Simplifying AI at the Edge.
Accelerating Computer Vision.

Accelerating Edge AI with RISC-V
Florian Zaruba
Hardware for AI - November 22nd, 2023
About Me

• Working with RISC-V since 2015
• Mostly worked on open-source HW projects: PULPino, Ariane/CVA6, Snitch
• PhD about energy-efficient computer architectures at ETH Zurich from 2017 – 2021
• Director at OpenHW Group since 2020
• RISC-V Ambassador since 2021
• Founding member of Axelera since November 2021: Responsible for CPU subsystem (and many things more)

florian.zaruba@axelera.ai
@be4web
linked.in/zarubaf
github.com/zarubaf
About Axelera

• Founded in July 2021 by core team from IBM, IMEC, ETH, Google, Qualcomm

• Accelerate **computer vision** with proprietary digital in-memory computing (D-IMC)

• In November 2021 we were around a total ~15 people (total!)

• One test-chip with D-IMC taped-out

• Runway of ~12 months to prototype TO (fully functional)

• Team spread (Netherlands, Switzerland, Belgium, remote)

• Onboarding of new members

• **Essential to hit the ground running!**
Architecture
Compute Datapath

L1
Local Memory
- Input, output data
- DMA transfers and explicit data management

4MB L1

D-IMC
Matrix-Vector Multiplication
- Large sets of weights
- Load and execute
- Computational cells in memory array

VEC
Vector Engine
- Element-wise vector operations and activations

AUX
Auxiliary Engine
- Pooling
- Depth-wise convolutions
Architecture

AI Core

- RISC-V
  - AI Core Controller
  - Networks change and evolve
  - AI field is rapidly evolving, don't bake structure into hardware
  - Flexible and programmable control flow

- Uncore
  - 4MB L1
  - D-IMC
  - VEC
  - AUX
Architectural Features:

**RISC-V System Controller**
- SoC Management
- Booting
- Workload dispatch
- Operating system support
- Tight timeline: Go for IP

**RISC-V Root of Trust**
- Models are the prime assets of our customers
- Authenticated booting and encryption are a must from a product perspective
- ✓ RISC-V based Root-of-Trust

Overview:
- 4MB L1
- D-IMC
- Uncore
- VEC
- AUX
- 32MB L2
- PCIe
- LPDDR4x

**RISC-V AX25**
- D-IMC
- Uncore
- VEC
- AUX

**Root of Trust**
- RISC-V AX25
- D-IMC
- Uncore
- VEC
- AUX

**AX25**
- 4MB L1
- D-IMC
- Uncore
- VEC
- AUX

**AX25**
- 4MB L1
- D-IMC
- Uncore
- VEC
- AUX

**AX25**
- PCIe
- LPDDR4x
RISC-V from an Integrator’s Perspective

Opportunities – The Market

• Very good IP landscape: Many IP providers in the low (to mid) end

• Choice creates a healthy market, this is especially important for startups

• Accelerated time-to-market: Contact to RTL in hand ~3 weeks

• Complete systems and end-to-end solutions

• IP comes with interrupt controller, core-local interrupt controller, debug module etc.

• Significant advantage to other ISA(s) and core providers
RISC-V from an Integrator’s Perspective

Opportunities – The Software Stack

- The software stack is expensive and complex!
- Impossible to change ISA from generation to generation with tight timelines
- (Practically) no vendor lock-in
- An open ISA means freedom to (re-)choose
- Creates a healthy ecosystem
- Keeps the door open for own developments in the future
RISC-V from an Integrator’s Perspective

Opportunities – The Software Stack

• Availability of open-source, understandable, and extendable reference models
• Fully system simulator: spike and QEMU
• Models complete system: Host and device

• From concept to tape-out in 1 year: You need to have a reference platform
• spike interacts with Axelera’s HW model
• Thanks to permissive license easy regression model

• SW bring-up before FPGA and emulation is ready
RISC-V from an Integrator’s Perspective

Opportunities – Hardware Availability

- The availability of open-source hardware is a game changer
- CVA6, CV32E40, Rocket, Boom etc.
- Maturing into real, production-ready cores

- Enables prototyping from day zero
- No license fees, no hurdles

- Software stays compatible
RISC-V from an Integrator’s Perspective

Opportunities – The Learning Curve

• Software and firmware teams were up-to-speed extremely fast
• Most did not have prior RISC-V experience, nevertheless thanks to:
  
  Simple ISA, internal introduction in one afternoon
  Good and understandable low-level software to learn from (OpenSBI, U-Boot, etc.)

  Getting up to speed was very smooth!
RISC-V from an Integrator’s Perspective

Challenges – Rate of Adoption

• Maturity of ISA increasing
• IP vendors still catching up

• Many problems have been solved in a vendor specific way because there wasn’t a specification ratified
• CMOs, ECC, etc.

• Other custom vendor extensions
• Vendor “lock-in” with low-level software
• Risk to the aggressive timelines of a start-up
RISC-V from an Integrator’s Perspective

Challenges – Interoperability

- Interoperability between vendors
- Semantic might be defined, reality provides challenges:
  - Atomic handlings
  - (Non-)coherency
  - PMAs
  - Bus interface

- Platform specification to the rescue?
- Reference platform? Integrators guide?

- The devil’s in the detail
RISC-V from an Integrator’s Perspective

Challenges – Growing Pain

• Keeping up with ecosystem developments is challenging

• Many different task groups, impossible to subscribe to them all just for tracking

• Main ISA document struggling to keep up

• Spec development lives in different places for different extensions (i.e., CMOs)

• Possible mitigation:
  • Tech newsletter?
  • Overview/status page?
RISC-V from an Integrator’s Perspective

Bonus

• Love the ecosystem:
The Metis Platform

- From concept to tape-out in ~12 months
- First RISC-V booting hours after chip-back
Axelera Today

• Little over two years old, now with 130+ people, 45+ PhDs and 20 nationalities, with offices in Eindhoven (NL), Zurich (CH), Leuven (BE), Milan (IT), Bristol (UK)

• Raised USD 50 million in just 20 months

• designed and taped out METIS, the most powerful AI Processing Unit for Edge AI based on proprietary Digital In-Memory Computing and RISC-V technologies

• RISC-V is ready for prime-time!
Get In Touch

• Pass by our Demo Booth (S17) to learn more about our technology

• Make sure to get in contact!

• Continued focus of vision applications

• We are further expanding our RISC-V footprint

• Come join us!

https://www.axelera.ai/careers/
Keep in Touch!

Address
HTC5, High Tech Campus
5656 AE Eindhoven
The Netherlands

E-mail
florian.zaruba@axelera.ai

www.axelera.ai