Update on Ara

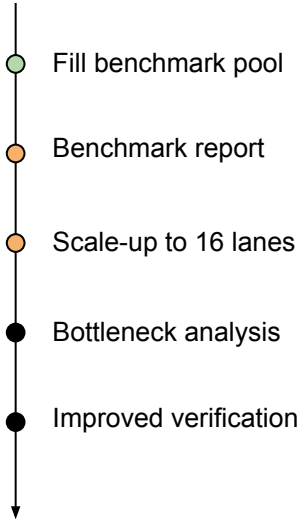
28/09/2022

Matteo Perotti
Matheus Cavalcante
Nils Wistoff

Professor Luca Benini
Integrated Systems Laboratory
ETH Zürich

Summary

- **Software**
  - Performance summary on GitHub
  - Implement ideal dispatcher (questa)
  - Gather performance data + Report
- **Hardware (RTL + Backend)**
  - Scale to 8 lanes
  - Merge FP Reductions

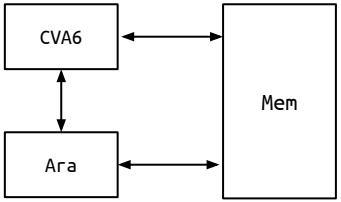


Software

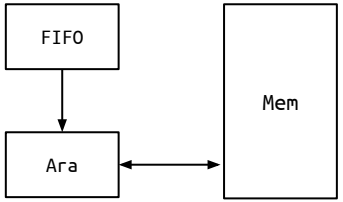
- **Software**
  - Extended “ideal dispatcher” analysis
  - Still no Verilator - Dynamic memory loading
  - Use QuestaSim offline

Software

Default System

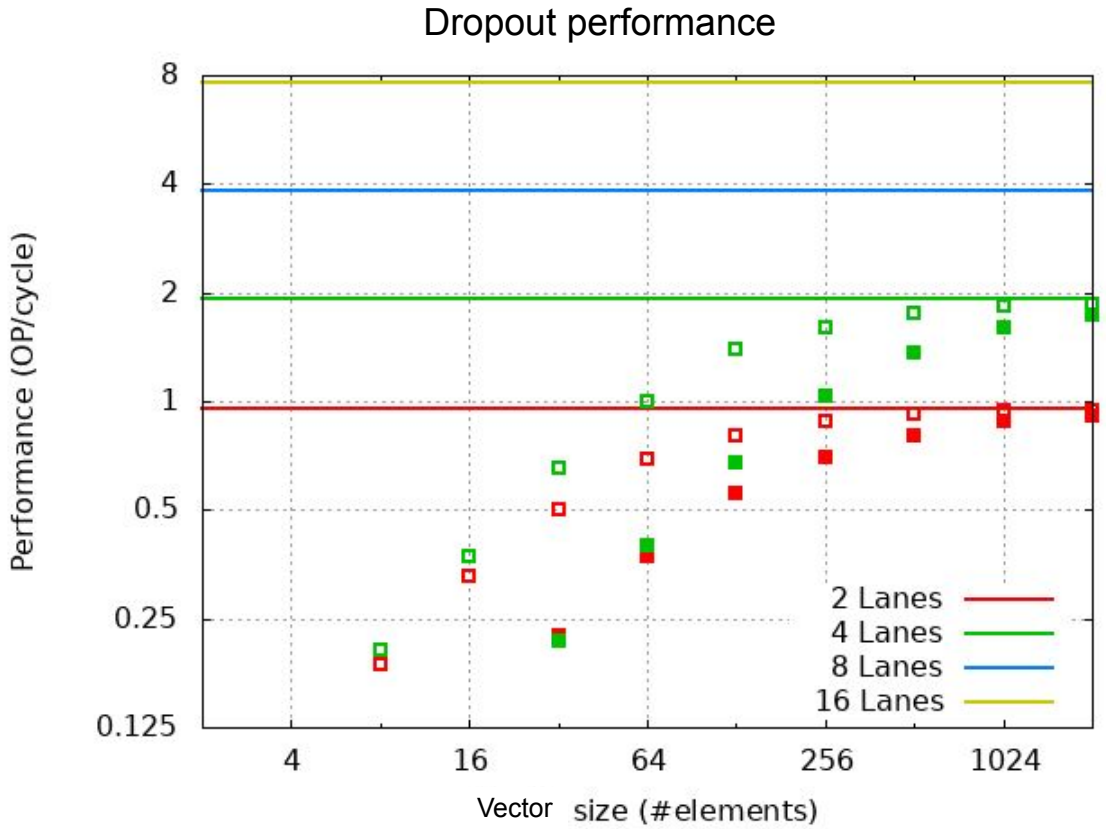


Ideal Dispatcher System



Ideal issue rate from FIFO to Ara

Software

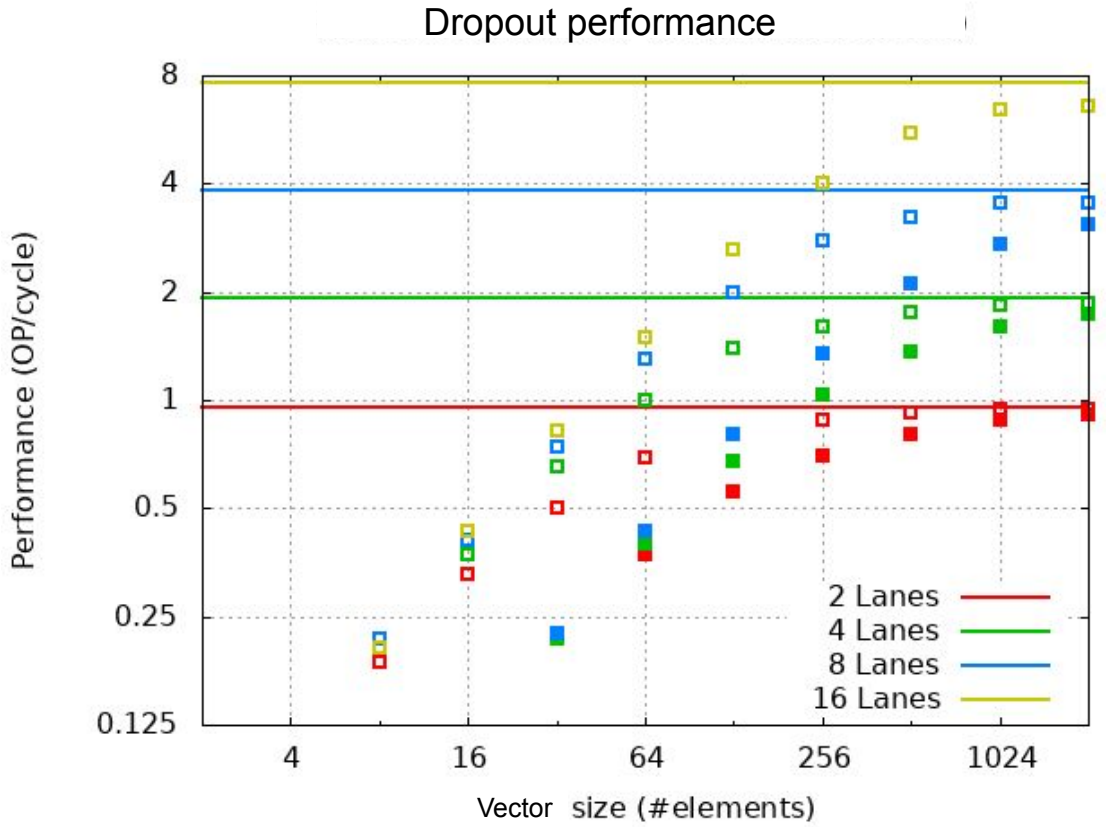


Always Memory Bound

- Ideal Dispatcher System
- Default System with CVA6

Constant arithmetic intensity!

Software

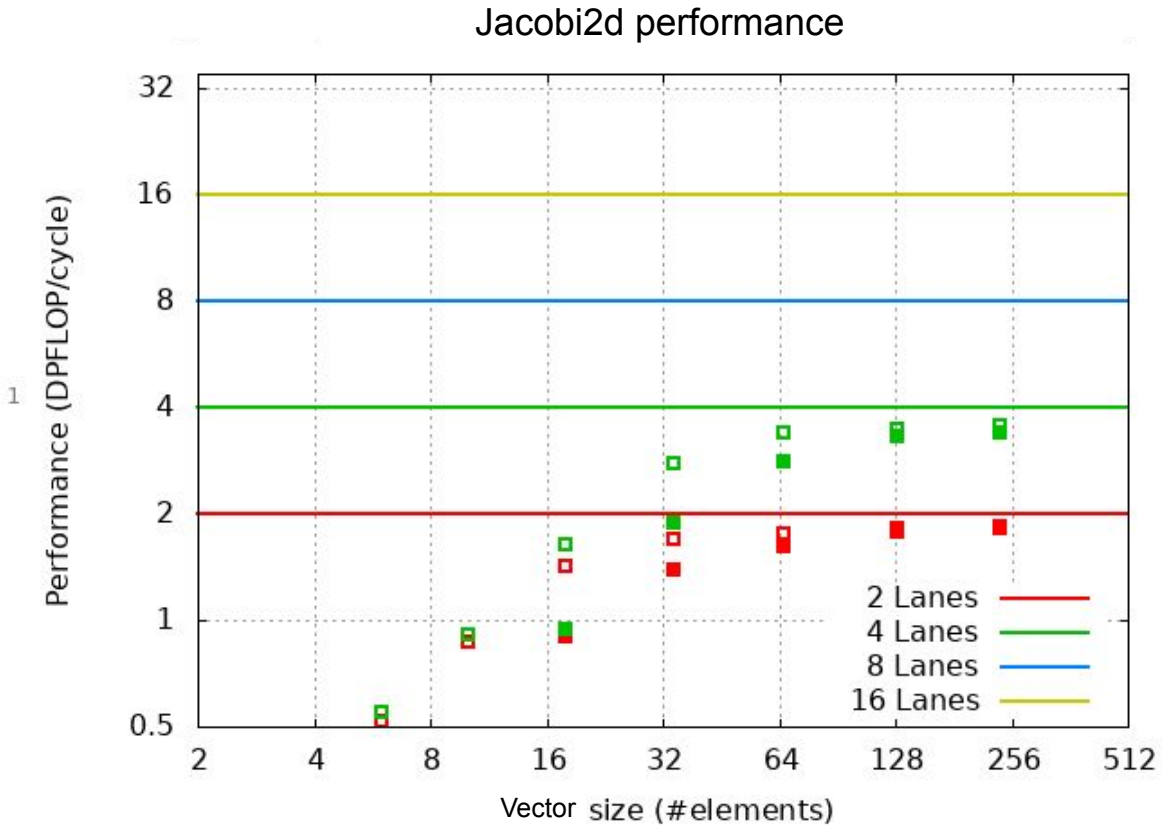


Always Memory Bound

Ideal Dispatcher System
Default System with CVA6

Constant arithmetic intensity!

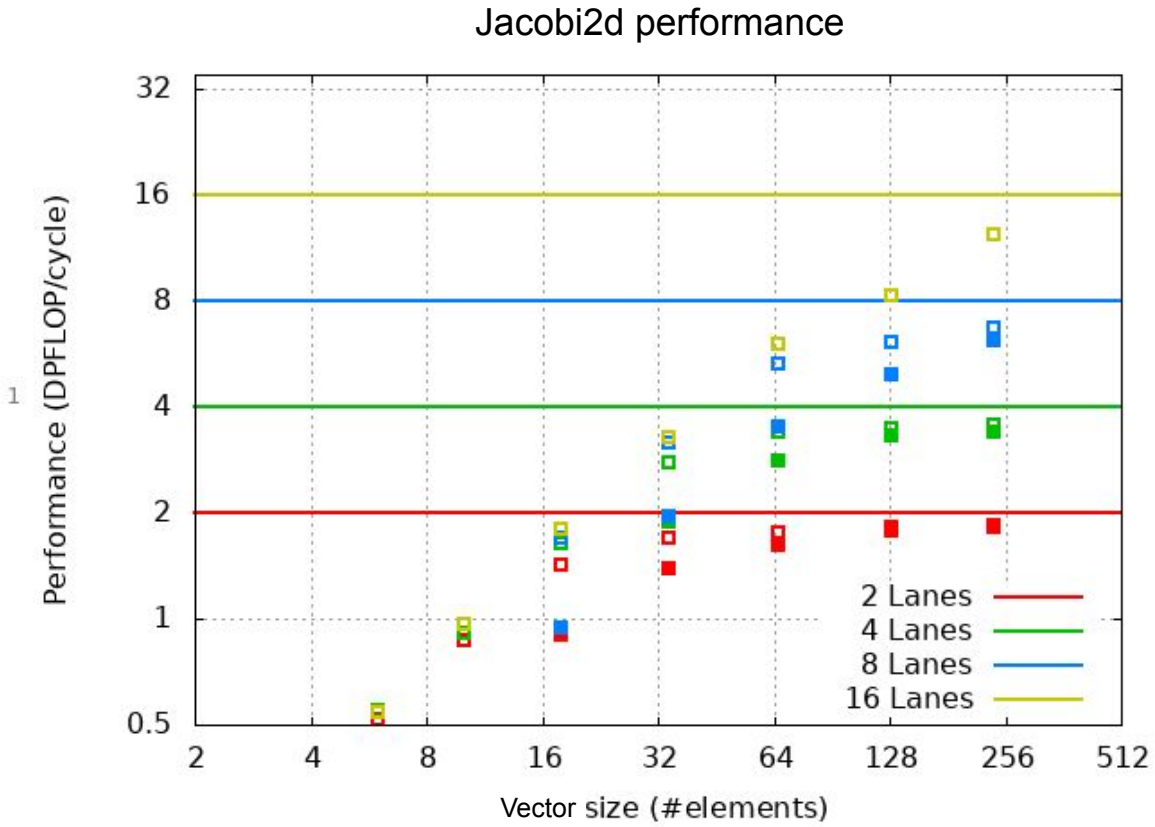
Software



Always Computation Bound

Constant arithmetic intensity!

Software



Always Computation Bound

Ideal Dispatcher System
Default System with CVA6

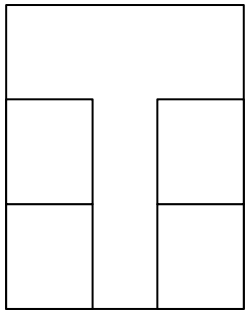
Constant arithmetic intensity!

Hardware

- **Scale up to 8 lanes**
  - Cannot close timing with hierarchical flow
  - Critical convoluted paths
    - From main sequencer to lanes
  - Different parallel trials:
    - Flat-system flow
    - Different floorplan shapes

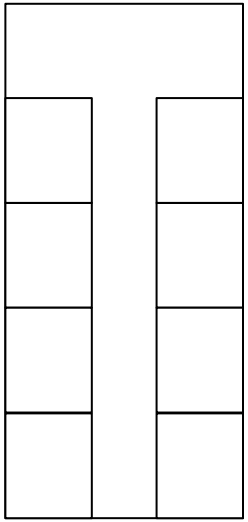
Hardware

Hier - 4 Lanes



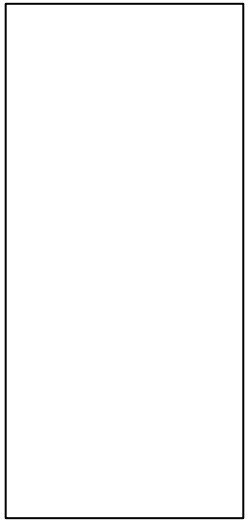
✓

Hier - 8 Lanes



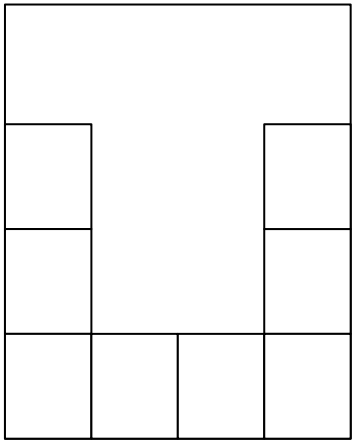
❌

Flat - 8 Lanes



?

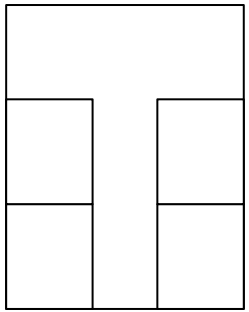
Hier - 8 Lanes



?

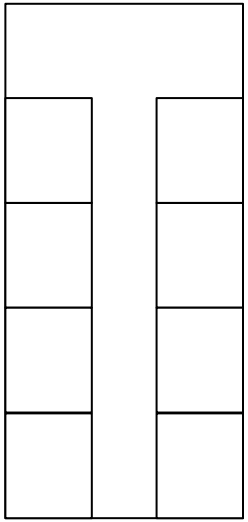
Hardware

Hier - 4 Lanes



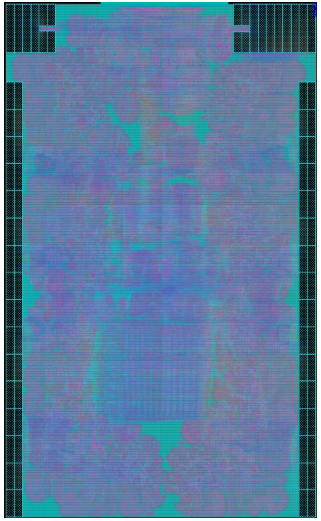
✓

Hier - 8 Lanes



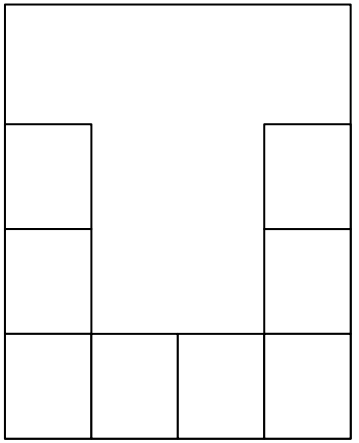
❌

Flat - 8 Lanes



✓

Hier - 8 Lanes



?

Hardware

- **8 lanes - Flat Flow**
  - ~950 MHz in SS
  - Some spurious DRC violations
  - Add space between the caches

- Ongoing:
  - Different die shapes
  - 16 lanes - Flat Flow

Further

- **Software**
  - Merge ideal dispatcher branch
  - Gather performance data + Report
- **Hardware (RTL + Backend)**
  - Try different die shapes
  - Close timing with 16 lanes