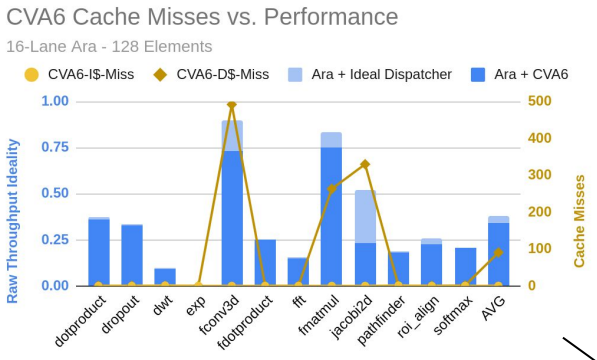
Update on Ara

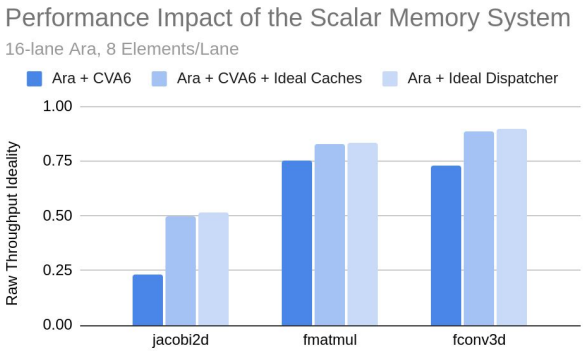
05/04/2023

Matteo Perotti
Matheus Cavalcante

Professor Luca Benini
Integrated Systems Laboratory
ETH Zürich

Correlation between Performance and CVA6's $-Misses



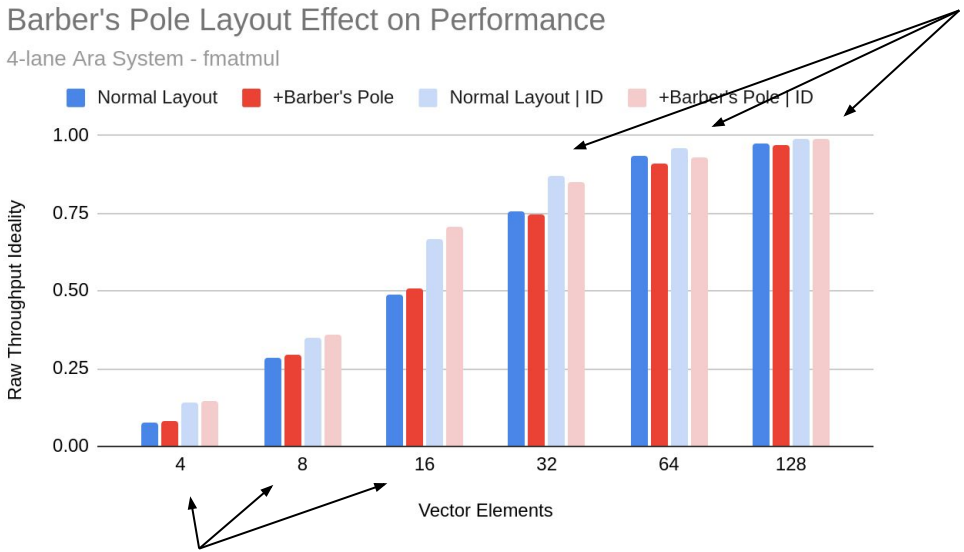


Performance drop mainly due to scalar cache size!

Barber’s Pole Layout can be detrimental for medium vectors

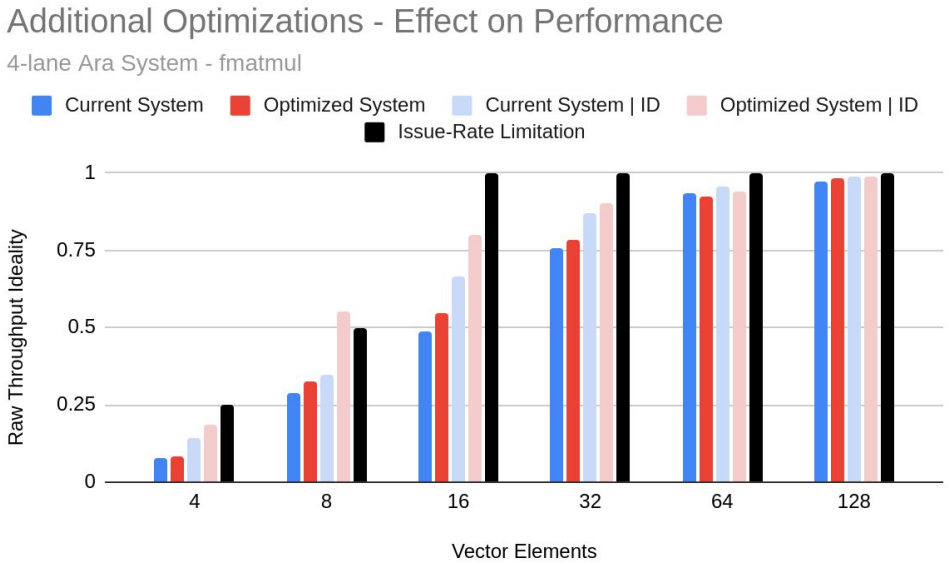
Ara has 8 banks per lane

Perturbate VRF access pattern



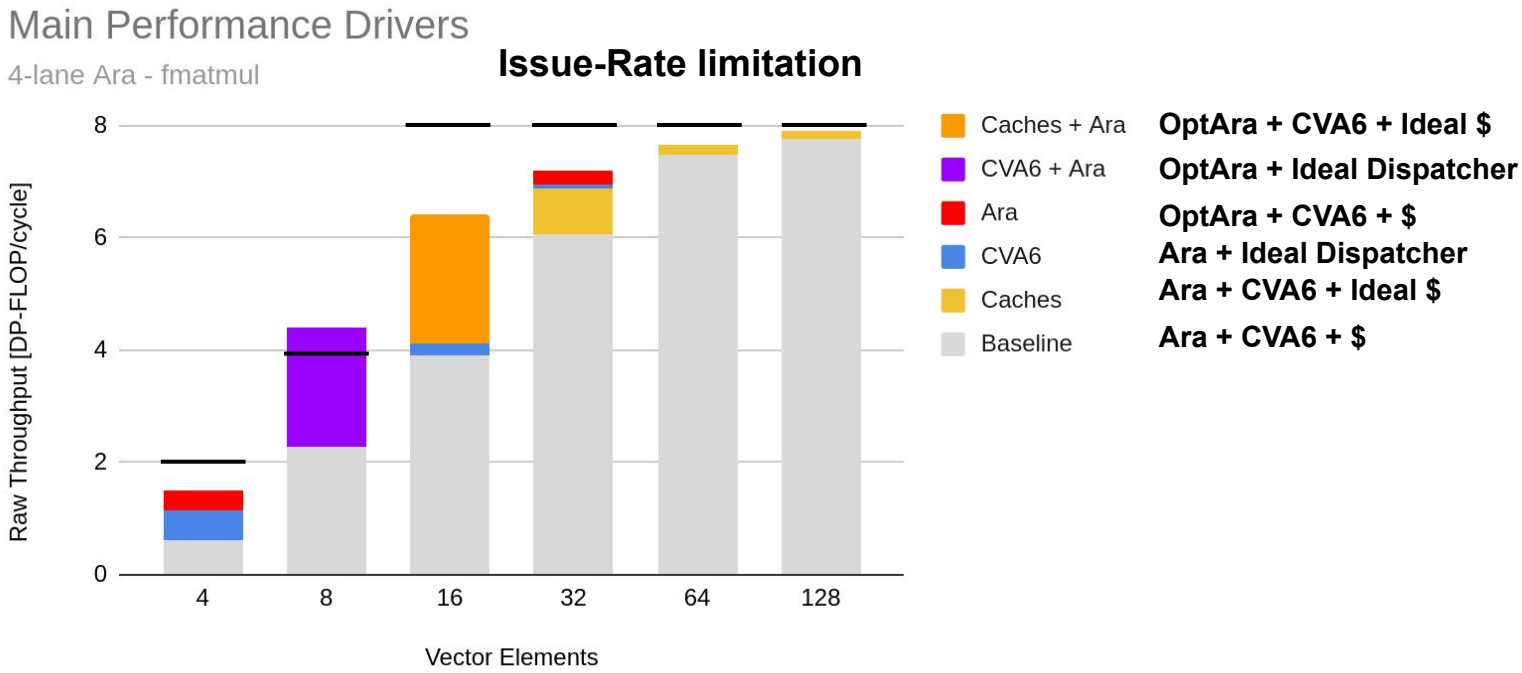
Increase the effective number of banks for short vector applications

Further Ara-optimizations



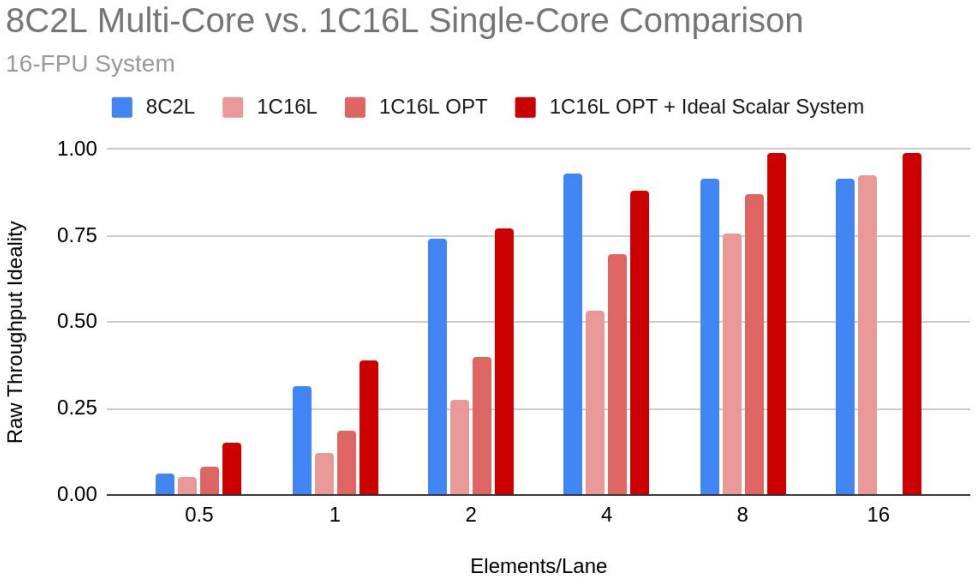
Further optimizing Ara does not boost performance much if not coupled with an improved CVA6 + Caches

Main Performance Drivers



CVA6 limits short vectors' performance, while scalar caches limit medium vectors'

16-FPU - Multi-Core System can overtake Single-Core ID-System



Multi-Core can overcome the issue-rate limitation