OCP HPC Sub Project

Minutes - 06/14/2022

Meeting Transcript :

https://docs.google.com/document/d/1rAfFUCfpbsznvTm9ZJmlVg3CpSrk-qUS/edit?usp=sharing&ouid=100358437598195537439&rtpof=true&sd=true

Attendees

Attendee	Affiliation	06/14/2022
Phillipe Boisvert	Boyd	x
Allan Cantle	Nallasway	x
Chris Chapman	Boyd	x
ILang Cheng	Lenova	x
Michael Choi	Samsung	x
Calum Devlin	CoolIT	x
Chris Graves	Samtec	x
Zaid Krisberg	SUNY Downstate Medical Center	x
Brent Massey	Individual	x
Min	SK Hynix	x
Nhat Nguyen	Rambus	x
David Ratchkov	Thrace Systems	x
Steven White	Tormem	x
Totals		13

Meeting Objectives

Proposed Agenda for this meeting :

1) Introductions to New Attendees.
2) Analysis of potential Energy Benefits of HPCM
    a) From a presentation I'm giving at HIP-CHIPS Workshop this coming weekend.
3) Discussion around feedback on
    a) The HPCM Wall of Compute
    b) The New direct Water Cooled HPCM concept

4) Review forward Agenda for topics of discussion
5) AOB

Meeting Discussion Notes

   1) Introductions to New Attendees.

David RatchKov - Help run CDX inside ODSA - background chip level power and thermal

   2) Ongoing discussion around
      a) The HPCM Wall of Compute
      b) The New direct Water Cooled HPCM concept

Feedback from Brad Reger at Ingrasys on wall of compute

I have a few random thoughts to share with you. Admittedly, I have a hard time thinking as far out of the box as you do.

- The wall concept instead of an OAI chassis is interesting and IMHO worth exploring
- Forcing a system based on HPCMs to fit on a horizontal board in a legacy chassis in a legacy rack is challenging and highly likely to lead to switches, retimers or optics that we're specifically trying to eliminate to minimize power & energy
- Before getting too deep in implementations, we might want to capture a handful of system requirements as bullets to guide the concepts. Something like:
    - Should be possible to configure and reconfigure systems (modules & cables) onsite, so end users can optimize domain-specific configs
    - Should be possible to configure in a factory, so it can be tested, packaged & shipped, and then installed onsite w/minimal configuration (buy vs make; pay a factory to do the work, and more importantly test the configuration before shipping; need to figure out what & how to package for shipping); while some like to build large objects from Lego blocks, others prefer to buy the finished object.
    - Must be possible to service/replace cables, NICs, memory & storage modules onsite
    - Must be possible to service/replace HPCMs onsite
    - Should be scalable from N to M HPCMs in some defined 3D volume (as a starting point)
    - Should fit more HPCMs per square-foot of floor space than using OAI chassis in an ORv3 rack?
    - Should be possible to mix & match…[need a lot more requirements around what's important here]
    - [10-100 more requirements…]

- A wall that's like a sliding "barn door" seems overly complex; requires a frame design that prevents tipping; requires long service loops on the power, cooling and any I/O that enters/leaves the wall; requires depth in the service aisle that doubles the equipment footprint
- Changing the motion so the wall slides left-to-right (like shifting books on a shelf) instead of front-to-back might reduce required space and improve service loops but still has the other issues to work through
- Whatever holds the wall will have to be strong enough to support the weight of all that metal & fluid
- Doesn't have to be an EIA or ORv3 rack, but fundamentally it needs some type of beefy rack to hold everything in place
- More speculatively and more implementation, I wonder if an ORv3 rack could be used sideways, with only one compute wall per rack, with cables facing the main service aisle, power & cooling connections on the same or opposite side (I'm purposely not saying front or back because those terms are too vague; cold & hot aisles don't apply if it's all liquid-cooled). Then create rows in the data hall of appropriate width; since the new racks aren't as deep as legacy racks, the pitch between rows can be tighter to increase the compute density
- This concept is loosely modeled off a telco-style rack with switches and patch panels; the compute wall is like the set of switches in a rack and the HPCM cables are like the switch to patch panel cables
- This concept isn't nearly as dense as your double-wall in a 24" width; I just don't see the double-wall concept being practical; YMMV
- Staubli, Parker and others have quick-disconnects that can be adopted for blind-mate use; see folks in the OCP Rack & Power group and the Advanced Cooling group for their ideas. Theoretically, the ORv3 rack supports dual, blind-mate liquid manifolds but I can't find any documentation on it. My company has an implementation.

Phillipe asked the question as to what improvements did the Wall implementation bring compared to the OAI Chassis version. Callum Devlin agreed highlighting that the wall brings additional challenges.

The OAI Chassis version supporting HPCM does bring a lot of the benefits of HPCM without the additional complexities that the wall introduces. The fabric expansion port should support up to 24 or more HPCM modules with the same power saving benefits as the Wall offers for 80 HPCM Modules.

It was agreed to refocus efforts back into the OAI Chassis implementation and put the Wall version on the back burner for now.

3) **Review forward Agenda for topics of discussion**

Ideas for future meeting presentations.

a) 03/01/2022 - Flex & Progranalog Presenting Flex Power Tree for VP1802 HPCM

b) ~~03/08/2022 - How we can help make OCP HPC efforts more visible and support a presence at SC22, workshop, etc. - Cliff Grossner & Dirk Van Slyke~~
c) ~~03/08/2022 - Lattice Presenting LibreBMC / Sideband Management Solution for POWER10 HPCM~~
d) ~~03/22/2022 - Samsung - CXL update, including Samsungs CXL.mem E3.S module~~
e) Infineon on PSUs?
f) NVidia?
g) Define the Cooling and Power stuff - Standardize?
h) Optical Interconnect?
i) Turn into a spec - what's locked down and hard and what's flexible?

4) **AOB**