
Dhamesh Jani, Meta
James Kelly, OCP
Rich Lappenbusch, Supermicro
Bijan Nowroozi, OCP
Rob Ober, NVIDIA
Amy Short, PhD, Denvr Dataworks
Berhanu Wondimu, Intel
Kasper Wszolek, Intel

Special thanks

OCP AI Steering Committee Members
Brian Aspnes, Intel

Version Table

Note, refer to the OCP Contribution Versions, Revisions and Errata best practices documentation. Versions and revisions are made to this document and logged here. Errata is a separate document such that the contribution specification document it refers to was not revised. Ex: Version 1, Errata E1

Date	Version #	Author	Description
28 JUN 2024	0.1.0 Initial Draft	Bijan Nowroozi	Starting point
30 JAN 2025	1.0.0 Release Candidate 1	Bijan Nowroozi et al	Content by contributors
01 FEB 2025	1.0.0 Release Candidate 2	Bijan Nowroozi et al	Edits/corrections
03 FEB 2025	1.0.0-RC 2 to SC	Bijan Nowroozi et al	Final edits
17 FEB 2025	1.0.0-RC 3	Bijan Nowroozi et al	Incorporate Review comments from Steering Committee Review
18 FEB 2025	1.0.0-RC4	Bijan Nowroozi	Additional feedback incorporation

Current Template Version:

Whitepaper/Document Submission Template v1.2.0
Effective January 2024

Executive Summary

Integrated Strategic Vision for Next-Generation AI Infrastructure and Economic Development

The world is entering a new era driven by artificial intelligence (AI), with computational demands doubling every six months. Meeting this challenge requires a fundamental transformation of data center infrastructure—one that embraces openness, collaboration, and innovation. The Open Compute Project (OCP) is leading this effort through its Open Systems for AI (OSAI) Strategic Initiative, bringing together hyperscalers, technology vendors, and engineers to develop scalable, high-performance AI infrastructure built on open standards and interoperability.

At the same time, AI presents an unparalleled economic opportunity. Silicon Valley has historically led the world in unveiling general-purpose technologies, fostering a global innovation network driven by public-private partnerships. To sustain and expand AI's potential, continued investment in technology R&D, workforce development, and global AI infrastructure deployment is essential. By aligning open designs with large-scale private investment, OCP's work can accelerate AI adoption, reduce costs, and strengthen open supply chains for trusted, scalable, and secure AI solutions.

OCP's proven success in open hardware and software standards—evident in networking, servers, and edge deployments—provides a strong foundation for addressing the technical and sustainability challenges of next-generation AI clusters. Through open interfaces and modular system architectures, the OSAI initiative fosters a diverse vendor ecosystem that enables high performance, interoperability, and cost-effectiveness at any scale. This work extends beyond servers to include data center power delivery, cooling, and management. Ultra-high-density AI clusters, consuming up to one megawatt per rack, require advanced liquid cooling, high-voltage DC power distribution, and automated orchestration frameworks to operate efficiently and sustainably.

Building a robust AI ecosystem will depend on the ability of industry and research collaborators to attract continued investment in technology R&D. This investment will drive the expansion of

an AI-ready workforce, ensuring trustworthy, sustainable, and high-performance AI infrastructure is available to users worldwide. The long-standing synergy between public funding for basic research and private investment in product development has been a key driver of innovation, bringing ideas from universities into the marketplace. This model will be critical in the coming years, as AI fuels job creation across industries, including software, data science, construction, manufacturing, and energy. OCP's open designs, paired with large-scale private investment in AI-driven data centers, can form the backbone of continued leadership—especially when supported by effective public policies that expand STEM programs, apprenticeships, and community-college partnerships to ensure AI literacy across all demographics.

Equally important is the international context. As global market competition intensifies, success in AI deployment may not be determined by the first mover or the highest-performing system, but by the provider of the most trusted, scalable, efficient, and secure solutions. End users will ultimately shape both market demand and technical strategies, favoring open, adaptable ecosystems. The collaborative nature of the OSAI initiative and the open resources available to the OCP Community provide a foundation for market growth, positioning adopters as preferred partners for organizations seeking reliable and cutting-edge AI infrastructure.

Meeting AI's infrastructure demands requires collaboration. OCP's community-driven model unites data center operators, chipmakers, developers, manufacturers, service providers, academics, and standards organizations around common development practices, best practices, and open specifications. These shared standards fuel interoperability, sustainability, and efficiency, reducing both environmental impact and total cost of ownership. Meanwhile, universities, technical institutes, and training programs can cultivate the skilled workforce needed to sustain innovation, foster entrepreneurship, and spread AI capabilities across industries. Together, these efforts will help local economies—from rural regions to industrial hubs—transition toward high-value jobs and long-term growth.

The convergence of OCP initiatives, public policy, and private innovation presents a unique opportunity to shape the future of AI infrastructure. By applying OCP's modular, collaborative approach and leveraging the collective expertise of the OCP Community, we can drive economic growth, create new careers, and strengthen international partnerships. AI is the electricity of our age, and with strategic investments in infrastructure, skills, and global AI deployment, its transformative benefits will be realized worldwide.

The following sections outline a technical, non-prescriptive vision for OCP Open Systems for AI, providing a blueprint for scalable, efficient, and sustainable AI infrastructure.

Mission Statement

Our mission is to transform AI through open, interoperable systems that democratize access, foster innovation, and accelerate positive societal impact. Drawing inspiration from cloud computing and white-box server architectures, we eliminate proprietary barriers to create an ecosystem where collaboration drives progress.

We pursue this mission through several interconnected approaches. At our foundation lies a commitment to open-source principles and collaborative development, which accelerates AI progress by enabling knowledge sharing and modification across the community. We strictly adhere to OCP's core tenets of efficiency, impact, openness, scalability, and sustainability, ensuring all developments align with the community's established requirements.

Our work emphasizes creating robust interoperability standards that promote modularity and seamless integration among AI systems, leading to a more vibrant and innovative ecosystem. We actively cultivate participation from individuals and organizations across all sectors, bringing together diverse expertise and perspectives to enrich the AI landscape. By making open AI system specifications, tools, and resources widely discoverable and available, we remove barriers to entry and enable broader participation in AI development.

The systems we create are designed with scalability and adaptability at their core. Through careful attention to composability and modularity principles, we ensure our solutions can evolve alongside changing requirements and technological advancements, while efficiently scaling to meet increasing demands and complexity.

Table of Contents

Contributors **2**
Special thanks **2**
Version Table **3**
1. Introduction **11**
The AI Infrastructure Challenge 11
OCP's Vision and Strategic Initiative: Open Systems for AI 12
A Community-Driven Approach 14
Next-generation Challenges for Open Systems for AI 15
2. Compliance with Open Compute Project Tenets **16**
3. System View **17**
3.1. Introduction 17
3.2. History 17
3.3. Physical System Design 20
3.4. Summary 34
4. Datacenter Physical Infrastructure **35**
4.1. AI/ML landscape and trends 35
4.2. Work Items Considered 37
4.3. Current Work and Contributions 38
4.3.1. Sample pathfinding technology and solutions 38
4.4. Identified Gaps, Emerging R&D Areas, OCP Opportunities for Leadership & Impact 47
5. Data Center IT Infrastructure **52**
5.1. Components of "Data Center IT Infrastructure" 52
5.2. Work Items Considered 60
5.3. Identified Gaps 61
6. System Management **64**
6.1. Overview 64
6.2. Relationship to IT and Physical Infrastructure 64
6.3. Manageability Interface and Standards 65
6.4. OCP and Redfish Profiles 65
6.5. Work Items 66
6.6. Identified Gaps 68
6.7. Scoping System Management 69
6.8. Manageability domains 71
Conclusion **72**
OCP Glossary **74**
References **96**
License **99**

Creative Commons 99
About Open Compute Foundation **100**

1. Introduction

In the rapidly evolving world of artificial intelligence (AI), the availability of scalable, robust, and efficient computational infrastructure has reached an unprecedented level of focus to the markets. As AI technologies continue to expand and permeate every sector, the imperative for a standardized, open approach to AI systems development resulting in a competitive and resilient supply chain is becoming increasingly clear. The Open Compute Project (OCP), with its dynamic and innovative community, is at the forefront of this movement, aiming to lead the charge in building open, sustainable, and cost-effective AI systems designed to serve a broad spectrum of users and applications. This white paper aims to describe the vision of how to begin this process and includes some discussion of key tactical (in-progress) development efforts.

The AI Infrastructure Challenge

Current AI computational requirements are straining the limits of traditional IT infrastructure and data center designs. Today's approach to building infrastructure for AI is often fragmented, leading to higher costs, deployment delays, and inefficiencies throughout the ecosystem. As AI workloads grow in complexity and variety, the need for a cohesive, community-driven approach to AI infrastructure has never been more critical.

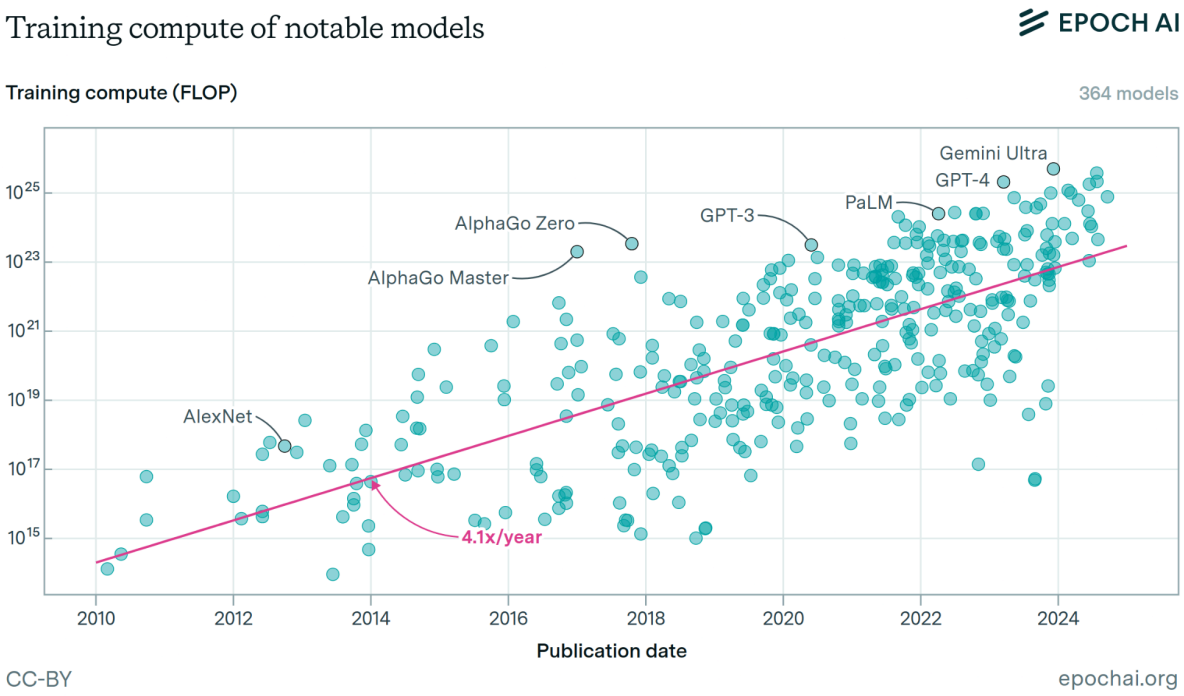


Figure 1: Training Compute of Notable Models Over Time

The chart, Figure 1 above, highlights a consistent **annual 4X increase in compute requirements** for training Generative AI Models. To keep pace with this rapid growth, the Open Systems for AI initiative must align and accelerate all OCP projects, ensuring contributions not only meet but exceed this trajectory. This is essential for maintaining the relevance of Open Systems for AI within the broader community and ecosystem.

OCP's Vision and Strategic Initiative: Open Systems for AI

Within the OCP mission focus on data centers, AI is today's most prominent data center use case driving innovations, followed by HPC and the emerging edge. Moreover, AI clusters are pushing technology innovation to new heights when implemented at ever larger scale by hyperscalers—foundational community constituents of the OCP. The OCP is, thus, well aligned with current AI trends. Its Community brings together hyperscale data center operators, vendors, solution providers, and colocation providers as OCP members, offering sustainable at-scale computational infrastructure and best practices to deliver AI for hyperscalers who also lead the way for enterprises.

AI and HPC workloads are also distributed and diverse, much more so than other virtualized and cloud-native applications. In the pursuit of greater performance, infrastructure specialization for some workload categories is burgeoning, but risks fracturing the supply chain into silos, raising costs and reducing efficiencies. This is exacerbated by the accelerated drive for innovation to meet new requirements in the hardware ecosystem and rush for all players in the market to provide solutions.

To prevent these downsides, the OCP facilitates open-sourcing hardware specifications and standardizing building blocks that are forged for the most demanding hyperscale solutions. The OCP's trusted IP model allows multiple vendors to safely collaborate and specialize their products with greater interoperability and composability which benefits those vendors as well as the customers, moving the entire market forward. Spanning the OCP Projects, the OCP "Open Systems for AI" Strategic Initiative brings together the work of the Community to meet AI's scale and workload diversity.

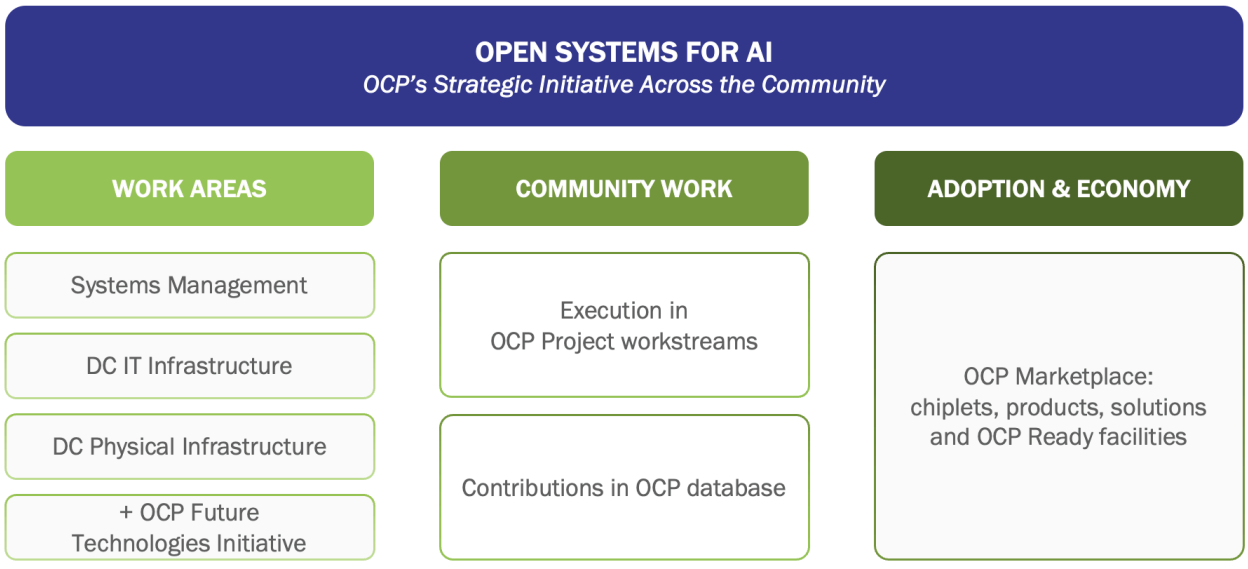


Figure 2: Open Systems for AI Collaboration Areas

A Community-Driven Approach

OCP's greatest strength is its community-driven model. By uniting leaders, innovators, and experts from across the technology spectrum, OCP is tackling the multi-dimensional design challenges of AI infrastructure. From advanced cooling technologies and high-density power solutions to cutting-edge networking protocols and hardware management, OCP is addressing the full stack of AI infrastructure requirements. Inside the OCP Open Systems for AI (OSAI) Strategic Initiative, acting to collect efforts across OCP Projects, Workstreams and the Future Technologies Initiative too, we aim to bring together holistic full stack and subsystem cross-stack collaboration necessary to solve the challenges introduced by AI clusters and the demands they place on data center facilities. For example the new requirements on power, don't only change things in the Rack and Power Project, but also impact larger facility Cooling Environments and finer aspects in the IT hardware down to the server and even chip power design. Many such areas of innovation require co-design and cross-stack collaboration that the Open Systems for AI effort aims to bring together.

As we embark on this journey, we invite you to join us in shaping the future of Open Systems for AI. This whitepaper will explore the key challenges in AI systems and present a roadmap for creating open, efficient, and sustainable AI systems that will power the next wave of data center innovations at the heart of the OCP's mission.

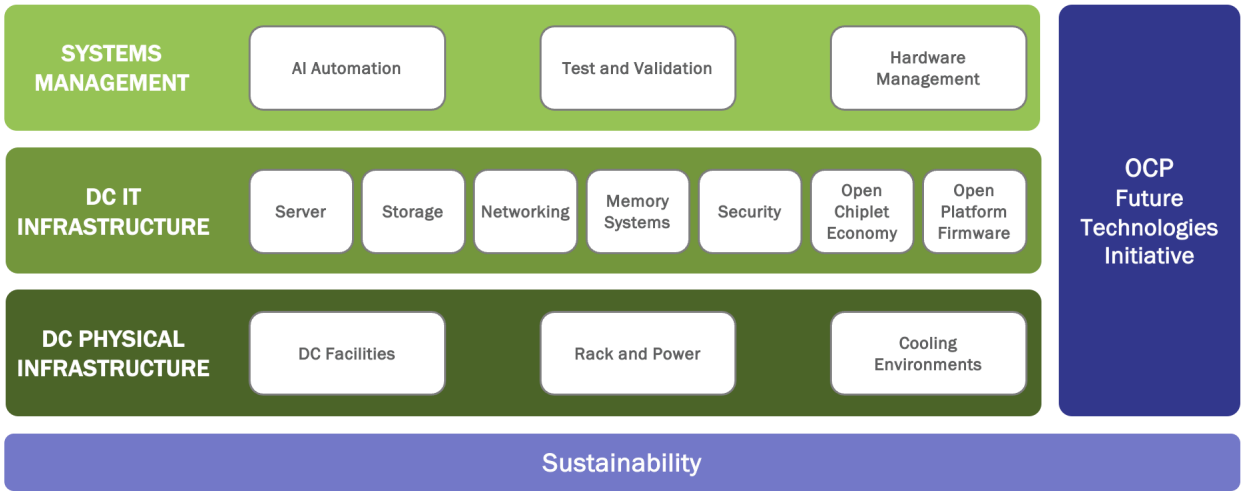


Figure 3: *Predominant work areas of the Open Systems for AI Strategic Initiative*

Next-generation Challenges for Open Systems for AI

As AI and machine learning (ML) workloads continue to grow, the demand for high-power, high-density compute racks is becoming critical for data centers. To go beyond today's specifications, new technologies must be explored and adopted across multiple domains: cooling, power delivery, energy efficiency, interconnectivity, rack design, system management, and data center facilities.

As we prepare for next-generation high-power racks—scaling from 250kW to 500kW and eventually 1MW per rack—transformational advancements in cooling, power delivery, energy efficiency, and interconnectivity will be essential. Cutting-edge solutions such as advanced micro liquid loops, nanoparticle fluidics, phase-change cooling (up to cryogenics), high-voltage power delivery and storage, photonics-based compute and communications in systems, and evolved such as superconducting materials will need to be further developed and integrated to overcome further thermal and power challenges.

Sustainability will be central, integrating heat recovery and renewable energy to minimize environmental impact. As these systems become critical infrastructure, robust security measures will be imperative to ensure resilience and reliability.

Facilities designed to accommodate ultra-high-density racks will require a significant overhaul compared to traditional data centers. Advanced systems for power, cooling, physical layout, and management will be necessary to meet the extreme demands of AI-driven infrastructure. New frontiers such as heat reuse, energy storage, carbon labeling and management, smart grid integration, circularity, automation and more emerge as the next sets of challenges for the OCP Community to solve.

Integration of these efforts and technologies with an OCP Community designed System Architecture, is necessary to facilitate this evolution of OCP Open Systems for AI.

2. Compliance with Open Compute Project Tenets

Because this paper doesn't look at any one contribution or single technology area, but rather overviews them, aligning with the nature of the Open Systems for AI Strategic Initiative—an umbrella over projects—the implicit assumption is that OCP Community development **will continue to reflect the OCP tenets**. Though it can be said that it's fairly obvious today that AI clusters are pushing the boundaries of the scale, sustainability, efficiency and impact aspects of our tenets, and the Strategic Initiative name itself emphasizes the importance of openness as described in the introduction of this paper.

3. System View

3.1. Introduction

This section provides a high-level overview of how we arrived at our current state in AI/HPC infrastructure and suggests possible directions for future system-level AI/HPC implementations. It is not intended as a definitive roadmap, but rather as a starting point for ongoing discussion within the OCP Community.

3.2. History

By early 2021, Data Center/Cloud/Enterprise markets predominantly used single-CPU or multi-CPU systems with accelerators, showing minimal standardization across vendors. Just before Generative AI's emergence, OCP Community introduced GPU-accelerated AI/HPC systems using Open Rack v3 (12kW per rack) and the newly introduced Open Accelerator Infrastructure (OAI), featuring Universal Base Boards (UBB) and Open Accelerator Modules (OAI). This successful launch resulted in multiple vendors consuming these specifications and producing these designs and NVIDIA launching their HGX chassis and contributing this specification to OCP in 2022. These architectures prioritized interconnected accelerator density—specifically 8 accelerators per 4RU.

Over the course of 2022 and 2023, the OCP Server Project contributed the Data Center Modular Hardware System (DC-MHS) set of specifications to the OCP community. DC-MHS introduced a modular platform architecture capable of supporting traditional 19" rack form factors as well as emerging 21" Open Rack standards. This new framework opened the door to more flexible scaling strategies.

This is not an exhaustive list of OCP server architectures that supported AI in that timeframe, but a subset based on leading market trends for high density AI architectures that are germane to a discussion of developing OCP Open Systems for AI.

In May 2023, NVIDIA introduced its MGX platform, which, while developed independently, aligned with the vision of the MHS platform. In late 2024, NVIDIA announced the NVL72 extension of the DGX/HGX platform, supporting up to 120 kW per rack. Subsequently, at the 2024 OCP Global Summit, NVIDIA contributed the MGX platform to the OCP Community, including an MGX-based NVL72 system. This announcement (see Figure 4 below) introduced an evolved density of accelerated Host Processor Modules (HPM) for a total of 72 interconnected accelerator modules per rack.

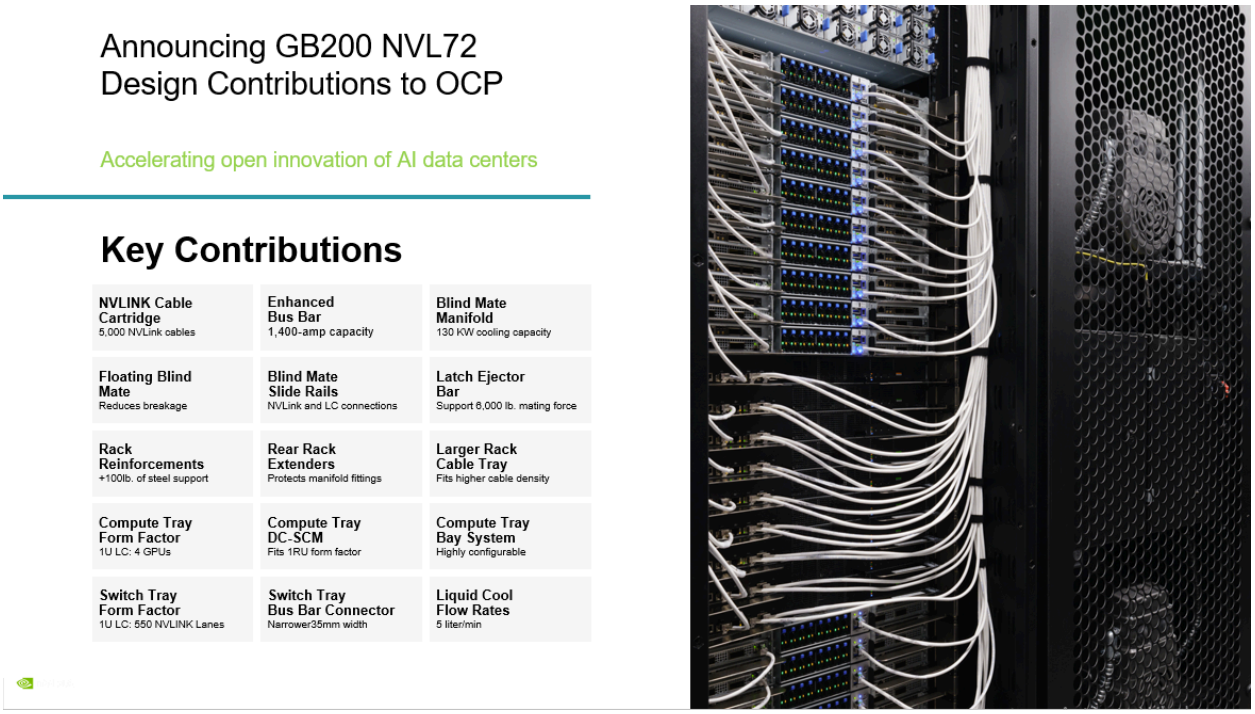


Figure 4: Ian Buck, NVIDIA, Keynote Announcement of GB200 NVL72 Contribution from 2024 OCP Global Summit

This contribution, illustrated in Figure 4, includes the physical rack, chassis, and integrated cooling and power infrastructure supporting NVIDIA's NVL72 architecture for Hopper and Blackwell accelerators. By contributing this platform, NVIDIA enables suppliers to engage in open hardware through OCP Open Contribution Database, enabling those with the ability to build from design specifications, the capability to build and deploy the same MGX-based AI-optimized systems.

Similarly, at the 2024 OCP Global Summit, Meta announced their upcoming contribution of the Catalina rack- an ORv3 based rack supporting up to 140kW including newly contributed Meta Wedge fabric switches for the NVL72 architecture.



Figure 5: Slide from Omar Baldando, Meta, Catalina Announcement, Keynote Presentation from 2024 OCP Global Summit

The key observations from the contributions are:

- Shifting from the often used cloud strategy of massive arrays of resources, similar to microservices, now multiple forms of parallelism can be combined and utilized to scale cluster level processing systems for AI workloads.
- The first order limitations of this scaling strategy are thermal, power, and interconnection related. The result is specific design decisions such as cold plate/liquid cooling, high power racks, and increasingly dense interconnection.
- The power of commonality. If a family of systems share key component specifications, the supply and logistics chains are more likely to keep up with demand due to reduced SKUs and economies of scale.
- Beyond performance only, economic optimization is necessary.

It's also worth noting there was high market interest at 2024 OCP Global Summit as most vendors on the show floor and contributors presenting on related topics discussed and/or displayed NVL-72 MGX systems.

We will begin to contemplate the potential architecture of an optimized platform leveraging the previous and new contributions as a starting point.

3.3. Physical System Design

Conceptually, an AI cluster comes together like the diagram below.

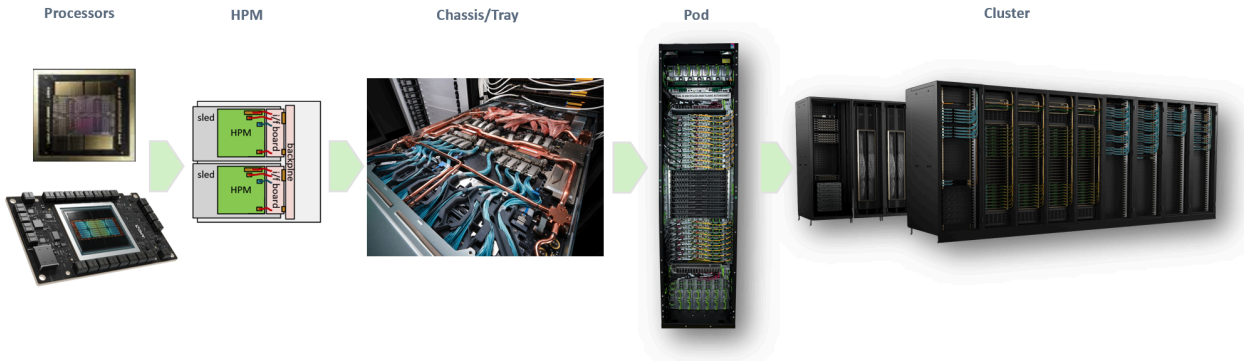


Figure 6: Conceptual AI Architecture from Component to Cluster

The need for synergy and participation from the entire ecosystem

AI/HPC systems require close collaboration among component suppliers, system integrators, data center operators, and software developers. Such synergy ensures that each layer—processing, chassis, networking, cooling, and management—aligns with open and modular design principles. What follows are key items for consideration and collaboration.

Component - Processing

For large-scale AI training and inference, an optimal platform must accommodate a wide array of processors and accelerators (CPUs, GPUs, DPUs, custom ASICs, etc.). Today's AI training accelerators typically integrate high core count processing along with high-bandwidth memory (HBM) to support growing inference and training algorithms. Many of these components are chip-on-substrate or 'chiplets'. Over the past year, the industry has seen a surge of

announcements for both inference and training accelerators. Given the rapid pace of innovation, a modular design philosophy either at the board level or even with the OAI/UBB approach offers the greatest flexibility to integrate current and future accelerator types without extensive redesign.

Chassis - Tray

A chassis (or tray) is the physical enclosure that houses key computing components—such as processors, storage drives, and memory—consolidating them into a compute, storage, or memory node. These enclosures are generally designed to fit one of two primary rack types: the EIA-310-based 19-inch (482.6 mm) rack or the OCP Open Rack v3 21-inch (539.4 mm) rack.

In practice, 19-inch EIA racks are commonly used in enterprise environments, while 21-inch OCP Open Rack systems dominate in hyperscale deployments. During early open AI system development, no single rack format universally suits all needs. Nonetheless, the additional width afforded by a 21-inch Open Rack can facilitate greater density and potentially simplify thermal management and networking.

Server - Compute Node

For the purposes of an OCP Open Systems for AI, it is worth examining the recent evolution of architecture for AI acceleration. One of the predominant focus areas is density scaling and this is being achieved through utilizing parallelism.

OCP OAI and NVIDIA's HGX were platforms built specifically for accelerator density. Figures 7 and 8 illustrate the density of the OAI/UBB/OAM architecture. The OAI system consists of multiple parts including Open Accelerator Module [OAM] and Universal Base Board [UBB]) and features a 4U chassis (or tray) supporting up to 8 accelerators, designed to support 3200W.



Figure 7: Closeup of OAI OAM/UBB System with heatsinks removed for clarity[1]

As you compare figures 7, 8, and Figure 9, it's clear to see the resemblance of OCP Open Accelerator Initiative [OAI] systems and HGX systems. They are highly similar but not the same.



Figure 8: Aivres OAI OAM/UBB System 4OU[2]

[1] https://www.anandtech.com/show/17054/amd-announces-instinct-mi200-accelerator-family-cdna2-exacale-servers

[2] https://aivres.com/product/oam/



Figure 9: Image of an NVIDIA HGX board from The Register.[3]

NVIDIA MGX and Meta's upcoming Catalina system for NVL-72 rely on a compute tray density similar to OCP's MHS. The MGX and MHS platforms both are popular in the market, and support multiple 1U, 2U, and 4U form factors with high composability for a wide variety of configurations. NVL-72 systems support up to 2 CPUs and 4 GPUs per compute node in 1U with a budget of approximately 6325W. It's worth noting that both of these systems fit EIA 19 inch and OCP Open Rack 21 inch racks.

Pod Density- Compute Nodes

Parallelism has emerged as a key means to accelerate AI/HPC computation. Types of parallelism employed have evolved in response to workload demands, resulting in system platform architecture that has reflected these trends. A pod is a group (domain) of resources such as compute nodes that in a parallelism context are scaled up to act as a single logical unit.

[3] https://www.theregister.com/2024/03/21/nvidia_dgx_gb200_nvk72/

Thus a workload may be distributed on multiple physical resources but logically to the application layer appears to be processed on a single node. A typical consideration is the shortest communication path maintaining a low latency profile for maximum memory coherence. Scaling up resources is to increase resources within a single pod and this is done with high performance networking fabrics. OAI/HGX accelerators in 3OU chassis (or trays) with 8 accelerators may be scaled up to a pod of 4 chassis per rack for a total of 32 accelerators per rack.

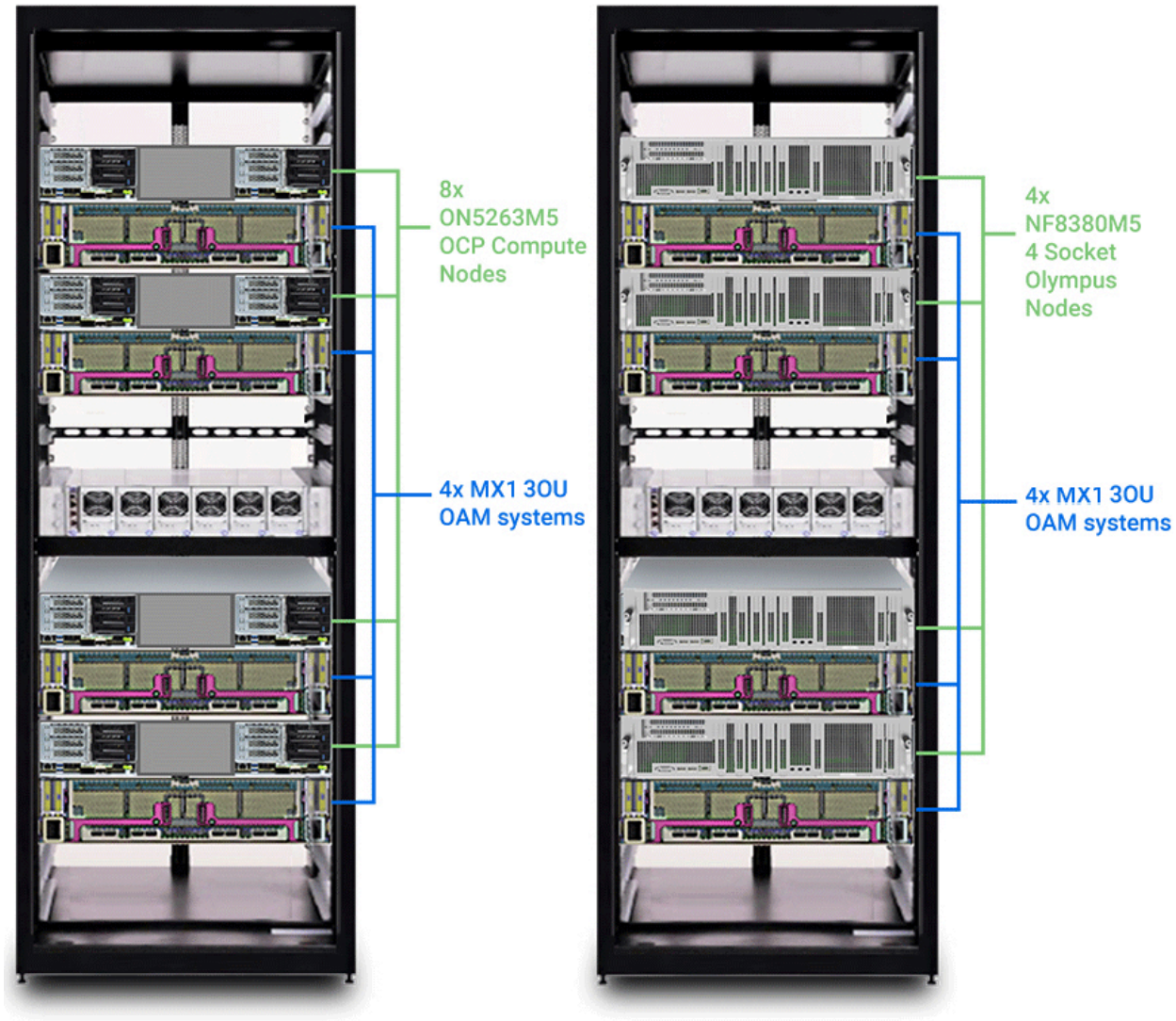


Figure 10: Image of a typical OAI UBB/OAM Accelerator Pod (in Blue)[4]

[4] https://aivres.com/product/ocp-rack-ai/

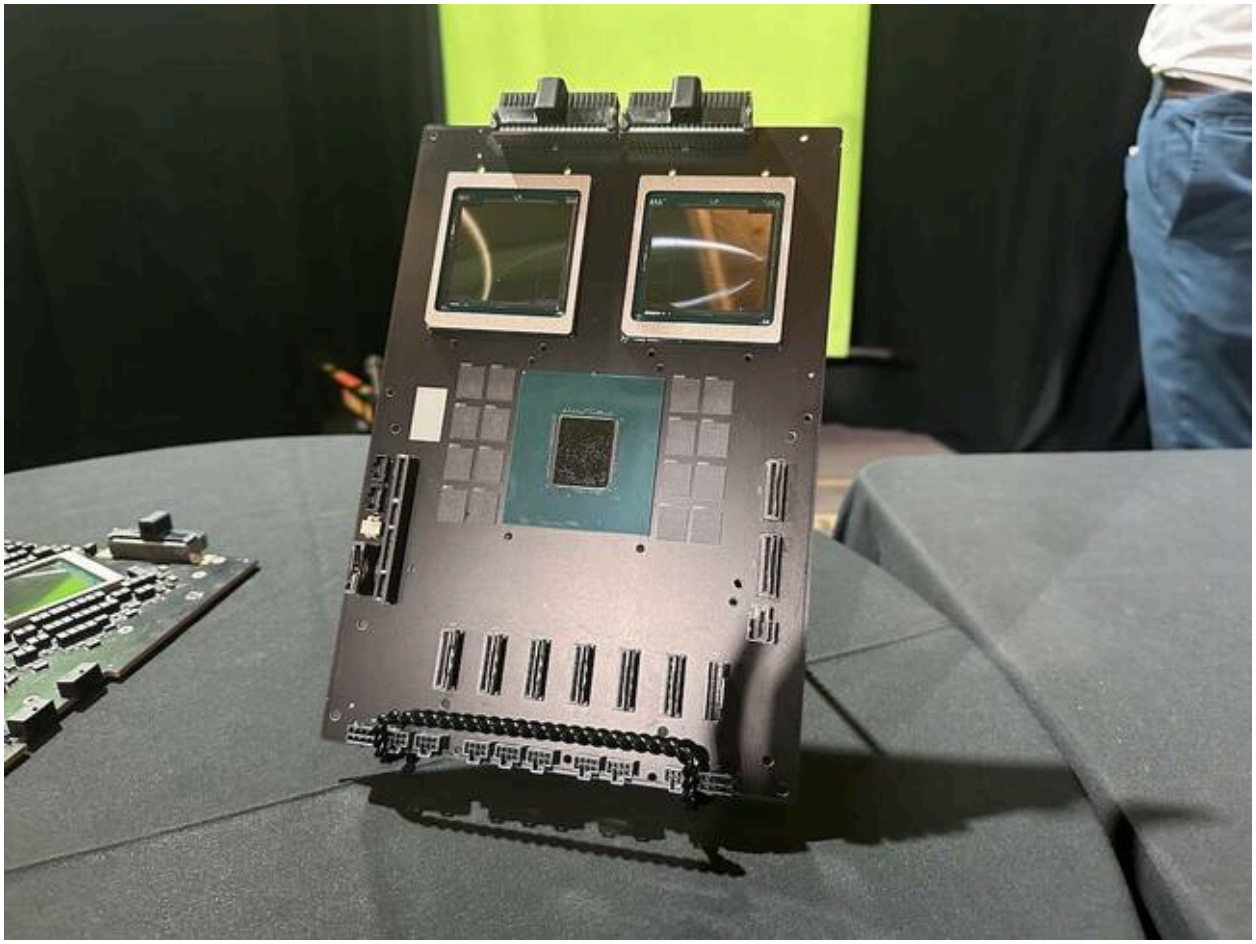


Figure 11: Image of NVIDIA's GB200 Host Processor Module *(there are 2 per chassis)* from The Register.[5]

Considering Figure 11 above, a MHS or MGX tray supports 2 of these types of boards or Host Processor Modules (HPM). The extensibility of the platform provides useful modularity to accommodate multiple accelerator types. For more details see the MHS specification for M-DNO which clearly defines these form factors.

As of the publication of this white paper, NVIDIA has been the first to contribute to the OCP Community, a 3rd generation density accelerator platform. Meta's Catalina NVL-72 based contribution is expected 1H 2025.

[5] https://www.theregister.com/2024/03/21/nvidia_dgx_gb200_nvk72/

From a density standpoint, the existing NVL-72 based MGX contribution is designed around a pod of 18 AI compute node servers supported by 4 accelerators per 1U tray, connected via various fabrics through 9 network switch trays. It is not difficult to imagine that subsequent iterations would scale density moving beyond 72 accelerators per pod.

It's easy to imagine an OSAI architectural goal of a roadmap of increasing density with HPM and fabrics that can be configured to physically scale. For instance, starting with an HPM design supporting more than two accelerators per HPM or 4 per S-DNO for example, geometric optimization and reconfiguration could achieve an 8X density increase, reaching 576 accelerators per rack[6]—a figure that could expand further with advancements in fabric technology.

Kicking off this OCP OSAI effort, starting with existing contributed platforms naturally allows for a shorter development cycle. The community can rapidly begin evolving these concepts, utilizing OCP's Project expertise and existing OCP Contributions such as MHS, OAI, Grand Teton and many others along the way to reduce total development effort (versus a green field approach) and to maximize commonality.

Other Platform Considerations

There are multiple networking aspects to facilitate the necessary parallelism (for acceleration) for AI workloads. Conceptually these carry traffic between components like CPU to GPU, or to memory, to data center management systems, or between clusters and each different performance capabilities. This has resulted in network fabrics for:

- Frontend Networking (Normal Ethernet)
- Backend (Scale out) Networking (InfiniBand/RoCE Ethernet or other high performance fabrics)
- Scale Up Accelerator Interconnect (NVLink/ UALink / Ethernet)
- Out of Band Networking for Manageability etc…

[6] Moving beyond 144 requires increasingly complex reconfigurations of the layouts and is only conceptual to show potential evolutionary paths

At the 2024 OCP Global Summit, the Ultra Ethernet Consortium (UEC) presented a view of a layered networking model (see Figure 9: “Ultra Ethernet Introduction”) and their vision of utilizing ethernet for scaling out.

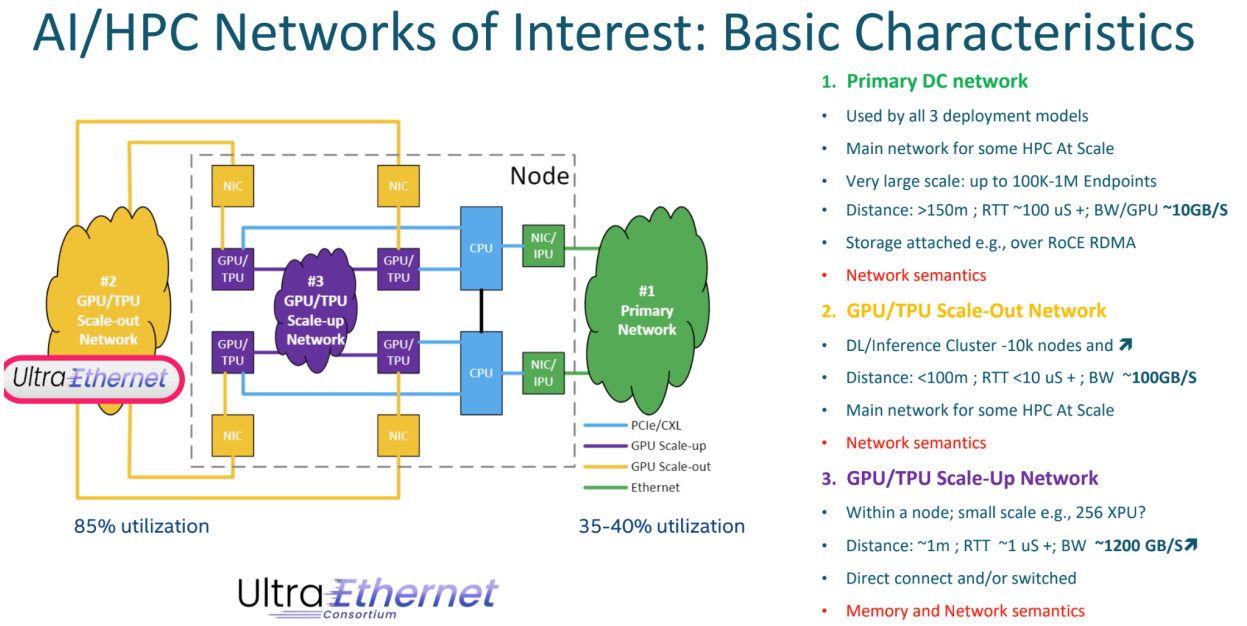


Figure 12: UEC presentation [“Ultra Ethernet Introduction”] at 2024 OCP Global Summit

Scale-Up Networking

A key to acceleration is the ability to scale up the capacity of a single server to a pod via parallelization. The scale-up interfaces themselves are closely coupled with accelerator silicon interface support. In layman’s terms, this type of parallelization effectively virtualizes multiple server trays to act as a single server to the AI workload.

Firstly examining the contributed NVL72 platform, it implements multiple fabrics and networks as described above. The scale-up fabric uses NVLink (roughly comparable to PCIe). NVLink is a direct GPU-to-GPU interconnect that scales (up) multi-GPU input and output (IO) within a server. There are fabric trays in the current pod configuration that interconnects these links. These trays are developed around the NVIDIA NVLink Switch chips that in this generation connect 72 NVLinks to provide all-to-all GPU communication fabric at full NVLink speed (~1800Gbps) within a pod. This is the current market leading architecture for AI clusters and the

OCP Community use this as a benchmark and focus on bringing highly performant open networking and platforms to scale up fabrics.

There was a presentation at 2024 OCP Global Summit by Ultra Accelerator Link Consortium, whose goal is to develop an open specification for a performant GPU-GPU type fabric for scale up networking. The presentation, (refer to the overview slide shown in Figure 10 below), described their vision of scale up for a pod and later in the presentation gave a high level timeline for upcoming specifications. Publicly stated technical points of interest are the initial target is support for 1024 accelerator scale up leveraging packet based 802.3dJ (200Gbps) Ethernet PHY. Once these specifications are made publicly available, the OCP OSAI community should plan to include this type of scale up fabric as a design requirement. Aspirationally, The OCP Community can subsequently begin developing related fabric components for scale up in AI Clusters interchangeably with existing scale up network fabrics found in OCP Contributions and OCP Marketplace products today. This may be the longest lead work item for the OCP OSAI Community development.

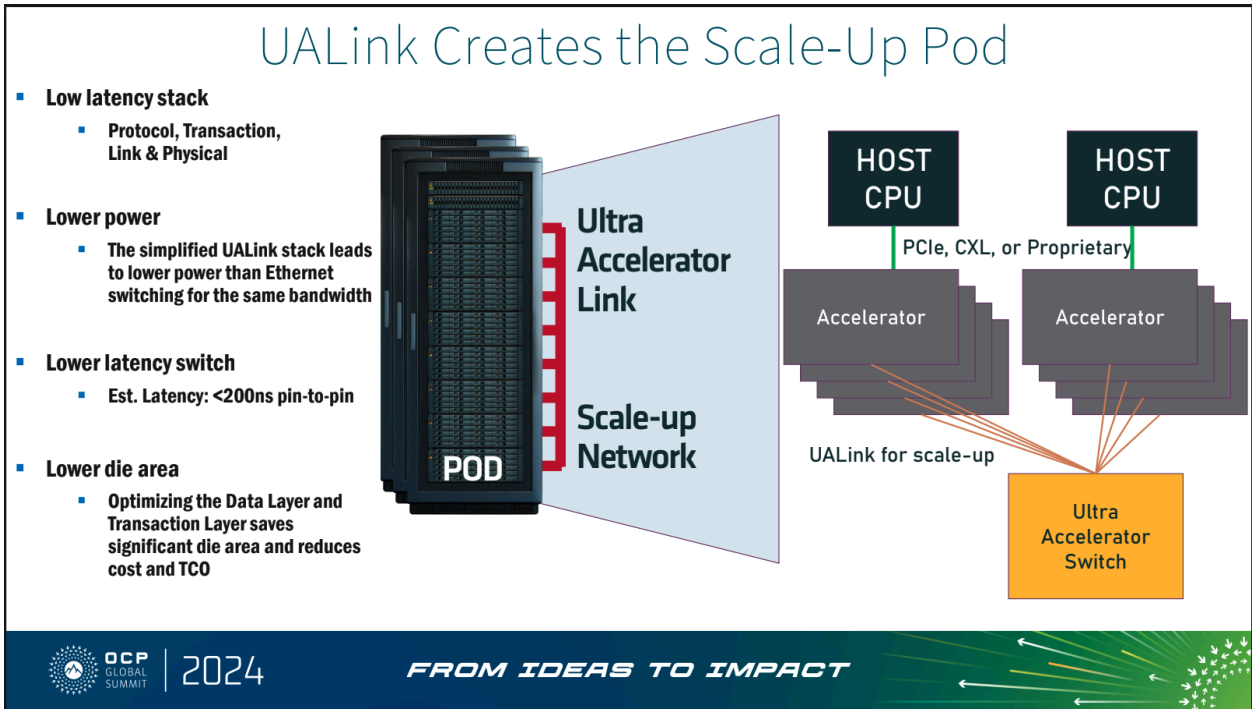


Figure 13: Image from "UALink: Pioneering the AI Accelerator Revolution" presented at 2024 OCP Global Summit.

Scale-Out Networking

In broad terms of an emerging requirement, OCP Open Systems for AI (OSAI) scale-out is an essential element of acceleration and presently is not coupled with support in the accelerator silicon, only the networking and software stacks. If you look at Figure 11, there is an overview of the relationship of pods scaling via parallelization techniques with some suggestions on scaling rules.

The NVL72 contribution is defined in terms of supporting NVIDIA Infiniband. The current state of the art is Infiniband NDR with a per lane PHY rate of 106.25Gbps. The next step is XDR which is defined at 200 Gbps per lane PHY rate.

By default, currently non optimized IEEE 802.3ba Ethernet is available at a PHY Rate of 100 Gbps and there is an upcoming release the merchant silicon providers will support, IEEE 802.3dj with a per lane PHY rate of 200Gbps.

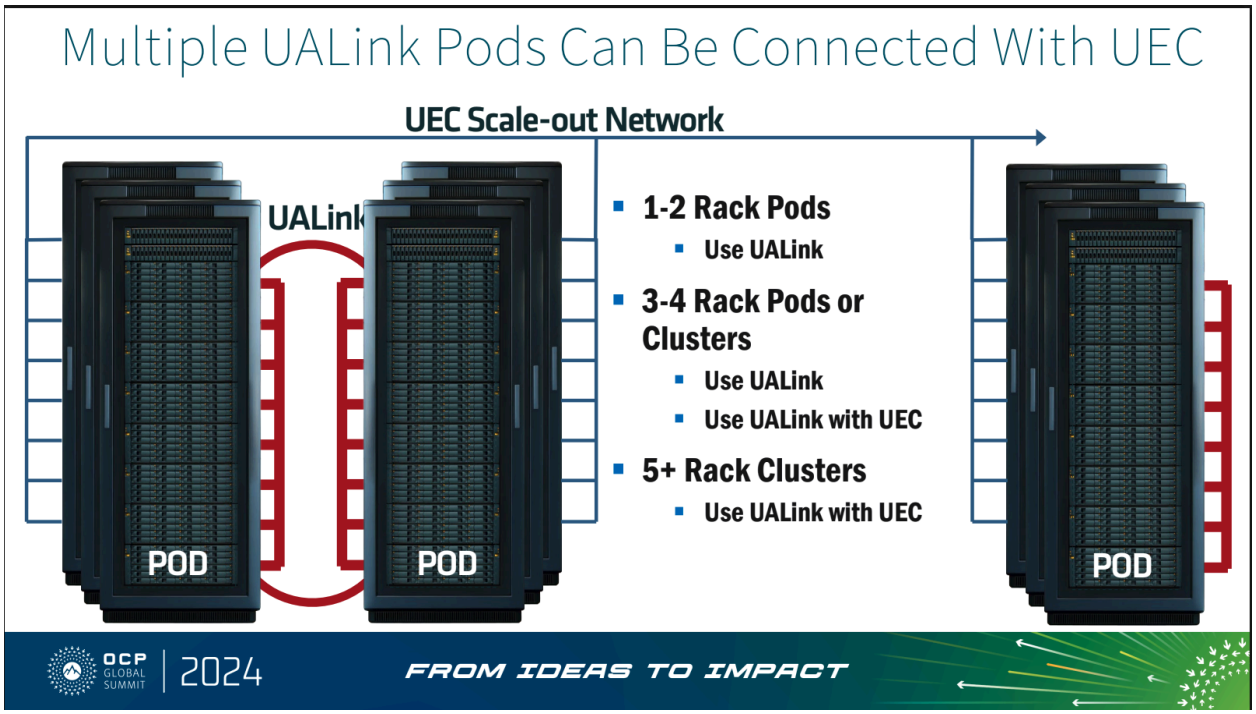


Figure 14: Image from "UALink: Pioneering the AI Accelerator Revolution" presented at 2024 OCP Global Summit.

The OCP Community is presently working with the UEC for optimized scale-out Ethernet based fabrics. The public facing information is this will be compatible with both IEEE 802.3 ba PHY

and 802.3 dj PHY. OCP and the UEC (and a few OCP Community members) have a few joint efforts working toward specific contributions related to enabling the use of this approach by the OCP Community and ecosystem. For more information, the OCP Networking Project typically has updates on this work.

The design requirement is to accomplish this through modular networking—such as commonality with switches, cables, and connectors—allowing flexible integration and configuration across major fabric types.

It is expected that future significant parallelization gains will continue via the scale out fabric.

Other Networking

Front end networks are fairly established and may include Remote Direct Memory Access (RDMA), RDMA over Converged Ethernet (RoCE), storage networking etc… Implementation is well supported from merchant silicon and the various stacks, and also is incorporated into multiple OCP Contributions.

Front-end networks are well-established and may include Remote Direct Memory Access (RDMA), RDMA over Converged Ethernet (RoCE), and storage networking. These implementations are widely supported by merchant silicon and software stacks and are already integrated into multiple OCP Contributions.

Beyond traditional networking, additional interconnect and server design innovations can be leveraged from the broader ecosystem. Examples include chiplet interconnects such as OCP's Bunch of Wires (BoW) and Universal Chiplet Interconnect Express (UCIe), as well as Compute Express Link (CXL) for cache-coherent interconnects. While these technologies enhance system capabilities, they are not mandatory at the foundational level. Instead, modular design principles should ensure flexibility and compatibility, enabling seamless integration as needed.

The OCP Base Specifications for modular Open Systems for AI (OSAI) should be architected to support these technologies appropriately, allowing implementers to choose and refine them into Design Specifications that accommodate advanced capabilities.

Platform Name

Internally, some of the community has been informally referring to a 72-accelerator pods as “OPEN-72” to conceptually align with NVIDIA’s NVL72. This is a logical start point for naming conceptually, but is subject to change as it may not accurately describe the system with enough clarity. This requires further discussion and development, however OPEN-72 is the current place holder.

Data Center Integration Considerations

The power consumption and heat production of a single AI accelerator ranges from hundreds of watts to over 1.6kW. Scaling up, OPEN-72/NVL-72 pods currently have a total power and heat load of approximately 140kW per rack (pod), based on NVIDIA’s published literature and Meta’s Global Summit announcement of an NVL-72 capable system. A 32,000-accelerator cluster would require approximately 444 pods, consuming 62.2MW in total. Given this density, air cooling alone becomes impractical, making liquid cooling solutions essential.

For specific information on hardware thermal management refer to OCP Cooling Environment project community and or other resources such as American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) etc.

As Figure 15 illustrates, AI compute clusters—labeled as Data Center IT (dashed area)—generate high heat loads that require specialized thermal management strategies. To accommodate these systems, data centers optimize thermal exchange capacity at the point of use, adapting cooling loops as heat density increases:

- Mass air handling: Using hot and cold aisles to manage airflow.
- Localized cooling loops: Focusing thermal exchange on specific clusters, reducing inefficiencies.
- Rack-level cooling: Implementing rack-mounted air heat exchangers, immersion cooling, or cold plate/direct-to-chip liquid loops. (often with hot spots)
- Heat reuse: Advanced facilities integrating common Technology Cooling Systems (TCS), per ASHRAE, cooling infrastructure to repurpose heat efficiently.

As heat density increases, loops may become smaller as thermal exchange capacity is directed towards specific areas such as cluster rows down to the chip level.

The OSAI platform will initially align with ORv3, a rack system already in the market and designed for liquid cooling. Existing OCP Specifications for distribution manifolds ensure efficient integration between equipment and data center thermal management systems. As compute density increases, the heat flow design—including interfaces, tubing, pipes, and fittings—will continue to evolve. Cooling loops may directly connect to resources or route through intermediate subsystems such as Cooling Distribution Units (CDUs). The OCP OSAI initiative must synchronize thermal performance development with the ongoing advancements in compute density.

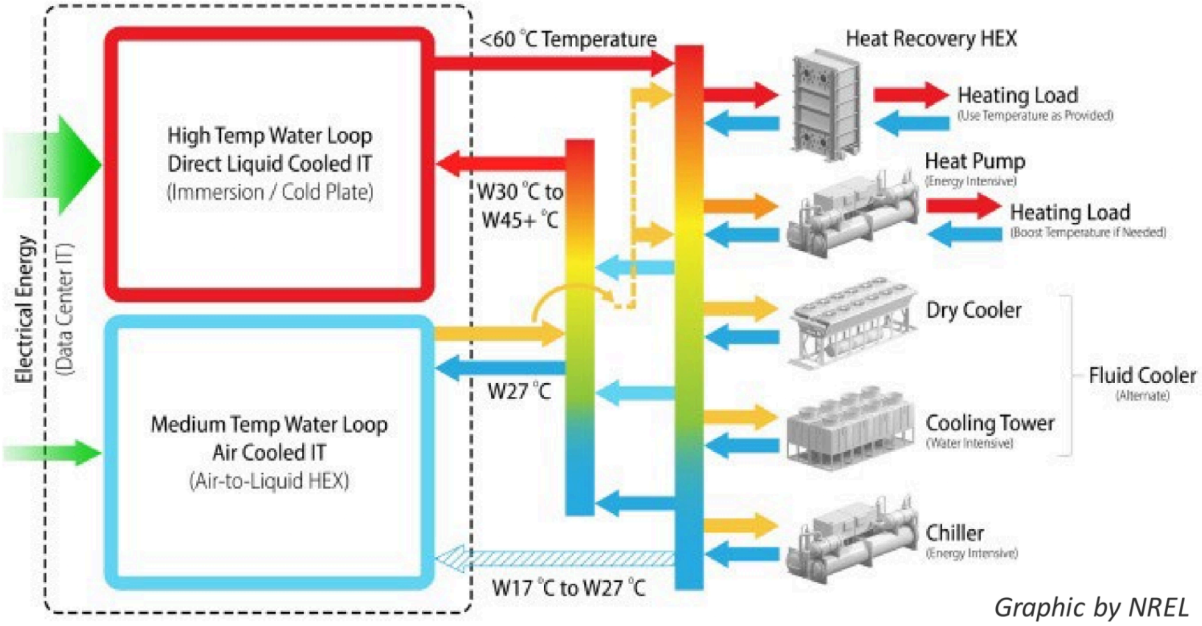


Figure 15: In the diagram above provided by NREL and presented at OCP Global Summit 2024, the illustration gives a representative view of how data center cooling loops move heat from the components.

AI infrastructure, consuming megawatts of power, increasingly faces power availability constraints. An Open System for AI (OSAI) architecture must prioritize evolving efficiency as a core design requirement, as energy limitations can restrict total compute capacity within a given data center.

Consideration of components and subsystems that may be affected by high density AI compute system design may include:

- The data center interfaces (power/cooling/networking/synchronization)
- Cooling Distribution Units (CDU)
- Rack (OpenRack) and Conduit systems
- Security and Hardware Root of Trust
- Platform firmware, board/rack management controllers
- Power systems (either via Rectifier or DC Converters)
    - Bus bars
    - Power connectors
    - AI accelerators (such as NVidia GB200, AMD Mi3xx etc…)
    - The boards that host the processor modules
    - The chassis that hosts the boards and provides management services
    - Scale up and scale out network connectivity
- Rack/Network switches

This list highlights the broad range of subsystems involved, requiring collaboration across multiple OCP projects to establish common requirements that enable a cohesive, system-level AI platform.

Full Stack Development

Much of the focus on the vision is around the speeds and feeds and capabilities of the platform. Equally important is the enablement of the platform itself. There are several layers in the stack that must be co-developed with the hardware. Firstly, the lowest layers need firmware and drivers. Drivers are typically enabled by component vendors, while firmware, security/Hardware Root of Trust, and board management has active projects at OCP that develop these areas and much of this will be inherited by leveraging the OCP platforms above as a starting point. The next layer worthy of discussion is abstraction layers. There is a layer of abstraction provided by accelerator vendors that may include acceleration APIs. Supporting multiple vendor APIs enabled by OCP developed abstraction layers will enable the platform to be universal in terms of reducing developer complexity to optimally serve these workloads for a given set of hardware. This abstraction layer will be directly useful to enable composability, manageability and RAS features, plus parallelization via collectives for example. This is necessary development from the start of this OSAI effort. The OCP Server Project has a subproject named Composable Memory Systems (CMS) and they are working towards enabling real composability of systems. Another sub project, the AI HW/SW Co-design effort is working towards developing

a polymorphic architecture. This is the next level of abstraction and would enable things like heterogeneous compute, dynamic performance optimization, RAS benefits and more. This will not likely be ready for the first iterations of the OSAI platform but development should begin immediately and likely be available for the next generation of accelerators.

3.4. Summary

The OCP Open Systems for AI (OSAI) initiative seeks to develop and coordinate an open, modular platform that optimizes performance and energy efficiency for AI/HPC workloads in data centers. Recognizing the complex interplay between compute, power, cooling, and networking, the OCP Foundation has organized its 12 active Projects into three primary areas (Data Center IT Infrastructure, Data Center Physical Infrastructure, and System Management) to best coordinate on developing an OCP OSAI system.

The content in this section is intentionally high-level, to provide a conceptual framework for the high level shape of the effort and to include some directional statements on anticipated ongoing developments. Subsequent sections (or companion documents) will provide further detailed technical specifications, reference designs, best practices, and future proposals across the diverse areas of OCP's AI/HPC efforts.

4. Datacenter Physical Infrastructure

Datacenter Physical Infrastructure refers to the essential foundation that supports the advanced AI Clusters supporting AI workloads. This includes the physical facilities, power distribution, cooling systems, and the sustainability practices that ensure efficient operation.

It's worth special mention that projects such as OCP Cooling Environments pioneered efforts to develop liquid cooling that has proven necessary to AI clusters. The OCP Community has been developing cold plates, door heat exchangers, immersion systems and others, and is specifically focused on solutions for high TDP hardware environments (e.g., relevant tracks at 2024 OCP Global Summit: Immersion Solutions/Requirements definition, Design Spec for targeted flow heat sinks, fluid thermophysical properties/performance, ORV3 immersed power distribution base spec, RedFish interoperability profiles/hardware management, and more, etc.).

The scope of effort within Open Systems for AI focuses on developing infrastructure capable of supporting the increasing demands of AI, including high-density power and cooling solutions, enhanced heat reuse techniques, and sustainable building materials. By fostering openness and standardization, the objective is to enable scalable, efficient, and resilient physical infrastructure that can accommodate the evolving needs of AI technology while maintaining flexibility for multi-vendor interoperability.

4.1. AI/ML landscape and trends

Figure 1 showed the trends for computational requirements for AI/ML systems based on market demands and emerging workloads and use cases demonstrating a 4X annual increase in computational performance. These system trends are directly related to corresponding trends in the following Physical Infrastructure attributes:

- Chip power rise
- Increasing number of chips co-located within a system (e.g. rack)
- Higher volume consumed by IT inside the rack
- Reduced volume inside rack for power, battery, networking and thermal infrastructure
- Increase in rack power and rack power density
- More stringent serviceability requirements related to Mean-Time-To-Repair (MTTR)

Figure 16 shows trends from Nvidia on both GPU and GPU System powers that was recently published at the 2024 OCP Global Summit, Inherent in the data displayed in Figure 16, is an increase in GPU and GPU System powers over an 8 Year period, even accounting for a massive 45,000 improvement in Performance/Watt over this time. Per the above listing of attributes, the increasing number of chips co-located inside a rack/system is typically related to the Copper interconnects and the distances that govern Signal Integrity (SI). greater number of GPUs within such an interconnected domain leads to higher Performance/TCO and greater Throughput/GPU [Y1]. The more stringent requirements around MTTR arise from very large LLM (Large Language Model) workload runs and a need for faster repair to enable shooter downtime.

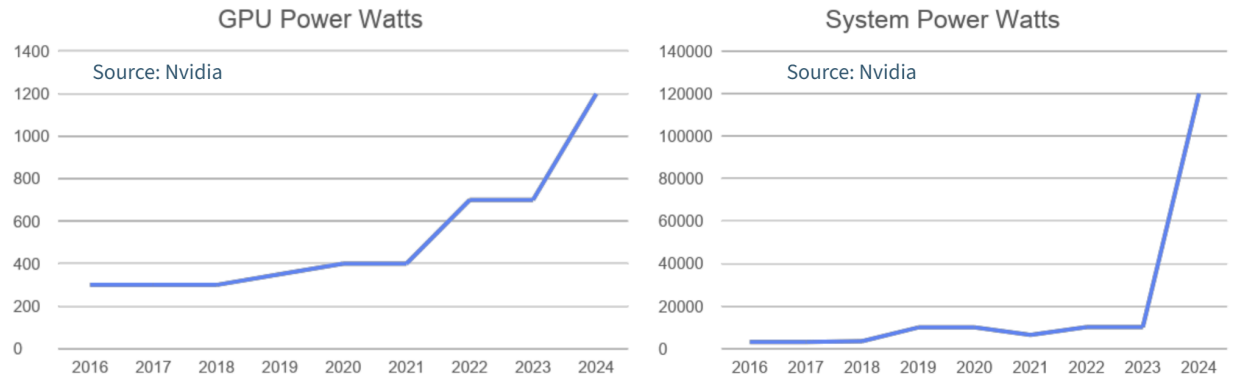


Figure 16: Trends in Nvidia GPU & System Powers [Y1] relevant to Physical Infrastructure

Based on the above trends of increasing chip power gen to gen as well as number of chips closely interconnected (per rack), we foresee a need to address the following challenges:

- Power delivery to the chip
- Heat extraction from the chip, system/racks
- ML rack power is outpacing battery modules and PSU power density improvements. Next gen ML racks will not have space to accommodate power and battery components

4.2. Work Items Considered

DC Facilities	OCP Ready AI Version
	DCF Sustainability (Green Concrete)
	Total Cost of Ownership (TCO) workstream (TCO Model V1.0)
	AI Driven Facility Management
	Power Co-Generation (Alliances)
	Enhanced Heat Re-Use
	TCS
Rack and Power	Power Distribution in Immersion (Base Specification)
	AI Rack (200 - 250KW) (Power)
	AI Rack (200 - 250KW) (Cooling)
	AI Rack (200 - 250KW) (Power Rail)
	AI Rack (200 - 250KW) (Frame)
	AI Rack 48VDC Power
	Advanced Power Architecture AI Rack 400V
	Advanced Power Architecture AI Rack 800V
	AI Rack (300 - 500 KW) (Power)
	AI Rack (300 - 500 KW) (Cooling)
	AI Rack (300 - 500 KW) (Power Rail)
	AI Rack (300 - 500 KW) (Frame)
	AI Rack (300kW - 1MW) (Cooling) (OCP Global Summit MS- Mt.Diablo)
	AI Rack (300kW - 1MW) (Cooling)
	AI Rack (300kW - 1MW) (Cooling)

Table 2: OCP Open Systems for AI Work Items: Data Center Physical Infrastructure

The items shown above in green have work identified and/or are in progress with links to existing OCP efforts. Orange represents unstarted efforts.

4.3. Current Work and Contributions

- Multiple parallel but non OSAI specific efforts are in progress in OCP Projects (not listed) will be used with AI platforms such as High power Rack, Cooling Environments liquid cooling systems, heat reuse, Modular Data Centers and many more.
- Specific tasks, such as those in the table 2 above, are found at: https://docs.google.com/spreadsheets/d/1XbrknCo34VrDO85sasxtDwG44H7NPw-t95zUTDv4aQc/edit?usp=sharing Public - Open Systems for AI SI Work List and Tracking

4.3.1. Sample pathfinding technology and solutions

While there exists a wide array of technology and design variants (and yet to surface inventions), that we expect to be explored and even productized for AI/ML in the upcoming years, we present herein a sampling of a few known significant solutions that we consider to be an important sample set. We will also include a discussion of some alternative solutions that are not known to be confirmed for AI/ML productization, but are prominent in the industry discussions, as an acknowledgement of industry efforts in this area.

A sample set of technology based solutions to address the challenges discussed in the preceding text of this section include:

- Higher power and efficiency power delivery to the chips including Vertical Power Delivery (VPD).
- Customizing IT Rack design for increased density
- Liquid cooling for greater number of devices within the IT rack for higher thermal performance and greater compactness of thermal management solutions. Liquid cooling approaches are inclusive of water- and dielectric-based solutions, e.g., direct to chip (DTC), immersion and cold plate.
- Enabling faster service for uptime by using blind mate trays/blades including power and liquid cooling blind-mate interconnects
- Re-architecture of ML rack, row power and battery systems for 250kW -> 500kW -> 1MW using higher Voltage (+/-400VDC, 800 VDC) power delivery to reduce losses.
- Removal of the battery backup infrastructure from the rack and inclusion at the data center level.

- Reimagining the data center power distribution with a possibility of higher voltage power delivery at the data center (+/- 400VDC, 800 VDC[7]).

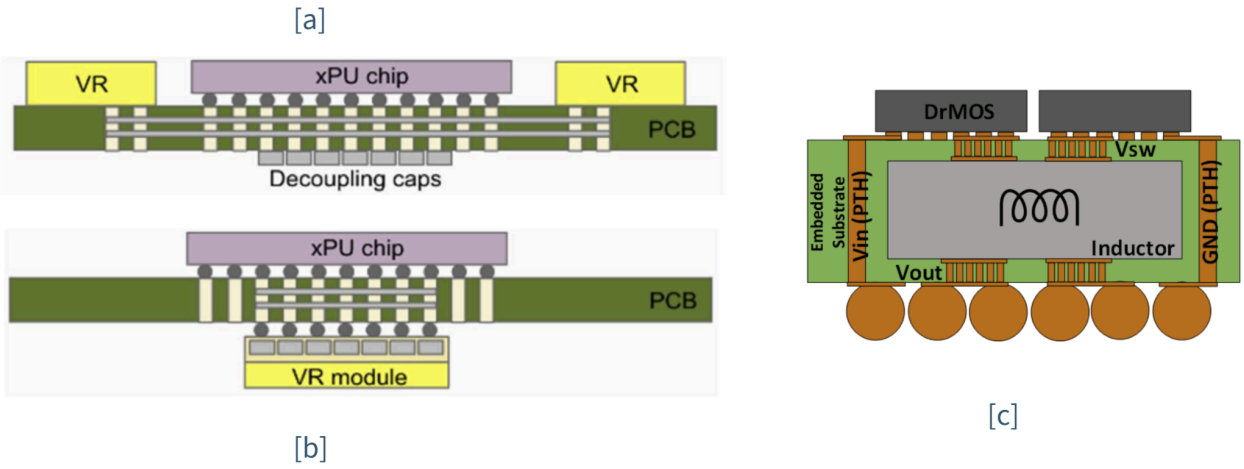


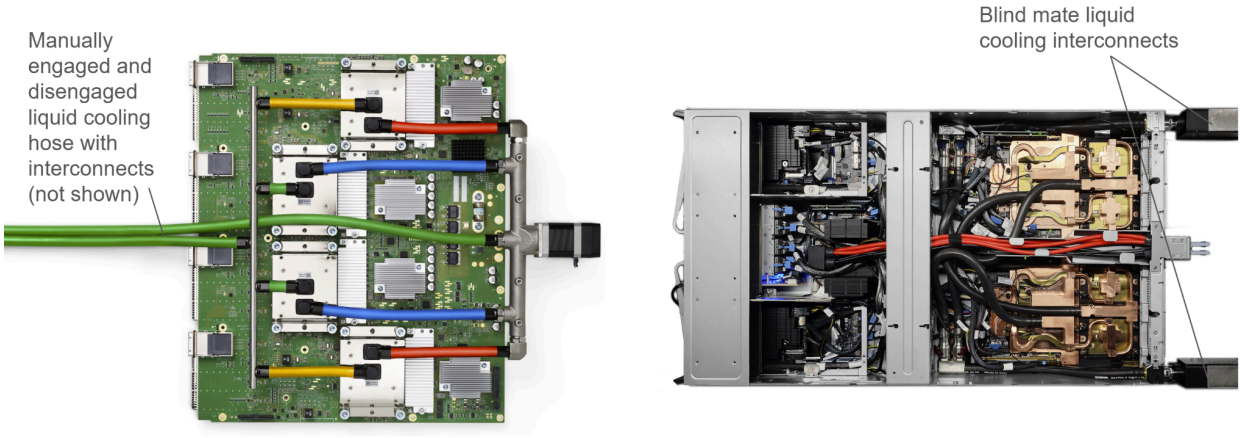
Figure 17: Vertical Power Delivery (VPD) for higher power/efficiency power delivery [Y2, Y3]

Figure 17 (a) depicts a traditional cartoon for lateral power delivery to a chip, while Figures 17 (b) [Y2] and (c) [Y3] show a Vertical Power Delivery option which substantially reduces the losses in distributing high levels of power to the chip. AI ASICs, network switches, and high-performance CPU/GPUs, are reaching the inflection point where IR loss in the "last-mile" lateral distribution path exceeds 10% of the chip's TDP. VPD solution is beneficial not only by substantially reducing the distribution power loss, but also by enabling better system floor plan, improving signal integrity channels, and better thermal performance. Industry [Y2] has successfully deployed for the first time in the industry, a mass-produced, high-density, high-efficiency 1kA Vertical Power solution, with 70% reduction in power losses, and foresee a commitment to the technology to have multiple generations of VPWR solutions implemented in our future platforms.

Figure 18 illustrates two distinct examples of liquid cooled AI/ML chips and systems which have been or are being scaled into high volume manufacturing and data center deployment. The thermal management is achieved through the mechanical attachment of a copper cold plate to the top side of a chip module using a Thermal Interface Material for thermal conduction coupling and heat flow path. The power dissipated by the chip flows through the chip and the TIM into

[7] 400/800VDC technology has been developing rapidly with electric vehicles charging infrastructure, providing a window of opportunity to adopt components from a growing ecosystem and take advantage of the same efficiency gain using higher voltages.

the cold plate copper structures and is transferred to the liquid coolant flowing through the copper structures via solid surface to liquid flow contact. Figure 18 (a) [Y1] and (b) [Y1] both show 4 devices each that are liquid cooled.



(a) Google TPU (ML) version 4 (2021) Liquid cooled Google server

(b) Nvidia GB200 NVL72 Tray (2024) with Liquid Cooled Blind Mate Interconnects

Figure 18: Industry AI/ML examples of liquid cooled chips and systems

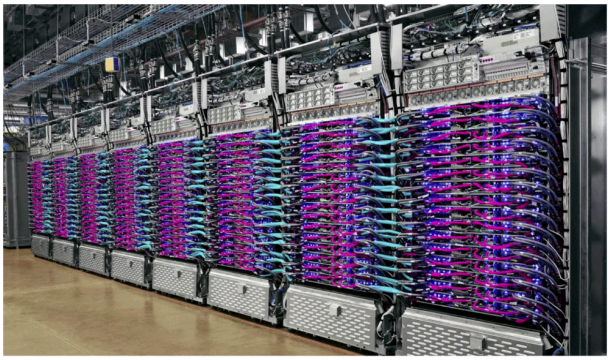
The liquid connection from the liquid cooled tray to the rack coolant distribution manifold can be carried out in different ways, i.e. through flexible hoses and manually engaged/disengaged couplings as shown in Figure 18 (a) example, or through “blind-mate couplings shown on the right side of Figure 18 (b), which is engaged automatically when the server is pushed into the tray and disengaged when it is pulled out.

With respect to the cold plate cooling solution shown in Figure 18 above, there does exist a lot of pathfinding that has been and to be done “under the hood”. Much of this work will be covered as a part of the chip related government funded research efforts that will be referenced in the next subsection (4.3). However, there is also medium term work (more relevant to OCP) related to the type of Thermal Interface material (TIM) to provide higher cooling performance while satisfying the needs for mechanical compliance and structural integrity (reliability), or the design of the package with respect to the use of a lid (which is conventional in CPU space) or baredie config (emerging trend in accelerator chip space). Additional thermal management studies are required related to the AI/ML packages which include both ASICs and High Bandwidth Memory stacks, which each have different power levels and constructions. Also, in the future, industry trends do indicate that ASIC may transition to 3D stacking which is expected to make thermal

management even more challenging.

Figure 16 shows groupings of liquid cooled AI/ML racks in a data center facility cooled by a Coolant Distribution Unit (CDU). A CDU facilitates the liquid cooling of smaller domains of IT racks enabling higher purity and chemistry controlled liquid coolant that will pass through the fine dimensions inside the copper micro-structures of the cold plates shown in Figure 16. The CDU also allows for customized water flow, pressure, and temperature control of the liquid flow to the coldplates cooling the AI/ML servers. Figure 16 (a) shows liquid cooled AI/ML racks that cooled with a CDU [Y1] and Figure 19 (b) depicts a configuration of liquid cooled racks supported by a CDU that is part of an ongoing OCP study [Y1, Y4].

(a) Grouping of AI/ML liquid cooled racks at the data center



(b) Use of CDU (Coolant Distribution Unit) to provide high quality water supply to chip cold plates with controlled pressure & flow

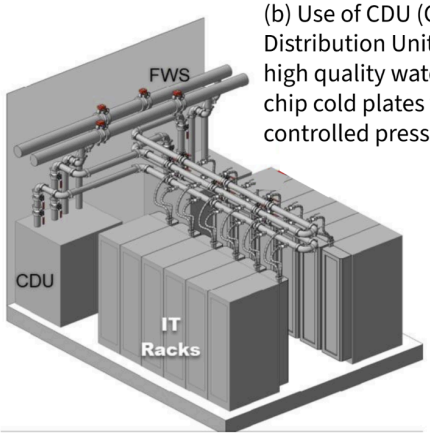


Figure 19: Liquid cooled racks in a data center facility

It's worth noting that the OCP liquid cooling efforts are very advanced, as they were initiated in 2018 with the cold plate, door heat exchanger and immersion projects and have directly and/or indirectly enabled an industry to flourish. This better aligns the thermal transfer performance roadmap with the heat potential of upcoming technologies and products. These projects together are focused on the Technology Cooling System (TCS) loop that is the cooling loop that goes from the CDU to the "cold side" rack manifold and through the liquid connectors, into the IT equipment cooling solution, through the "hot side" liquid connectors and rack manifold, and back to the CDU again. The cooling fluid is of course also a focus in the TCS loop cooling solution.

Several contributions have been made by the OCP community that supports AI/ML liquid cooled

rack solutions related to the TCS cooling loop. Examples of such contributions are:

- Cold plate cooling loop requirements,
- Rack manifold requirements and design,
- Connector specifications: Universal Quick Disconnects (UQD) blind mate/hand mate, and ORV3 blind mate quick connectors,
- Cold plate development and qualification,
- Guidelines for (various) system liquid cooling,
- Guidelines for using water-based and glycol-based cooling fluids, and
- Leak detection and intervention.

There are also additional focus areas within the liquid cooling community that is either ongoing or in the planning phase on addressing the following TCS related topics:

- Liquid-to-liquid CDU test methodology,
- CDU requirements and qualifications,
- Fluid velocity requirements for pipes & flexible hoses,
- 2-phase pumped refrigerant based cold plate technology, and
- Leak sensor specification.

Figure 20 (a) [Y1, Y5] shows one version of an Industry rack power architecture (also represented in OCP Open Rack) via a 48V DC busbar for distributing DC power to the UT and with 416 VAC power supply to the rack from the data center. Figure 20 (a) also shows the UPS functionality housed within the IT rack in the form of BBUs (Battery Backup Units). As there is a greater need to fit more accelerator chips (for example, GPUs) inside the rack domain owing to a desire for larger numbers of chips within a copper based interconnected system, there is a need to remove some of the rack supporting infrastructure to be located outside that rack (as a Side rack or via the data center). Figures 20 (b) and (c) [Y1] show a transition from the configuration shown in Figure 19(a) to configuration without BBU (Figure 17 (b) and with both BBUs and the AC to DC power conversion functionality (Figure 17 (c)).

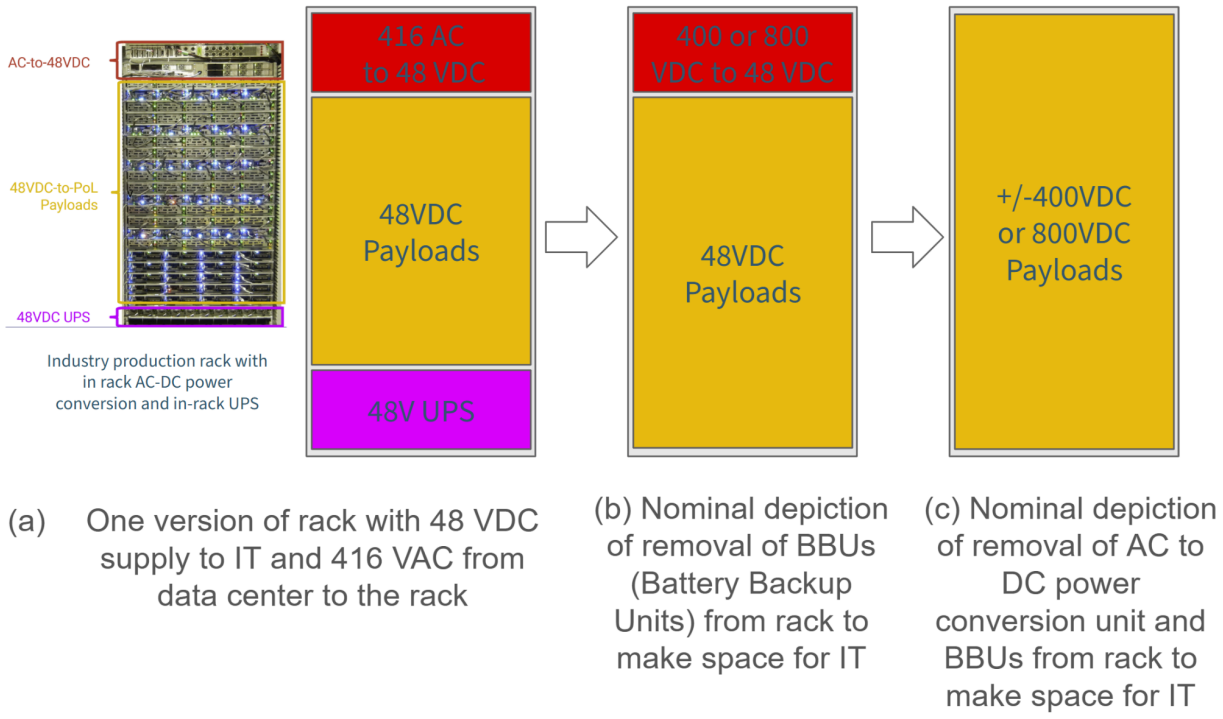


Figure 20: Transitioning from AC to DC power conversion and UPS functionality from inside the IT Rack to outside of the IT Rack to make room for more interconnected chips

While Figure 20 (c) assumes that the data center power will be distributed at higher voltage DC (e.g. +/- 400VDC or 800 VDC), the industry is not quite there with respect to a large-scale adoption of such a configuration. Also, there will exist a large amount of data center footprint with the legacy configuration distributing 416 VAC that would need to be potentially utilized by next gen AI/ML systems. Figure 21 [Y5] addresses this transitional need to support a higher voltage DC current IT rack while housing such a rack in a more conventionally configured AC distribution based data center. This Sidecar rack [Y5] can support multiple AI/ML IT racks while receiving power at the data center scale through legacy AC power.

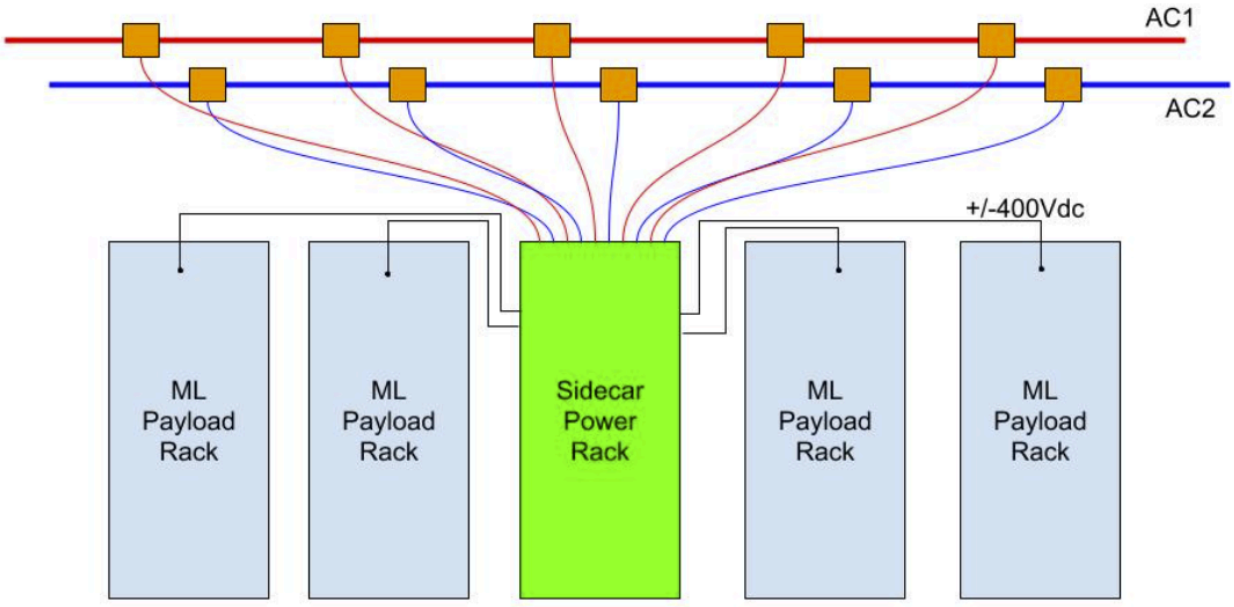


Figure 21: Use of AC to DC conversion Sidecar Rack to supply high voltage DC to the IT rack while being housed in a AC based data center [Y5]

Such a Sidecar rack [Y5] can be configured with or without BBUs depending on facility backup power configuration, and could be capable of supporting different payload rack configurations depending on payload rack power requirements.

While the discussions using Figures 20 and 21 explicitly mentioned the need to create more space in the IT rack for more accelerator chips, they also implicitly depict higher voltage DC distribution as a requirement. The need to move from 48 VDC as is currently utilized in both OCP and industry racks (hyperscalers) to a higher voltage (+/- 400 VDC or 800 VDC) is driven by reducing the current required to supply high power to reduce the electrical losses, the cooling requirements of the bus bars (cooling the power dissipation from losses), and the cross-sectional area of the bus bars required to distribute the power (which grows in area and volume as the power and current increase).

There continues to be a discussion in the industry regarding what the higher voltage DC distribution is most optimal for high power AI/ML racks, i.e. +/-400 VDC or 800 VDC. This white paper will not discuss the pros and cons and will rather identify this as a potential area of

convergence opportunity to be driven by the OCP community.

Immersion cooling is an alternative approach to achieve improved power density and thermal management of power distribution equipment. In this approach, the rack is often turned on its side so that the busbar is located at the bottom of the tank, with the compute and power supplies installed above it, and support from the buoyant fluid and tank/product design elements to hold the equipment in place (Figure 22). Multiple iterations of power supplies and connectors have been validated so far in the community, including 3.6 kW supporting 54V (preferred) or 51V. The approach is complementary to the existing OCP ORv3 specification for air-cooled designs.

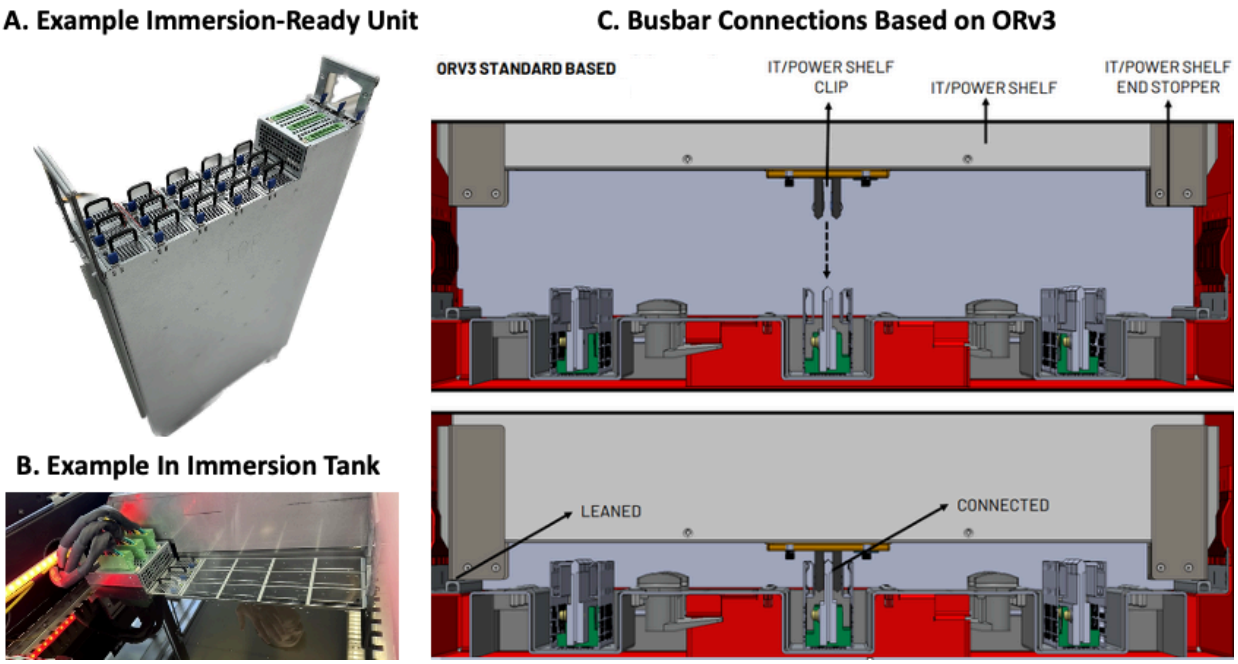


Figure 22. Adapting power distribution equipment in a single phase immersion environment.

Meeting the needs of specific use cases and thermal profiles, OCP Immersion and Power Distribution communities have proven out several key advantages when single phase immersion is implemented.

For AI-focused data center deployments, the improvements to power density and thermal profiles are of notable importance:

- Busbar capable of 2-3x power at equivalent temperature rise vs. air
    - ➢ Why it matters: improved power density on a standard busbar without requiring a high power rack (HPR) or similar solution
- Hot spot temperatures reduced by ~50% at equivalent workloads (e.g., ~70 °C in immersion vs. ~150 °C in air; Figure 22 below)
    - ➢ Why it matters: significantly improved thermal management, with all waste heat transferring to liquid rather than facility ambient (a win-win for efficiency)
- Immersion fluid tolerates elevated temperatures, even at maximum PSU loads
    - ➢ Why it matters: enables dry cooler usage in warmer climate deployments. This unlocks the potential for a zero water data center.
- Greater than 100% PSU loads are achievable. At 130% PSU load and a relatively hot fluid environment (55 °C), only 18 °C increase in component temperature was observed and all observed equipment/system temperatures remained <90 °C.
    - ➢ Why it matters: additional performance delivered with same PSU configuration
- Bulk fluid provides a more homogenous thermal environment around hot spots
    - ➢ Why it matters: fewer extreme temperature swings during routine usage; reduced severity of material expansion/contraction cycles may improve reliability
- Fan removal reduces the number of high-failure rate parts at scale
    - ➢ Why it matters: fewer moving parts, reduced outages due to component failure
- Failure mode considerations proven out, including facility outages and cooling loop outages; more opportunities to explore safeguards and failure modes in community
    - ➢ Why it matters: critical to ensure safe function in this new operating environment
- Reduction in overall power consumed by AI workloads: Multiple examples of scaled use with significantly improved server efficiency
    - ➢ Why it matters: less power is required to operate an AI data center
- In a "bath" style immersion enclosure, a great degree of scalability and adaptability exist to support diverse equipment types and form factors
    - ➢ Why it matters: tank system is largely agnostic to how it is populated, with components being swappable and adjustable after deployment to support hardware upgrades. Demand for next generation compute refreshes is particularly high for users that support AI workloads.



Figure 23. Test results for per-component thermal benchmarking for a 55 °C fluid environment. Component maxima were 100-175 °C in the air cooled environment and 40-70 °C in the immersion environment.

Immersion cooling provides a path to enhanced data center efficiency, density and sustainability - addressing the most significant challenges associated with AI at scale. Additional considerations do have to be made to successfully implement power distribution equipment in an immersion environment. These include unique consideration to materials compatibility, firmware, and system- and equipment-level safeguards. Significant opportunities exist to optimize design and standardize validation for compute and power equipment in immersion applications. The group is currently writing a Power Distribution in Immersion Base Specification to outline product design & validation requirements.

4.4. Identified Gaps, Emerging R&D Areas, OCP Opportunities for Leadership & Impact

There is an explosion of industry and university R&D in the area of high power and power density related to power delivery and thermal management. These efforts include those that have gradually evolved over the last two decades of increasing chip and system powers in the IT market, and those that have been sparked off by recent awareness AI/ML related chip and system requirements in the power and thermal space as enablers (or limiters) for advancement of technologies critical for the US and Global economy. This is illustrated by 2019-2024 IEEE Roadmap efforts in Heterogeneous Integration [Y7], the ongoing ARPA-E Coolerchips program [Y8] focused on “transformational, highly efficient, and reliable cooling technologies for data

centers", the DARPA Microsystems technology Office initiatives on 3D chips and heterogeneous integration [Y9], and the upcoming and substantial US Government Chip Act funding opportunity [Y10]. All these initiatives have dedicated efforts focused on power and/or thermal management. We expect all these industry activities to yield longer term progress in power and thermal management research and technology that will serve as pathfinding foundations for AI/ML advancement. While many of these research areas are not yet mature enough for broader industry standardization, we expect solutions to emerge that can benefit from OCP advocacy and engagement in the AI/ML space.

In addition to these chip-centric initiatives, there are other industry efforts such as those driven by ASHRAE (American Society for Heating Refrigeration and Air-Conditioning Engineers) focused on various data center areas [Y11] including almost a two decade emphasis on liquid cooling (See books in, [Y12]). In case of the ASHRAE efforts, they are expected to yield broader understanding of data center requirements, guidelines, and standardization that leverage the broader outreach that is possible through ASHRAE (data center service providers, A&E forms, IT OEMs, vertically integrated hyperscalers, components and sub-system suppliers). OCP and ASHRAE have already been engaged in discussions and through overlapping participation which we expect to continue.

We believe that the emphasis of our work at OCP in AI/ML Physical Infrastructure is best served with a focus on short and medium term requirements and the articulation of sample reference designs and case studies that address this space (versus fundamental R&D that overlaps with existing efforts). Here are some sample areas where OCP short- and medium- term (~2025+) contributions can serve to better prepare and converge the AI/ML Physical Infrastructure industry for rapid advancement of solutions and a supply ecosystem:

1. Reference case studies for cooling of the Accelerator chip with single-phase cold plates using different options of Thermal Interface Materials, and with different types of package construction (e.g. bare die chips in contact with the cold plate or lidded chips contacting the cold plates), and a narrative AI/ML chip thermal attributes with ASICs and HBMs.
2. High power rack reference designs for mechanical and power delivery using Open rack V3 (ORV3) to support 250, 500, 750, and 1 MW rack powers, yielding CAD and supplier prototypes that can be made available to the industry for experimentation.

3. Understanding trade-offs between different high voltage power delivery options at the rack, e.g. +/-400 VDC vs 800 VDC, encouraging industry convergence for solution sets with options through data-driven discussions. Such studies can drive greater streamlining of supplier solutions sets for components and subsystems.
4. Reference designs for Side-Rack voltage step down units (e,.g. 416 VAC to 400 VDC) which also include UPS functionality to serve as a transitional step to enable high voltage powered IT in legacy and in-flight data centers (which commonly provide 416 VAC). These solutions allow the removal of some of the power and all of the UPS functionality from the rack thus increasing the number of GPUs inside the rack and of course the rack power (as discussed in the trends section herein).
5. Creation of guidance documents that enable design and qualification of power distribution equipment with improved density and thermal management. A base specification is currently in progress to support application in a single phase immersion environment (OCP Immersion Power Distribution workstream). Of particular importance in that work is the incorporation of design elements for performance and safety in a liquid environment (rather than air). Community collaboration opportunities exist to ensure safety, proper overlap of existing ORv3 guidance, and proof of performance in a production environment.
6. High power rack liquid cooling reference designs including liquid cooling distribution manifold sizing and placement/support/serviceability, and interconnect options (quick disconnect coupling - blindmate or otherwise) to IT trays that house the chips.
7. Potential for open reference designs for liquid cooling Coolant Distribution Units (CDUs) supporting multiple high power liquid cooled racks inclusive of manifold distribution and interconnect optionality. Based on the 2024 OCP Global Summit Expo, there exist several existing or in-flight industry solutions which can be documented to serve as a solution set for the industry and empower users to make trade-offs more transparently.
8. Modular data center reference design for high voltage and liquid cooled IT with battery backup functionality provided at the data center level (e.g. BESS - Battery Energy Storage Systems). While industry might not yet be fully ready to take the leap towards high voltage (+/- 400VDC or 800 VDC) distribution at the data center scale, reference designs in this area can serve to drive faster market adoption and supplier offerings availability.
9. A framework for Data Center Sustainability Metrics for AI/ML based IT and data centers that can be applied with case studies so as to proliferate the use of such methodologies

and metrics in a consistent manner (and providing guidance to the industry on tradeoffs and options). The topic of Waste-Heat Recovery can also be addressed through a treatment of this area.

10. OCP Ready Data Center engagement with service providers for enablement of unified checklists and criteria for AI/ML.
11. Data center automation reference solutions for faster deployment/serviceability velocity of AI/ML racks into the data center.
12. Cost modeling tools that enable users to benchmark different approaches to implementing IT equipment given their deployment's unique constraints. With a variety of hardware and supporting equipment options to support AI workloads, validated total cost of ownership (TCO) tools are of high value. OCP CE Total Cost of Ownership workstream has created a V1.0 tool, but is seeking vendor and user validation to refine assumptions, inputs and overall experience. Common approaches to performance benchmarking across different infrastructure approaches would further improve the tool

Each of the above twelve areas could be driven (e.g. in 2025) as a workstream and/or a white paper for consumption through OCP by the broader industry. These twelve areas can be assimilated into a broader vision for a next generation of OCP AI/ML efforts (e.g. in 2026).

There also exist several emerging areas that are either deployed on a much smaller scale or have not quite found a market foothold as solutions for the AI/ML space. We are listing a few such emerging areas, each of which have substantial market interest:

A. Direct to Chip with CDU - The conventional liquid cooling system utilizes a Coolant Distribution Unit (CDU) to transfer heat from the secondary liquid cooling (Technology Cooling System [TCS]) at the IT to the primary liquid loop (Facility Water Loop [FWS]) at the facility level. However, there are data centers which do not possess a supply of water at the facility level. The coolant liquid typically contains a blend of glycol and water to reduce risk of microbial growth. For such infrastructure, an air-cooled rack that supplies liquid coolant to a liquid-cooled rack can be employed as a transitional solution. (Typical capacity per rack can range from 10–30 kW, depending on the design and airflow parameters; see [X1], [X2].)
B. Direct to Chip with Two Phase Fluid - This cooling technique relies on two-phase boiling of a liquid coolant inside the cold plate mounted on the chip to effect heat transfer. The coolant performance relies on appropriate fluid boiling point and latent heat of

vaporization, as compared to natural or forced convection seen in single phase approaches. (Two-phase cold plates can achieve very high local heat flux dissipation, in the order of 100–500 W/cm², with typical server-level capacities reaching 15–60 kW depending on rack density; see [X5], [X6].)

C. Microchannel-Assisted Direct to Chip - This advanced cooling technology represents the use of micro-structures embedded or carved into the ASIC itself, with liquid flowing through these microstructures in direct contact with the silicon chip. (Early research prototypes suggest heat flux densities of 300–1,000 W/cm² are possible, though commercial adoption is still emerging; see [X7], [X8].)

D. Immersion Cooling - This thermal management technique can take the form of immersing the IT inside a tank filled with single- or two-phase dielectric coolant or the form of encapsulating the IT server in the dielectric coolant. The liquid coolant is non-conductive and so unlike many Direct to Chip approaches, can make direct physical contact with ITE hot spots and completely eliminate the need for water utilization. OCP has a very active community of workstreams organized in four technical committees: Fluids, Compatibility & Reliability, IT Equipment, and holistic Solutions. (Immersion solutions can handle 50–200+ kW per rack, often limited by pump and fluid circulation design; see [X3], [X4], [X2] - see references section.

5. Data Center IT Infrastructure

Data center IT infrastructure encompasses server, storage, networking, and security domains. OCP Open Systems for AI in this space aims to eliminate performance bottlenecks, enhance architectural density and computational capacity, and foster a multi-vendor ecosystem through openness and interoperability. By collaboratively developing specifications, best practices, and white papers, we drive standardization that strengthens AI-ready IT infrastructure, improving performance while leveraging economies of scale.

Several OCP efforts will not go it alone and will reach out and be collaborative with external industry partners such as Ultra Ethernet Consortium (UEC), Ultra Accelerator Link Consortium (UAL), Google's Falcon[8], Storage Networking Industry Association (SNIA), Optical Internetworking Forum (OIF), IEEE, and many others to increase effort synchronization.

5.1. Components of "Data Center IT Infrastructure"

The rapid advancement of artificial intelligence (AI) and machine learning (ML) technologies has necessitated the development of specialized server racks designed to handle the immense computational and networking demands of AI workloads. The intended outcome is to define standards for AI server racks within the Open Compute Project (OCP) Open Systems for AI (OSAI) initiative, specifically focusing on the "Data Center IT Infrastructure" part of OSAI.

The primary components of the "Data Center IT Infrastructure" might include, but are not limited to, four elements: 1) Compute Trays, 2) HBD Switch Trays, 3) Power Supply Trays, and 4) System Management Trays. Each section of this whitepaper will provide a detailed description of these trays, their functions, and the standards that should be established to ensure interoperability, scalability, and efficiency.

5.1.1. Introduction to High Bandwidth Domain (HBD)

The HBD is the critical specialized network within the AI "system" that provides high-bandwidth low-latency communication between accelerators in Compute Trays at the required scale for an AI workload. The HBD often has electrical channel limitations and latency, power, and cost targets which are a significant driving force behind the system architecture typically leading to

[8] https://cloud.google.com/blog/topics/systems/introducing-falcon-a-reliable-low-latency-hardware-transport

dense flat switched or glueless networks. The HBD may utilize standardized interfaces like UALINK™, which is designed to meet the demanding requirements of AI workloads.

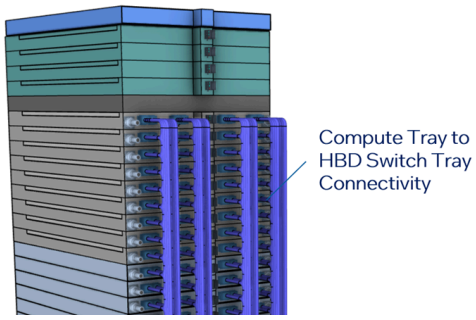


Figure 24: Example Compute Tray to HBD Switch Tray

An illustration of this HBD concept in Figure 25 below demonstrates a single rack implementation. This example is flexible, as the number of Compute Trays and HBD Switch Trays in the rack can vary. In this specific example, 200G passive copper channels are achieved by centering HBD Switch Trays between the Compute Trays.

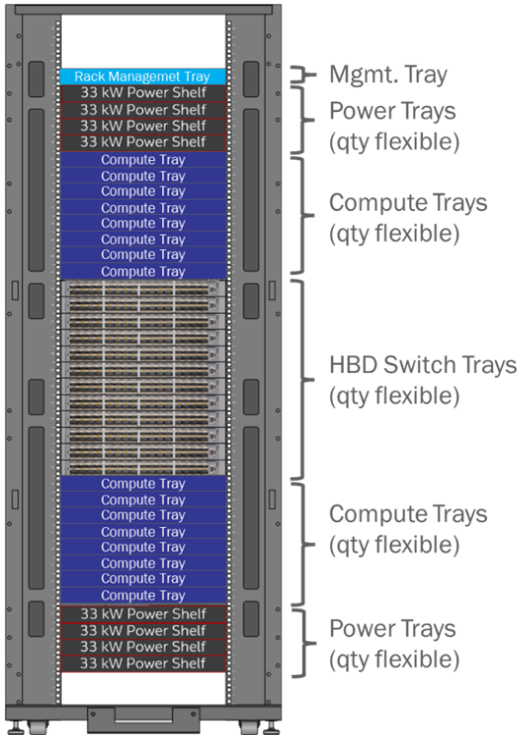


Figure 25: Data Center IT Infrastructure Ingredients in HBD Rack

Two racks can be interconnected to create a Dual Rack HBD, and four racks can be combined to form a Quad Rack HBD. Further details on the Compute Tray, HBD Switch Tray, Management Tray, and Power Trays are provided in the next sections.

5.1.2. Compute Trays

Compute Trays are the heart of the AI server rack, housing the processing units responsible for executing AI and ML workloads. These trays are designed to accommodate high-performance GPUs, CPUs, and other accelerators, providing the necessary computational power to handle complex AI tasks.

Design and Specifications

- **Form Factor**: It is recommended that Compute Trays adhere to a standardized form factor to ensure compatibility with various rack configurations. A common form factor, such as the Open Rack v3 (ORv3) standard or future derivative, is suggested.
- **Cooling**: Efficient cooling mechanisms, including liquid cooling and advanced airflow designs, are encouraged to manage the heat generated by high-performance components.
- **Interconnects**: Standardized high-speed interconnects, such as PCIe Gen5/6+, are recommended to facilitate rapid data transfer between processing units. Additionally, supporting standardized interfaces like Compute Express Link (CXL) from the CXL Consortium can provide high-speed, low-latency interconnects between CPUs, GPUs, memory, and other accelerators, enabling efficient data sharing and memory coherence within the system. This is crucial for AI workloads that require rapid data access and processing.
- **Modularity**: Compute Trays are designed to be modular, allowing for easy upgrades and replacements of individual components, such as GPUs, memory modules, and network interfaces. This modularity accelerates AI computing needs by enabling quick adaptation to new technologies and scaling requirements.
- **Power Efficiency**: Power-efficient designs, including dynamic voltage and frequency scaling (DVFS), are recommended to optimize energy consumption.
- **Storage**: Compute Trays are capable of housing large amounts of NVM Express® (NVMe®) storage to support data-intensive AI workloads. This storage is easily accessible and upgradable to accommodate growing data requirements. It is also a

security concern as large amounts of data at rest is an attack surface. Efforts need to continue and to expand security and privacy protection for data storage. OCP Storage Project[9] has a number of specifications that may be key to implementation of a secure and performant storage solution for AI.

- **Scale-Up HBD Connection**: Compute Trays should include a standardized scale-up connection to the HBD Switch Tray. One such option could be UALINK™ which ensures high-speed, low-latency communication between the Compute Trays and the HBD Switch Trays, facilitating efficient data transfer and processing within the rack.

Standards

- **Compatibility**: Compute Trays are compatible with a wide range of AI accelerators, including GPUs, TPUs, and FPGAs, to provide flexibility in hardware selection.
- **Scalability**: The design supports scalable configurations, enabling the addition of more Compute Trays to meet increasing computational demands.
- **Reliability**: Redundant components and fault-tolerant designs are incorporated to ensure high availability and reliability.

Scale-Out Networks

Compute Trays may interface to one or more scale-out networks, which allow connectivity beyond the HBD. It is recommended that these connections be based on Ultra Ethernet (UEC) from the UEC Consortium as an example. UEC provides high-speed, low-latency networking capabilities that are essential for large-scale AI deployments. The scale-out network is located on the user/cold aisle.

- **Interoperability**: The scale-out network adheres to industry-standard networking protocols, such as UEC, to ensure compatibility with a wide range of devices and systems.
- **Scalability**: The design supports scalable networking configurations, allowing for the addition of more racks and network connections as the AI infrastructure grows.
- **Security**: Security is designed to provide confidentiality, integrity and availability of the system to ensure sensitive data is protected from unauthorized access and remains

9 https://www.opencompute.org/contributions?contributions%5BrefinementList%5D%5Bproject%5D%5B0%5D=Storage

accurate and unaltered. It is aligned with OCP and NIST Cybersecurity Framework (CSF) security features, including encryption and access control features, are implemented to protect data integrity and confidentiality across the scale-out network.

- **Relaxed Ordering**: UEC supports relaxed ordering to improve performance by allowing out-of-order data processing. This can reduce bottlenecks and increase throughput, which is crucial for AI workloads that involve large-scale data processing.
- **Tail Latency**: UEC is designed to minimize tail latency, which is the latency experienced by the slowest requests in a system. Reducing tail latency is essential for maintaining consistent performance and ensuring that AI workloads are processed efficiently without significant delays.
- **Multi-Pathing Packet Spraying**: UEC supports multi-pathing packet spraying, which distributes packets across multiple paths to balance the load and improve network utilization. This approach contrasts with flow-level multi-pathing, where entire flows are sent along a single path. Packet spraying can reduce congestion and improve overall network performance, making it ideal for AI workloads that require high throughput and low latency.
- **Simplified RDMA**: UEC supports simplified RDMA with a streamlined API, native workload interaction, minimal endpoint state, and not based on traditional IBTA verbs like legacy RDMA. This simplified RDMA approach reduces complexity and overhead, making it more efficient and easier to integrate with AI workloads. The benefits include lower latency, reduced CPU overhead, and improved performance for data-intensive AI applications.

5.1.3. HBD Switch Trays

HBD Switch Trays are responsible for the high-bandwidth networking required to connect the various components within the AI server rack. The HBD Switch Trays facilitate data transfer between Compute Trays, ensuring seamless communication and data flow.

Design and Specifications

- **High Bandwidth Domain (HBD)**: The HBD is a critical aspect of the AI server rack, providing a high-speed, low-latency communication channel between the Compute Trays and the HBD Switch Trays. A standardized interface, such as UALINK™ from the

ULTRA ACCELERATOR LINK CONSORTIUM™, is recommended to ensure interoperability and performance.

- **UALINK™ Specification**: UALINK™ is designed to provide flexible bandwidth and high-speed connections between system nodes (in the Compute Tray) and the UALINK™ Switch (in the HBD Switch Tray). UALINK™ supports the concept of UALINK™ Station, where ports can be ganged together to provide additional bandwidth as needed. This flexibility allows for efficient scaling of network resources to match the computational demands of AI workloads.
- **Scalability**: The HBD is designed to support scalable configurations, allowing for the addition of more Compute Trays and HBD Switch Trays as the network grows. This scalability ensures that the AI server rack can accommodate increasing computational and networking demands.
- **Low Latency**: Low-latency networking protocols and hardware are employed to minimize communication delays between components.
- **Port Density**: A high port density is provided to support the numerous connections required in a large-scale AI server rack.
- **Redundancy**: Redundant networking paths and failover mechanisms are implemented within the HBD to ensure continuous operation in the event of a network failure. This redundancy enhances the reliability and availability of the AI server rack.
- **Management**: Advanced Ethernet based network management features, including software-defined networking (SDN) capabilities, are integrated to enable efficient network configuration and monitoring.

Standards

- **Interoperability**: HBD Switch Trays adhere to industry-standard networking protocols, such as Ethernet and InfiniBand, to ensure compatibility with a wide range of devices.
- **Scalability**: The design supports scalable networking configurations, allowing for the addition of more HBD Switch Trays as the network grows.
- **Security**: OCP and NIST Cybersecurity Framework (CSF) security features, including encryption and access control, are implemented to protect data integrity and confidentiality.

5.1.4. Power Supply Trays

Power Supply Trays provide the necessary electrical power to the components within the AI server rack. These trays are designed to deliver reliable and efficient power, ensuring the continuous operation of the rack's hardware.

Design and Specifications

- **Open Rack v3 Standard**: Power Supply Trays adhere to the Open Rack v3 standard, which defines the physical and electrical specifications for power distribution in data center racks.
- **High Efficiency**: High-efficiency power supplies, such as 80 PLUS Platinum or Titanium certified units, are recommended to minimize energy loss and reduce operational costs.
- **Redundancy**: Redundant power supplies and power distribution units (PDUs) are implemented to ensure continuous power delivery in the event of a failure.
- **Scalability**: The design supports scalable power configurations, allowing for the addition of more power supplies as the rack's power requirements increase.
- **Monitoring**: Advanced power monitoring and management features are integrated to enable real-time tracking of power consumption and efficiency.

Standards

- **Compatibility**: Power Supply Trays are compatible with a wide range of power distribution systems to provide flexibility in power infrastructure.
- **Scalability**: The design supports scalable power configurations, enabling the addition of more power supplies to meet increasing power demands.
- **Reliability**: Redundant components and fault-tolerant designs are incorporated to ensure high availability and reliability.

5.1.5. System Management Tray

The System Management Networking Tray is responsible for the management and monitoring of the AI server rack's components based on Open RMC. It is recommended that the System Management Tray is both the Open RMC appliance and a network switch. This tray provides the necessary interfaces and tools to manage the hardware, monitor performance, and ensure the

efficient operation of the rack. Each component/tray in the rack that needs to be managed shall have an Ethernet port as a management interface.

Design and Specifications

- **Management Interfaces**: The System Management Networking Tray provides a variety of management interfaces, including IPMI, Redfish, and SNMP, to enable remote management and monitoring of the rack's components.
- **Monitoring Tools**: Advanced monitoring tools are integrated to track the performance, health, and status of the rack's hardware. These tools provide real-time data and alerts to facilitate proactive maintenance and troubleshooting.
- **Security**: Security is an essential feature in system management, and it is designed to provide confidentiality, integrity and availability of the system management to ensure sensitive data is protected from unauthorized access and remains accurate and unaltered. It is aligned with OCP and NIST Cybersecurity Framework (CSF) security features, including authentication, encryption, and access control, which are implemented to protect the management interfaces and data.
- **Redundancy**: Redundant management paths and failover mechanisms are implemented to ensure continuous management and monitoring capabilities in the event of a failure.

Standards

- **Interoperability**: The System Management Networking Tray adheres to industry-standard management protocols to ensure compatibility with a wide range of management tools and systems.
- **Scalability**: The design supports scalable management configurations, allowing for the addition of more management interfaces and tools as the rack's complexity increases.
- **Security**: OCP and NIST Cybersecurity Framework (CSF) security features are implemented to protect the management interfaces and data from unauthorized access and cyber threats.

5.1.6. Summary

The development of standardized AI server racks within the Open Compute Project Open Systems for AI initiative is essential to meet the growing computational and networking

demands of AI workloads. By defining standards for Compute Trays, HBD Switch Trays, Power Supply Trays, and System Management Networking Trays, we can ensure interoperability, scalability, and efficiency in AI server rack designs. These standards will facilitate the deployment of high-performance AI infrastructure, enabling organizations to harness the full potential of AI and ML technologies.

The integration of UALINK™ for high-bandwidth, low-latency communication within the High Bandwidth Domain (HBD) and the use of Ultra Ethernet (UEC) for scale-out networking between racks will provide the necessary networking capabilities to support large-scale AI deployments. Additionally, adhering to the Open Rack v3 standard for Power Supply Trays and implementing robust management and monitoring features in the System Management Networking Tray will ensure reliable and efficient operation of the AI server racks.

By establishing these standards, the OCP Open Systems for AI initiative will create a foundation for the development of interoperable, scalable, and efficient AI server racks, driving innovation and enabling organizations to meet the demands of modern AI workloads.

5.2. Work Items Considered

The Table below shows a snapshot of the current work items within the Server, Cooling Environment, Networking, Storage, and Security projects related to AI.

Project	Workitem
Server	MHS Open System for AI
	NIC 3.x - AI form factors
	AI HW/SW Codesign
	Composable Infrastructure
	Quantization (OFP8)
	Quantization (MX)
	Chiplets for AI
	System in Package Security (SiP)
	Common/Standardized Hardware Abstraction layers
	In-Memory Computation
Cooling Environment	ASIC Cooling Techniques
	Immersion
Networking	FTI Short Reach Optical Interconnect

	CSIG (with UEC)
	Falcon + NVME + RDMA
	AI Driven Load balancing
	Co-packaged Copper and/or Co-packaged Optics
Storage	FTI Data Centric Computing
Security	Chiplets (Chiplets as a Security Device)
	In-memory encryption
	SAFE Outside Storage

Table 1: OCP Open Systems for AI Work Items: Data Center IT Infrastructure

The items shown above in green have existing work in progress in the relevant OCP project.

Important note, non OSAI specific efforts such as OCP High Performance Compute (HPC) are highly important to the future roadmaps, however are not reflected in the OSAI work items for clarity and brevity.

5.3. Identified Gaps

The items highlighted in blue on the tracking worksheet are not yet instantiated efforts.

The gap in data center IT infrastructure for AI systems is a key concern as organizations increasingly adopt artificial intelligence (AI) for various applications. While data centers traditionally support general IT operations, AI systems have distinct requirements that often reveal gaps in infrastructure capabilities. Below are some key gaps identified in the context of AI system integration in data centers:

1. Insufficient Processing Power: Traditional data center infrastructures are typically optimized for handling workloads like web hosting, storage, and transactional systems. Nevertheless, AI applications, especially ML and DL, need a lot more power because they frequently deal with huge data sets and intricate calculations[1]. AI workloads demand high-performance hardware, such as Graphics Processing Units (GPUs) or specialized hardware like Tensor Processing Units (TPUs). These components are not commonly part of traditional IT infrastructure, which generally relies on CPUs and standard servers.

2. Storage and Data Management: AI models often require vast amounts of data for training and inferencing, which means data centers must accommodate high-performance storage solutions. Traditional storage systems, typically designed for transactional or relational data,

are not optimized for the massive volumes and high-throughput demands of AI workloads. Hence, data centers need to adopt distributed storage architectures and high-speed storage networks to efficiently store and access datasets, as well as handle the continuous flow of data required for AI tasks.

3. Networking and Bandwidth Limitations: AI systems often require fast, low-latency networking to process real-time data for applications like autonomous driving, industrial automation, or smart cities. Traditional data centers may lack the required networking bandwidth and low-latency connectivity to support these advanced AI applications[3]. Hence, high-speed networks, such as fiber-optic or 5G connectivity, are often necessary to support AI's real-time data requirements and enable data centers to meet performance demands. Also, Domain Specific Network configurability (is an essential option vs generic switched networks, for the most efficient implementations of a given application. Optical circuit switches will ultimately provide this domain specific reconfigurability, but manual configurations will be a cost effective solution to compute that is built for a targeted application 100% of the time.

4. Power and Cooling Requirements: The traditional IT infrastructure is often not designed to handle the heat dissipation and power consumption needs of AI hardware, leading to potential issues with thermal management and the need for advanced cooling solutions[4]. AI systems, particularly those involving deep learning, demand substantial amounts of electrical power and efficient cooling systems due to the high energy consumption of GPUs and specialized hardware.

5. Scalability and Flexibility : Traditional data centers are often not designed with the scalability required for AI systems, which need the flexibility to rapidly expand or contract based on the demands of processing massive datasets and running complex models. AI systems require flexible and scalable infrastructure that can dynamically adapt to changes in processing power and storage needs, which traditional data centers may not be able to efficiently support.

The gap in data center IT infrastructure for AI systems is significant and necessitates improvements in a number of critical areas, such as scalability, power and cooling, networking, storage, and processing capacity. Traditional data centers must be upgraded or supplanted with

specialized infrastructures that can accommodate the distinctive requirements of AI workloads in order to provide comprehensive support for AI applications.

6. System Management

6.1. Overview

Systems Management plays a critical role in ensuring the smooth operation of both IT and physical infrastructure within data centers. As AI workloads scale, the complexity of managing these systems increases, requiring advanced automation, monitoring, and orchestration capabilities. The scope of effort in Open Systems for AI focuses on creating open and interoperable management frameworks that enable efficient control over compute, storage, networking, and power resources. By standardizing telemetry interfaces and establishing common protocols such as Redfish for dynamic infrastructure management, the goal is to support autonomous operation across both IT and physical environments, ensuring that AI workloads can be dynamically optimized and adjusted to changing conditions. These innovations help to reduce manual intervention, improve system uptime, and increase the overall efficiency of AI-centric data centers.

6.2. Relationship to IT and Physical Infrastructure

Figure 26 shows the relationship of system management to IT infrastructure and physical infrastructure. The two infrastructures form the axes with each axis representing the level of automation within each infrastructure. The origin would be a status datacenter in which one manually alters the as-built datacenter (hot aisle/cold aisle).

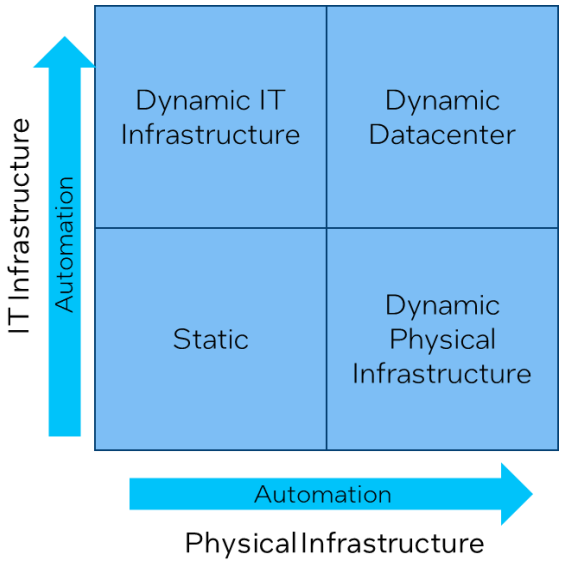


Figure 26: Relation between Physical Infrastructure and IT Infrastructure

Along the IT Infrastructure axis, automation has increased as the platforms (compute, storage, network) have exposed interfaces for remote manageability.

6.3. Manageability Interface and Standards

These manageability interfaces can be in-band (require the workload processor is running) or out-of-band (executing on dedicated hardware available regardless of the state of the workload processor(s). Standards for the out-of-band manageability interface existed for decades. Today, out-of-band manageability of compute and storage is performed via Redfish/Swordfish. Platforms still exist which expose older manageability standards (e.g. IPMI, SNMP). Network platforms are managed via the YANG data model, in the form of RESTCONF or NETCONF. OCP has specified Redfish as the manageability interface in this domain.

Along the Physical infrastructure axis, automation is less mature. The industry is interested in a standard for enhanced (automated) management using emerging DCIM system capabilities. Redfish has been proposed.

In the IT Infrastructure and Physical Infrastructure quadrants, the orchestrator collects telemetry and performs action within their domain in order to perform autonomous action. The autonomous policy in one quadrant may use telemetry from the other quadrant, but it cannot initiate action in the other quadrant. The orchestrator in the dynamic datacenter quadrant collects telemetry from both the IT and physical infrastructure and can perform actions in each. With this scope, the orchestrator can pursue the optimal path to fulfill a policy or achieve a goal.

6.4. OCP and Redfish Profiles

OCP has specified Redfish as the manageability interface for OCP platforms and have posted OCP Profiles on the Profiles repository. An OCP profile prescribes the required manageability on an OCP platform.

The OCP baseline profile prescribes the common manageability across OCP platforms. The OCP server profile prescribes the manageability specific to a server platform. The OCP rack manager profile prescribes the interface for a rack manager which abstracts the complexities of managing a group of platforms within a rack infrastructure.

There are platforms that do not have BMCs. In order for these platforms to conform to the baseline profile, the out-of-band aspect may be relaxed. For example, on a switch, a Redfish software agent can execute on top of the NOS. This has been done on a couple of product contributions to OCP. The code for the Redfish software agent is in the OCP repository See OCP Hardware Management on this link as a starting point.

The other profiles shown in Figure 27 are being developed by the subject matter experts in the other OCP projects. The profiles are currently in draft state and welcome review and feedback. The Sustainability Project is specifying the prescription in sustainability profile (calculating the new-PUE). The Cooling Environment Project's HW Management for Liquid Cooling workstream prescribes the CDU management profile and RearDoorHX management profile. The closed GPU Management CLA-workstream recently contributed their GPU management profile to OCP.

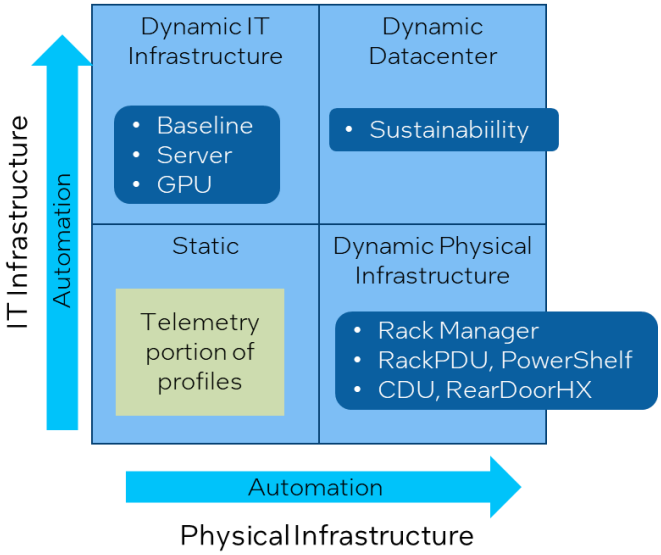


Figure 27: Manageability Profiles across Physical and IT Infrastructures

6.5. Work Items

Table 3 shows a snapshot of the current work items within the management of a system for AI. The work items include those within the Hardware Management Project:

"Test and Validation" would be considered as a subsection in each domain..

Systems Management	Hardware Management	RAS (Silent Data Corruption)
		RAS (GPU & Accelerator RAS Requirements)
		GPU HW Management (GPU & Accelerator Management Interfaces)
		GPU HW Management (Firmware Update Specification)
		Rack Management - Composable Disaggregated Infrastructure
		Rack Management - PDU
		Liquid Cooling Management
		Standardize Hardware Abstraction layers
		Fleet Management
		GPU HW Management Stack (Tooling & Telemetry)
		AI Driven Automated Management (Digital Twins)
		AI System Orchestration (Alliance with Linux Foundation)
		~~Standarized Hardware Abstraction layers (OAI)~~
	Test and Validation	AI Hardware System Benchmarking

Table 3: OCP Open Systems for AI Work Items: System Management

Current work and contributions

The following items have a work effort associated with it. The status, plans and deliverables for the work effort are described.

The Server Component Resilience Specification v1.0 was contributed in 2023 by a CLA-group. The scope of the document includes CPUs, GPUs, and other hardware accelerators. The CLA-group expects to continue work on additional revisions.

The GPU & Accelerator Management Interfaces Specification v0.9.1 and a GPU Profile v0.9.1 was contributed in Q2 2024 by a CLA-group. The v1.0 will be contributed in Q3 2024.

The GPU & Accelerator RAS Requirements v0.9.1 was contributed in Q2 2024 by a CLA-group.

The GPU Firmware Update Specification v0.9.1 was contributed in Q2 2024 by a CLA-group.

The Rack Manager Interface Usage Guide v1.1 and Rack Manager Profile v1.1 are being updated to add support the composition and telemetry to support the Composable Disaggregated Infrastructure (an SC Initiative)

6.6. Identified Gaps

The following items enumerate gaps in system management that have been identified. The description includes an explanation of the gap and required deliverable.

GPU HW Management Stack (Tooling & Telemetry) - GPUs are expected to generate a plethora of telemetry that needs to be exported out of the system for external analysis. The amount of telemetry data needs to be quantified and a mechanism proposed to export and consume the telemetry data.

Standardized Hardware Abstraction Layers - Software layers to abstract specific components from AI technical stack such that developers may have heterogeneous acceleration components for AI without developing bespoke interfaces for each (like system device drivers)- this will grow the ecosystem. Collaborative with OCP Open Platform Firmware and OCP server group. High interest in standardizing the OPF/HW MGt aspects such as firmware etc…

Open Accelerator Interface abstraction: It may be an extension of RAS definitions that have already been defined or are in the process of definition. This underlying capability is the potential to define an accelerator hardware independent API such that developers only need to focus on their model implementation and not the underlying primitives of things like collective communication to hardware or even adding a new mode of operation: heterogeneous acceleration. This would extend and make common manageability aspects of accelerators.

AI Driven Automated Management (Digital Twins) - Digital Twins to better manage the DC both for deployment and operations. (Session held at OCP Global Summit 2024, see the presentation here.)

AI System Orchestration (Alliance with LF) - Integration of FOSS orchestration layers with OCP hardware. FOSS = Free and Open Source.

6.7. Scoping System Management

The definition of management of the system is driven by the definition of the system. For this document, the **system** may span multiple-racks (i.e. AI cluster).

The **system management hierarch**y will include the cluster manager, the rack manager, and the node/platform manager. The node manager will be an out-of-band entity, independent of the component used to handle the workload (processors, memory). Hence the entire system management hierarchy will be out-of-band. What is out-of-scope is in-band management and managing application and workloads.

The entities which are presently managed are:

- Physical or virtual computer system
- Trays (chassis level) of components: GPUs, CPUs, memory, storage
- Interconnect fabrics and interfaces between components and networks
- Power distribution unit, batteries, etc.
- Cooling infrastructure (CDU, TCS, fans, liquids, et al)

Figure 25 shows the out of band management entities.

- The DC manager manages one or more cluster or rack managers
- The cluster manager manages one or more racks
- The rack manager manages one or more nodes
- The node manager manages a node or platform

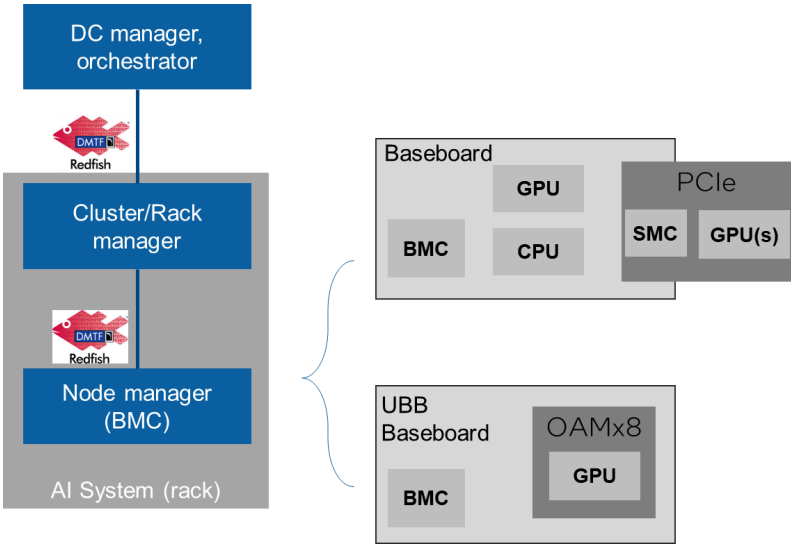


Figure 28: Out of Band Management

The node or platform exposes a Redfish interface for manageability. At a minimum, the interface shall conform to the OCP Baseline Hardware Management profile. For platforms with AI processors the interface shall conform to the GPU Management profile (draft). The node/platform runs a Redfish service which exposes the Redfish interface. The Redfish service can execute on the BMC or as a software agent.

The node/platform may or may not have a BMC (baseboard management controller). The BMC is a dedicated out-of-band management entity with its own memory and storage powered by trickle power. So its state is independent of the in-band entities. On such platforms, the BMC hosts the Redfish service. An open source implementation of the firmware which runs on the BMC is available in the OpenBMC repository.

If the BMC exists on a hardware management module (ala DC-SCM or RunBMC), the Redfish interface shall conform to the OCP Hardware Management Module profile. If the baseboard is a DC-MHS modular baseboard, the Redfish interface shall conform to the OCP Modular Baseboard Management profile.

On platforms without a BMC, the Redfish service can be a software agent. If the software agent runs on in-band entities, the mechanism is not strictly out of band. An open source

implementation of such a software agent is available on the OCP repository. The interface conforms to the OCP Baseline HW Management profile.

The cluster/rack manager manages one or more nodes. The rack manager can exist as a standalone platform, or with a switch. For the rack manager, the nodes are semantically in the same rack. The cluster manager implies no such physical constraints on the nodes. For managing racks, OCP has specified the OCP Rack Management profile. An open source implementation of this Redfish service is available on an OCP repository (in GoLang).

The DC manager manages the entire data center. There may be other managers in the management hierarchy. For this document, the DC manager or orchestrator means the management hierarchy above the cluster/rack manager.

6.8. Manageability domains

A system goes through many phases during its lifecycle. Each phase can be managed. The section below scopes the phases which will be managed for the effort.

- Design [not included]
- Deploy
    - Compose
        - From disaggregated components
    - Provision
        - Update firmware(s)
        - Install OS
    - Config
        - IP addresses
        - Boot path
- Operate
    - Monitor (power, temperature, cooling, status, logs)
    - Remediate (policy)
- Repair (locate, diagnose, isolate) [not included]
- Retire (retrieve information, purge) [not included]

Conclusion

The Open Compute Project's Strategic Initiative for Open Systems for AI is a decisive response to the rapidly evolving demands of artificial intelligence in modern data centers. Throughout this white paper, we have introduced a vision for a framework that coordinates all of the OCP's 12 projects within three interdependent domains—Data Center IT Infrastructure, Data Center Physical Infrastructure, and System Management—each crucial for delivering open, efficient, and scalable AI deployments.

Our analysis underscores a striking reality: AI workloads are growing so quickly that they now demand a fourfold increase in compute power each year. This surge compels a fundamental rethink of system design and implementation. The OCP initiative addresses this challenge through several transformative strategies:

- Standardized rack architectures supporting power envelopes of 250 kW to 1 MW
- Advanced cooling solutions (e.g., liquid cooling) for high-density nodes
- High-voltage, high-efficiency power delivery systems
- Support for multiple, evolving scale-up and scale-out fabrics for performance
- Full stack development to enable a universal and truly open system for AI
- Comprehensive management frameworks for near-autonomous operations

We aim to set clear requirements for compute trays, switch trays, power distribution, networking, and management interfaces. These standards enable a long term roadmap for vendor innovation and maintain interoperability. Open standards and modular design allow quick adaptation to new AI needs.

The foundations here drive next-generation products and services. The OCP model invites contributions from component designers, silicon designers, system integrators, and data center operators. It cuts wasted investment and builds a pipeline from idea to solution.

Success will be measured by technical gains and the democratization of AI. Open systems will empower organizations from hyperscalers to small businesses.

The OCP Open Systems for AI Initiative will evolve with new technologies and market changes. It preserves community-driven project development and supports systems-level collaboration. Open source AI models show the power of a community. We build a future where interoperable systems are the norm, driving industry change.

Next, the OCP will unite its community for an architectural design process. This design will synchronize work items and contributions to produce final platform specifications for universal, performant, and efficient Open Systems for AI.

OCP Glossary

A

A (Ampere): The SI unit of electric current.

ACH (Air Changes per Hour): A measure of the rate at which the air in a space is replaced. It is defined as the volume of air added to or removed from a space per hour divided by the total volume of that space.

Active solar heating system: A system that uses solar energy to heat a fluid (either liquid or air) and then transfers that heat directly into a building or stores it for later use.

AFR (Annual Failure Rate): The estimated percentage of failures a device or system is expected to experience over the course of one year.

AI (Artificial Intelligence): A field of computer science devoted to creating systems that can perform tasks normally requiring human intelligence—such as understanding language, recognizing images, reasoning, learning, and making decisions.

AI Cluster: A collection of tightly interconnected compute nodes (often containing GPUs or other accelerators) that work together to train or serve AI models. Such clusters may range in power from a few kilowatts up to hundreds of kilowatts per rack.

AI/ML Workloads: Tasks or applications involving the training of machine learning (ML) models or performing inference with them (including deep learning, natural language processing, etc.).

Air Conditioner: A self-contained unit designed to cool, circulate, filter, dehumidify, or otherwise process air.

Airflow Management: The process of directing and controlling the movement of air within an IT system or data center to efficiently remove heat from electronic components.

ASTM (American Society for Testing and Materials): Now known as ASTM International, this is a globally recognized organization that develops and delivers international voluntary consensus standards.

Author (OCP contribution context): An individual who primarily originates the content of a document or contribution, including core ideas, concepts, designs, and text.

Auto Ignition Point: The lowest temperature at which a substance will spontaneously ignite in a normal atmosphere without an external ignition source.

B

BBU (Battery Backup Unit): A component that provides short-term power during a utility power failure. In many high-density AI data centers, batteries that were once rack mounted are increasingly placed in centralized or sidecar racks.

BEC / Boiling Enhancement Coating (BEC): A microstructured surface coating designed to improve heat transfer by enhancing boiling characteristics.

Bill of Materials (BOM): A comprehensive list detailing all raw materials, components, assemblies, and quantities needed to manufacture a product.

Building (Data Center context): A physical structure designed to house IT equipment and/or the associated power and cooling infrastructure.

Bureau International des Poids et Mesures: The French organization responsible for maintaining the International System of Units (SI).

Burst pressure: The minimum internal fluid pressure at which a component (such as a fluid coupling) will fail catastrophically.

Buyer: The individual or organization that purchases a product. (Note: In some contexts, the Buyer may differ from the eventual Owner, which can affect warranty and support issues.)

C

Cable Routing: The planning and organization of cable pathways within an IT system or data center to ensure efficient operation and ease of maintenance.

Capillary Action: The ability of a liquid to flow in narrow spaces against gravity due to adhesive and cohesive forces.

Carbon Usage Effectiveness (CUE): A metric that quantifies the amount of carbon emissions produced per unit of computing energy consumed by a data center.

CDU (Coolant Distribution Unit): A specialized system that conditions and distributes a liquid coolant (such as water or a dielectric fluid) from a facility's central cooling system to individual racks or IT equipment.

Centistokes (cSt): A unit of kinematic viscosity equal to one square millimeter per second (mm^2/s).

Central Processing Unit (CPU): The primary electronic circuitry within a computer that performs the mathematical calculations and logical operations.

CFC (Chlorofluorocarbons): Compounds targeted by environmental regulations (EPA, UN) because of their ozone-depleting properties.

Chassis (Tray): A physical enclosure or frame that houses components (CPUs, GPUs, networking, power supplies, etc.) in a server or other IT equipment. (Often “chassis” may also be called a “tray” in certain designs.)

Chemical stability: The ability of a chemical system or coolant to resist decomposition or corrosion under operating conditions.

Closed-loop System: A cooling system that continuously circulates a coolant within a sealed circuit with no direct exchange with the external environment.

CMH (Cubic Meter Per Hour): A measure of volumetric flow rate, indicating how many cubic meters of air or fluid pass a given point each hour.

Cogeneration (Combined Heat and Power – CHP): The simultaneous production of electricity (or mechanical power) and useful thermal energy from a single fuel source.

Cold Plate: A thermally conductive plate (often copper) that directly contacts a high-power component (such as a CPU or GPU) to absorb heat and transfer it to a circulating coolant.

Combined heat and power (CHP): See “Cogeneration.”

Computer Room Air Conditioning (CRAC): A cooling device—typically large and self-contained—that regulates temperature, humidity, and air distribution in data centers or server rooms.

Computer Room Air Handler (CRAH): Similar to a CRAC unit but typically using chilled water to remove heat from the air instead of a built-in compressor.

Conduction: The transfer of heat through a material or between materials in direct contact.

Containment strategy: Measures implemented to prevent the leakage or spillage of liquid (such as coolant) into the data center environment.

Contaminant: Any foreign substance or particle present in an environment or material that may harm performance or cause damage.

Contributor (OCP contribution context): An individual who provides input or content to a document or project but is not the primary author.

Convection: The transfer of heat by the movement of a fluid (air or liquid) caused by temperature differences.

Coolant: A fluid used to absorb and transport heat away from electronic components.

Cooling Capacity: The ability of a cooling system to remove a specified amount of heat from a

Cooling Tower: A device that uses the evaporation of water to remove heat from a cooling system and release it to the environment.

Coordinated Universal Time (UTC): The primary time standard by which the world regulates clocks and time.

Counterflow: A design in heat exchangers where two fluids flow in opposite directions to maximize the temperature gradient and heat transfer.

Crosstalk: Unwanted interference between adjacent electrical signals, which can degrade signal quality.

Crude- or natural gas-derived hydrocarbon: Organic compounds, primarily composed of hydrogen and carbon, extracted from crude oil or natural gas.

Cubic Meter: The SI unit of volume defined as the space occupied by a cube with one-meter sides.

D

D2C or DTC (Direct-to-Chip): A cooling method in which liquid coolant is delivered directly to a chip (such as a CPU or memory module) to remove heat efficiently.

Data Center (DC): A facility used to house servers, storage, networking equipment, and supporting infrastructure (power, cooling, etc.) for IT operations.

DBT (Dry-Bulb Temperature): The temperature of air measured by a freely exposed thermometer that is shielded from radiation and moisture—essentially the ambient temperature.

Deionized Water: Water that has had most or all of its mineral ions removed, often used in liquid cooling applications.

DHx (Door Heat Exchanger): A specialized rear door for IT cabinets that incorporates a heat exchanger to assist in cooling the equipment within the rack.

Dielectric Constant (Dk): A measure of a material's ability to store electrical energy in an electric field.

Dielectric Liquid: A liquid that not only cools electronic components but also provides electrical insulation.

Dielectric Strength: The maximum electric field that a material can withstand without breaking down electrically.

Differential Impedance: The difference in electrical impedance between two conductors, important for characterizing high-speed digital signal quality.

DIMM (Dual Inline Memory Module): A type of computer memory module with separate electrical contacts on each side for connection to a memory bus.

Direct Liquid Cooling (DLC): A method in which a liquid coolant is brought into direct contact with heat-generating components to remove heat efficiently.

Direct-to-Chip (D2C or DTC) Cooling: See "D2C or DTC."

Dripless connectors: Pluggable liquid connectors designed to prevent drips and ensure long-term reliability in liquid cooling systems.

Dry Coolers: Heat exchangers that use air (without direct contact with the liquid coolant) to remove heat from a circulating fluid.

Dual Inline Memory Module (DIMM): See "DIMM."

Dynamic viscosity: A measure of a fluid's resistance to flow (often expressed in pascal-seconds or centipoise).

E

ECS (Environmental Control System): A system (or set of guidelines) used to monitor and regulate the environmental conditions (e.g., temperature, humidity) within a facility such as a data center.

Electrochemical potential: A thermodynamic measure that includes both the chemical and electrical potential energies, used to assess a material's reactivity.

Enclosed chassis: A sealed server chassis design (often used in immersion cooling) where the coolant circulates within a closed enclosure.

End of Life (EOL) (Manufacturer/User): The point at which a manufacturer ceases support (updates, parts, etc.) for a product (Manufacturer EOL) or when a user decommissions a product (User EOL).

Energy Reuse Effectiveness (ERE): A metric that measures the fraction of energy from a data center that is reused (for example, to heat nearby buildings) as opposed to being wasted.

Energy Reuse Factor (ERF): The ratio of energy reused in other applications to the total energy consumed by the data center.

Environment, Social, and Governance (ESG): A framework for evaluating an organization's ethical impact and sustainability practices—including environmental stewardship, social responsibility, and governance.

Environmental Protection Agency (EPA): A U.S. federal agency responsible for protecting human health and the environment by creating and enforcing regulations based on laws passed by Congress.

Esters: Organic compounds (often derived from plant oils or synthesized) that can be used in certain cooling fluids.

Eutectic: A mixture of substances that has the lowest melting point for any combination of those substances, sometimes used in phase change cooling applications.

Evaporation temperature: The temperature at which a liquid begins to vaporize at its surface.

Evaporative Cooling: A cooling method that uses the natural process of liquid evaporation to remove heat from a surface.

Examination: A post-sales inspection process (often for warrantied products) to assess whether returned products meet the conditions for warranty claims.

Express warranty: A written guarantee provided by a seller or manufacturer about the quality, performance, or other characteristics of a product.

Extended warranty: An optional service contract offering additional warranty coverage beyond

Eye Diagram: A graphical tool used in high-speed digital communications to evaluate signal quality.

F

Facility Water System (FWS) [or FWS loop]: A closed-loop water (or other liquid) circuit that transports heat from data center equipment to a central cooling source (such as chillers) and back again.

Fan Control: The process of monitoring and adjusting the operation of fans to optimize airflow and cooling performance.

Fan Emulator: A device that replicates the electrical signals of a functioning fan, often used for testing or when actual fan operation is not possible.

Fire point: The lowest temperature at which a fuel's vapor will continue to burn for at least 5 seconds after ignition.

Firmware: Software permanently programmed into hardware (such as a server's motherboard or BMC) to control core functions.

Flash Point: The lowest temperature at which the vapors of a liquid can ignite when exposed to an ignition source, under standard test conditions.

Flow rate: The volume of fluid passing through a given point per unit of time.

Fluid Dynamics: The branch of physics that studies the behavior and movement of fluids (liquids and gases) and the forces acting upon them.

Fluid Solution Vendors: Companies that develop and supply liquids (often dielectric fluids) used in advanced cooling systems.

Fluorocarbon: A family of compounds composed of carbon and fluorine used in various applications including refrigerants and dielectric fluids.

Fluorochemicals: Chemical compounds that contain fluorine; in cooling they are notable for their thermal stability and insulating properties.

Force Majeure: A contract clause that relieves parties of liability in the event of unforeseeable catastrophic events (e.g., natural disasters).

Forced Convection: Heat transfer in a fluid that is actively moved by external means (like a fan or pump) rather than by natural buoyancy effects.

G

Galvanic Corrosion: An electrochemical process in which one metal corrodes preferentially when electrically connected to another metal in the presence of an electrolyte.

GHGs (Greenhouse gases): Gases in the atmosphere that trap heat, contributing to the greenhouse effect and global warming.

GHS (Globally Harmonized Systems of Classification and Labelling of Chemicals): An internationally agreed system for standardizing chemical classification and labeling.

Gigawatt-hour (GWh): A unit of energy equal to one billion watt-hours.

Global Warming Potential (GWP): A measure of how much heat a greenhouse gas traps in the atmosphere relative to carbon dioxide.

GPU (Graphic/Graphics Processing Unit): A specialized processor designed to handle parallel computations, commonly used for rendering graphics and accelerating AI or HPC workloads.

H

Heat Exchanger: A device that transfers heat between two or more fluids without mixing them, as used in many cooling systems.

Heat Load: The total amount of heat energy produced by equipment that must be removed by a cooling system to maintain safe operating temperatures.

Heat Pipe: A sealed heat-transfer device that uses the evaporation and condensation of a working fluid to transfer heat efficiently from one location to another.

Heat Rejection Unit (HRU): A system or device designed to expel heat from a cooling circuit to the environment.

Heat Sink: A component (often made of metal) designed to absorb and dissipate heat from electronic components by increasing surface area contact with the cooling medium.

Heat Spreader: A device made of highly conductive material (like copper) used to evenly distribute heat over a larger area.

Heat Transfer Coefficient: A parameter that quantifies the effectiveness of a material or fluid in transferring heat.

Heat Transfer Effectiveness: A measure of how efficiently a cooling system transfers thermal energy from a heat source to a heat sink.

HFCs: Hydrofluorocarbons used as refrigerants in air conditioning and refrigeration systems.

High-Speed I/O Interfaces: Interfaces (such as PCIe or high-speed Ethernet) designed to support rapid data transfer in computing systems.

Hot-Swap Controller (HSC): A device that enables components to be replaced or added to a live system without shutting it down.

HPC (High-Performance Computing): The use of supercomputers and parallel processing techniques for solving complex computational problems.

HVAC: Heating, Ventilation, and Air Conditioning systems that control the indoor environment.

Hydrocarbons: Organic compounds consisting primarily of hydrogen and carbon, such as those derived from crude oil or natural gas.

Hydrophilic: A property of a material that attracts water, facilitating its spread or absorption.

Hydrophobic: A property of a material that repels water.

I

IC (Integrated Circuit): A set of electronic circuits on one small flat piece (chip) of semiconductor material.

IEC (International Electrotechnical Commission): An international standards organization that prepares and publishes standards for electrical, electronic, and related technologies.

Immersion Cooling: A cooling technique in which servers or electronic components are fully (or partially) submerged in a dielectric fluid that absorbs and removes heat. (Variants include single-phase—where the fluid remains liquid—and two-phase—where the fluid undergoes phase change.)

Indium Foil: A soft, thermally conductive metal foil used as a thermal interface material in some immersion cooling applications.

Information Technology (IT): The use of computers, networks, software, and other electronic systems to store, process, transmit, and retrieve information.

Information Technology Equipment (ITE): The physical hardware (servers, storage, networking devices, etc.) used in IT systems.

Ingress Protection (IP): A rating system (defined by IEC 60529) that specifies the degree of protection provided by an enclosure against intrusion of solid objects, dust, and water.

Ingress Protection (IP) Rating: The specific numerical rating (such as IP65) that describes the sealing effectiveness of an electrical enclosure.

Insulation: Materials or techniques used to reduce the transfer of heat or electricity between objects or environments.

ISO (International Organization for Standardization): An independent, non-governmental international organization that develops and publishes standards for a wide range of industries.

J

Joule Heating: The process by which the passage of an electric current through a conductor produces heat.

K

K (Kelvin): The SI unit of thermodynamic temperature.

Key performance indicator (KPI): A measurable value that indicates how effectively an organization is achieving its key objectives.

kg (Kilogram): The SI unit of mass.

Kilopascal (kPa): A unit of pressure equal to 1,000 pascals.

Kilowatt (kW): A measure of power equal to 1,000 watts.

Kilowatt-hour (kWh): A unit of energy equal to the consumption of 1,000 watts for one hour.

L

L (Liters): The SI unit of volume for liquids.

L-A (Liquid-to-Air Heat Exchanger): A device that transfers heat from a liquid coolant to the surrounding air.

L-L (Liquid-to-Liquid Exchanger): A device that transfers heat between two liquid streams without mixing the fluids.

L/m (Liters per minute): A unit measuring the volumetric flow rate of a liquid.

Laminar Flow: A smooth, orderly flow regime in which layers of fluid move parallel to each other with minimal mixing.

Latent Heat: The heat absorbed or released during a phase change (e.g., from liquid to gas) without changing temperature.

Latent Heat of Vaporization: The specific amount of heat required to convert a liquid into a

Liquid: A state of matter that has a fixed volume but adapts to the shape of its container.

Liquid Cooling: A cooling strategy that uses a liquid (water, dielectric fluid, etc.) to remove heat from components more efficiently than air cooling.

M

m (Meter): The SI unit of length.

Maintenance Window: A scheduled period during which systems are taken offline for maintenance, updates, or repairs.

Manufacturer: The company or entity that produces a product.

Manufacturing bill of materials (MBOM): A detailed list of parts, assemblies, and materials (including packaging and accounting for production waste) required to produce a sellable product.

Material Compatibility: The ability of different materials to coexist or function together without adverse reactions or degradation.

Material Safety Data Sheet (MSDS): A document that provides safety, handling, and hazard information about chemicals used in products or processes.

Megawatt (MW): A unit of power equal to one million watts.

Megawatt-hour (MWh): A unit of energy equal to one million watt-hours.

Microchannel Heat Sink: A heat sink design featuring tiny channels that increase the surface area in contact with a coolant for improved heat transfer.

Microchannels: Small channels or passages within a heat exchanger or cold plate that enhance heat transfer by increasing the contact surface area.

Mineral Oil: A petroleum-derived dielectric fluid used in some immersion cooling applications.

Mist: Tiny droplets of liquid suspended in air.

Motherboard: The main printed circuit board (PCB) in a computer that connects and allows communication between key components.

MTCO2e: Metric Tons of carbon dioxide equivalent—a unit for comparing the global warming potential of different greenhouse gases.

N

Nanofluids: Engineered fluids that contain suspended nanoparticles to enhance thermal conductivity and improve heat transfer.

National Fire Protection Association (NFPA): A global non-profit organization dedicated to reducing the risk of fire, electrical, and related hazards.

NCG (Non-Condensable Gas): A gas that does not condense under conditions present in a cooling system and that may reduce heat transfer efficiency.

NEMA (National Electrical Manufacturers Association): An association that develops standards for electrical equipment and systems.

Network Interface Card (NIC): A hardware component that provides a physical connection between a computer and a network.

Network Switch: A device that connects multiple network devices and uses packet switching to forward data to its destination.

NGO (Non-Governmental Organization): A nonprofit organization that operates independently of any government.

O

OBI (Open Bath Immersion): A type of immersion cooling in which multiple pieces of IT equipment are submerged together in an open, shared liquid bath.

OAM (Open Accelerator Module): A standard accelerator card form factor defined under the Open Accelerator Infrastructure initiative to enable interoperability across vendors.

OAI (Open Accelerator Infrastructure): An initiative to establish standard interfaces for accelerators (e.g., GPU cards) that promote multi-vendor interoperability in high-density systems.

OCP (Open Compute Project): A global community and foundation dedicated to designing and sharing open-source hardware and data center designs that emphasize efficiency, scalability, sustainability, and multi-vendor interoperability.

OCP Tenets: The core principles guiding OCP designs: Efficiency, Impact, Openness, Scale, and Sustainability.

Open Rack v3 (ORv3): A rack specification (21-inch width and standardized power bus bars) designed to support high-density, liquid-cooled, and high-power systems.

Open Systems for AI (OSAI): An initiative that coordinates efforts across IT infrastructure, physical infrastructure, and system management to support the rapid growth of AI.

ODM (Original Design Manufacturer): A company that designs and manufactures a product that is later marketed under another company's brand.

OEM (Original Equipment Manufacturer): A company that produces products or parts that are used in another company's final product and may be rebranded.

Open Bath (OB): An immersion cooling configuration in which multiple pieces of IT equipment are submerged together in an open liquid bath.

Open-loop System: A cooling system in which the coolant is exposed to the environment during heat exchange, as opposed to being in a sealed loop.

Operating Humidity Range: The acceptable range of humidity levels in which equipment can function without performance degradation.

Operating pressure: The pressure at which a system or component is designed to operate under normal conditions.

Operating Temperature Range: The safe temperature range within which a device or system can operate reliably.

Overclocking: The practice of running a processor or component at a higher speed than its rated specification, often increasing heat output.

Owner: The entity (individual or organization) that legally possesses a product. (May differ from the original buyer.)

Oxidation Stability: The resistance of a fluid or material to degradation through oxidation.

P

P-T (Pressure-Temperature): A notation or relationship describing how the pressure and temperature of a fluid relate to its phase and behavior.

PAO (Polyalphaolefins): Synthetic hydrocarbons commonly used as base fluids in lubricants and sometimes in cooling applications.

PAO x (Polyalphaolefins variant): A designation indicating a specific nominal viscosity (in centistokes at 100°C) for a PAO fluid.

Parallel Flow: A flow arrangement where multiple fluid streams flow in the same direction to enhance heat transfer in a system.

Partial Immersion: A cooling technique in which only part of the electronic components is submerged in a dielectric fluid.

Partial vacuum: A cooling system condition where the operating pressure is maintained below ambient pressure to, for example, help contain leaks.

Passively Cooled: A cooling method that relies on natural convection or radiation without active components like fans or pumps.

PCB (Printed Circuit Board): A flat board used to mechanically support and electrically connect electronic components.

PCBA (Printed Circuit Board Assembly): A PCB that has been populated with electronic components through assembly processes such as soldering.

PCM (Phase Change Material): A substance that absorbs or releases thermal energy during a phase transition (e.g., melting/freezing) used for thermal management.

PDU (Power Distribution Unit): A device (often located at the rack or row level) that distributes electrical power to IT equipment. In high-power systems, PDUs may be supplemented by high-voltage bus bars or dedicated sidecar racks.

PEEK (Polyether Ether Ketone): A high-performance polymer noted for its strength and thermal stability.

Peltier Effect: A thermoelectric phenomenon where applying a voltage across two dissimilar materials creates a temperature difference, used in solid-state cooling.

Perfluorocarbons (PFCs): Compounds made up of carbon and fluorine that serve as dielectric coolants in immersion cooling, though they often have a high global warming potential.

Personal Protective Equipment (PPE): Clothing or gear (such as helmets, gloves, or goggles) designed to protect individuals from injury or infection.

PFAS: Per- and polyfluoroalkyl substances—a group of highly stable chemicals used in various industrial applications including some cooling fluids.

Pour point: The lowest temperature at which a liquid coolant will still flow.

Power factor: The ratio of real power (in kilowatts) to apparent power (in kilovolt-amperes) in an electrical system.

Power Supply Unit (PSU): A device that converts AC mains power into regulated DC power levels required by computer components.

Power usage effectiveness (PUE): A data center efficiency metric defined as the total facility power divided by the power used by IT equipment.

Pressure Drop: The loss of pressure in a fluid as it flows through components (such as pipes or heat exchangers) due to friction and other factors.

Pressure per square inch (psi): A non-SI unit for measuring pressure.

Pressure-Temperature (P-T) Relationship: The correlation between the pressure and temperature of a fluid, which can influence its phase behavior and heat transfer properties.

Primary Coolant: The fluid that is in direct contact with heat-generating components in a cooling system.

Product: An item manufactured or assembled for sale.

Pump Control: The process of regulating the operation of pumps to maintain the desired coolant flow rate.

Pumped Liquid Cooling: A liquid cooling method where pumps actively circulate the coolant through a closed loop.

Pumped Refrigerant Cooling: A liquid cooling approach that uses a refrigerant (instead of water or dielectric fluid) circulated by pumps for heat removal.

R

Radiator: A heat exchanger designed to dissipate heat from a fluid (typically by convection and radiation) into the surrounding air.

RDHx (Rear Door Heat eXchanger): A door-mounted heat exchanger that is integrated into the back of a computer rack to extract heat from the enclosed equipment.

Redfish: An open, RESTful management standard that uses JSON to facilitate out-of-band management, telemetry, and automation of data center hardware.

Redundancy: The inclusion of extra components or systems to ensure continued operation in the event of a failure.

Refrigeration Cycle: A thermodynamic cycle used in cooling systems to transfer heat from a lower-temperature source to a higher-temperature sink.

Reliability: The ability of a system or component to perform its required functions under stated conditions for a specified period.

Reservoir Pumping Unit (RPU): A module in liquid cooling systems that houses both the coolant reservoir and the pumps used to circulate the fluid.

Retailer: An individual or business that sells products directly to the public.

Return Merchandise Authorization (RMA): A process by which a customer obtains permission to return a product for repair, replacement, or refund under warranty or other agreements.

Reviewer (OCP contribution context): An individual responsible for evaluating and ensuring the quality, clarity, and accuracy of a document or contribution.

Risk: The potential for an undesirable event or outcome to occur.

S

s (Second): The SI unit of time.

Safety Data Sheet (SDS): A document that details the hazards, safe handling, and emergency measures for chemicals used in products or processes.

Sales bill of materials (SBOM): A detailed list of all components, parts, and materials—as well as packaging and warranty information—as they appear in a sales order.

Sales Contract: A legally binding agreement under which a seller agrees to transfer ownership of a product to a buyer for a specified price.

Secondary Coolant: In an indirect cooling system, the fluid that transfers heat between the primary coolant and the heat sink or exchanger.

Seller: The individual or organization that offers products for sale.

Server: A computer, device, or program dedicated to managing network resources and providing services to other computers.

Server Cooling: The process or systems used to maintain safe operating temperatures for servers and their components.

Server Rack: A structure or cabinet used to house servers, switches, and other IT equipment in an organized manner.

Server Room: A dedicated space within a building designed to house servers and associated IT equipment.

SF6 (Sulfur hexafluoride): A colorless, odorless, non-flammable, and potent greenhouse gas used as an electrical insulator.

SI (System Integrator): A company or individual that specializes in assembling subsystems into a complete, functioning system.

SI (Système international d'unités / International System of Units): The globally accepted system of measurement based on decimal multiples.

Single Phase (electric): An electrical power distribution system using a single alternating current (AC) voltage waveform.

Single Phase (fluids): In cooling, a process in which the coolant remains in the same state (usually liquid) throughout circulation.

Single-phase Immersion Cooling (1PIC / SPIC): An immersion cooling technique in which the dielectric fluid remains in the liquid state throughout the cooling process.

SKU (Stock Keeping Unit): A unique alphanumeric code used to identify and manage a product in inventory.

SoW (Statement of Warranties): A document outlining the specific warranty guarantees provided by the seller or manufacturer.

SoW (Statement of Work): A formal document that defines the work activities, deliverables, and timelines required for a specific project or service engagement.

Specific Heat: The amount of heat required to raise the temperature of one kilogram of a substance by one degree Kelvin.

Specific Target Organ Toxicity (STOT): A classification that indicates which organs are most likely to be affected by exposure to a toxic substance.

Spillage: The accidental leakage or overflow of a coolant or other liquid from its containment.

Spray Cooling: A method of cooling that uses a fine spray of liquid to absorb and remove heat directly from components.

Supplier: A company or individual that provides products or components over an extended period.

Sustainability (Data Center context): Practices and initiatives designed to make the construction and operation of data centers more energy efficient and environmentally friendly.

System Operator: An individual or team responsible for the daily operation, maintenance, and monitoring of a system.

T

TC (Technical Committee): A group of experts assembled to address specific technical issues or develop industry standards.

TCE (Total Cost to the Environment): A comprehensive metric that assesses the overall environmental impact of a product or operation throughout its life cycle.

TCO (Total Cost of Ownership): A financial estimate that includes both direct and indirect costs associated with purchasing, operating, and maintaining a product or system.

TCS (Technology Cooling System): A closed liquid circuit designed to transfer heat from an IT system (often via immersion or direct liquid cooling) to the facility water system.

TDP (Thermal Design Power): The maximum amount of heat a chip (such as a CPU or GPU) is expected to generate under normal operating conditions, used to guide cooling system design.

TDS (Technical Data Sheet): A document that provides detailed technical information and specifications for a chemical, material, or product.

Temperature Gradient: The rate of change in temperature over a certain distance, which drives heat transfer.

Terawatt (TW): A unit of power equal to one trillion (10^12) watts.

Terawatt-hour (TWh): A unit of energy equivalent to one terawatt of power sustained for one hour.

TGG (The Green Grid): An industry consortium that promotes energy efficiency in data centers and IT infrastructure.

Thermal Conductivity: A measure of a material's ability to conduct heat.

Thermal Expansion: The tendency of a material to change in volume or shape when heated.

Thermal Grease: A viscous material applied between a heat source (like a CPU) and a heat sink to improve thermal conductivity.

Thermal Interface Material (TIM): A material used to enhance the thermal coupling between two surfaces (e.g., a chip and its cooler).

Thermal Loss to Environment: The amount of heat energy that escapes from a cooling system into the surrounding environment.

Thermal Resistance: A measure of a material's resistance to heat flow.

Thermal Test Vehicle (TTV): A test device that mimics the form factor and heat characteristics of actual electronic components for thermal testing purposes.

Thermocouple: A temperature sensor consisting of two dissimilar metal wires that generate a voltage proportional to a temperature difference.

Thermodynamic Equilibrium: The state in which all macroscopic properties (temperature, pressure, etc.) in a system are uniform and unchanging over time.

Thermoelectric Cooling: A cooling method based on the Peltier effect, using semiconductor devices to create a heat flux between the junctions of two materials.

Thermosyphon: A passive cooling system that circulates a coolant by natural convection without the need for mechanical pumps.

Three-phase power: An AC power system using three separate alternating currents offset in phase, widely used for efficient power distribution.

Top of Rack (ToR): A data center network design in which a dedicated network switch is installed at the top of each server rack for localized connectivity.

TPIC (Two-phase Immersion Cooling): Also known as Two-phase Cooling: an immersion cooling method in which the dielectric fluid undergoes a phase change (liquid to vapor and back) to remove heat.

Turbulent Flow: A fluid flow regime characterized by chaotic, irregular motion that enhances mixing and heat transfer.

Two Phase: In cooling, a process where the coolant undergoes a phase change (typically liquid to gas) during heat removal; in electrical power, it may refer to a two-phase system.

U

UL (Underwriters Laboratories): An independent safety science organization that tests, certifies, and develops standards for products.

UN (United Nations): An international organization composed of member states that works on global issues including peace, security, and sustainable development.

Underclocking: The practice of running a processor or component at a lower speed than its maximum specification to reduce power consumption and heat generation.

Uninterruptible power supply (UPS): A system that provides emergency backup power to IT equipment when the main power source fails.

Universal Quick Disconnect (UQD): A standardized connector that enables fast and secure connection/disconnection of fluid lines in cooling systems.

V

V (Volt): The SI unit of electric potential or electromotive force.

Vapor: The gaseous phase of a substance that is typically liquid or solid under standard conditions.

Vapor Pressure: The pressure exerted by a vapor in equilibrium with its liquid (or solid) phase at a given temperature.

Vapor Quality: The ratio (by mass) of vapor to the total mixture (vapor plus liquid) in a two-phase flow system.

Variable Frequency Drive (VFD): A device that adjusts the speed and torque of an electric motor by varying the frequency of the power supplied.

Viscosity: A measure of a fluid's resistance to flow or deformation.

Volatile organic compound (VOC): Organic chemicals that easily vaporize at room temperature and can contribute to air pollution.

Voltage Regulator (VR): A system that automatically maintains a constant voltage level to electronic equipment.

Volumetric expansion: The increase in volume of a material per unit temperature increase, often expressed as a coefficient of thermal expansion.

W

W (Watt): The SI unit of power, equal to one joule per second.

Warrantied Product: A product that is covered under a warranty provided by the seller or manufacturer.

Warrantor: The entity (individual or company) that issues a warranty assuring that a product will meet specified standards.

Warranty: A written guarantee that a product will be repaired or replaced if it fails to meet certain conditions within a specified time.

Warranty Period: The length of time during which the warrantor is obligated to honor the warranty.

Warranty Statement: The detailed written terms and conditions outlining what is covered under the warranty.

Water Cooling: A cooling method that uses water or a water-based coolant circulated in direct contact with components to remove heat.

Water Usage Effectiveness (WUE): A metric for assessing the efficiency of water use in a data center, defined as the total water usage divided by the energy consumed by IT equipment.

Wetted Surface Area: The area of a component that is in direct contact with a cooling fluid, affecting heat transfer performance.

Workload: The amount or type of computational or processing tasks performed by a system.

World Semiconductor Council (WSC): A global organization composed of leading semiconductor industry associations.

X

xPU(s): A generic term referring to any type of processing unit—including CPUs, GPUs, and specialized accelerators.

This glossary is intended to serve as a single comprehensive reference for terminology related to open systems for AI, data center infrastructure, cooling technologies, and associated standards.

References

a. Books

b. Web resources

i. Blog

[1] Paul Martynek, (2023), Training AI models require an enormous amount of data center storage and power. As the demand for AI escalates, can data centers keep up?, https://builtin.com/artificial-intelligence/ai-data-center-shortages

[2] PreRack IT, (2023). The impact of Artificial Intelligence on storage Infrastructure. https://prerackit.com/the-impact-of-artificial-intelligence-on-storage-infrastructure/

[3] Matt Liebowitz, (2024). Optimize your AI Network to Keep Your AI Flowing. https://www.dell.com/en-us/blog/optimize-your-ai-network-to-keep-your-ai-flowing/

[4] Pavel Durst, (2024). The Emerging Challenges for the Data Center Industry in the Age of AI. https://www.datacenterfrontier.com/sponsored/article/55136906/comap-the-emerging-challenges-for-the-data-center-industry-in-the-age-of-ai

[5] Kaladhar Voruganti, John Duren, (2024). Deploying State-of-the-Art AI Infrastructure at Speed. https://blog.equinix.com/blog/2024/10/24/deploying-state-of-the-art-ai-infrastructure-at-speed/

ii. OCP Whitepapers

[6] Guidelines for Connection of Liquid Cooled ITE to Data Center Facility Systems https://www.opencompute.org/documents/ocp-acs-connection-guidance-acf-wp-05032022-pdf

[7] Data Center Liquid Distribution Guidance & Reference Designs https://www.opencompute.org/documents/ocp-acf-reference-design-guidance-white-paper-r1-pdf

[8] Modular Technology Cooling System for Cloud Scale https://www.opencompute.org/documents/ocp-wp-submittal-modular-tcs-at-cloudscale-design-delivery-selection-guidance-dlm1-docx-pdf

Embedded References

[X1] OCP Advanced Cooling Solutions (ACS) Project – Liquid Cooling Architecture:

- OCP ACS Liquid Cooling Documentation
- Provides guidelines and design considerations for facilities using liquid-cooled racks, including some references to air-assisted setups.

[X2] ASHRAE Technical Committee 9.9 – Liquid Cooling Guidelines for Datacom Equipment Centers (2019):

- https://www.ashrae.org/File%20Library/Technical%20Resources/Bookstore/WhitePaper_TC099-WaterCooledServers.pdf
- Discusses recommended limits for liquid-to-air cooling system capacities and design best practices.

[X3] OCP Immersion Cooling Sub-Project resources::

- Sub-Project homepage: OCP ACS Immersion Cooling Specifications
- Sub-Project wiki with links to individual workstreams and guidance documents: https://www.opencompute.org/wiki/Cooling_Environments/Immersion
- Immersion Cooling Requirements: https://www.opencompute.org/documents/ocp-acs-immersion-requirements-rev-2-1-pdf
- Contains power density recommendations and example deployments of single-phase and two-phase immersion solutions.

[X4] Green Grid & GRC (Green Revolution Cooling) Whitepapers on immersion cooling efficiency:

- Provide field data indicating rack-level capacity ranges and PUE benefits for immersion-cooled systems.

[X5] OCP Immersion Community - approaches to validate & improve immersion performance

- Hardware Design optimization: OCP Design Specification - Targeted Flow Efficiency Heat Sink https://www.opencompute.org/documents/ocp-spec-targeted-flow-efficiency-heat-sink-v1-0-pdf
- Application of support equipment in immersion: OCP Power Distribution workstream, e.g., Power Distribution in Immersion Environments Challenges and Opportunities (OCP Global Summit 2024) https://youtu.be/wHOeahX1rE0
- Fluid validation and design optimization, e.g., Exploring the potential of single phase immersion cooling through quantitative thermal performance testing (OCP Global Summit 2024), https://youtu.be/ho4h1jvZdLc

[X5] OCP HPC Sub-Project – Cold Plate Cooling of High-TDP Systems:

- COLD PLATE COOLING LOOP REQUIREMENTS
- Industry examples of cold-plate solutions handling 15–60 kW per rack, including discussions on two-phase designs.

[X6] Intel & Submer Joint Publications on Two-Phase Liquid Cooling (various whitepapers):

- Detail real-world performance metrics and highlight the increased chip-level cooling efficiency of two-phase methods.

[X7] T. Brunschwiler et al., "Direct Liquid Jet-Impingement Cooling with Integrated Micrometer-Scale Fluid Channels," in IEEE Transactions on Components, Packaging and Manufacturing Technology, 2019.

- https://ieeexplore.ieee.org/document/1645343
- Describes prototype direct chip microchannel cooling approaches and demonstrates very high local heat flux capabilities.

[X8] E.G. Colgan; B. Furman; A. Gaynes; W. Graham; N. LaBianca; J.H. Magerlein et al,, "A practical implementation of silicon microchannel coolers for high power chips," in Semiconductor Thermal Measurement and Management IEEE Twenty First Annual IEEE Symposium, 2005.

- https://ieeexplore.ieee.org/document/1412151
- Showcases embedded microchannel structures within silicon for effective high-density chip cooling.

[Y1] R. Ober and M. Iyengar, "*The Exponential Demands AI Places on the Rack & Datacenter*", OCP Summit 2024, Special Session AI/ML, https://www.youtube.com/watch?v=cu_3GMWl-rE

[Y2] H. Gan et al, 2023, "*Vertical Power Delivery for 1000 Amps Machine Learning ASICs*", Google paper at IEEE APEC Conference.

[Y3] S. Kudva et al, Challenges to Enabling Vertical Power Delivery in High-Power GPU Applications , 2023, Nvidia paper at the IEEE APEC Conference.

{Y4] OCP White paper under preparation on Modular Liquid Cooled Racks Cooled by a CDU.

[Y5] X. Li, K. Ravikumar, C. Peabody, OCP Summit 2024, "400Vdc Rack for AI/ML Applications", Google presentation, https://www.youtube.com/watch?v=l8ChVDv5aoo

[Y7] IEEE Electronics Packaging Society - Heterogeneous Integration Roadmap, https://eps.ieee.org/technology/heterogeneous-integration-roadmap.html

[Y8] ARPA-E - Cooler Chips, "Cooling Operations Optimized for Leaps in Energy, Reliability, and Carbon Hyperefficiency for Information Processing Systems", https://arpa-e.energy.gov/technologies/programs/coolerchips

[Y9] Y. Joshi, DARPA efforts on 3D chips and Heterogeneous Integration, “DARPA Thermal Management Initiatives in Support of ThreeDimensional Heterogeneous Integration”, https://arpa-e.energy.gov/sites/default/files/2023-11/Day1_05_DARPA_Joshi.pdf

[Y10] US Government Chips Act - R&D Funding, https://www.chipsact.com/ (See reference to Power Delivery and Thermal Management)

[Y11] ASHRAE TC9.9 Technical Committee on Mission Critical Facilities, Data Centers, Technology Spaces and Electronic Equipment, https://tpc.ashrae.org/?cmtKey=fd4a4ee6-96a3-4f61-8b85-43418dfa988d

[Y12] ASHRAE TC9.9 Datacom Book Series (including Liquid Cooling), https://www.ashrae.org/technical-resources/bookstore/datacom-series

License

Creative Commons

OCP encourages participants to share their proposals, specifications and designs with the community. This is to promote openness and encourage continuous and open feedback. It is important to remember that by providing feedback for any such documents, whether in written or verbal form, that the contributor or the contributor's organization grants OCP and its members irrevocable right to use this feedback for any purpose without any further obligation.

It is acknowledged that any such documentation and any ancillary materials that are provided to OCP in connection with this document, including without limitation any white papers, articles, photographs, studies, diagrams, contact information (together, “Materials”) are made available under the Creative Commons Attribution-ShareAlike 4.0 International License found here: https://creativecommons.org/licenses/by-sa/4.0/, or any later version, and without limiting the foregoing, OCP may make the Materials available under such terms.

As a contributor to this document, all members represent that they have the authority to grant the rights and licenses herein. They further represent and warrant that the Materials do not and will not violate the copyrights or misappropriate the trade secret rights of any third party, including without limitation rights in intellectual property. The contributor(s) also represent that, to the extent the Materials include materials protected

by copyright or trade secret rights that are owned or created by any third-party, they have obtained permission for its use consistent with the foregoing. They will provide OCP evidence of such permission upon OCP's request. This document and any "Materials" are published on the respective project's wiki page and are open to the public in accordance with OCP's Bylaws and IP Policy. This can be found at http://www.opencompute.org/participate/legal-documents/. If you have any questions please contact OCP.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Li

About Open Compute Foundation

The Open Compute Project (OCP) is a collaborative Community of hyperscale data center operators, telecom, colocation providers and enterprise IT users, working with the product and solution vendor ecosystem to develop open innovations deployable from the cloud to the edge. The OCP Foundation is responsible for fostering and serving the OCP Community to meet the market and shape the future, taking hyperscale-led innovations to everyone. Meeting the market is accomplished through addressing challenging market obstacles with open specifications, designs and emerging market programs that showcase OCP-recognized IT equipment and data center facility best practices. Shaping the future includes investing in strategic initiatives and programs that prepare the IT ecosystem for major technology changes, such as AI & ML, optics, advanced cooling techniques, composable memory and silicon. OCP Community-developed open innovations strive to benefit all, optimized through the lens of impact, efficiency, scale and sustainability. Learn more at www.opencompute.org.