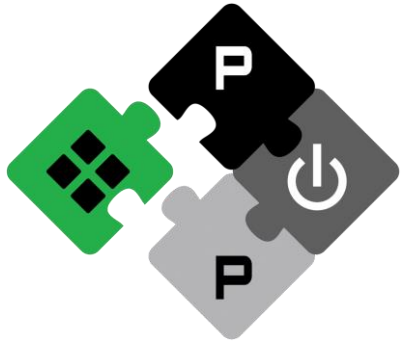


ETH zürich

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Ara: A RISC-V Vector Co-Processor for CVA6

ETH Zurich

Luca Bertaccini lbertaccini@iis.ee.ethz.ch

@pulp_platform

pulp-platform.org

youtube.com/pulp_platform

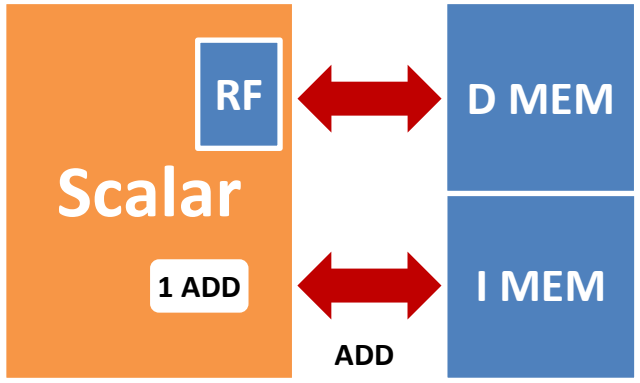
ChipsJU

The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Why Vector Co-Processors?

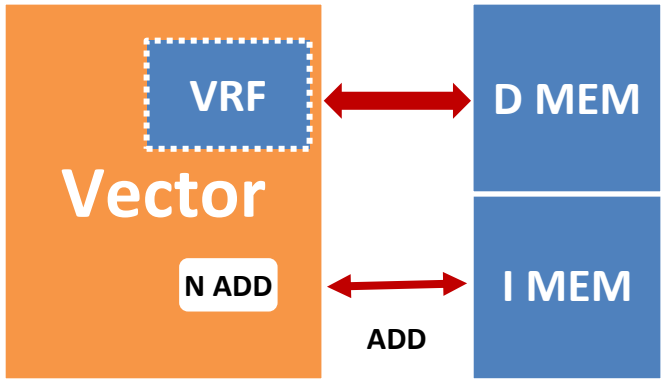
SCALAR CORE

- One instruction – one operation
  - **Higher I-MEM power consumption**
- Either memory or compute operation each cycle on single-issue cores (CVA6)
  - **Lower FPU/ALU utilization**
- Scalar register file → small
  - **Lower data reuse – Higher D-MEM traffic**



VECTOR CORE

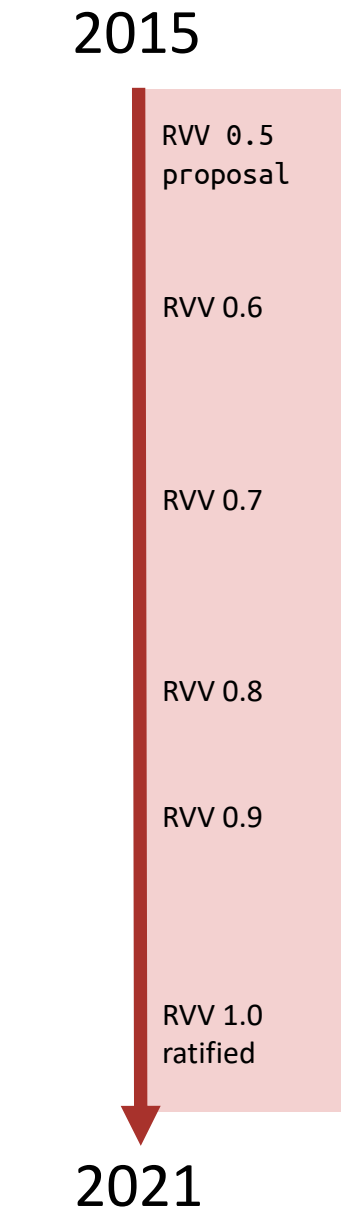
- One instruction – many operations
  - **Lower I-MEM power consumption**
- Concurrent memory and compute operations
  - **Higher FPU/ALU utilization**
- Vector register file → large
  - **Higher data reuse – Lower D-MEM traffic**
  - Programmable vector length (more flexible than SIMD)



The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Based on the RISC-V RVV ISA

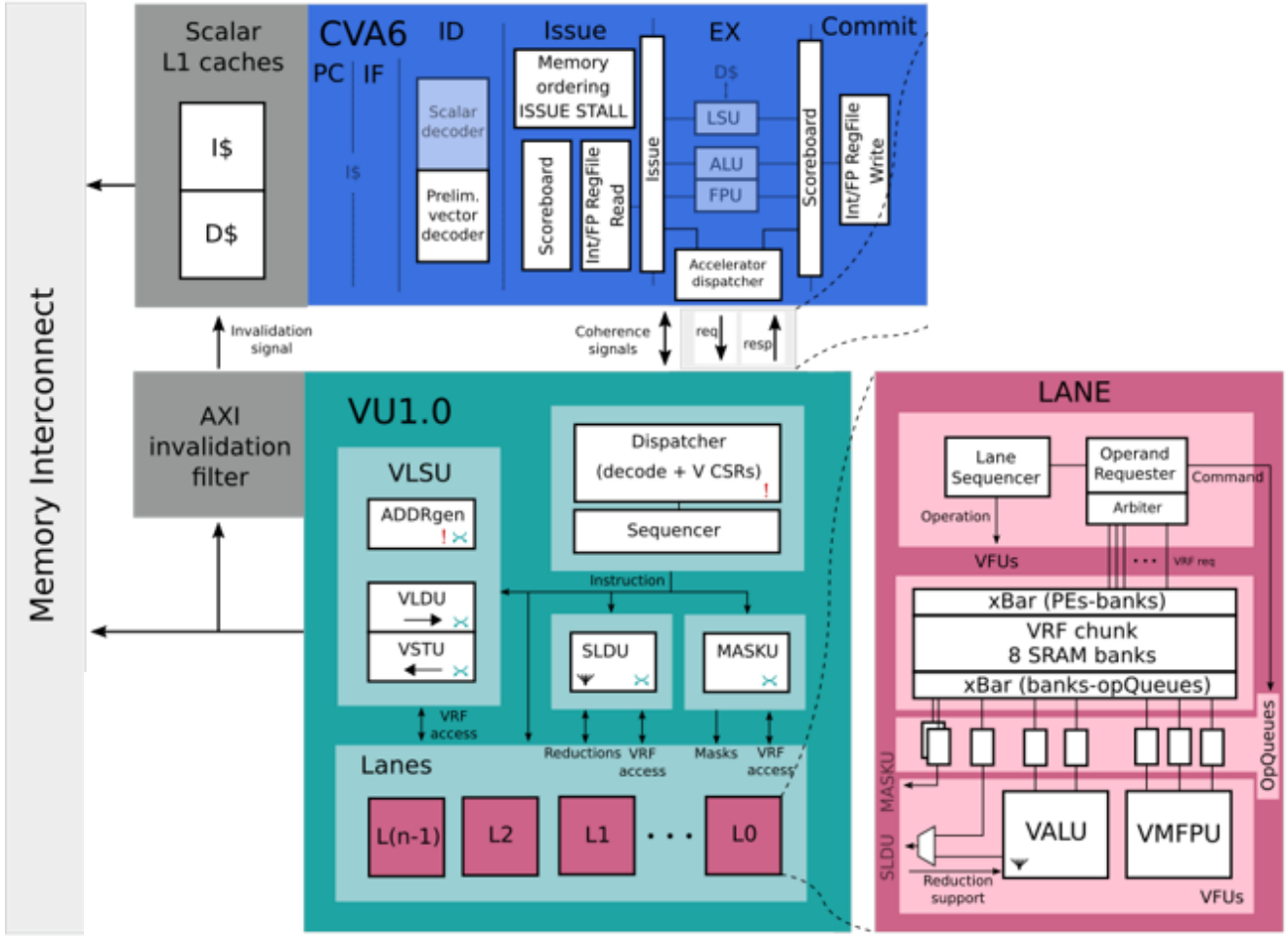
- Early draft in 2015
- Years of refinement. It was ratified in 2021.
- **Known programming model**
- Add **new VRF** with **parametrizable size**
- **Huge** and comprehensive ISA (+300 instructions)
  - VRF setup instructions (VL, element width, ...)
  - Arithmetic instructions (int, fixpt, fp)
  - Memory operations (unit-stride, strided, indexed)
  - Predicated execution (masks)
  - Permutation instructions (slide, reorder)
  - **Vector length agnostic** – High reusability!



The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Ara - PULP Application-Class Vector Processor

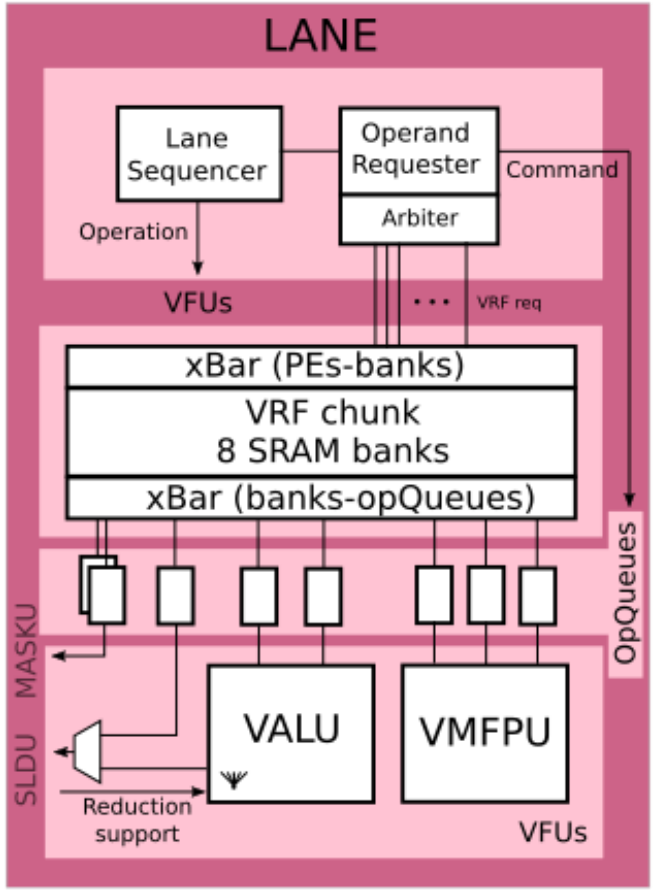
- **CVA6 + Ara**: RV64GCV
- **CVA6**
  - Access to I-MEM
  - Non-speculative V Dispatch
- **Ara**
  - Private VLSU (vload + vstore)
  - Slide Unit (permutations)
  - Mask Unit (predication)
  - **Lanes (computation + VRF)**



The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Ara System – The Lane

- **Computational** heart
  - One 64-bit SIMD FPU/lane
  - One 64-bit SIMD ALU/lane
- **Ara** default **VRF**: 4 KiB * #Lanes
  - 32 Vector Register
  - Vector Register can be merged
  - Split among the lanes
- Keep a **chunk** of the **VRF**
  - Eight 1R/W SRAM Banks
- **Maximize locality**
  - Element 0 remains in lane 0
  - Minimize inter-lane communication
- VRF in-lane reads/writes

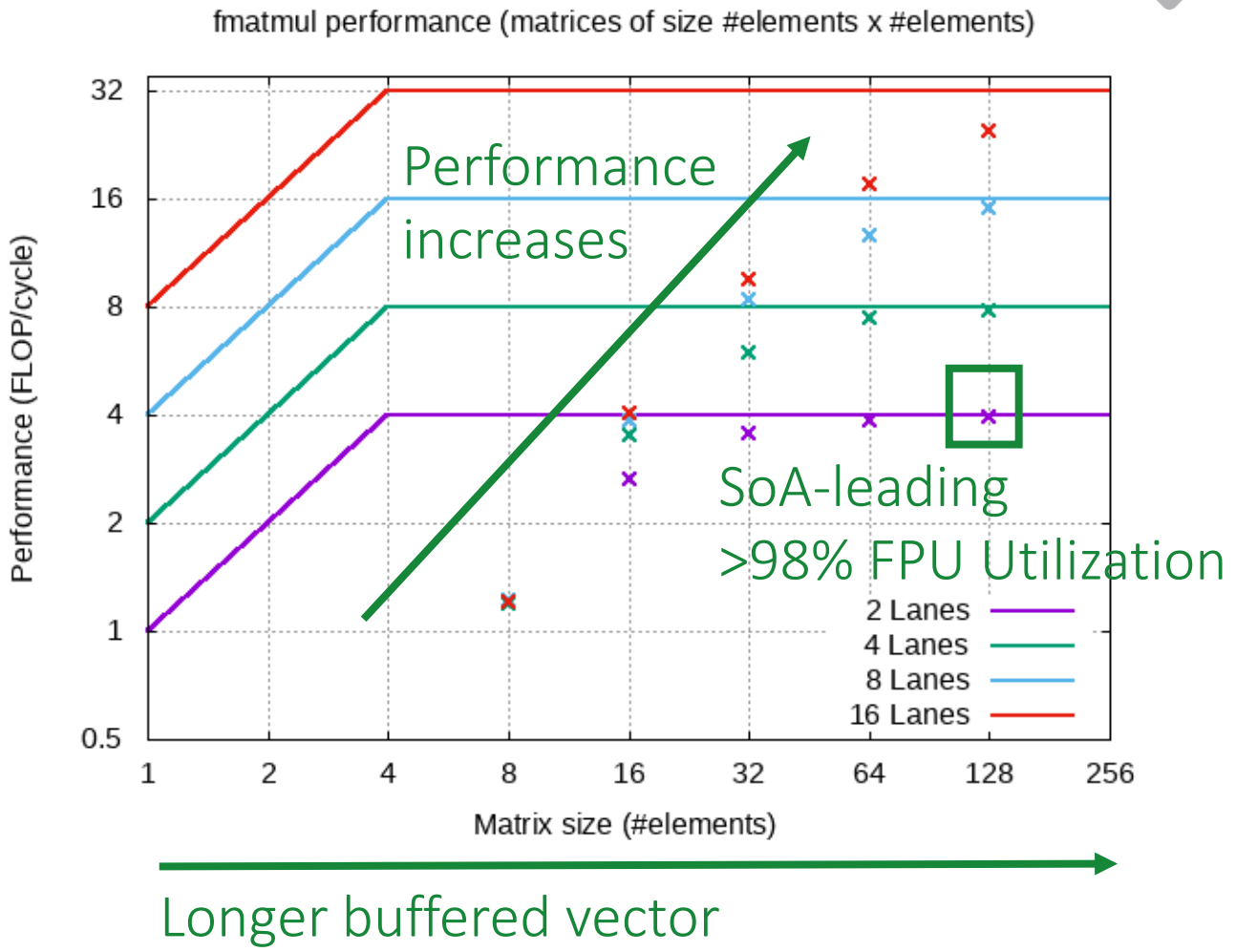


4 Lanes: 8 DP-FLOP/Cycle – 16 SP-FLOP/Cycle!

The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

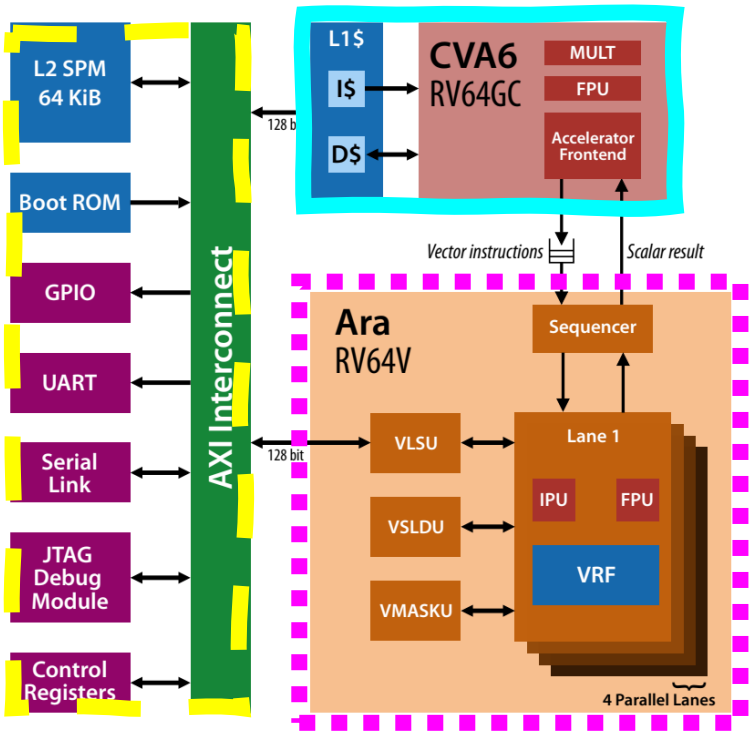
Ara on the Benchmark

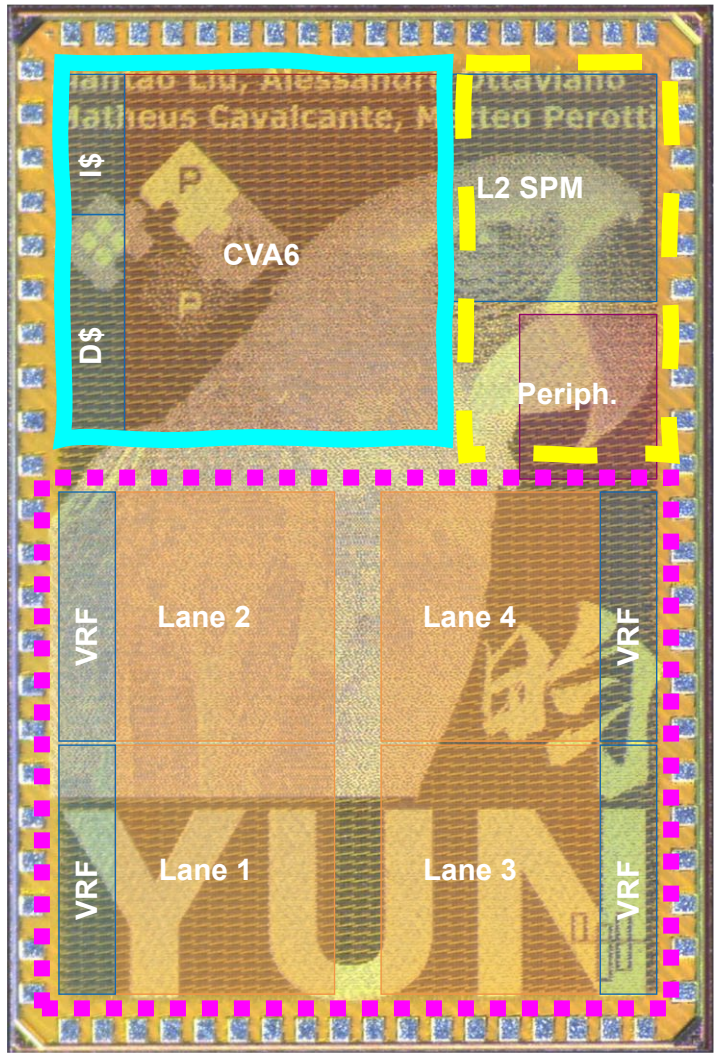
- Up to **32 DP-FLOP/cycle**
  - With 16 Lanes
- FP-matmul
  - A[N][N] * B[N][N]
  - X-axis: Matrix Size N
  - Always computation-bound
- Longer vector, better performance
  - Amortize setup time for V instruction
  - Amortize CVA6 + $ non-idealities
- For large problem sizes Ara achieves close-to-ideal FPU utilization



The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Yun: Silicon Implementation



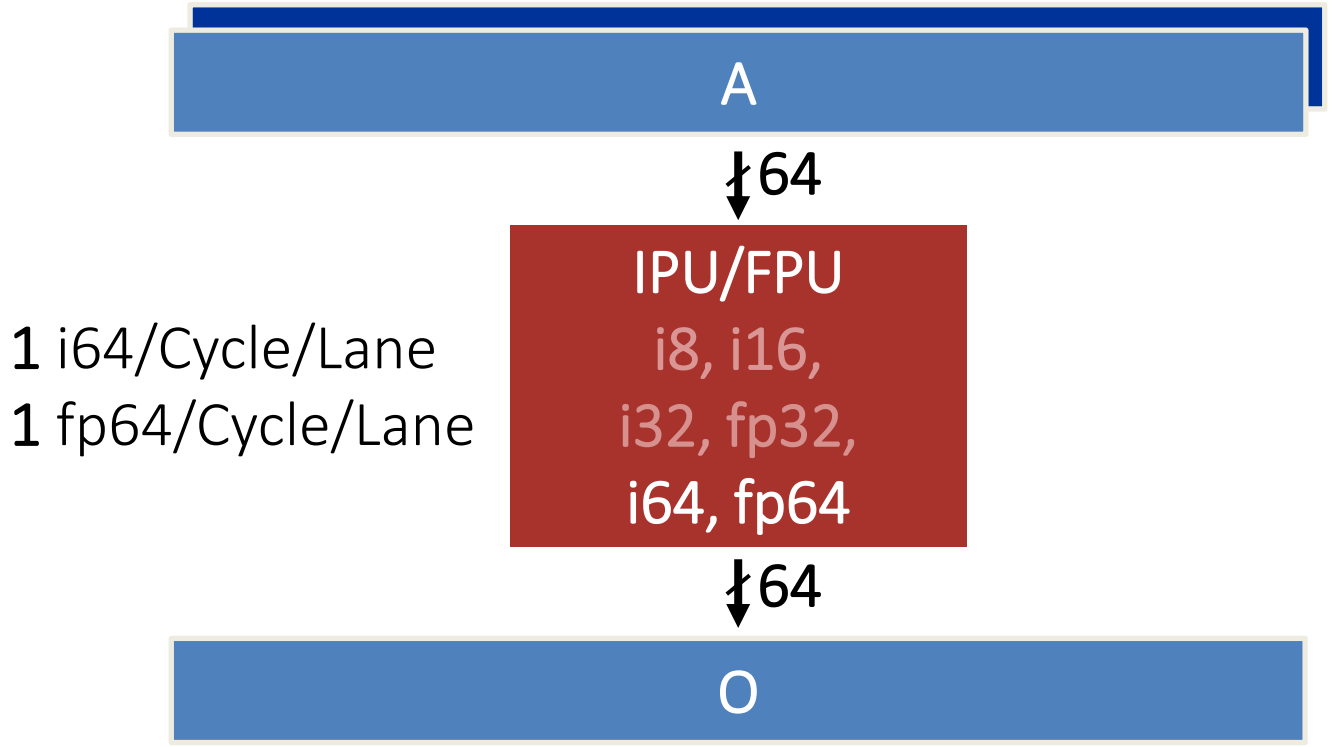


- Ara has been taped-out in the YUN chip
  - CVA6 + 4-lane ARA + Memory + Peripherals
- 6mm^2 chip in 65nm technology
  - Maximum frequency at 1.2V → 280MHz

The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Multi-Precision Support

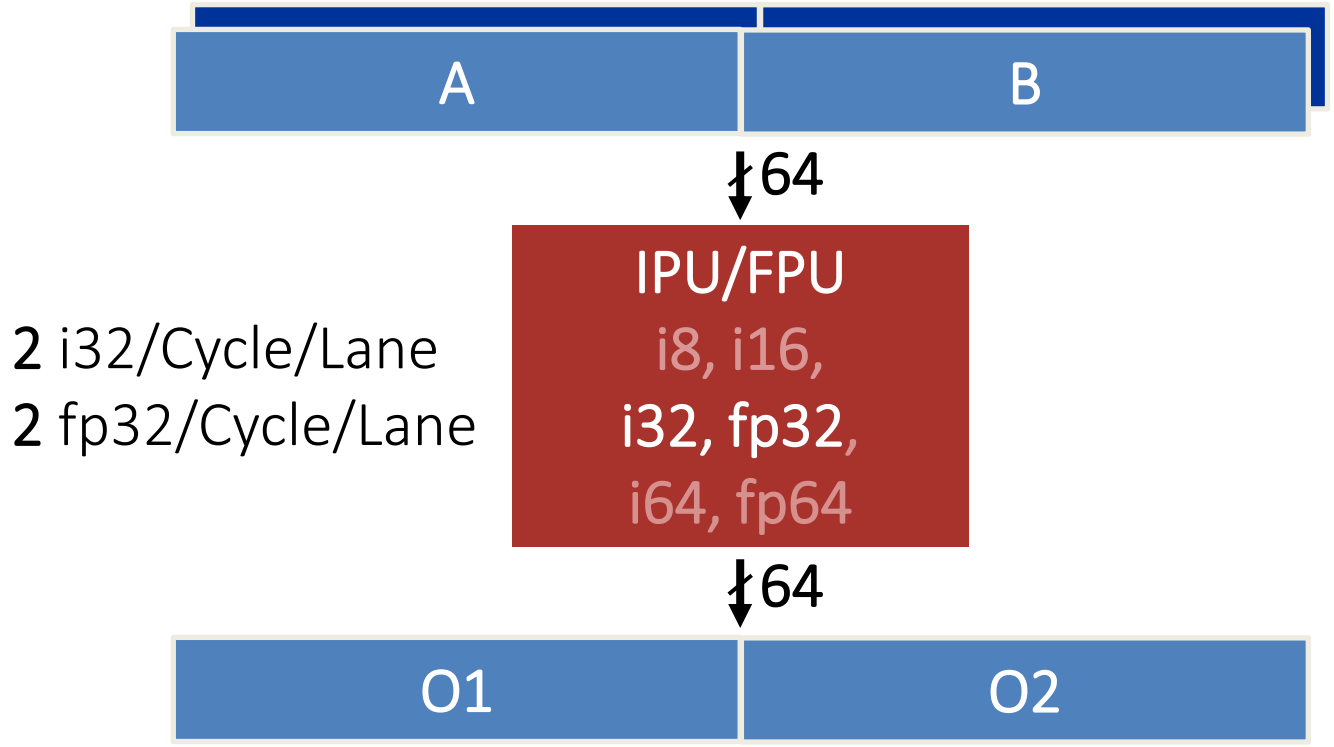
- **Multi-Precision SIMD** IPUs and FPUs
  - SIMD **IPU**: from **8**-bit to **64**-bit elements
  - SIMD **FPU**: from **32**-bit to **64**-bit elements



The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Multi-Precision Support

- **Multi-Precision SIMD** IPUs and FPUs
  - SIMD **IPU**: from **8**-bit to **64**-bit elements
  - SIMD **FPU**: from **32**-bit to **64**-bit elements

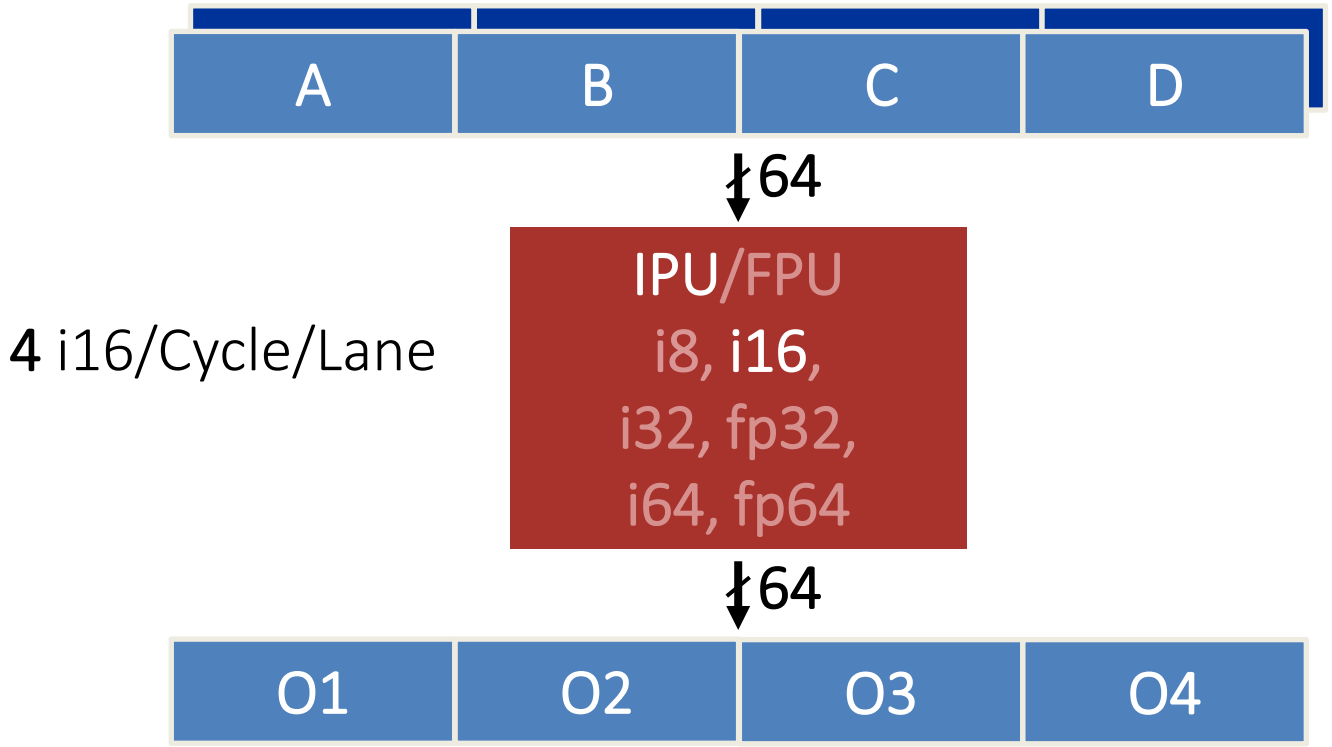


2 i32/Cycle/Lane
2 fp32/Cycle/Lane

The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Multi-Precision Support

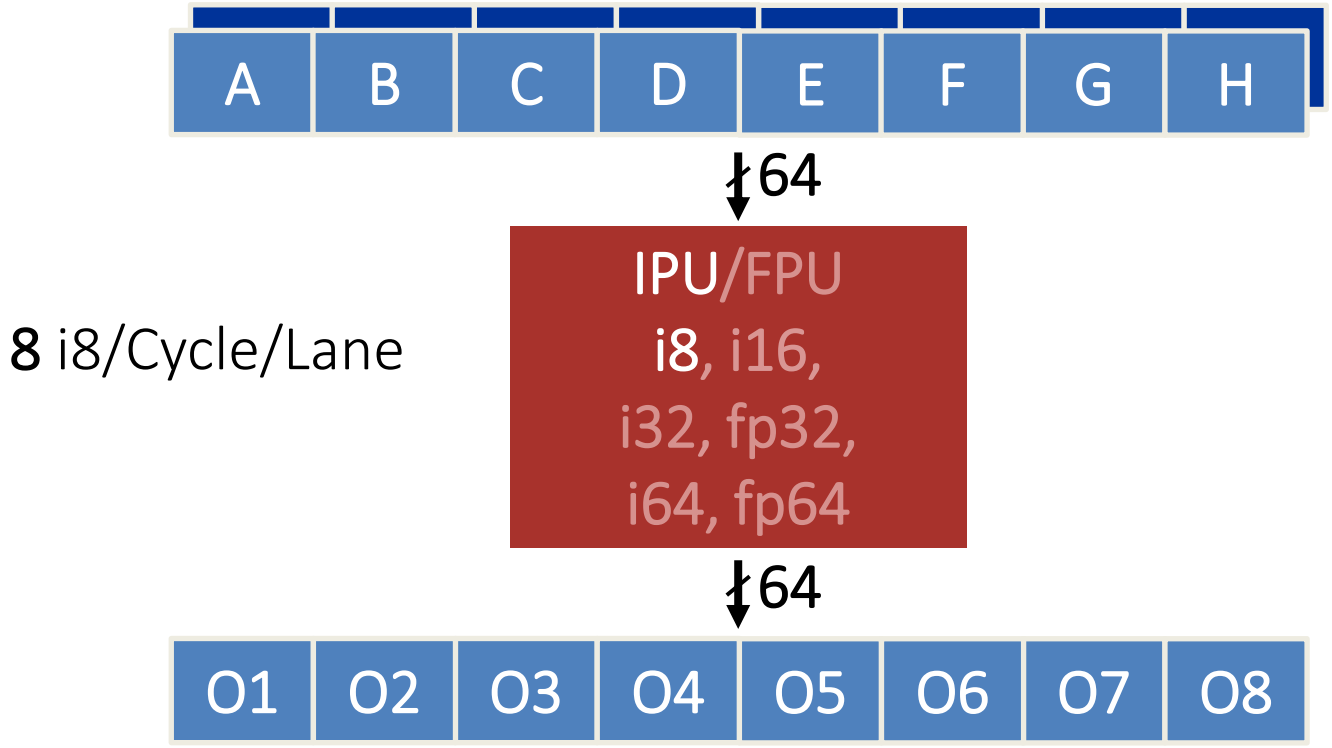
- **Multi-Precision SIMD** IPUs and FPUs
  - SIMD **IPU**: from **8**-bit to **64**-bit elements
  - SIMD **FPU**: from **32**-bit to **64**-bit elements



The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

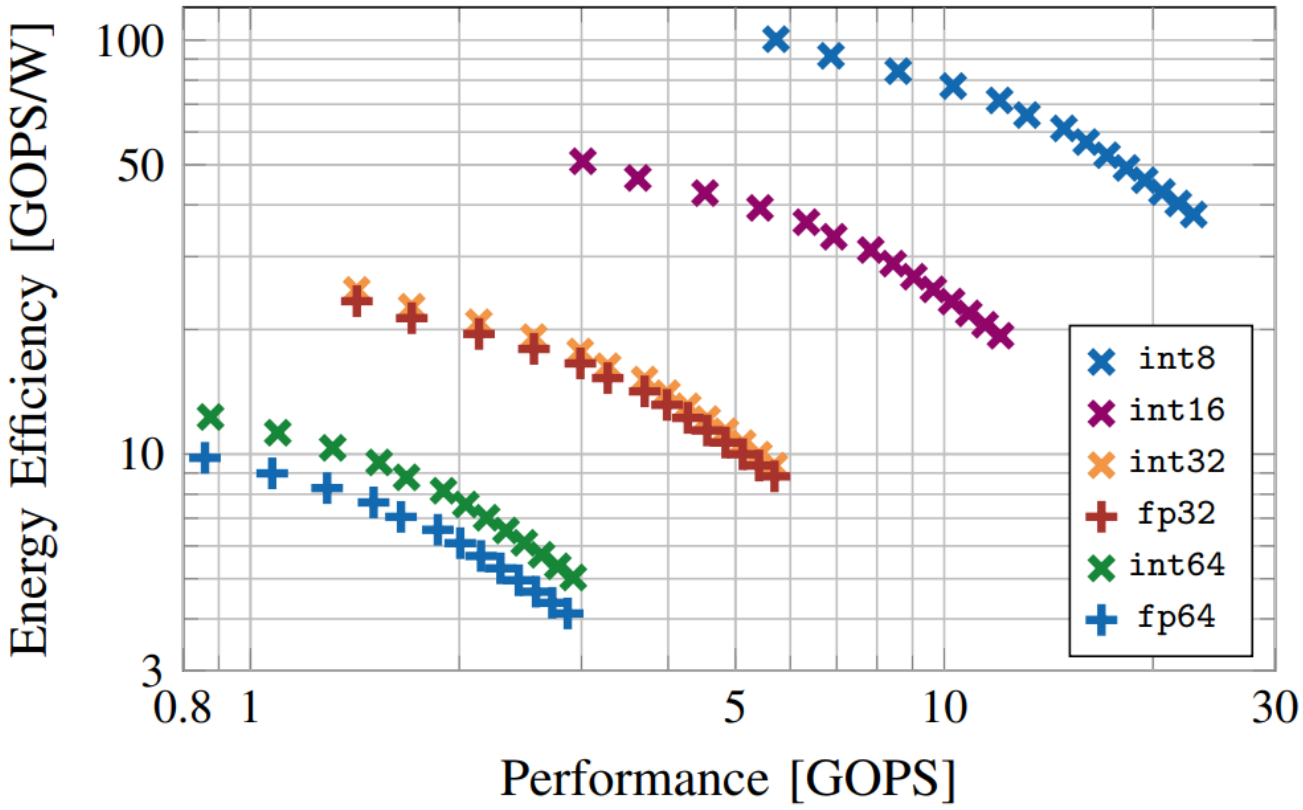
Multi-Precision Support

- **Multi-Precision SIMD** IPUs and FPUs
  - SIMD **IPU**: from **8**-bit to **64**-bit elements
  - SIMD **FPU**: from **32**-bit to **64**-bit elements



The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Yun: Energy Efficiency vs. Performance



- Multi-precision support 8-64b integer + 32-64b floating-point support
- Up to 100 GOPS/W @ int8

The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Ara is Available Open-Source!

- Available open-source on GitHub
  - RTL
  - C applications
  - Testbenches
- Exhaustive README covering details on how to compile applications with LLVM and run RTL simulations with Verilator

`github.com/pulp-platform/ara`

The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Ara: Next Steps

- Ara + MMU for OS support
- Ara scale-up (increasing lane count)
- Support for matrix instructions
- Microscaling (low-precision FP) formats

The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Spatz: A Compact and Efficient Vector Processor

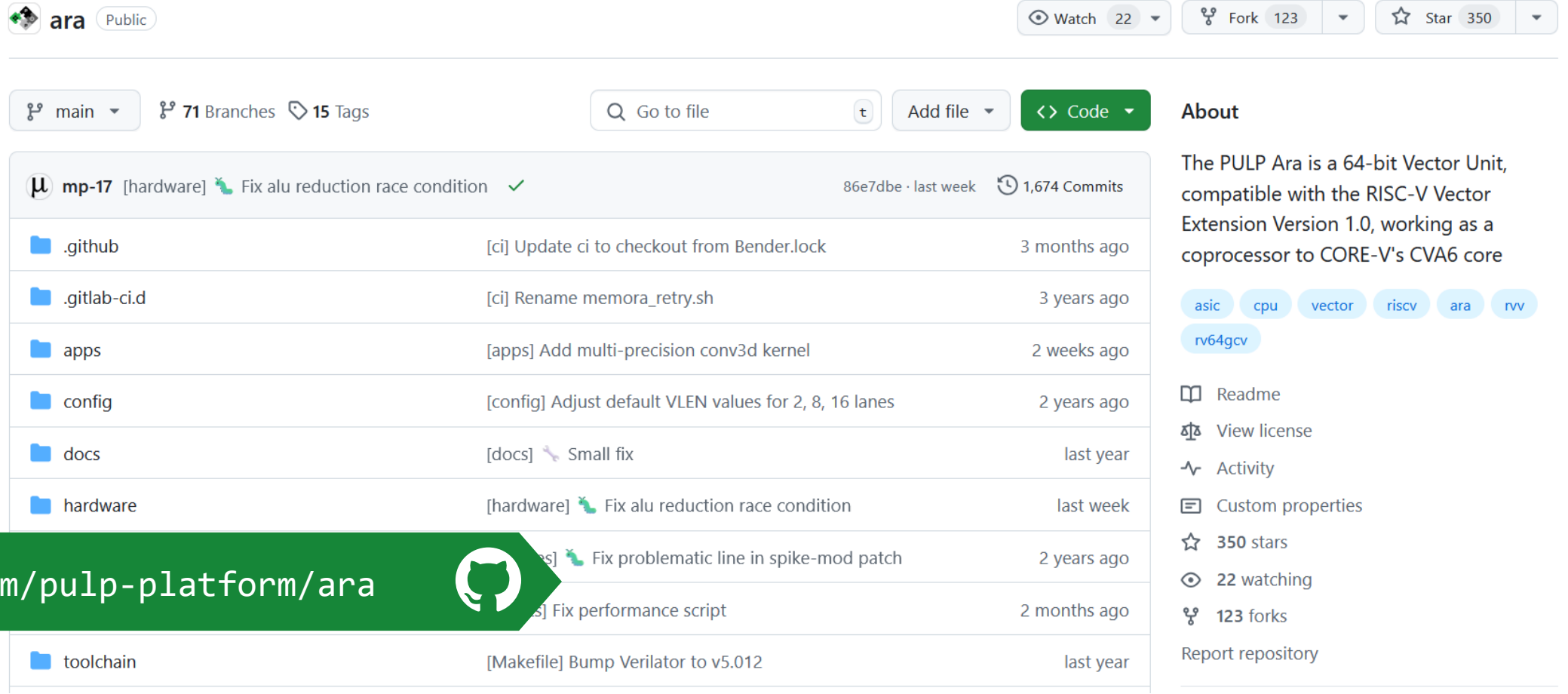
- Check out also Ara's little sibling: **Spatz**

- **Spatz** is a compact and efficient vector processor based on a subset of the RVV extension

`github.com/pulp-platform/spatz`

The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .

Ara is Available Open-Source!



github.com/pulp-platform/ara

The TRISTAN project, nr. 101095947 is supported by Chips Joint Undertaking (CHIPS JU) and its members Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Israel, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, Turkey .