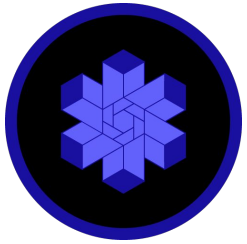
semidynamics

Building IP for the Next-Gen European Cloud

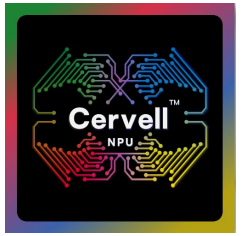
Josep Sans i Prats


OOO CPU Core


Gazzillion™ Technology


OOO Vector Unit


Tensor Unit



Cervell NPU

About Semidynamics

Semidynamics, founded in 2016, is a **100% European** supplier of RISC-V IP cores, HQ in **Barcelona**, specializing in **optimized processor IP** for **high bandwidth high performance SoC applications.**

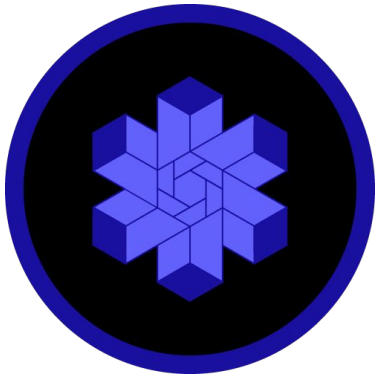
Semidynamics' IP

Atrevido

64b out-of-order CPU
RISC-V
AXI and CHI

Gazzillion™ Technology

Hiding the
memory wall

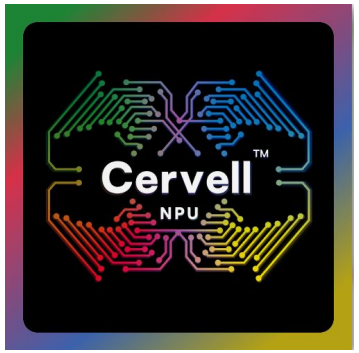


Vector Unit

RVV1.0
Out-of-order

Tensor Unit

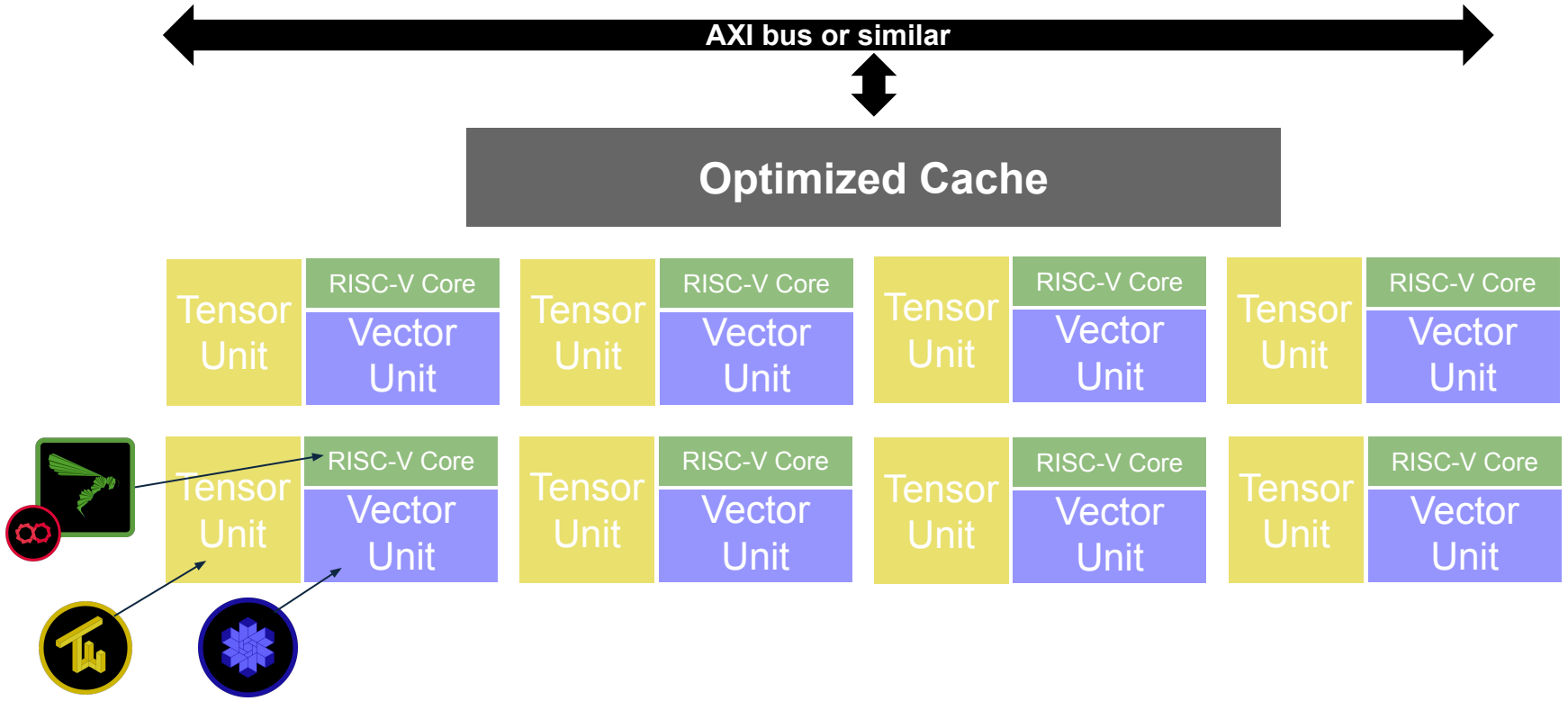
BF16, FP16,
INT8, INT4



Cervell

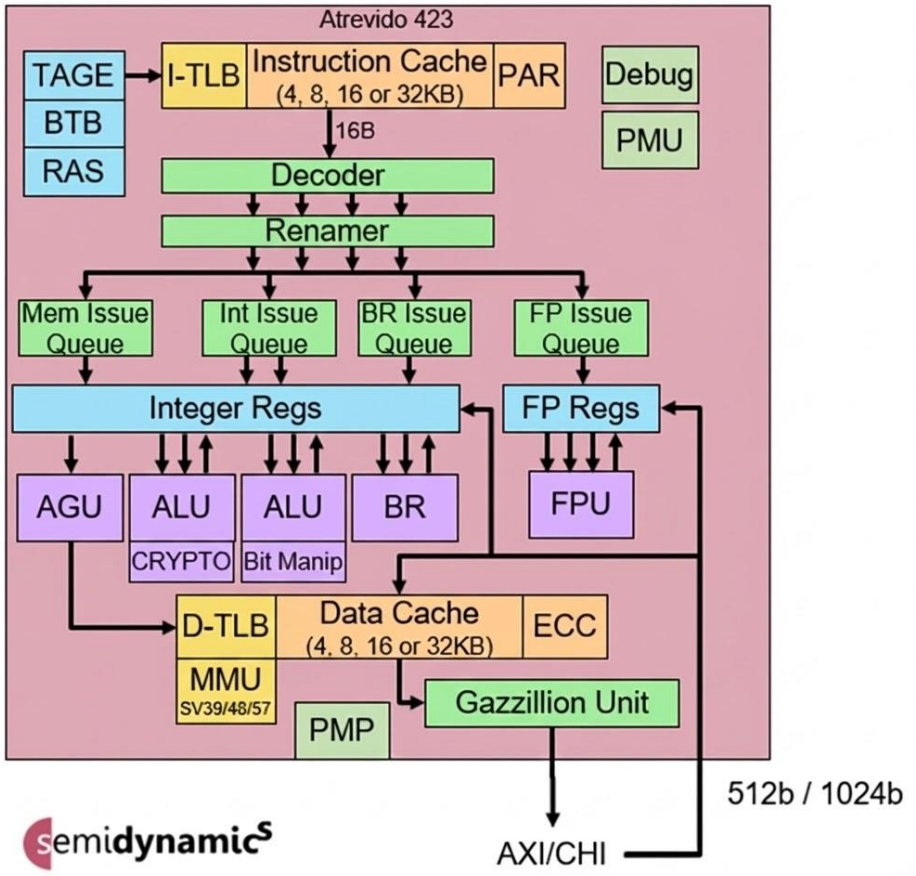
All-in-One
RISC-V NPU

Cervell NPU

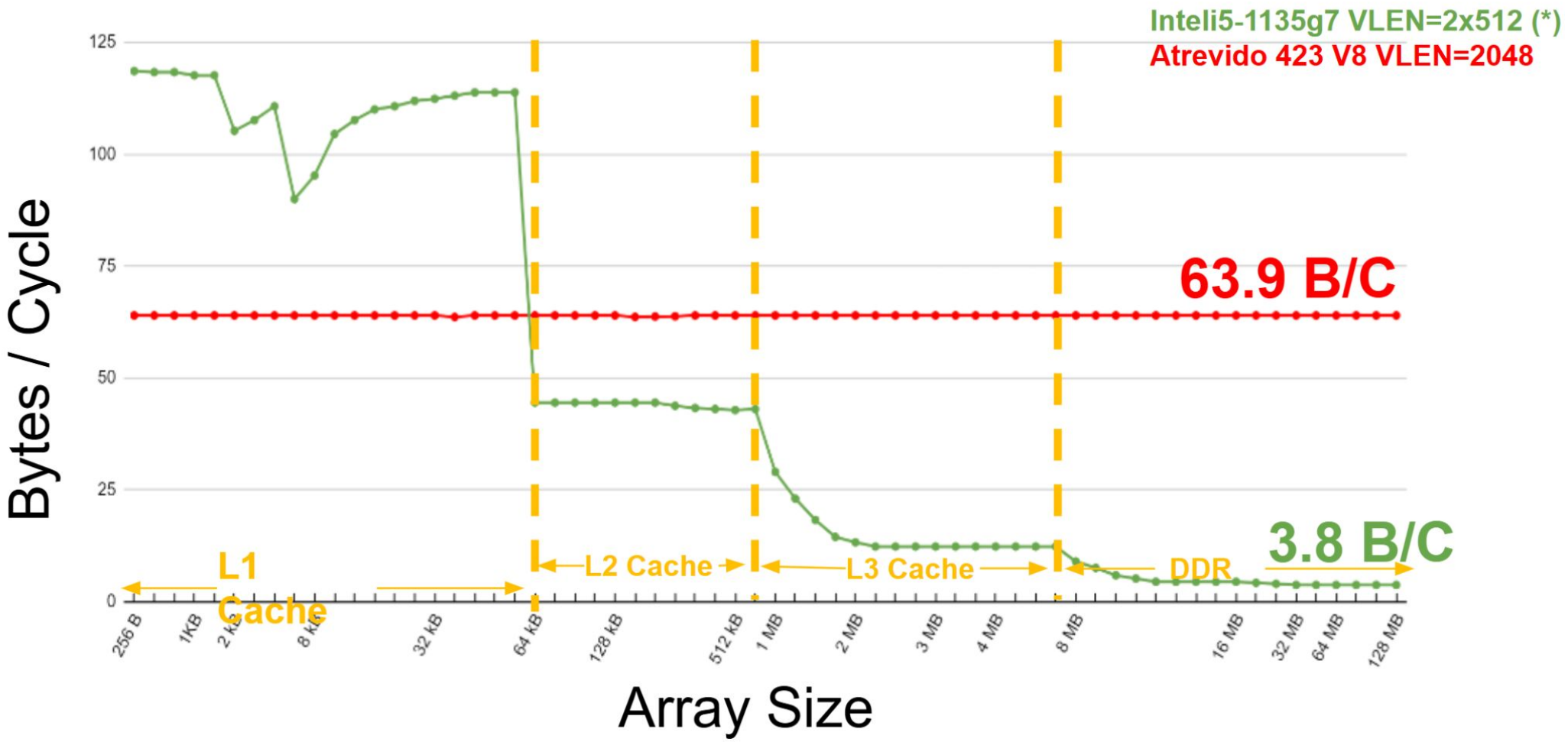


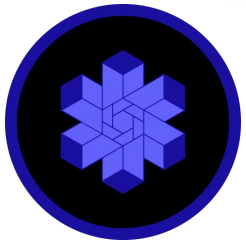
Atrevido

- 64-bit RISC-V core
- 4 way out-of-order execution
- AXI / CHI
- Linux ready
  - SV39,48,57
- Gazzillion Misses™
  - Effective memory bandwidth usage



Read Bandwidth: Atrevido + Gazzillion



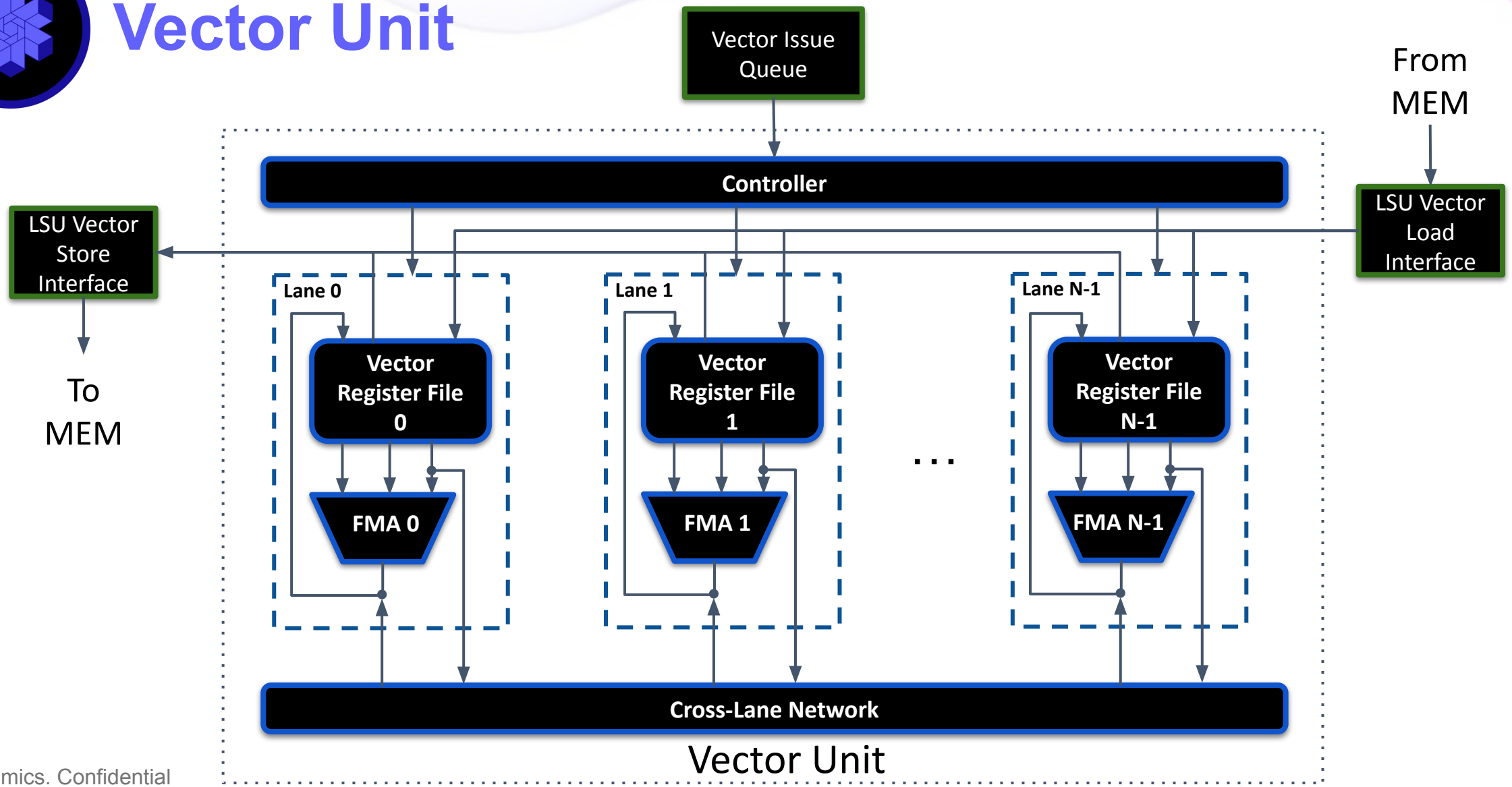


Vector Unit

- RISC-V Vector Extension 1.0
- Out-of-order execution
- INT64, INT32, INT16, INT8
- FP64, FP32, FP16, BF16
- Custom extensions
  - Brain Float (BF16)
  - FP8 converts

Name	VLEN	DLEN	FP64 GFLOPS at 1 GHz	FP32 GFLOPS at 1 GHz	FP16/BF16 GFLOPS at 1 GHz
V8x1	512	512	16	32	64
V8x4	2048	512	16	32	64
V16x2	2048	1024	32	64	128

Vector Unit

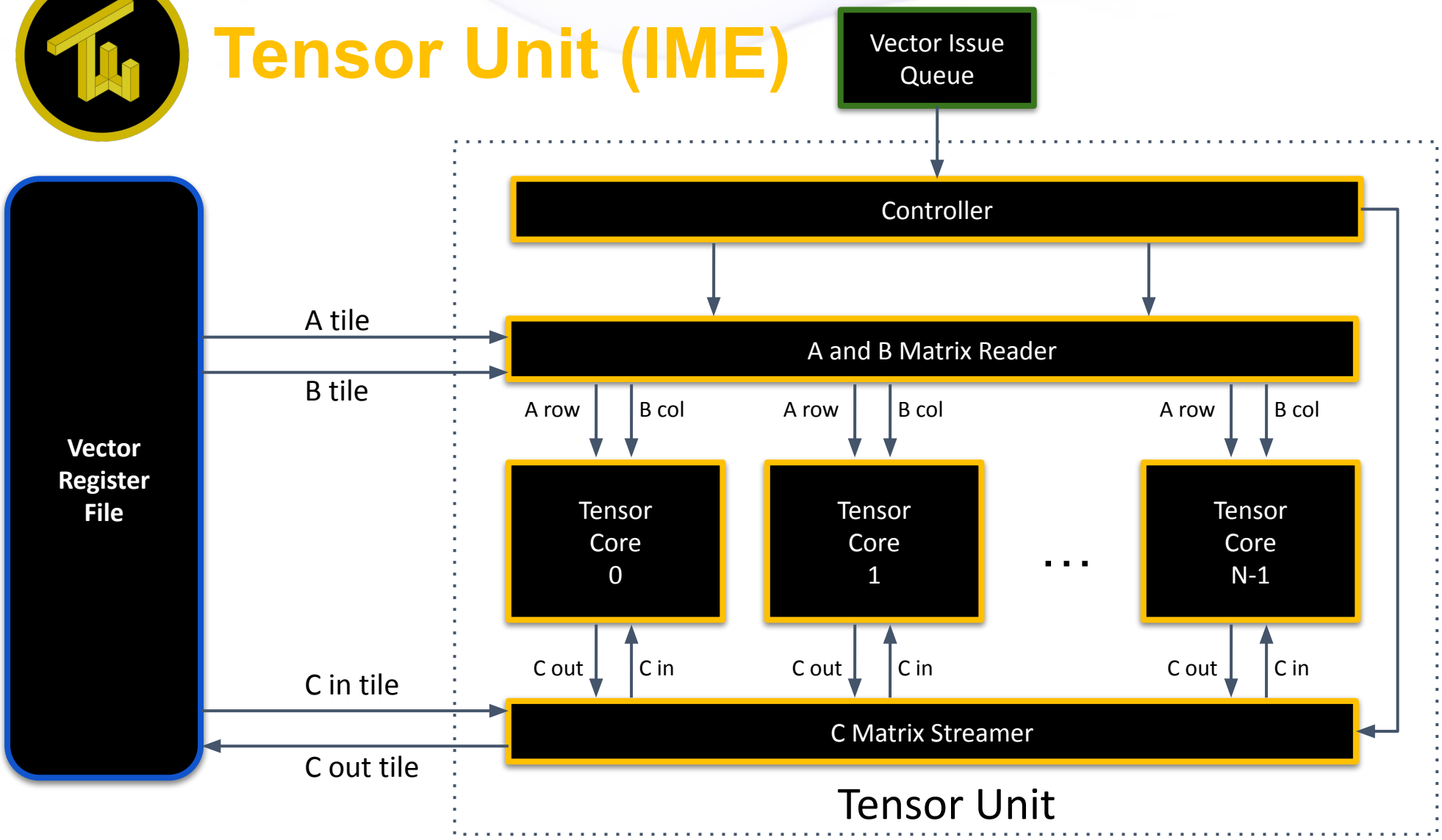


Tensor Unit (IME)

- Augmenting the **Vector Unit** with AI capability
- Custom matrix instructions
  - Load matrix: **vlrse** vd, (rs2), rs1
  - Store matrix: **vsrse** vd, (rs2), rs1
  - Mat multiply: **vmxm** vd, vs1, vs2
    - C = A x B + C
  - Matrices in vector registers (IME)

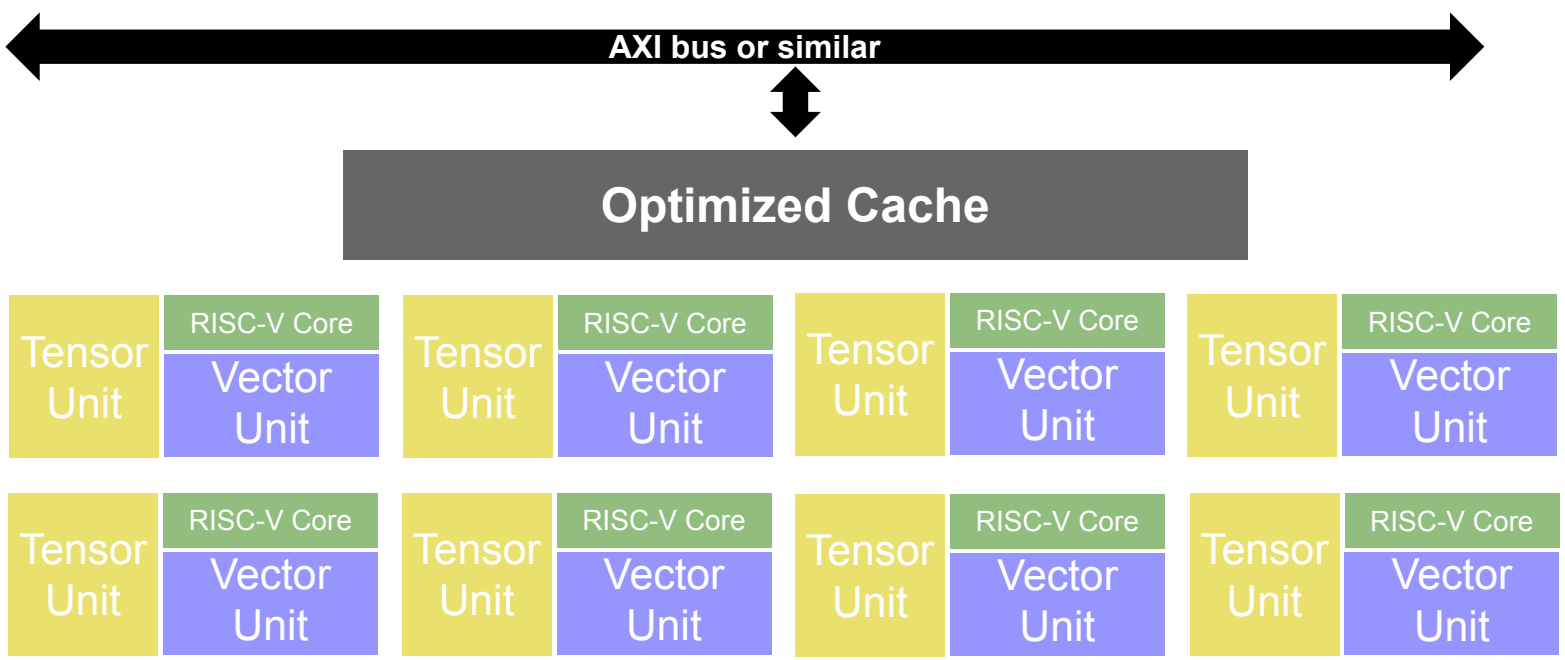
Data Types		
Matrix A	**Matrix B**	**Matrix C**
FP16	FP16	FP32
BF16	BF16	FP32
INT16	INT16	INT32
INT8	INT8	INT32
INT4	INT4	INT32

Tensor Unit (IME)

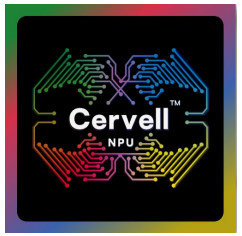


$$C_{out} = A \times B + C_{in}$$

Cervell NPU



- **Easy** to program
- **High Performance**
- **Zero** Latency
- **Better PPA**
- **Resilient** to new AI
- **Unified** programming

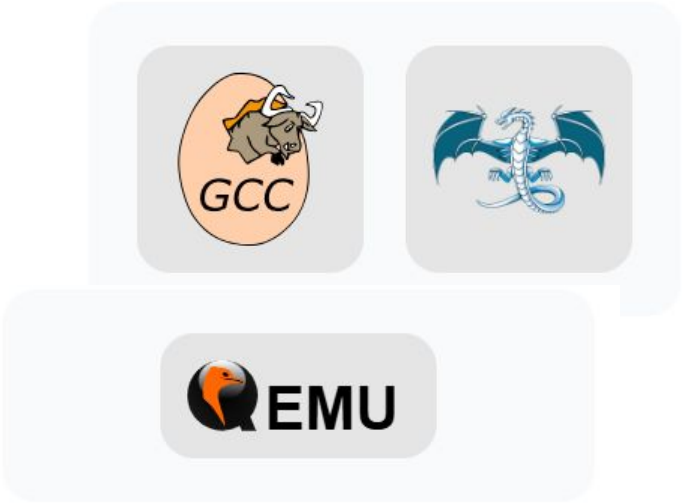


Cervell NPU

- Powerful **Out Of Order CPU** based on RISC-V ISA
- Creating a powerful AI capable processing element by tightly integrating heterogeneous **CPU, Vector** and **Tensor** functionalities **into a single coherent instruction flow**!
- Use of **Gazzillion™ Technology** to efficiently manage large data sets

Name	INT8 TOPS at 1 GHz	INT8 TOPS at 2 GHz	INT4 TOPS at 1 GHz	INT4 TOPS at 2 GHz
C1	1	2	2	4
C8	8	16	16	32
C32	32	64	64	128

Software: Aliado SDK



Standard Risc-V Tool chains

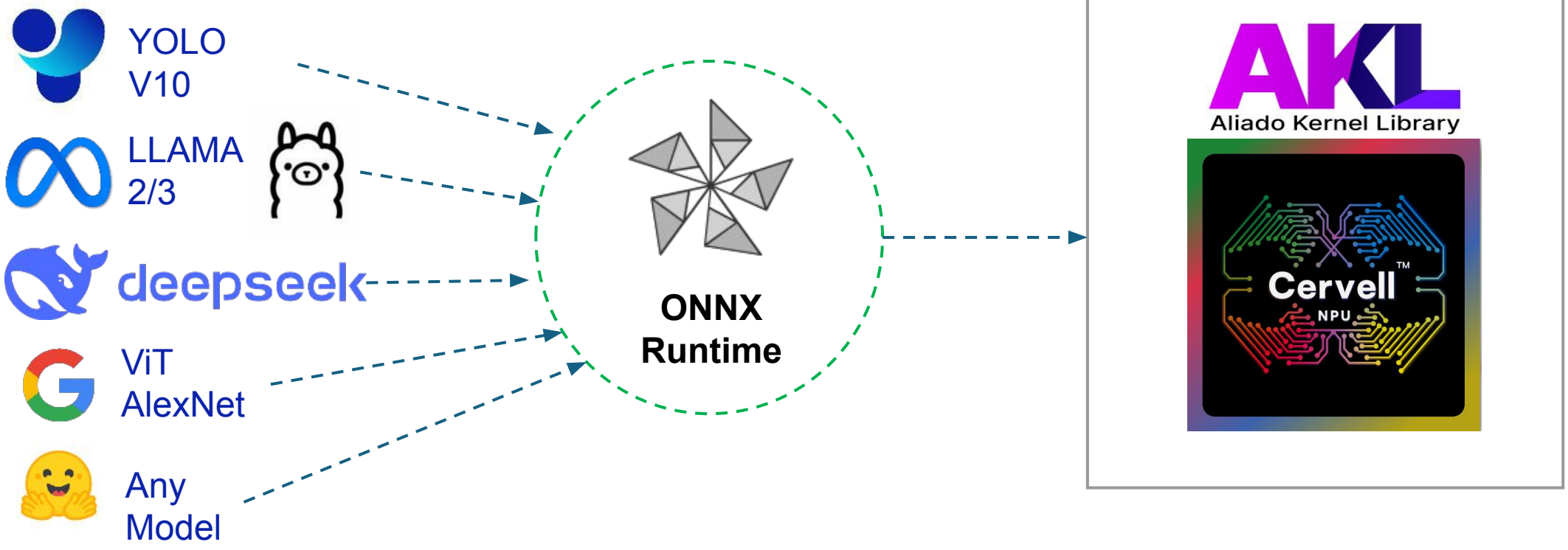
ONNX RUNTIME

No proprietary machine learning Compiler – Risc-V ISA based library approach

ONNX

Proven and tested AI model Zoo

Software: Aliado SDK



Download any model from Huggingface.com and run

semidynamics

Thank you