

Simplifying AI at the Edge. Accelerating Computer Vision.

# Accelerating Edge AI with RISC-V

Florian Zaruba

Hardware for AI – November 22<sup>nd</sup>, 2023

### About Me

- Working with RISC-V since 2015
- Mostly worked on open-source HW projects: PULPino, <u>Ariane/CVA6</u>, <u>Snitch</u>
- 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)









#### About Axelera

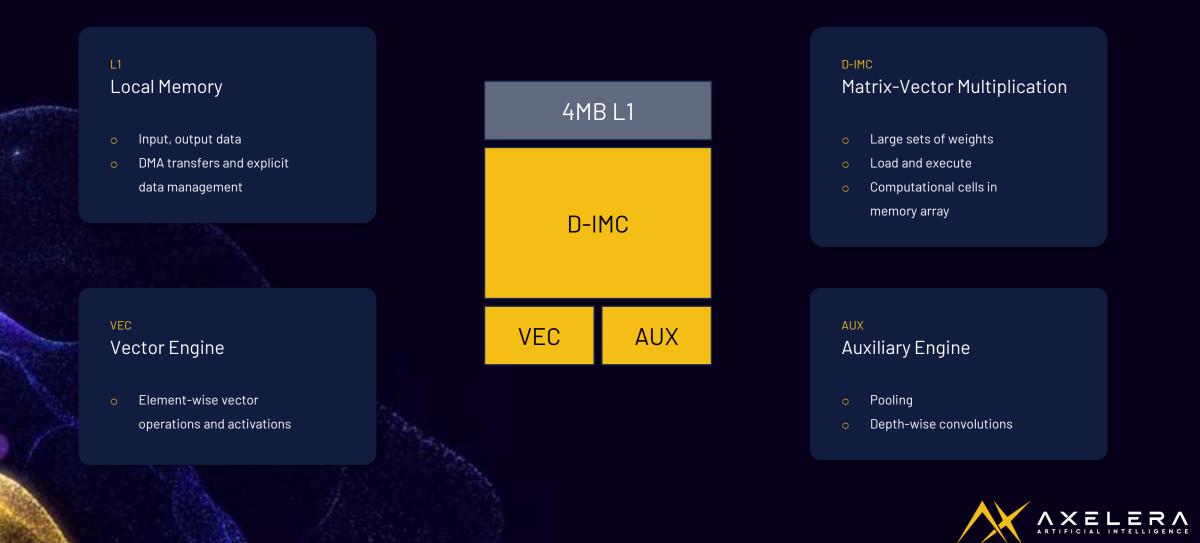
- Founded in July 2021 by core team from IBM, IMEC, ETH, Google, Qualcomm
- Accelerate computer vision with proprietary digital inmemory computing (D-IMC)
- In November 2021 we were around a total ~15 people (total!)
- One test-chip with D-IMC taped-out
- Runway of ~12 months to prototype TO (fully functional)
- Team spread (Netherlands, Switzerland, Belgium, remote)
- Onboarding of new members
- Essential to hit the ground running!





#### Architecture

#### Compute Datapath

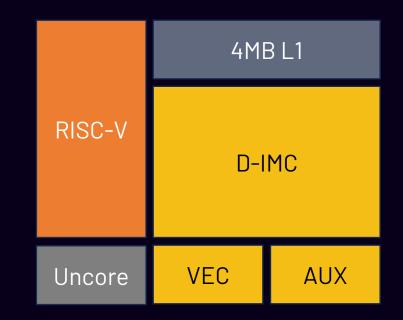


## Architecture

Al Core

#### RISC-V Al Core Controller

- Networks change and evolve
- Al field is rapidly evolving, don't bake structure into hardware
- Flexible and programmable control flow





#### Architecture

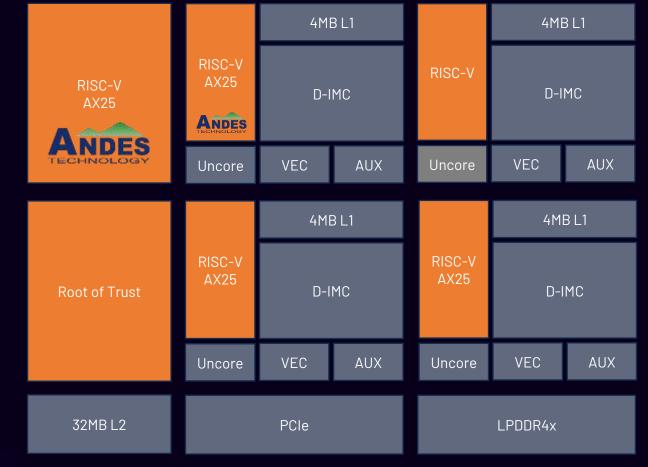
#### Metis

#### RISC-V System Controller

- o SoC Management
- o 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

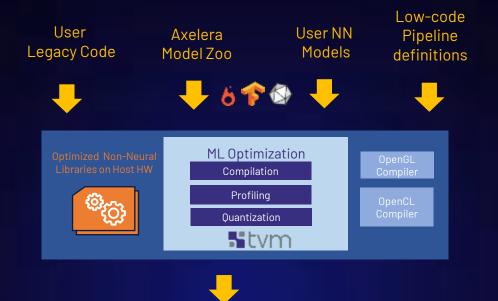




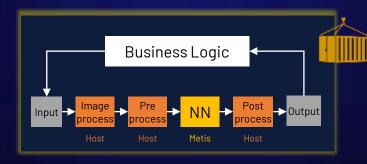
Opportunities – The Market

- 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





Ready-to-run inference pipeline



Deploy to Metis-enabled Edge



#### 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



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

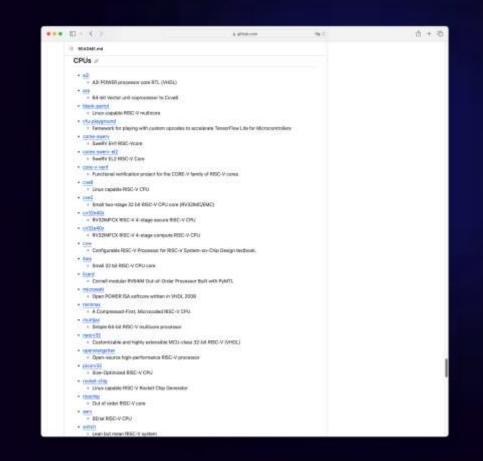






Opportunities - Hardware Availability

- 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







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!



# Zephyr'



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

#### ...

```
static inline unsigned long x pma_read_cfg(unsigned int pma_cfg_off)
#define switchcase_pma_cfg_read(__pma_cfg_off, __val)
    case __pma_cfg_off:
        __val = csr_read(__pma_cfg_off);
        break:
#define switchcase_pma_cfg_read_2(__pma_cfg_off, __val)
                                                            1
    switchcase_pma_cfg_read(__pma_cfg_off + 0, __val) \
    switchcase_pma_cfg_read(__pma_cfg_off + 2, __val)
    unsigned long ret = 0;
    switch (pma_cfg_off) {
    switchcase_pma_cfg_read_2(X_PMACFG_0, ret)
    default:
        sbi_panic("%s: Unknown PMA CFG offset %#x", __func__, pma_cfg_off);
        break;
    return ret;
#undef switchcase_pma_cfg_read_2
#undef switchcase pma_cfg_read
```



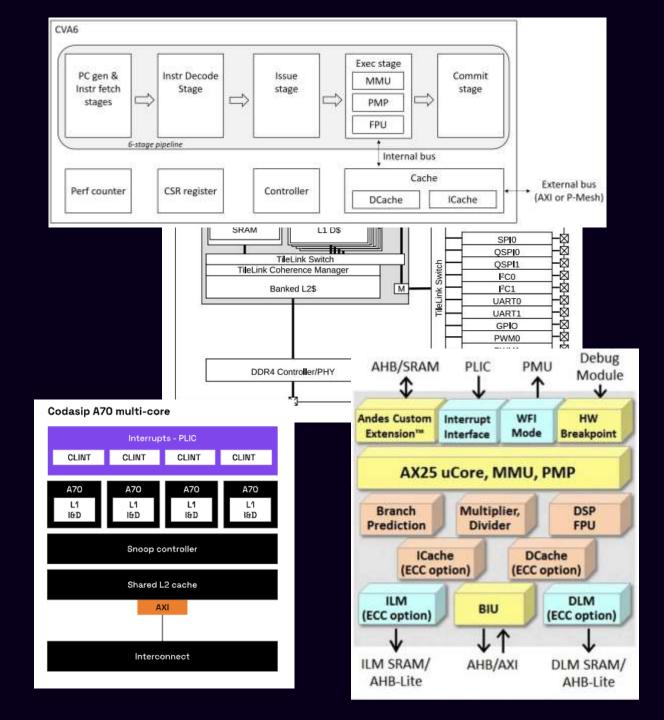


Challenges - Interoperability

- 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



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?

```
000 E - < >
                                                                                                                                                                                               白 + 〇
                                                                                               il hels-tacy.org
                                                                                                                                           Fe.C.)
       RISCV
                                                                                                                                                                        CEL- OHAD HUDAN
                                    A Man . C. Subgroups
 💏 Harre
 C Butanuar
                                     Main General
                                    RESC-Y MAIN GROUP
                                                                                                                  Welcome to the RISC-V Working Groups making list portal. RISC-V International is an open
                                                            "RESC-V Working Groups Portal" ***
                                    standard non-profit organization managing the IP and development activities for the RISC-V instruction Set Architecture (ISA), an open standard hardware initiative that is
                                    repidly transforming the way microprocessors are made. The primary website for the RISC V architecture is at https://www.org. The making lists on this portal are.
                                    Created: 12/14/18
                                    6,440 Members, 9 Taples, Avdraw is visible to members prio, Restricted
                                  Subgroups You Can Join
                                    Technical Task Groups & Committees
                                                                                                                                                                                                       4
                                    PESC-V Technical Task Groups & Committee
                                     anon-tools-software
                                    Application and Tools Horizontal Committee (HC) Welcome to the App & Tools HC. The goal of this task group is to coordinate efforts to build the REC-V anthease
                                     ecosystem and to standardize RISC-V activary interfaces.
                                    Created: 12/17/18
                                    378 Members, 154 Topics, Archive is visible to members only, Last Post: Oct 13
                                    CTO: Development Partners
                                    RISC-V Development Partners This group is deduated to providing support for specification development by providing engineering resources to help get the specification
                                    tasks done for nitification. Activities are tracked in the development pertner status spreadsheld (
                                    https://docs.google.com/spread/read/1_uXyC2WKgyeohd5oYT/8Aw207AvI5-6-NOPag0LTNa/1_If you're interested in becoming a development partner please email
                                    Created 9/90/20
                                    diff Members, 156 Topics, Archive is visible to members may Last Post: Col 20
                                    Jas-infra
                                    The ISA Infrastructure HC coordinates efforts of SIGs that are not ISA but are needed to ISA development. This includes: " Architectural Compatibility Tests to ensure vendor
                                    implementations are SA compatible. ' Simulators to provide references for compatibility. ' Continuous Integration to ensure charges to tests, simulators, and tools don't
                                    break exititing code. * Devops for computing resources needed for the above, and * Documentation to communicate ISA development.
                                    Creative: 2/24/22
                                    66 Mampers, 1 Rock: Arctive is visible to members with Last Post, 10/05/22
                                    priviaged-software
                                    Welcome to the Privileged Software list. This lists has content related to work groups that tail under the Privileged SW HC unterelis. Expect announcements and other
                                    communications about all things related to privileged software the, kernel, etcl.
                                    Created: 6/06/22
                                    118 Members, 46 Tablos, Archive is visible to members only. Last Post: Oct 4
                                    PESC-V Security Horizontal Committee Man Goals. • Promote RISC-V as an ideal vehicle for the security community • Laste with other Internal RISC-V committees and
                                    with external security committees + Qwate an information repository on new attack liveds, literals and countermeasures + Identify top 10 open challenges in security for the
                                    PESC-V community to address + Propose security committees (Menseling or Technical) to tackle specific security topics + Recruit security talent to the RSC-V ecosystem.
                                    Created 12/19/19
                                    543 Members, 362 Topics, Archive is visible to members why. Last Post: Oct 22,
                                    BISC-V becmants
                                    This making list is devoted to fostering insightful discussions on mapping diverse segments to their respective RISC-V profiles, with a focus on achieving optimal alignment.
                                    and impinization
                                    Created, May 8
                                    34 Mampers, 15 Tapics, Anthive is visible to members only. Last Post: A/ 5
                                     soc-infra
                                     Welcome to the SOC Infrastructure Horsontal Committee. The SOC infrastructure Horizontal committee contains but not limited to the components that straddle the
```



Bonus

• Love the ecosystem:

🥶 🗉 • K 🕗		🔒 kupleb.gitlab.io 🖒	凸 +
		🚺 0x38853023 - rvcodec.ja	
SC-V Instruction Encod	er/Deco		
		[ Instruction ]	
	Þ	x38853023	
		ed x8, 896(x10)	
		0011 1000 1000 0101 0011 0000 0010 0011	
		9x38853023	
		5-type	
Instruction		EV641	
		1/Master	

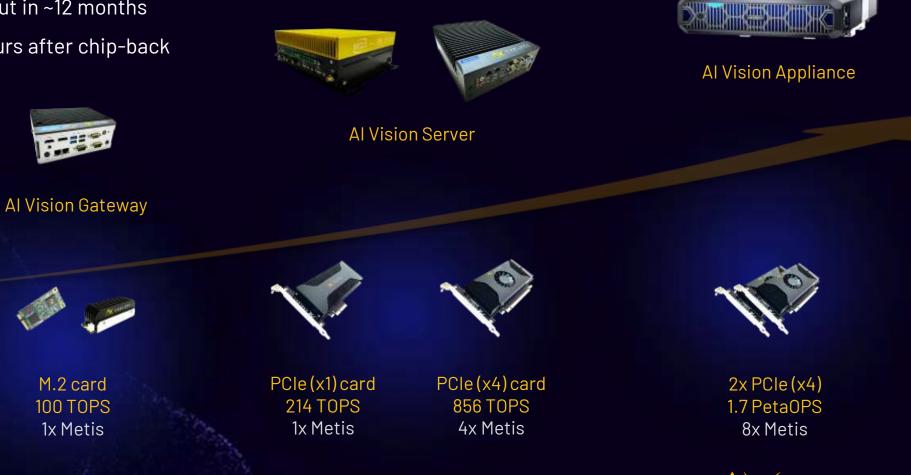
ļ	

<b>000</b> - C (	i nda	nia: 🔫			۵.
	<b>R</b> 8190-	VISA			
v64i2p1ma2p1f2p2d2p2g_zifences	i zk zvkn zvks <mark>SV39</mark> zVe6	4f1_sscofpnf8p1_xpn			
XLEN = 64					
1 v2.1 52 Base Integer Instruction Set 🔂					
m 52 13 Integer Multiplication and Division (	8				
v2.1 22 Atomic Instructions &					
1 v2.2 30 Single-Precision Floating-Point of					
d v2.2 32 Double-Precision Floating-Point					
C VI 33 Compressed Instructions &					
Elfenced 12 1 Instruction-Fetch Fence G					
sk 31 Scelar cryptography 6					
akn 31 Crypto, NEST Algorithm Suite 🖉					
zbkb 12 Bitmanip instructions for Cryptogra	ehy 🕢				
zbicz 2 Carry-less multiplication for Cryptogr	aphy 🚱				
zblox 2 Crossbar permutation instructions $\delta$	2				
zkne 4 NIST Suite: AES Encryption 🖉					
zknd 🗄 NIST Suite: AES Decryption 🎸					
zknh 🗉 N1ST Suite: Hash Function Instructio	ns <u>ór</u>				
zkr Entropy Source 🔗					
zkt Data Independent Execution Latency 🖉					
zvkn. Vector Crypto NIST Algorithm Suite 🔗					
zvikbi Vector Crypto Bitmanip 🔗					
zvkned Vector NIST Cloher AES encryption and	decryption of				
zyknihb Vector NIST SHA-256 and SHA-512 Sec	ure Hash 🔗				
zvka Viector Crypto Shang-Mi Algorithm Suite 🔗					
zvikb Vector Crypto Bitmanip 🖉					
zwish Vector Shang-Mi 3 Hash function (SM3)	9				
zvised Vector Shang-Mi 4 Cipher (SM4) encrypt	ion and decryption				
sv39 Page-Based 39-bit Virtual-Memory System					
zve64f v1 Embedded Vector Computation (64-	bit integer, 32-bit FP)				
escofpmf v0.1 Supervisor-level Count Overflow	v and Mode-Based Filtering				
Raufida			2023-06	-	0



#### The Metis Platform

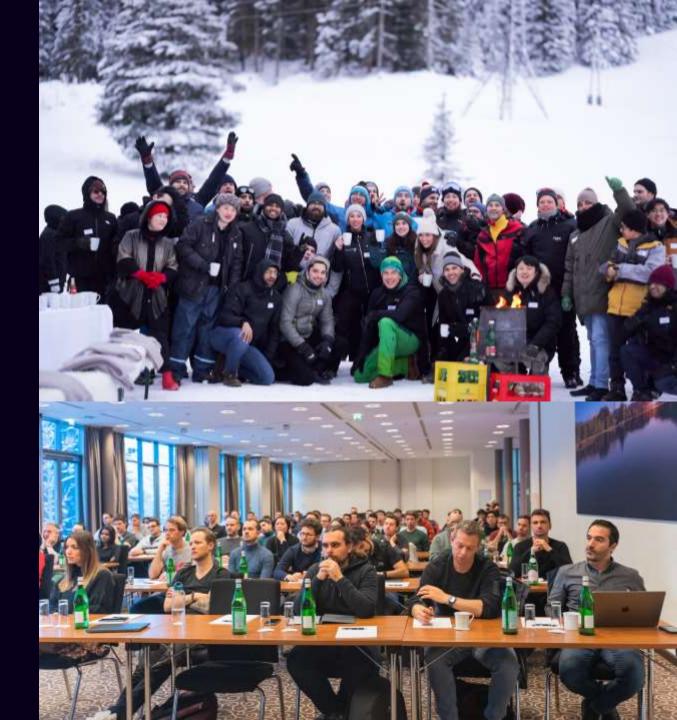
- From concept to tape-out in ~12 months
- First RISC-V booting hours after chip-back





#### Axelera Today

- Little over two years old, now with <u>130+ people, 45+</u> <u>PhDs</u> 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 Al Processing Unit for Edge Al 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!





# Keep in Touch!



#### Address

HTC5, High Tech Campus 5656 AE Eindhoven The Netherlands



#### E-mail

florian.zaruba@axelera.ai



