RISC-V Vectors & oneAPI

Accelerating the Future of Heterogeneous Compute

Stephano Cetola, Director of Technical Programs

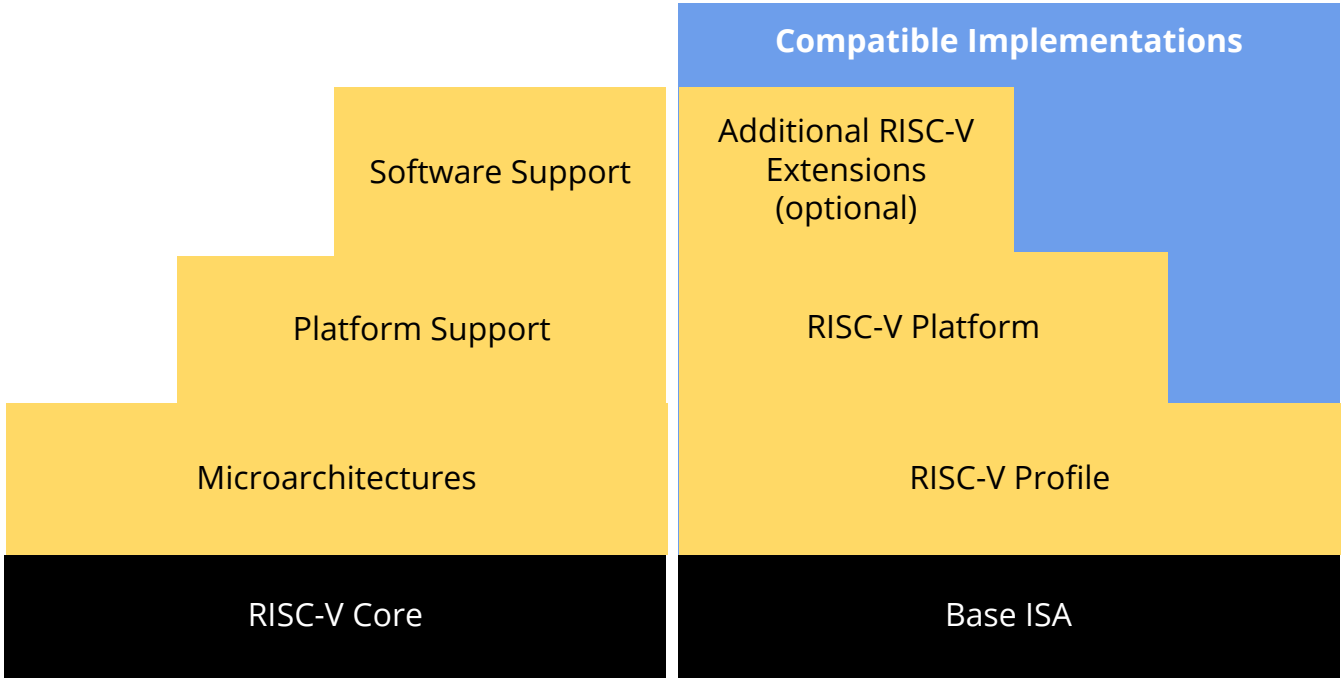
What is RISC-V?

Who is RISC-V

RISC-V International is a global nonprofit association based in Switzerland. Founded in 2015, RISC-V brings together 3k+ members in more than 70 countries across industries and technical disciplines.

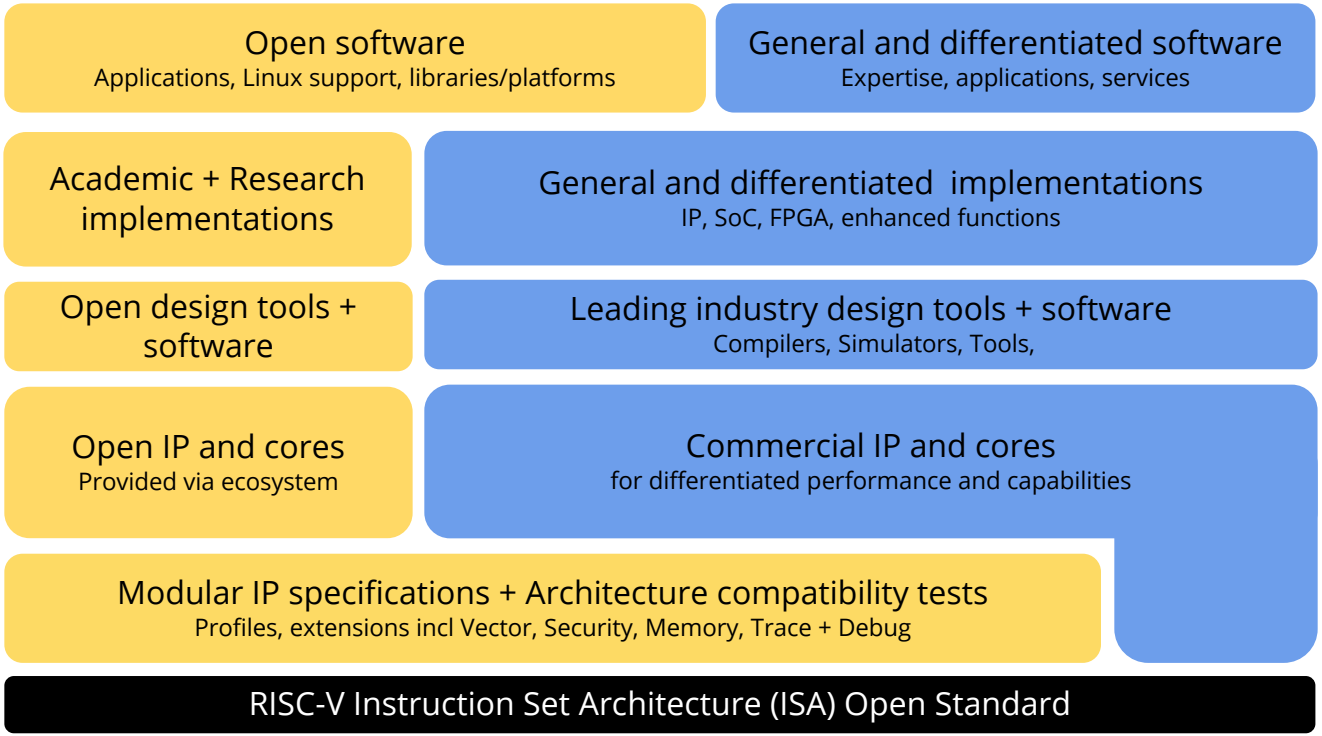
RISC-V supports the open RISC-V instruction set architecture, developing additional extensions, tools, and resources paving the way for the next 50 years of computing design and innovation. RISC-V also connects the community and industry through academia, commercialization, and strategic leadership.

Building Hardware using RISC-V



Standardization, Modularity, and Flexibility

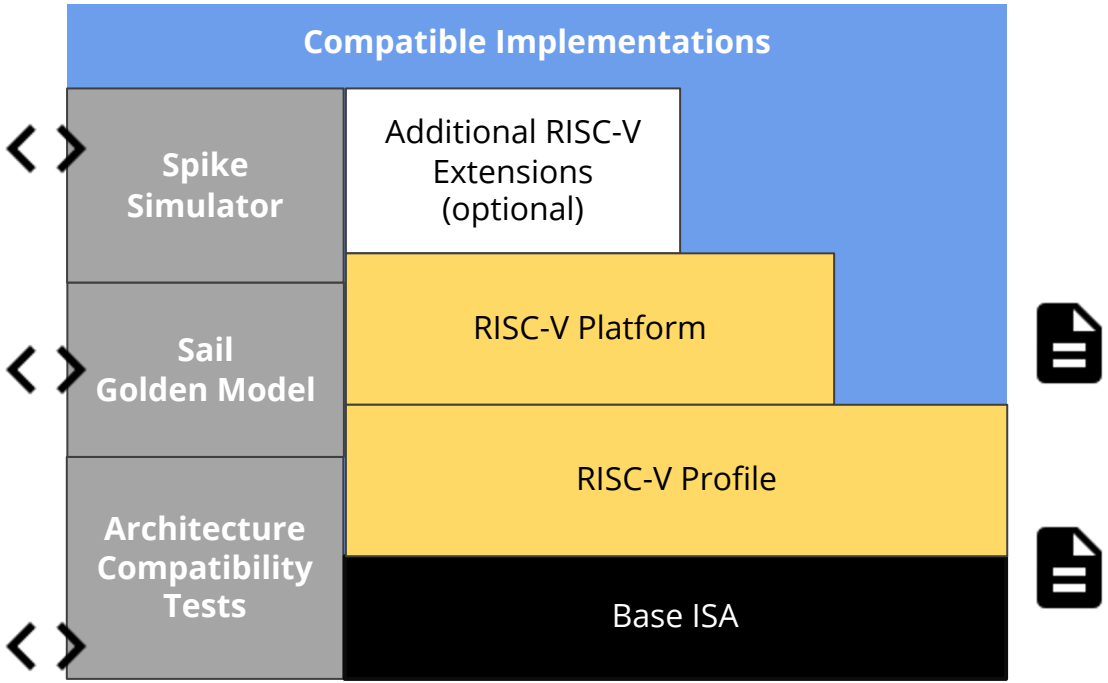
Open standard IP specifications for modular designs based on decades of investment and expertise



Commercial innovation and differentiation to solve full spectrum of compute challenges

Hardware Compatibility using RISC-V

- Profiles
- Platforms
- Sail Golden Model
- Architecture Tests
- Spike Simulator

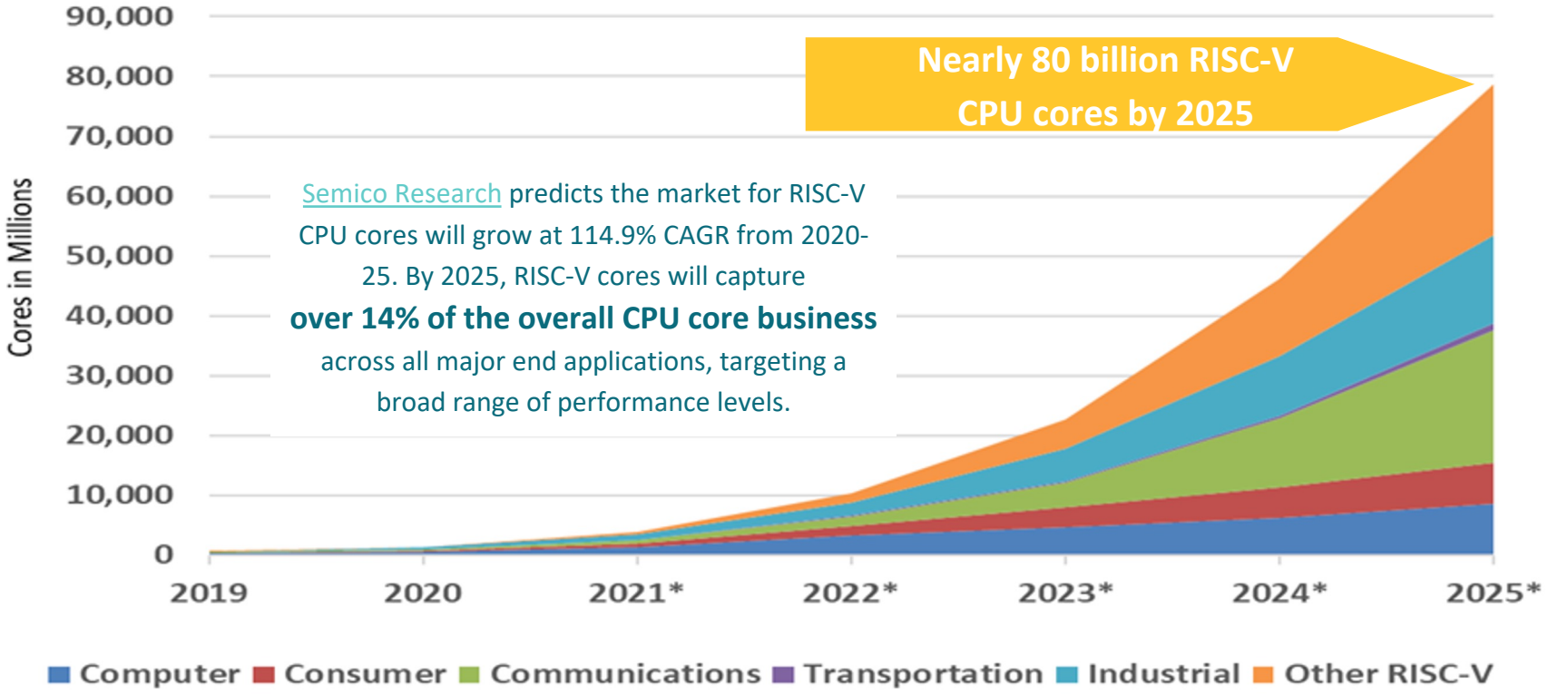


= Specification Document

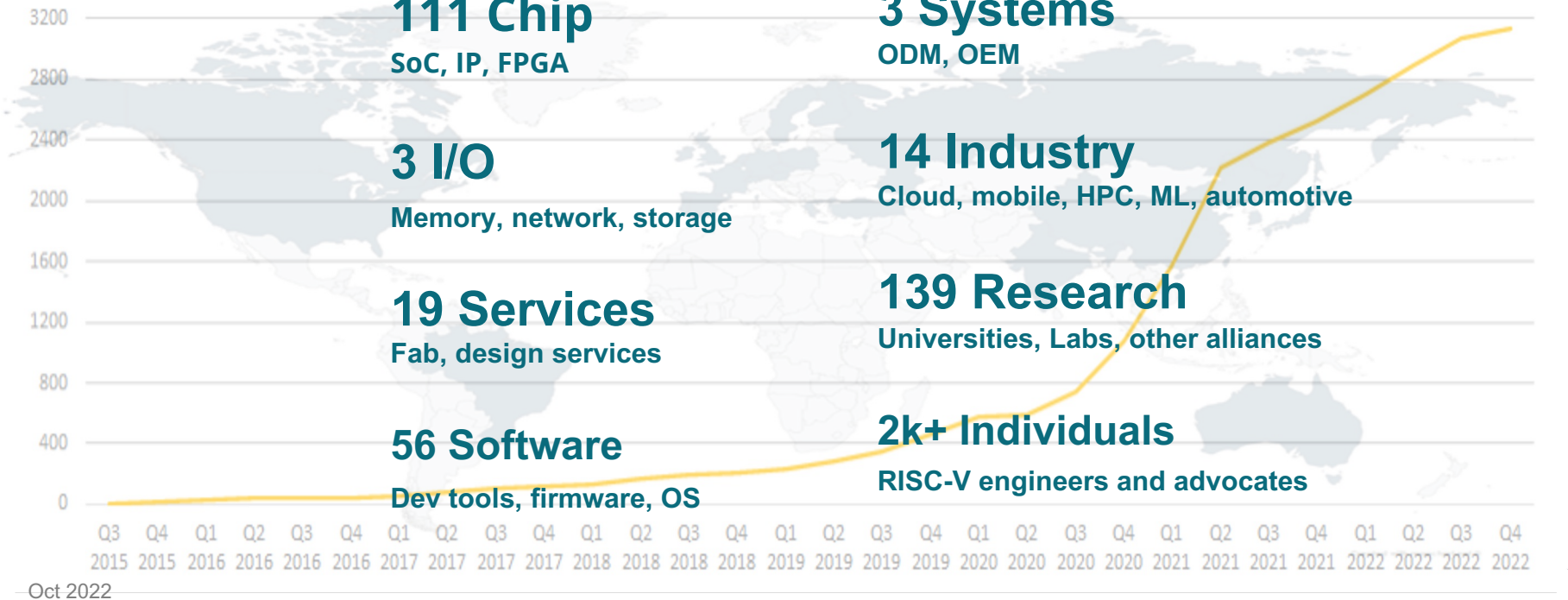
= Open Source Code

Who is using RISC-V?

Rapid RISC-V Growth



More than 3,100 RISC-V Members
across 70 Countries



111 Chip
SoC, IP, FPGA

3 I/O
Memory, network, storage

19 Services
Fab, design services

56 Software
Dev tools, firmware, OS

3 Systems
ODM, OEM

14 Industry
Cloud, mobile, HPC, ML, automotive

139 Research
Universities, Labs, other alliances

2k+ Individuals
RISC-V engineers and advocates

Oct 2022

RISC-V®

RISC-V membership rapid growth of 134% in 2021

RISC-V adoption spans industries

Data Center
Cloud, HPC
Top providers like Amazon and Alibaba are designing their own chips.
New features specific to HPC such as Vector and larger Virtual address space

Automotive
Transforming from autonomous vehicles to infotainment to safety, the whole vehicle relies on innovative electronics.

AI / ML
Artificial intelligence is incorporated across many areas including Industrial IoT, manufacturing, and financial

Telecom & Communications
Rapid evolution in 5G, handsets, and base stations grows with each generation of hardware and increased capability

Consumer and IoT devices
Incredible innovation is driving volume with billions of connected devices in the next 5-10 years.

Edge Computing
A distributed, open architecture decentralizes processing power, reduces latency, and supports IoT performance in low bandwidth environments.

RISC-V will capture 10% of the Automotive market by 2025

– Counterpoint, September 2021

- **MobileEye** vision-based advanced driver assist systems chips capable of 176 trillion ops per second with 12 RISC-V CPU cores.
- **Andes** ISO 26262 Functional Safety ASIL D Dev Process Certification for RISC-V embedded automotive safety with Andes processors
- **Renesas and SiFive** partner on next-gen, high-end RISC-V automotive applications. SiFive will license their RISC-V core IP to Renesas.
- **Imagination Technologies** GPU linked by a RISC-V core for ASIL-B level designs with ISO 26262 safety critical certification.
- **IAR Systems** extended functional safety of its Embedded Workbench sw tool chain to the RISC-V core of NSITEXE, subsidiary of automotive leader Denso.
- **Europe GaNext** simplifies power converters with GaN power semiconductors with better efficiency and compactness for EV chargers.

2020 RISC-V automotive opportunity 4M cores; growing to 150M cores in 2022 and 2.9B cores by 2025.

– Deloitte, December 2021

RISC-V Automotive Value Multicore SoC revenue to be $5.7B in 2022. Advanced RISC-V AI SoCs for High-End Passenger Cars to Reach $819M by 2027

RISC-V-based AI SoCs will grow **73.6% CAGR to 25B units and $291B in revenue by 2027**

– Semico Research, December 2021

- **Alibaba Cloud** tops MLPerf Tiny v0.7 Benchmark with its IOT processor
- **Esperanto** accelerating ML Recommendation With Over 1,000 RISC-V/Tensor Processors on ET-SoC-1 Chip
- **StarFive** released the world's first RISC-V AI visual processing platform
- **Andes** released superscalar multicore and L2 cache controller processors.
- **NVIDIA CUDA** support on Vortex RISC-V GPGPU enables scaling from 1-core to 32-core GPU based on RV32IMF ISA

RISC-V®

- **E4** Monte Cimone Cluster along with DEI-UNIBO contributing to architecture, software, and integration.
- **European Processor Initiative** RISC-V accelerator with first chip Sep 2021
- **Technical University of Munich** (TUM) quantum cryptography chip for quantum computing security demands
- **Tactical Computing Labs** HPC-centric software test suite for GCC and LLVM
- **Cortus** is developing a high-performance RISC-V Out-of-Order processor core for the European eProcessor project.
- **De-RISC** HW-SW platform for multi-core RISC-V SoC for safety critical aerospace

RISC-V®

Introduction to Vector on RISC-V

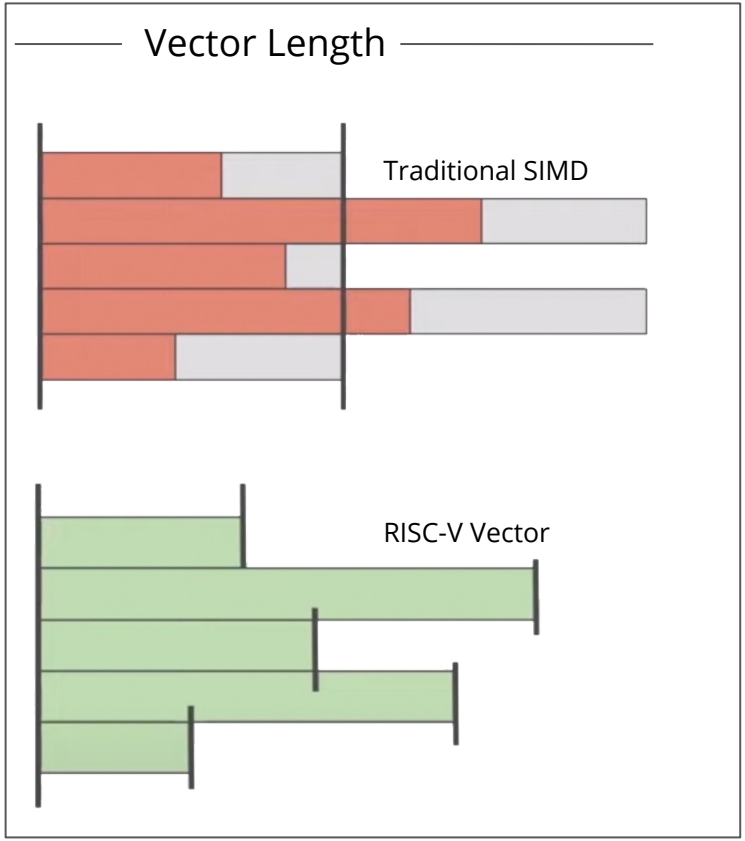
RISC-V Vector (RVV)

At the end of 2021, RISC-V International ratified version 1.0 of the Vector Extension

- Single RISC-V Extension
- Variable Vector Lengths
- Maps into existing neural network models (multiple maps)
- Allows for small code size and low power consumption
- Range of vector data types and sizes
- Enables open-source extensions

RISC-V Vectors vs Traditional SIMD

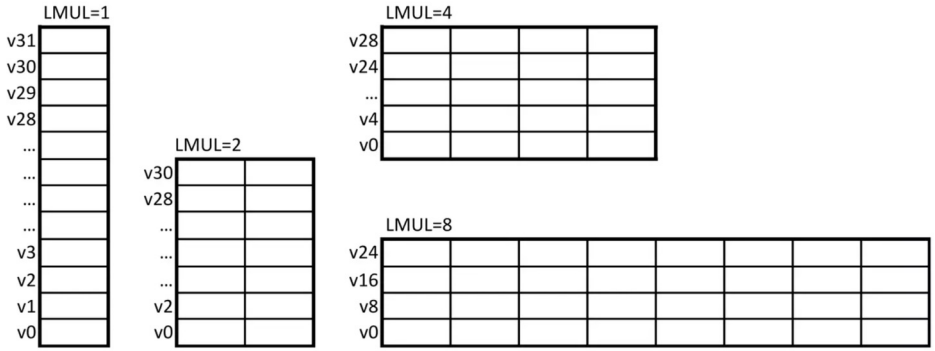
- Variable Vector Register Length
  - Vector width configurable at runtime
  - Up to maximum hardware length (VLEN)
  - Vector ISA is agnostic to VLEN
- Unified Code Base
  - ARM has two different ISAs (Helium & Neon)
  - IA-32 vector instructions were an order of magnitude larger when supporting SIMD vector lengths
- No requirement for dedicated vector data memory



	RVV	ARM SVE	x86 AVX512
Spec supports multiple vector sizes	Yes	Yes	No
VL register	Yes	No	No
Masks	Yes (1)	Yes (16)	Yes (8)
Narrowing/Widening	Yes, LMUL	Yes, 2 instructions	Yes, 2 instructions
Mixed Datatype Vectorization	Uses LMUL	Pack/Unpack	Pack/Unpack
Hardware Unrolling (LMUL)	Yes	No	No
Polymorphic Encoding	Yes, vtype	No	No
Strided Memory Access	Yes	No	No
Gather/Scatter	Yes	Yes	Yes
Structured/Segmented Loads	Yes	Yes	No
Forward Progress on gather/scatter	Vstart	Repeat full instruction	Mask state
Fixed Point Support	Yes	Partial	No
Complex Support	No	Yes	Partial
Reductions	Yes	Yes	No
Register Gather	Yes	Yes	No
Tail Element control	Merge, Agnostic	Merge, Zeroing	Merge, Zeroingº
Destructive destination	No	Yes	No

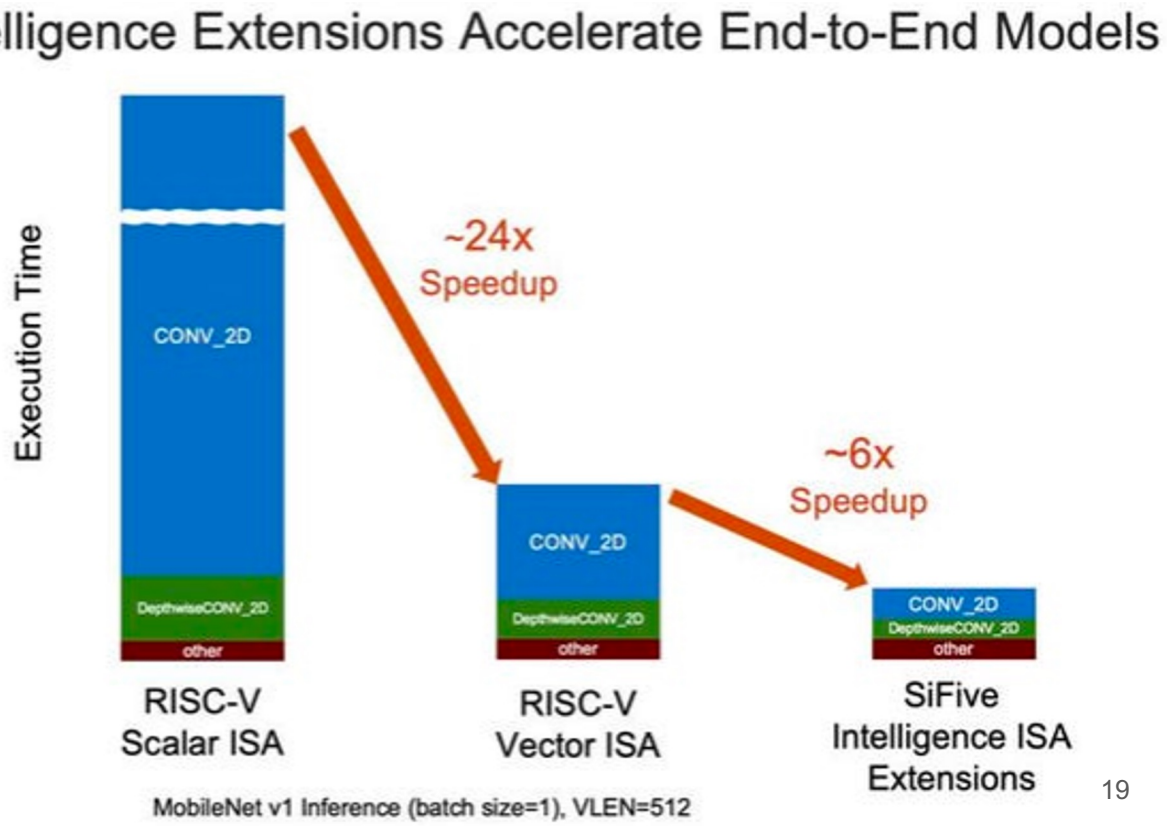
LMUL - Multiplying Vector Length

- LMUL - parameter that multiplies the number of VLEN registers operated on by each instruction
- Goal - Fewer Instructions
  - Large vector data computations
  - More efficient parallel computation (both scalar & vector)
  - Smaller code size, fewer memory accesses



RISC-V Custom Extensions

- Reserved opcode space for custom extensions
- Allows for layering of value-add custom extensions (open-source or proprietary)
- Flexibility to push the boundaries
- Cost effective model for R&D



The Future of AI & RISC-V

Vector SIG

ML: https://lists.riscv.org/g/sig-vector
Charter/Docs: https://github.com/riscv-admin/vector

- Finishing Zvfh / Zvfhmin extensions & tests
- Aligning efforts around Packed SIMD extension
- Matrix Multiplication extension

Floating Point SIG

Chair: Kenneth Rovers, Imagination Technologies
ML: https://lists.riscv.org/g/sig-fp
Charter/Docs: https://github.com/riscv-admin/fp
Goals:

- New FP formats and operations
- Move beyond IEEE-754 standard, Zfh/Zfhmin, Vector
- Develop the strategy, gaps, and prioritizations for FP & RISC-V

bfloat16 Task Group

Chair: Ken Dockser, Rivos
ML: https://lists.riscv.org/g/sig-fp
Charter/Docs: https://github.com/riscv-admin/bfloat16
Specification: https://github.com/riscv/riscv-bfloat16

- Extensions for vector and scalar floating-point
- Includes BF16 format and behaviors
- Basic instructions allowing BF16 as an interchange format
  - FP #'s converted between BF16 and other formats like single precision, FP32
- Add widening multiply-add type instructions (BF16 operands, FP32 result)
- Note: more instructions to come…

Software Committees

Platforms definition ⟵ ⟶ **(Software) applications enablement**

Platforms standardization
- Binary compatibility
- Source-code compatibility

ABIs and Firmware Services
Bootstrap and Secure Boot
Unified Discovery

Reduce the risk in adopting RISC-V

Toolchains and development tools
- Compilers and Debuggers
- Optimized runtime libraries
- Code-speed and code-size

Managed Runtimes
Simulators and Performance Models

Attract applications to RISC-V

Standardize solutions
Reach consensus and document other ideas
Avoid fragmentation

Drive enablement and optimization
Own and coordinate efforts
Enable innovation/differentiation

RISC-V Members Leading AI/ML

Coming in 2023

- RISC-V Vector C intrinsics Task Group
- AI/ML Special Interest Group
- Vector SIG
  - Matrix Multiply
  - Vector Phase 2
- LLVM Improvements
  - intrinsics
  - autovectorization
- oneAPI Engagement
  - oneDNN
  - OpenMP®
  - OpenCL™
  - SYCL™

RISC-V®

How can you get involved in AI and RISC-V?

RISC-V Technical Community

- **GitHub**
  https://github.com/riscv*
- **Mailing Lists**
  https://lists.riscv.org
- **RISC-V Wiki**
  https://wiki.riscv.org
- **Groups**
  https://lists.riscv.org/g/sig-ai-ml
  https://lists.riscv.org/g/sig-fp
  https://lists.riscv.org/g/sig-vector

Thank you!

stephano@riscv.org

RISC-V®