Update on Ara

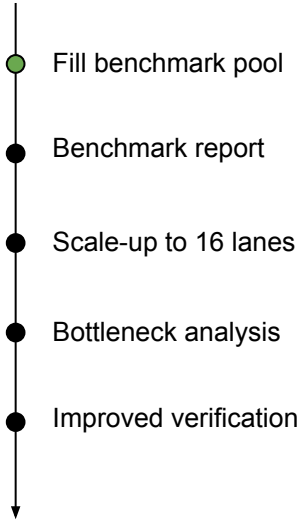
17/08/2022

Matteo Perotti
Matheus Cavalcante
Nils Wistoff

Professor Luca Benini
Integrated Systems Laboratory
ETH Zürich

Summary

- **Software**
  - Update SW environment
  - fp-dotp
  - Softmax
  - AWB
- **Hardware (RTL + Backend)**
  - Scalar Moves
  - FP Reductions
  - Bug Fixing



Update SW Environment

- Update to RVV1.0
- Some intrinsics were not properly working
- SPIKE golden model is also up-to-date now

Conversation 6 | Commits 6 | Checks 116 | Files changed 12

mp-17

Update LLVM to version `15.0.0` and Spike to `1.1.1-dev`.
This updates the RVV support to 1.0 frozen.

Changelog

Fixed

- Fix wrong variable in `vmerge` and `vmv` `riscv-tests`

Changed

- Update LLVM to version `15.0.0` (RVV 1.0)
- Update Spike to version `1.1.1-dev` (RVV 1.0)
- Update `newlib` from commit `84d068` to `5192d5`

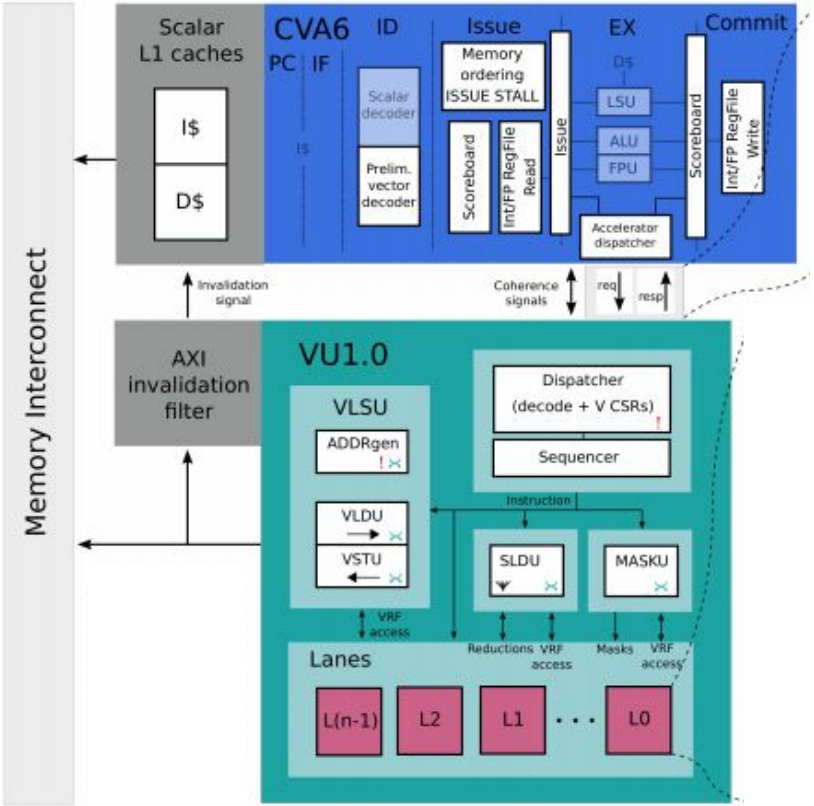
Checklist

- [x] Automated tests pass
- [x] Changelog updated
- [x] Code style guideline is observed

New Benchmarks

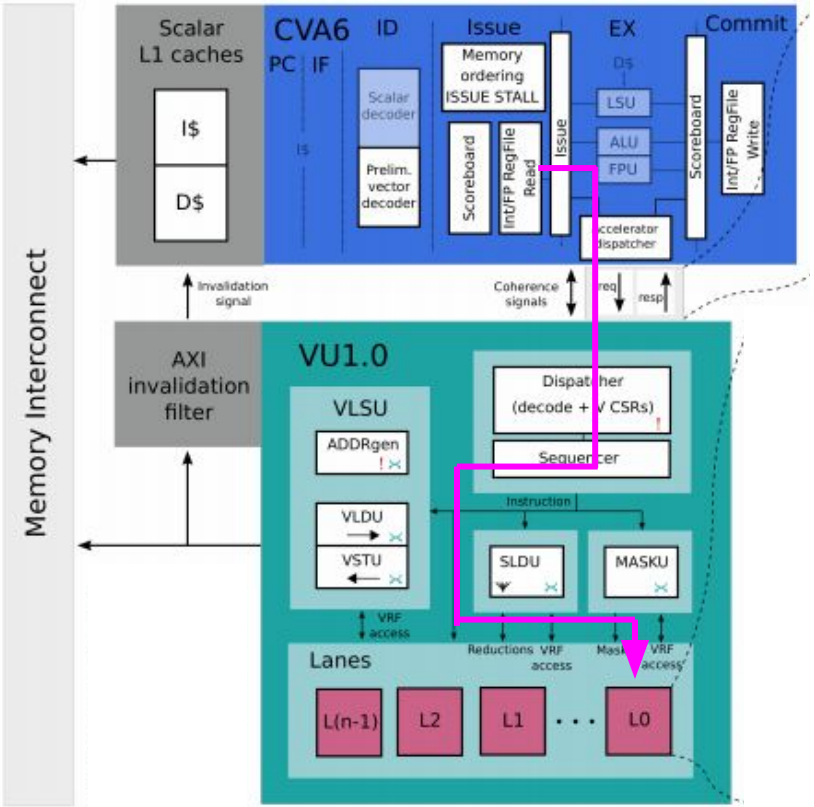
- fp-dotp
  - Fix from FP-reductions project
- Softmax
  - First version implemented
- AWB (Gray-World HP)
  - First version implemented

Hardware - Scalar Moves



Hardware - Scalar Moves

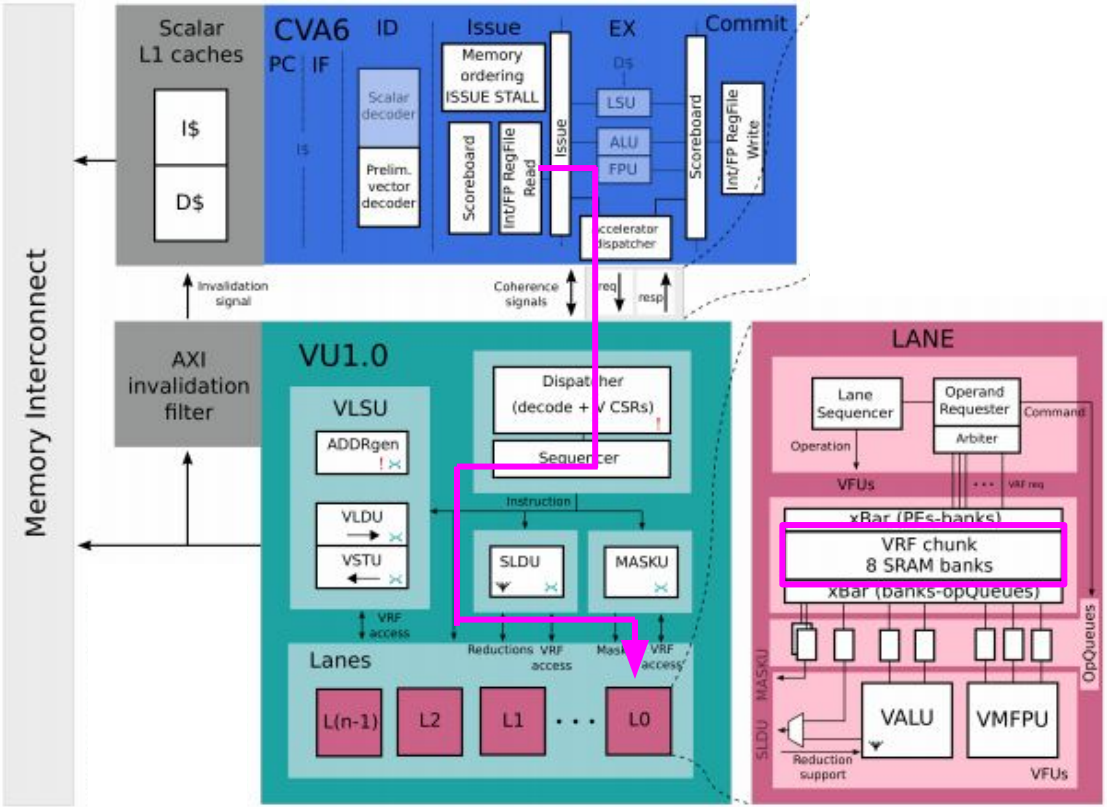
Move scalar from CVA6 to Ara



Hardware - Scalar Moves

Move scalar from CVA6 to Ara

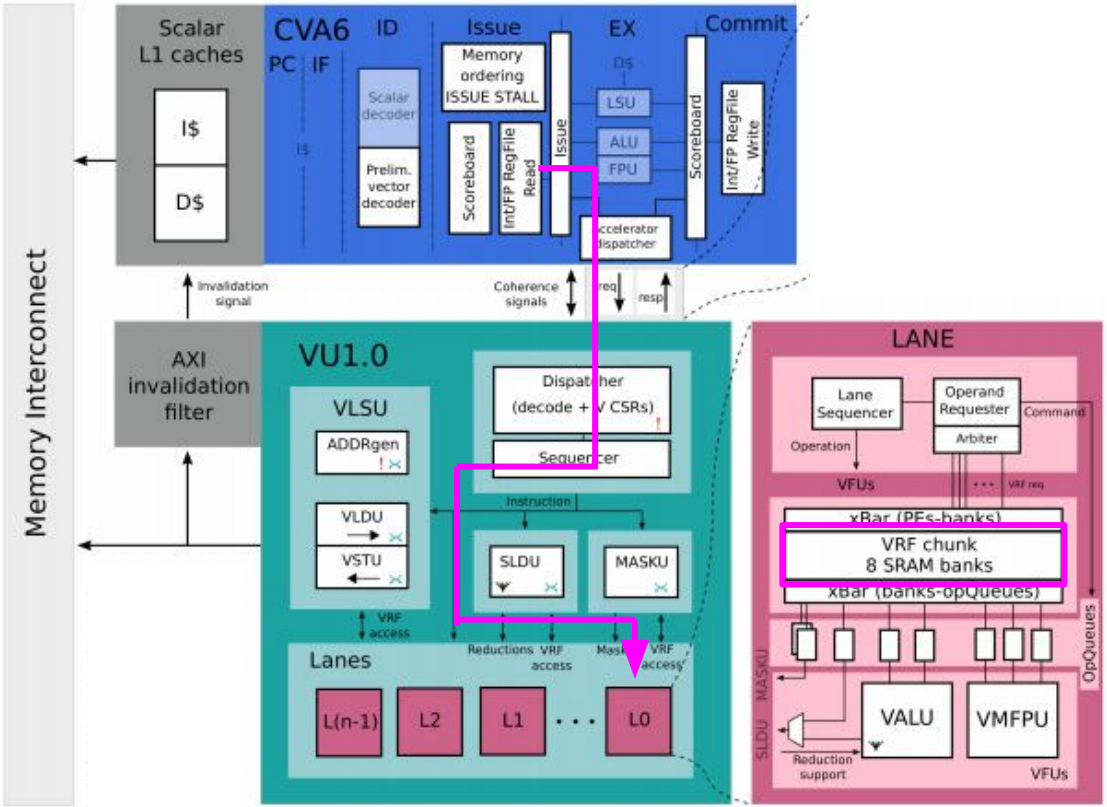
Scalar stored in Lane 0



Hardware - Scalar Moves

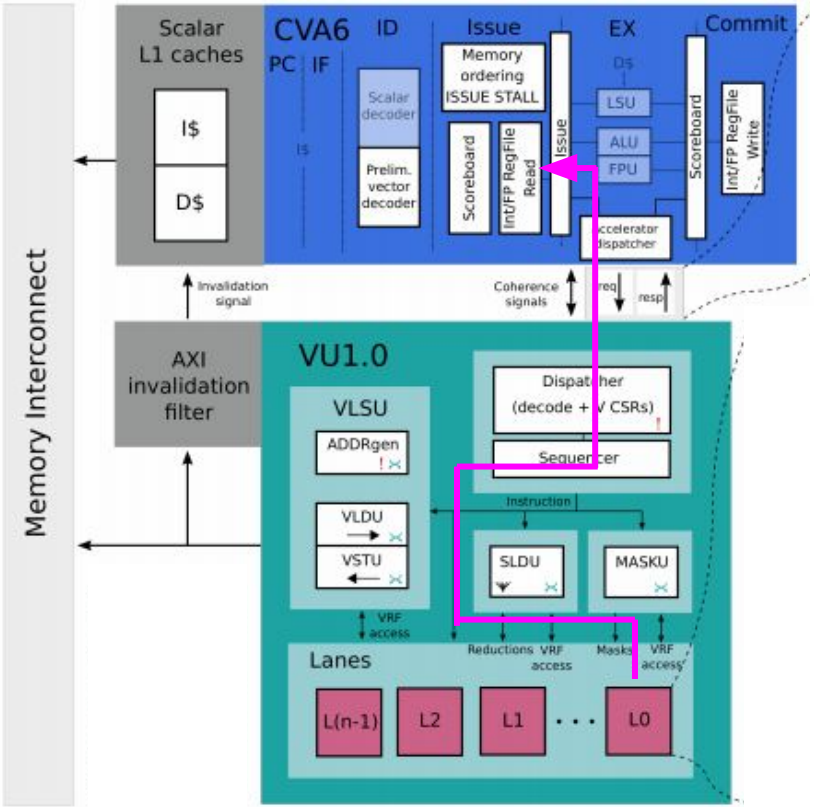
Move scalar from CVA6 to Ara

Scalar stored in Lane 0



Hardware - Scalar Moves

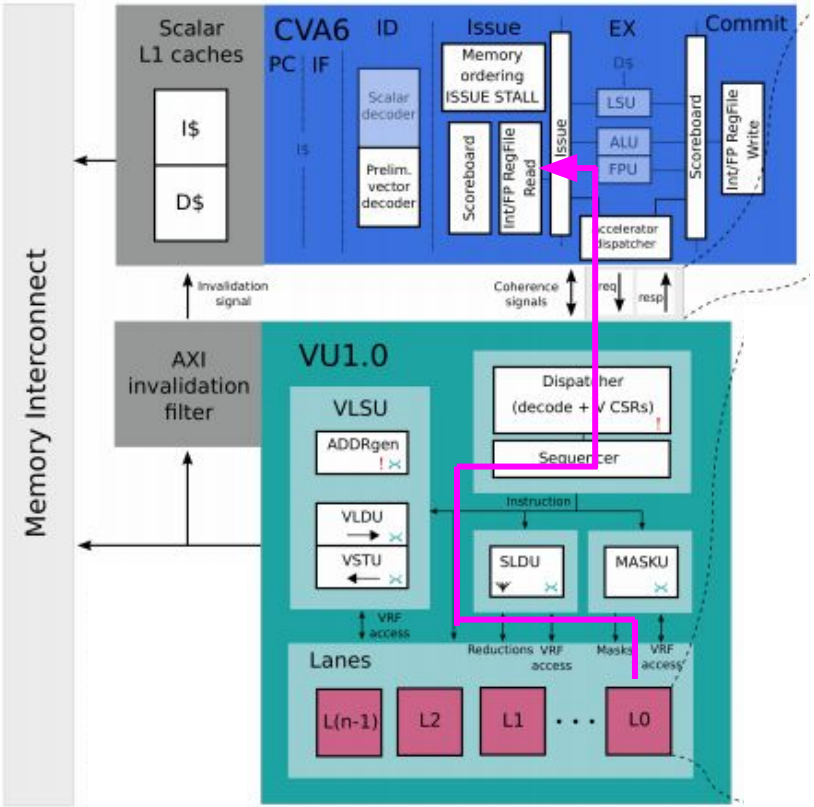
Move scalar from Ara to CVA6



Hardware - Scalar Moves

Move scalar from Ara to CVA6

CVA6 stalls until the result is received back!

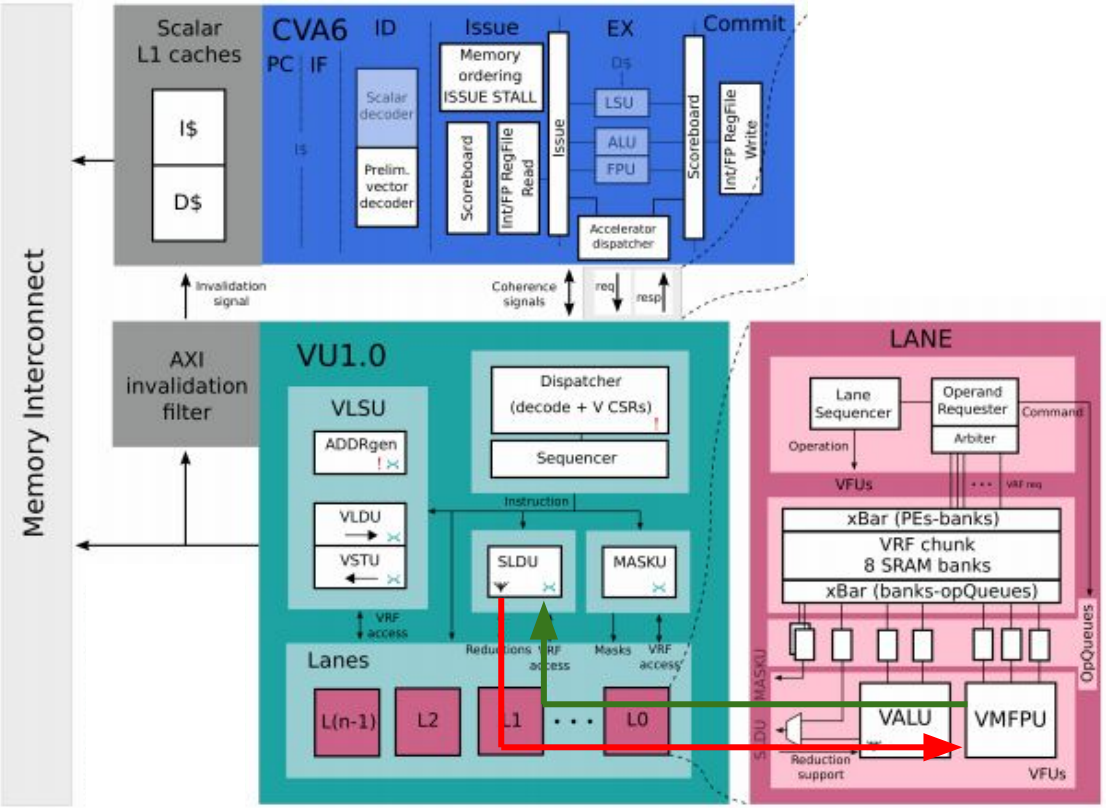


Hardware - FP Reductions

- Student project
  - Ordered sum
  - Unordered operations
- FPU is pipelined
  - Internal pipe regs -> Partial accumulators
- Non-negligible timing degradation

Hardware - FP Reductions

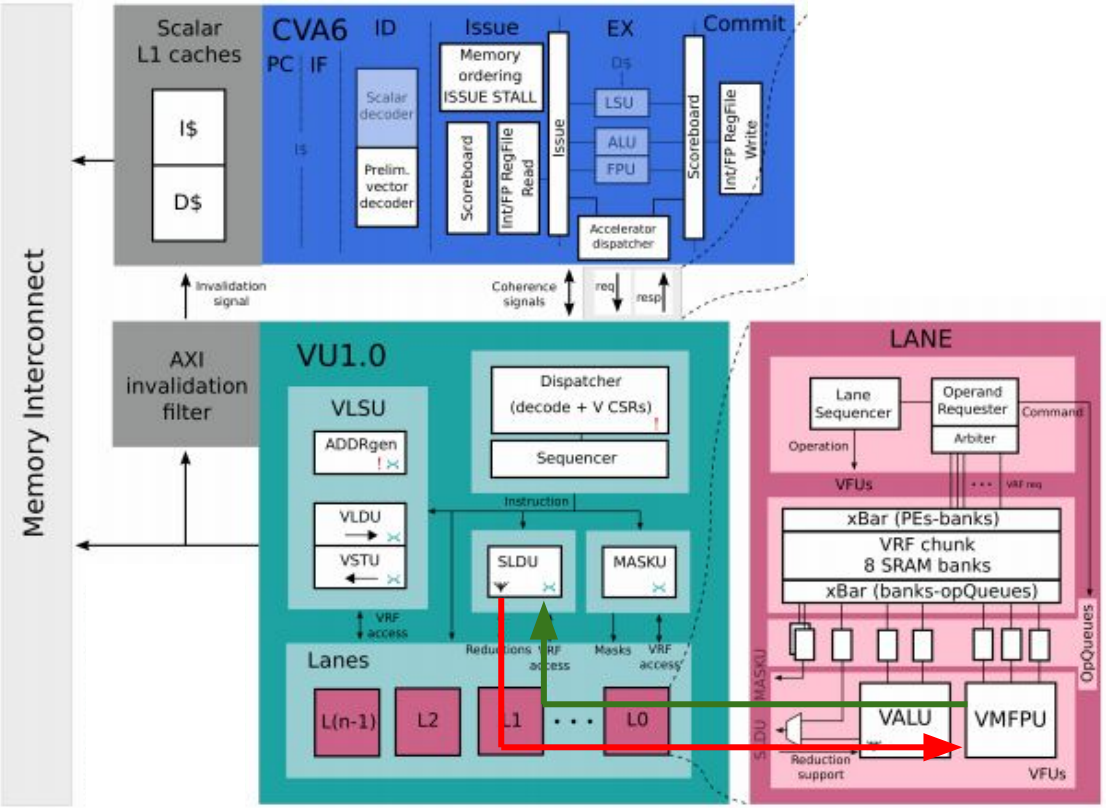
- Critical out2in path!



Critical path (valid-ready)

Hardware - FP Reductions

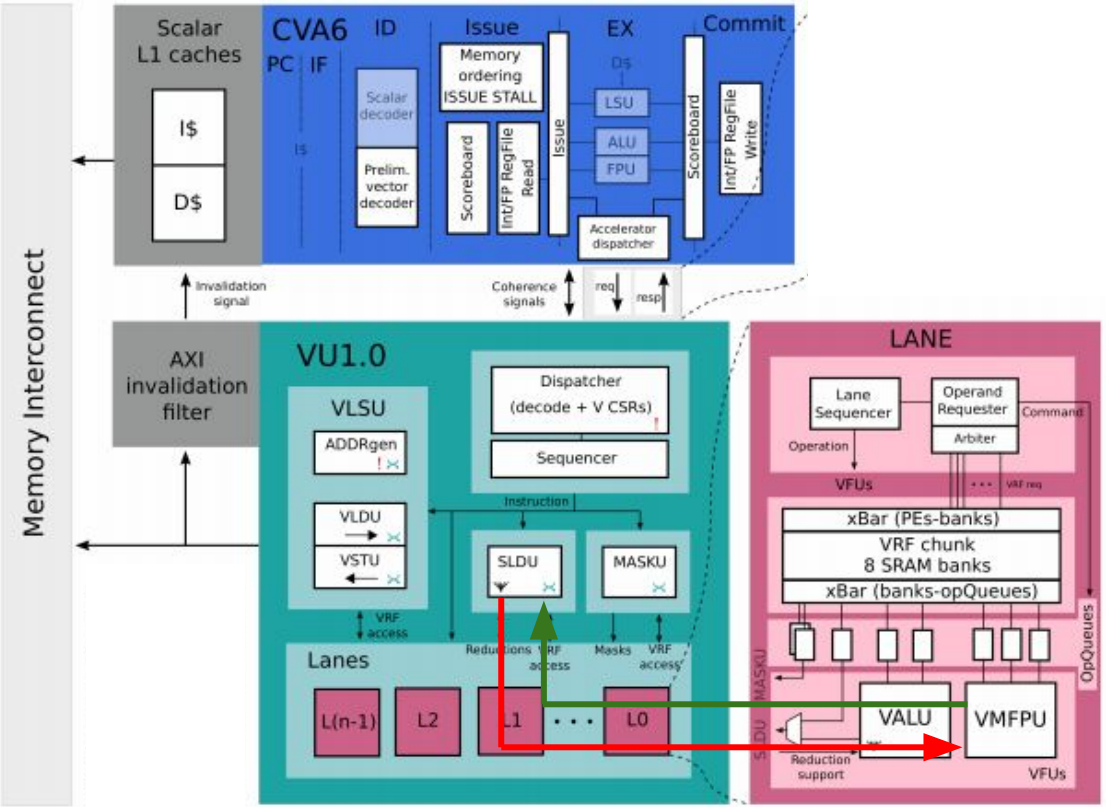
- Critical out2in path!
- Special tightened constraints



Critical path (valid-ready)

Hardware - FP Reductions

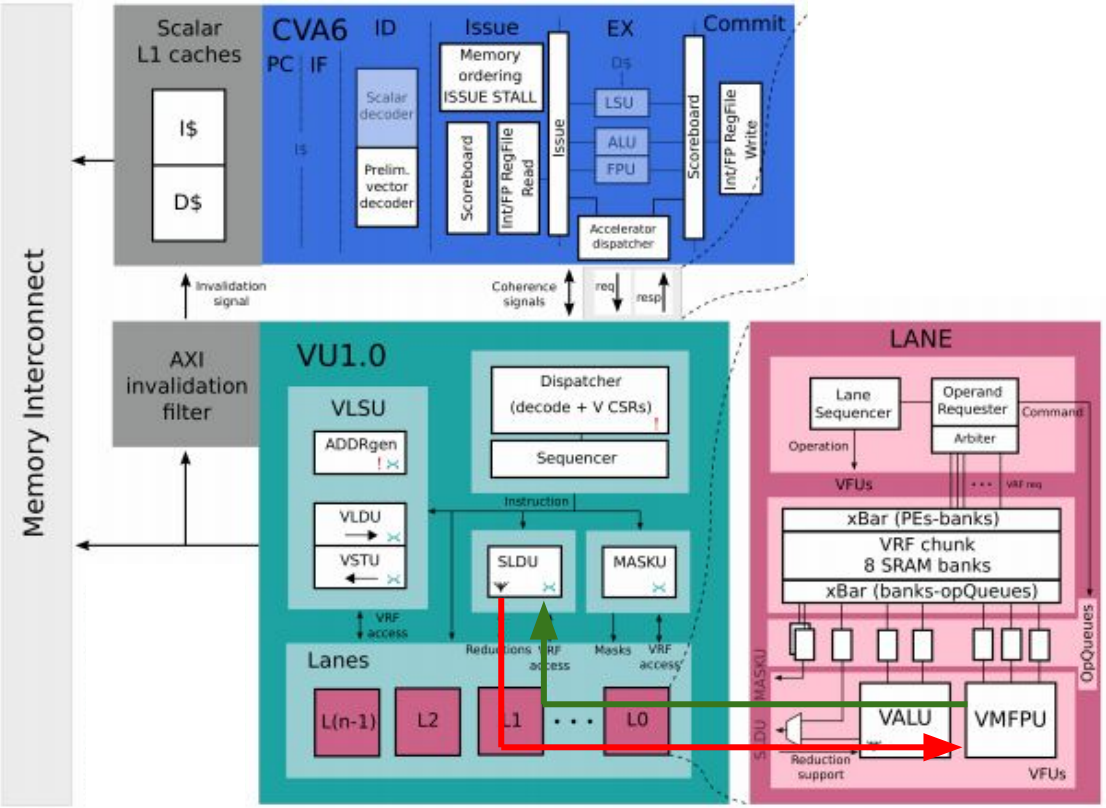
- Utilization > 85%



Critical path (valid-ready)

Hardware - FP Reductions

- Utilization > 85%
- 8% larger lane



Critical path (valid-ready)

Hardware - FP Reductions

- No frequency degradation
- Ready for merging

[hardware] Floating-Point Reductions ✓
#131 opened 13 days ago by mp-17 4 tasks done

[hardware] 🐛 Fix reductions + Rework the VALU ✓
#130 opened 13 days ago by mp-17 • Draft 3 of 4 tasks

Adding support for Fixed-Point vector instructions ✕
#106 opened on 11 Apr by hossein1387 • Changes requested 2 of 4 tasks

[hardware] Floating-Point classify, division, sqrt ✓
#100 opened on 19 Jan by mp-17 4 tasks done

Hardware - Bug fixing

- Stripmine conditions for very long vectors
- Misaligned bursts with more than 256 beats
- Hazard checks for LMUL > 1
- B2B integer reductions

Hardware - FP Reductions

- External efforts to implement fixed-point instructions!

[hardware] Floating-Point Reductions ✓
#131 opened 13 days ago by mp-17 4 tasks done

[hardware] 🐛 Fix reductions + Rework the VALU ✓
#130 opened 13 days ago by mp-17 • Draft 3 of 4 tasks

Adding support for Fixed-Point vector instructions ✗
#106 opened on 11 Apr by hossein1387 • Changes requested 2 of 4 tasks

[hardware] Floating-Point classify, division, sqrt ✓
#100 opened on 19 Jan by mp-17 4 tasks done

Benchmarks - Analysis ongoing

- Linear algebra:
  - [i,f]matmul - workhorse, b2b vmacc
  - [i,f]conv2d - vslides
  - jacobi2d - misaligned accesses
  - axpy - memory bound
  - spmv - indexed mem ops
  - [i,f]dotp - [i,f]reductions
- Machine Learning:
  - dropout - memory bound
  - softmax - fpred, fpdivisions
  - roi_align - different mem approaches
- DSP:
  - FFT- segmented, masked permutations
  - DWT - segmented or strided
  - AWB - Clip and conversions
- Others:
  - fp- cos, log, exp
  - memcpy, strncmp, strncpy
  - pathfinder

Next

- Benchmarks
  - ReLU
  - +Last benchmark (img processing? Biomedics?)
  - Performance report

- Hardware
  - Scale up to 16 lanes