RISC-V®

State of the Union

Krste Asanovic
Chief Architect, RISC-V International

RISC-V Summit
Bologna, Italy
June 10, 2026

The State of the Union is strong!

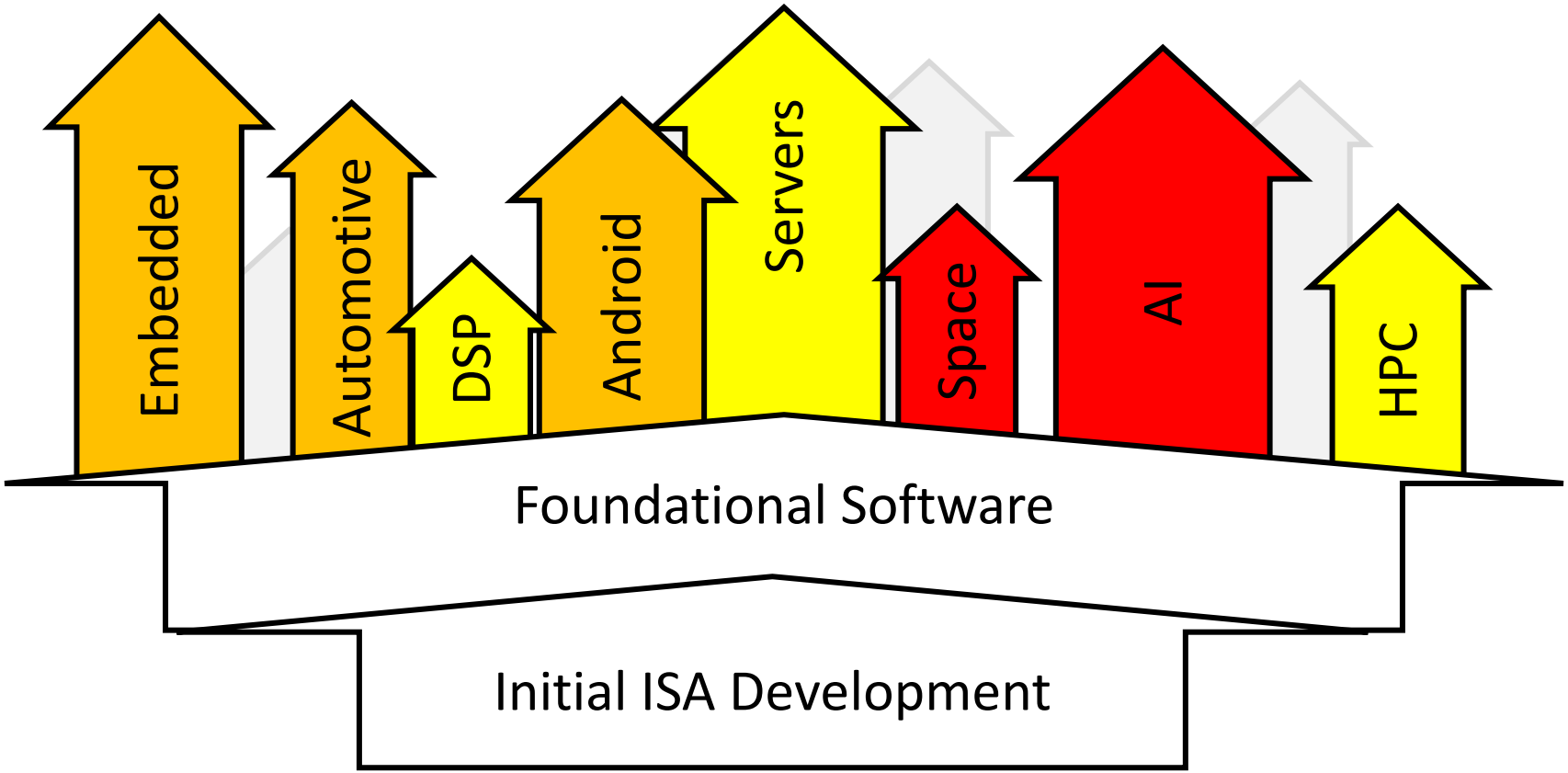
Solidly established in embedded space

The standard base for new AI accelerators

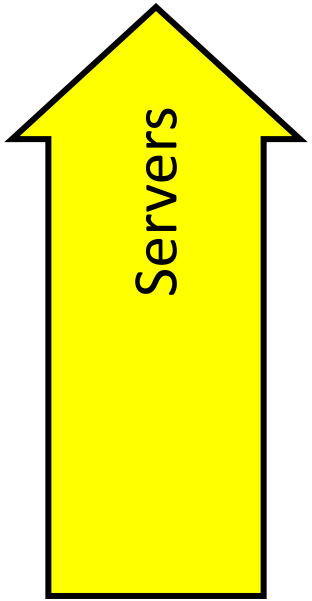
On verge of widespread adoption in application processors

In 2026, start of high-performance RVA23 silicon availability including server-class chips and boards

RISC-V move into industry verticals



RISC-V in the Datacenter



RISC-V Servers

- First RVA23 server-class systems appearing in 2026
  - Even higher performance cores and systems in development at multiple vendors
  - Main barrier to higher performance is economic, need virtuous cycle between supply and demand for sustainable improvements
- Server platform spec ratified
  - Community aligned around common modern spec
- This is just the starting point
  - Experience from other architectures is that it will take several generations to reach feature parity/maturity in general server space
  - But RISC-V is much more than just a replacement technology
- Community must work together to iterate on server ecosystem
  - This work will help all the other application-class verticals
  - Need large-scale deployments, cloud instances to support devs
  - Many software components need porting, then tuning

Optimization Guidance Options

- Current ISA strings encode functionality, not performance
  - e.g., Zilsm says misaligned will work, but not how fast.
  - Some hardware vendors might believe trap-and-emulate is acceptable
  - Software pessimistically codes around misaligned accesses, even though high-performance cores have fast hardware support.
- New optimization guidance options align hardware and software to raise the bar for RISC-V application processors
  - Intended to be mandatory for future RVA profiles, but will first appear as development options to guide the community forward

Initial Optimization Guidance Options

Oilsm: Software should assume misaligned scalar load/store accesses and misaligned element accesses within vector load/store instructions will be handled more efficiently than alternative code sequences that avoid the misalignment.

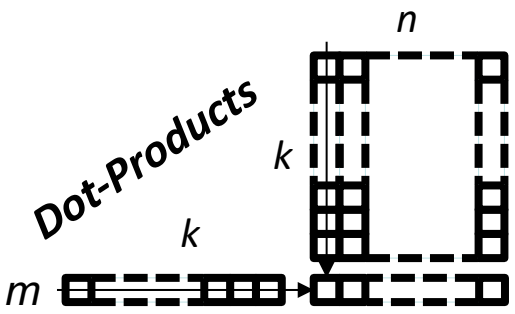
Ovlt: Software should assume microarchitectures will optimize execution of vector instructions based on the dynamic value in the "vl" register, i.e., execution time of a vector instruction will not be improved by reducing LMUL (and hence VLMAX) based on statically known or dynamically determined application vector lengths.

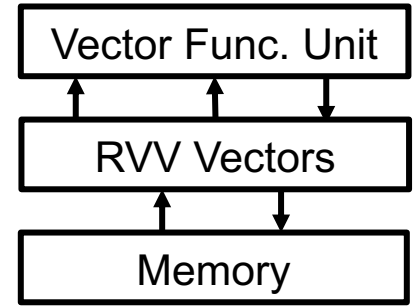
RISC-V Matrix Extensions, 1+3 Approaches

Element in existing RVV vector register

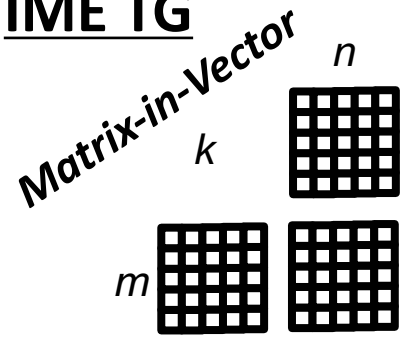
Element in added matrix registers

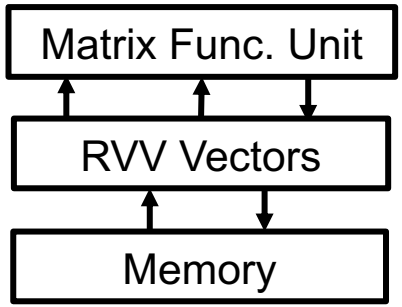
Zvbdota Fast-Track



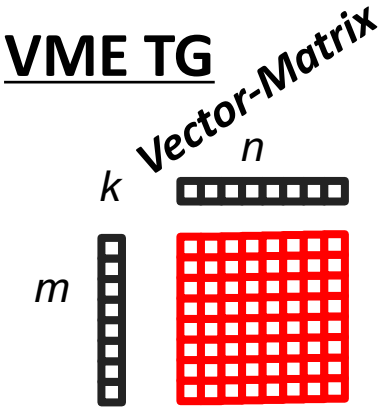


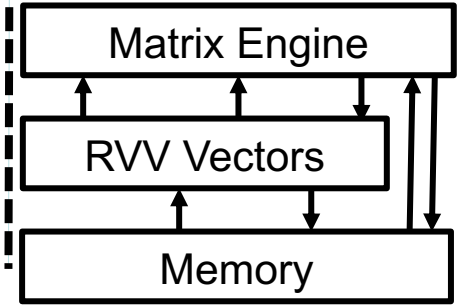
IME TG



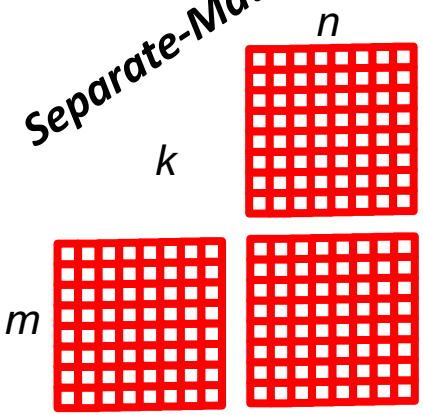


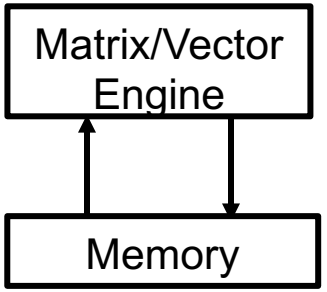
VME TG





AME TG





Alternative Matrix Instruction Extensions

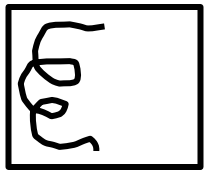
- **Dot-Products:** No additional matrix state. Small and effective extension at smaller vector lengths, also meets some DSP needs, but doesn't scale performance with larger vector lengths.
- **Matrix-in-Vector (IME):** No additional matrix state. Allows larger throughputs than dot-products for longer vector lengths.
- **Vector-Matrix (VME):** Add matrix state to processors that already have RVV. Maintains vector ISA and memory model.
- **Separate-Matrix (AME):** Unconstrained matrix and vector design.

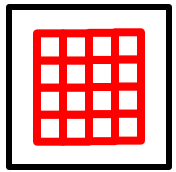
All extensions progressing well, expect freeze/ratifications for most this year.

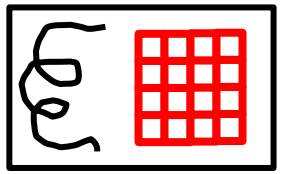
AI Compute Organizations

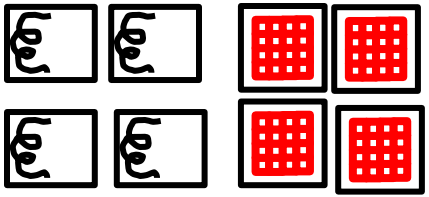

General-purpose compute

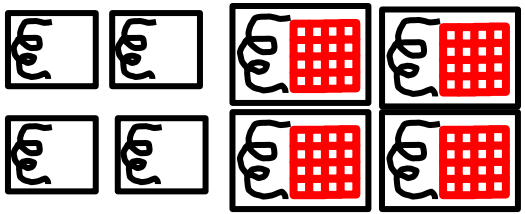

Number-crunching compute

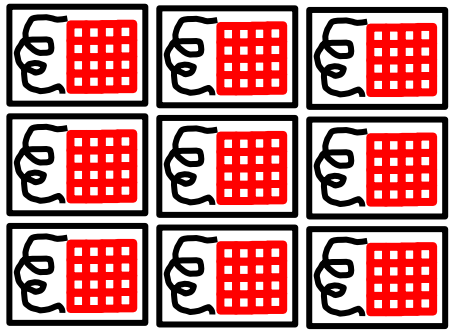

CPU


GPU


RISC-V scalar/vector/matrix core


Host Device


Host Device


Self-Hosted

Wheel of Reincarnation

Myer and Sutherland, “On the Design of Display Processors”, 1968

Why RISC-V is uniquely positioned for AI

- Among competing ISAs, only RISC-V works well for:
  - AI Host ISA (e.g., Nvidia port of CUDA stack to RISC-V)
  - AI Device ISA (e.g., many AI accelerators based on RISC-V)
  - Self-hosted AI ISA (e.g. edge AI, plus upcoming agentic CPUs)
- RISC-V ISA also scales down to MMU-less embedded devices while keeping same vector and matrix extensions
- RISC-V supports AI innovation and differentiation through custom extensions

- Share standard compute kernel and software stack costs across all possible AI compute environments

RISC-V New Security Extensions in Progress

- SPMP
  - Provide S-mode RTOS/supervisors with protection from U-mode tasks
- RISC-V Worlds
  - Provide secure global partitioning of devices on SoCs
- IOPMP
  - Embedded device security
- Supervisor Domains
  - Flexible support for confidential computing and other applications
- CHERI
  - New base ISAs (RV32Y/RV64Y) bringing capabilities to RISC-V
- Lightweight Memory Tagging
  - Improve memory safety, building upon pointer-masking support
- Additional Crypto
  - PQC, and other vector crypto enhancements

RISC-V DSP extensions

- RISC-V P extension
  - Long development time, but now stable on path to ratification
  - Packed integer/fixed-point (8b/16b/32b/64b) operations on integer **x** registers
  - Over 100 new instructions

- Upcoming RISC-V vector DSP extension, TG initiated
  - Add DSP instructions on **v** registers
  - Candidate new instructions for greater fixed-point support, permutations for FFTs, complex arithmetic support, …

Standardizing Microcontrollers

- While many microcontroller use cases benefit from extreme specialization, richer embedded software stacks (e.g. automotive) would benefit from standard profiles
- Richer microcontroller standard including:
  - RVM ISA profile mandating many useful extensions
  - Standard fast interrupts
  - Standard nested protection (sPMP)
  - Standard security features (RISC-V Worlds)
  - Standard debug /trace support
- Pieces mostly exist now, but need to keep working on common specs to drive ecosystem forward

Long (>32b) Instructions

- RISC-V ISA included variable-length instructions from beginning
  - Compressed instructions (16b) save code-size
- RISC-V designed for long-term success, won't disappear due to owner changing business model or folding.
  - Fixed 32b instruction format would be barrier to long-term evolution. Other fixed-width 32b ISAs already running out of encoding room.
  - Longer instructions (>32b) also help reduce code size, improve performance, and support ever-growing number of new operators and datatypes
- TG formed to finalize encoding format and first long instructions

Summary

- Milestone: 2026 sees multiple high-performance RVA23 systems delivered
  - Server ecosystem can now start to evolve rapidly
- RISC-V uniquely positioned to create portable open-standard AI platforms at all scales

- RISC-V is inevitable