Jul 25, 2024 | RV Performance Events TG

Attendees: tech.meetings@riscv.org Dmitriy Ryabtsev Beeman Strong

Notes

- **Attendees**: Dmitriy, Beeman, Bruce, MattT, Snehasish, Victor, Robert
- **Slides/video** here
- TMA MEMORY_BOUND description/definition
    - Discussed Beeman's non-normative text with guidance for micro-architects on how to count memory bound
        - RS-full stalls might be more common than ROB-full stalls, could add them to list
    - There may be implementation options that are simpler, if less accurrate
        - Many Intel implementations use a formula for MEMORY_BOUND
            - Though activity-stalls event counts cycles, but is otherwise fairly analogous to memory bound
            - Lower levels are typically imprecise formulas
        - Beeman will revise the text to include options for less precise HW events, or a formula
- Want more MEMORY_BOUND breakdown? L1/L2/L3/etc?
    - Perhaps not all implementations will include them, but for those that do will be nice to have standard events
    - Requires routing data source information on the system bus
        - Implementations that don't have that may just have L1_BOUND
    - MEMORY_BOUND could mean waiting for data (L1/L2/L3/DRAM/HBM/CXL/etc, depends on implementation), waiting for translation (DTLB, STLB, walk), or other delays (pipeline conflicts, memory ordering violations, etc)
        - Memory ordering probably falls more under BAD_SPEC
            - All the slots in the shadow of the clear would be bad spec, but any delay imposed on the load that blocks issue would be MEMORY_BOUND
        - Having an "OTHER_BOUND" event probably isn't useful, not clear what a developer would do to address that. Could simply be that MEMORY_BOUND sub-events don't add up to the MEMORY_BOUND count
    - Can discuss the MEMORY_BOUND sub-events on the mailing list
- **Out of time**, will push RVV events to next time

Action items

- [ ] Beeman Strong - May 23, 2024 - check on idea to count remote HITMs locally
- [ ] Beeman Strong - Apr 25, 2024 - check on how perf uses ScaleUnit