OCP HPC Sub Project

Minutes - 12/10/2024

OCP Local Recording :
https://drive.google.com/file/d/1OZPzJDtfO7egMjqDodb2ZakwfATSAvdy/view?usp=drive_link

Meeting Transcript :
https://drive.google.com/file/d/1gQMKgeknNmh3-TyWfH3sydd5mKc-lw53/view?usp=drive_link

Attendees

Attendee	Affiliation	12/10/2024
Howard Andrews	Lotes	x
Peter Born	Empower	x
Kevin Cameron	Individual	x
Allan Cantle	Nallasway	x
Zhineng Fan	Amphenol	x
Olivier Franza	Intel Foundry	x
Rashid Kukkady	Individual	x
Raymond Law	LOTES	x
John Lu	LOTES	x
Bijan Nowroozi	OCP Project	x
Chloe Pilon	Qarnot	x
Renu Raman	Terizza	x
Ariel Delos Reyes	LOTES	x
Mitch Wright	Liqid	x
Totals		14

Meeting Objectives

Proposed Agenda for this meeting :

1) Introductions to New Attendees.
2) Brief update on ORNL New Frontiers Proposal feedback

3) LOTES to present a proposed update to the pinout of the Universal Interconnect for easier manufacturability.
4) Discuss another OCP working group called Open Systems for AI. On a related note, there's some interesting discussions going on around OAI/UBB/OAM becoming part of the DC-MHS Working group.
5) AOB

Meeting Discussion Notes

1) Introductions to New Attendees.

Chloe - working at Qarnot computing in Europe. Part of OCP for a while.

Rashid - Silicon Valley - Ex Syonpsys, starting own in IP and looking into certain areas of HPC.

2) Brief update on ORNL New Frontiers Proposal feedback

3) LOTES to present a proposed update to the pinout of the Universal Interconnect for easier manufacturability.

4) Discuss another OCP working group called Open Systems for AI. On a related note, there's some interesting discussions going on around OAI/UBB/OAM becoming part of the DC-MHS Working group.

5) AOB

AI Generated Meeting Summary - NOTE this is an OK summary but is confused in several areas.

Meeting summary for OCP Server - High Performance Computing (12/10/2024)

Quick recap

Allan welcomed new members to the OCP High Performance Compute sub-project call and discussed various topics including a proposal for government funding, interconnect designs, and other OCP work. The team also discussed the potential issues with the cable PIN maps and proposed solutions, including the use of active cable design and the adoption of the LVDS tunneling protocol. The conversation ended with discussions on the inefficiencies of the current OAI system, the potential of the MHS approach, and the need for collaboration and innovation in the industry.

Next steps

• Allan to continue exploring funding options for the project, including potential European funding sources.

• Allan to follow up with Chloe from Carnot Computing for further discussions on collaboration.

• Lowe's team to update the HPCM cable pin map proposal based on the discussion with Raymond.

• Allan to sync up with Bijan offline to discuss how to introduce HPCM ideas into other OCP groups, particularly DCMHS.

• Allan to explore options for creating an interoperable UBB for startups within the OAM framework.

• Allan to investigate the possibility of developing a water cold plate solution for HPCM in the OCP Rack v3 format.

• Allan to consider folding the HPCM effort into DCMHS as a separate work stream or feeding ideas into the existing DCMHS group.

Summary

New Members Join OCP Sub-Project

In the meeting, Allan welcomed new members Chloe and Rashid to the OCP High Performance Compute sub-project call. Chloe, from Carnot Computing in Europe, expressed her interest in attending the meetings to stay updated on the community's activities. Rashid, based in Silicon Valley, shared his background in EDA and his interest in supercomputing architectures. Allan then outlined the three topics for the meeting: a proposal for government funding, Lots' interconnect designs, and other OCP work. He also mentioned the possibility of a separate call or sending an email with relevant videos for Chloe to get up to speed on the project.

New Frontiers Proposal and Emulation Platform

Allan provided an update on the New Frontiers proposal, stating that they will announce the winners by the end of February next year. He also mentioned that he had spoken to the Barcelona supercomputing center about their emulation platform, Meet MEEP, and that it could be a potential next-generation platform. Allan also discussed the possibility of getting funding from the Chip Act for advanced packaging and mentioned his new role at On Tevue AI. He then handed over to Raymond from Amphenol and Lotes to discuss the proposed PIN out update for the MCI connectors and the transition from 19-seventies to OCP LTPS on the cables. Raymond presented the potential issues with the cable PIN maps and proposed a solution.

HPCM PIN Map Proposal Challenges

Raymond discussed the differences between the current HPCM PIN map proposal and the standard PCIe connector map. He highlighted the issue of signal crossover in the HPCM design, which could cause difficulties in cable connections. Allan understood the issue and agreed that it could be solved, albeit making the cable design a bit more complex. Raymond also mentioned that the issue affected both the passive and active cable designs. Allan suggested that the active cable design could be used to separate the transmit and receive signals on the PCB to minimize crosstalk. Raymond agreed that there are various approaches to mitigate this issue in PCB layout.

Pedal Cart Design and Signal Repetition

Raymond and Allan discussed the design of a pedal cart with active components for signal repetition. They agreed on using two repeater chips per direction, with the possibility of using retimers if the industry moves towards a more chiplet approach. They also discussed the power requirements and the need to keep the middle portion of the ground untouched. Allan suggested that the double ground on certain pins might not be necessary due to the structured layout of the PCI-MCI pinout. The conversation ended with Raymond confirming that the transmit, clock, and LTP signals need to be arranged like the high-speed signals.

Progress and Collaboration in DC SCM

Allan discussed the progress of his work, particularly focusing on the DC Scm group and the adoption of the LVDS tunneling protocol. He proposed moving on from the 1970s standards and using LTPi in the universal topology interconnect design. Allan also expressed his desire to keep their work open and move at a faster pace, acknowledging the bureaucratic challenges they face. He suggested a need for collaboration with other groups, including the DC Scm and PCI Sig, to integrate their work. Bijan agreed with Allan's points and mentioned his efforts to

foster cross-project collaboration, particularly in the context of the AI open systems project.

AI Industry Shift and Modular Hardware

Allan and Bijan discussed the shift in the AI industry, particularly the sunset of Oam and Ocp. They noted that many startups are struggling to compete with established players like Nvidia and AMD due to the limitations of Oam. Allan suggested that the Open Chiplet Economy meeting had proposed the integration of Oai and Oem into the Modular Hardware System (MHS), which could provide a more flexible and adaptable solution for the industry. However, Allan expressed concern that the MHS might not be open to all, which could hinder innovation and limit the opportunities for startups.

Frustration With Closed-Door Practices

Allan expressed frustration with AMD and Intel's closed-door practices, suggesting that the community should take control of the situation. He proposed folding their effort into Dcmhs and emphasized the need for new blood in the industry. Allan also criticized the current system, suggesting the need for a department of data center efficiency to speed up progress. Kevin agreed with Allan's points, but suggested thinking harder about how to achieve their goals. The conversation ended without a clear resolution on how to proceed.

Exploring OAI and MHS Solutions

In the meeting, Bijan and Allan discussed the inefficiencies of the current OAI system and the potential of the MHS approach. Allan expressed interest in both avenues, but was concerned about the cost of developing a custom UBB. Bijan suggested that Allan could invest time in the OAI 3.0 to get an 8 pro 8 accelerator into RU, but acknowledged the challenge of someone having to invest in this. They also discussed the possibility of an MHS plus HPCM solution. Allan proposed a water-cooled plate solution for the HPCM format and mentioned the potential of new frontiers funding. The conversation ended with Allan expressing his desire to get involved in the DCMHS and seeking Bijan's guidance.