Outlook

[EXTERNAL] Fwd: Meeting Summary for Vector SIG Standing Meeting

From Rafael Sene <rafael@riscv.org>
Date Mon 12/9/2024 12:10 PM
To Jose Moreira <jmoreira@us.ibm.com>

FYI,

The Zoom summary.

Atenciosamente | Sincerely
Rafael Peria de Sene
Technical Program Manager, RISC-V
Phone: +55 19 98153 9778
E-mail: rafael@riscv.org

I sent this email when it was convenient for me.

Please reply whenever it's good for you!

---------- Forwarded message ---------
From: Meeting Summary with AI Companion <no-reply@zoom.us>
Date: 9 Dec 2024 at 14:04:25
Subject: Meeting Summary for Vector SIG Standing Meeting
To: rafael@riscv.org

Meeting summary with AI Companion now supports additional languages in preview. **Learn More**

Meeting summary for Vector SIG Standing Meeting (12/09/2024)

Quick recap

The team discussed the results of a survey on various topics, including video codec workloads, wireless workloads, and other technical aspects, with a focus on improving performance for video codecs and related workloads. They also proposed new instructions for video codec operations, with a focus on zip, absolute difference, and dot product operations, and discussed the potential for more general DSP capabilities. The team also addressed the Fb. 8 wireless issue and agreed to draft a fast track proposal for review and discussion.

Next steps

• Punit to take the video codec requirements to the Systems Libraries work group for feedback.
• Philip to follow up with Karen to share the Zip/Unzip proposal PDF with the Vector SIG.
• Punit to draft a fast track proposal for the absolute difference instruction and share for review.
• Krste to communicate with Andrew about creating a straw man proposal for FP8 conversion instructions.
• Andrew to develop a straw man proposal for conversions to and from FP8.
• Jose to provide an update on whether the December 23rd meeting will be cancelled.

Summary

Reviewing Meeting Summary and Policies

In the meeting, Jose led the discussion and shared the meeting summary generated by AI. He also mentioned the importance of reviewing the summary carefully to ensure it includes the right discussion points. The team discussed the concept of the Butlerian Jihad and Skynet, with Allen sharing his concerns about models being told to do something at any cost. Jose then began the meeting, sharing a disclaimer and highlighting three key policies: community collaboration, antitrust compliance, and export controls. He also mentioned the availability of Zoom captions. The conversation ended with Jose expressing his gratitude to Guy for organizing a poll and sharing the results.

Video Codec Workload Survey Results

Jose led a discussion about the results of a survey on various topics, including video codec workloads, wireless workloads, 8-bit floating point format, Olaf's mask shift instructions, multiply vector multiply accumulate instructions, OCP, incomplete partial architecture, smart matrix support, scale, memory to memory vector operations, GPU capabilities, 64-bit encoding, and subsetting. The survey showed a clear preference for video codec workloads, with half of the respondents ranking it as their number one priority. Other topics received fewer votes, with some not receiving any votes at all. The team discussed the implications of these results, with some suggesting that certain topics should be considered in relation to matrix-oriented ones. The discussion also touched on the potential for future developments in certain areas, such as 64-bit encoding and subsetting.

Improving Video Codec Performance

The meeting discusses next steps for proposing a set of new instructions to improve performance for video codecs and related workloads. Punit suggests the instructions seem relatively generic and mature, so they could move quickly to get broader community feedback. Jose notes there has already been substantial prior discussion on this topic. Krste recommends forming a task group to develop a charter and define the scope of

work, though the specific name is not crucial initially. Greg emphasizes the need to quantify performance benefits for relevant workloads to justify standardizing the instructions. Punit confirms they have evaluated some performance estimates, providing a starting point for the proposal.

Video Codec Operations Proposal
Punit discussed the proposal for video codec operations, which includes three classes of instructions: zip operations, absolute difference operations, and dot product operations. These operations are aimed at improving performance for video codecs. Punit also mentioned that the dot product and zip operations are being considered as fast track proposals. Krste suggested that the Sig should decide how to organize these proposals, considering whether they should be fast tracks or part of a larger task group. He emphasized the importance of ensuring that the proposals address the needs of video codecs effectively.

Generalizing Product for Wider Data Types
Dmitry proposed the idea of generalizing the product to include wider data types and accumulation to a single result. Krste clarified that the current focus is on unsigned 8-bit operations, which are primarily needed for video codecs. However, there was a discussion about the potential for more general DSP capabilities. Punit explained that the zip and unzip operations are useful for video codecs and other domains, such as complex arithmetic. The group agreed on the need for profiling data to justify the inclusion of these operations. There was also a discussion about the potential for segment load and store operations to be used for transpositions, but it was noted that these might not be as efficient as dedicated instructions. The group also discussed the need for specific instructions for video encoding and decoding, particularly for data center video processing.

Improving Performance With New Instructions
The team discussed the potential for new instructions to improve performance in video codecs and other applications. They considered the zip and zip instructions as a fast track, with the goal of engaging more people to avoid unnecessary duplication. The team also discussed the absolute difference instruction, which was seen as a potential addition to cover more requirements. The team agreed to move forward with these instructions, with the understanding that they would need to ensure they catered to the needs of video codecs and other applications. The team also agreed to split the discussion into two work streams, one for the zip and zip instructions and another for the absolute difference instruction.

Fb. 8 Wireless Issue Discussion
In the meeting, Jose led the discussion about the Fb. 8 wireless issue, which was not the most voted but still considered significant. Punit suggested that someone from Rivo should share the information with the vector. Sig. Philip took an action item to follow up with Karen to ensure the sharing happens. The team also discussed the absolute difference and decided to draft a fast track proposal for review and discussion. Jose then moved on to discuss the Fb. 8 issue, which was seen as a straightforward problem. Krste suggested starting with conversion instructions to and from Fb. 8, and Andrew was volunteered to work on this. The team agreed to share the proposal in January. The conversation ended with Jose wishing everyone a happy holiday season and a return in January.

AI-generated content may be inaccurate or misleading. Always check for accuracy.

Please rate the accuracy of this summary.

Best,

Zoom

Twitter LinkedIn Blog

+1.888.799.9666

Visit zoom.us

55 Almaden Blvd
San Jose, CA 95113