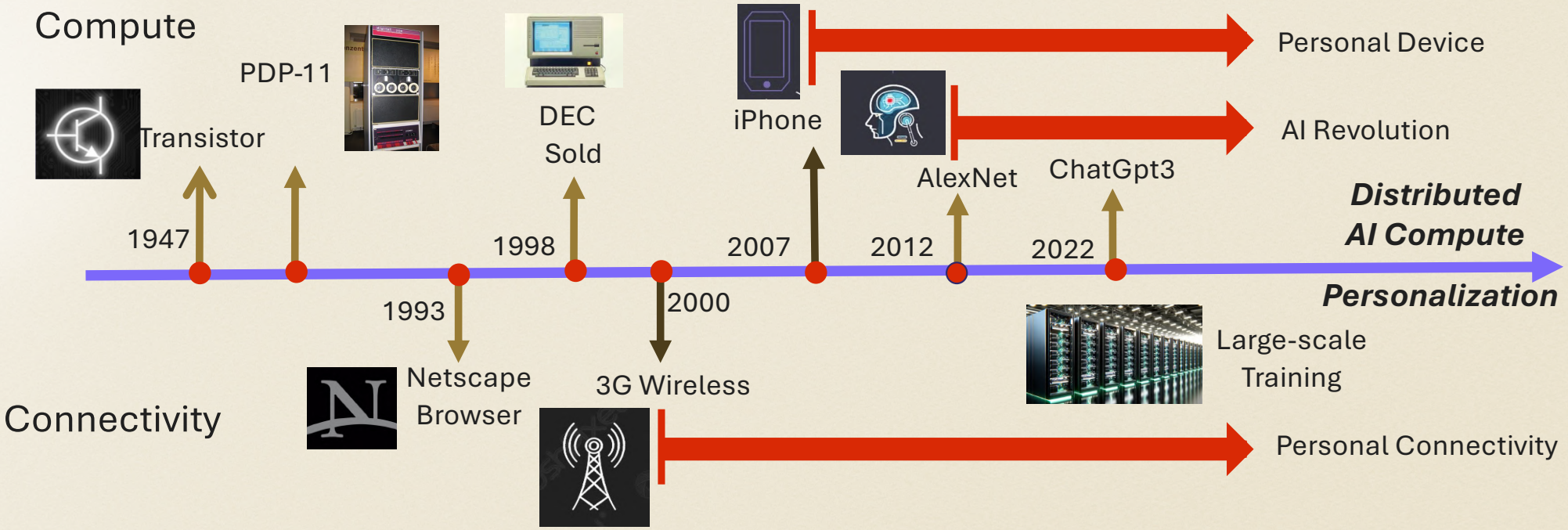
Beyond Innovation: RISC-V's Path to Mass Adoption with Mature IP

Wei-Han Lien
Chief Architect

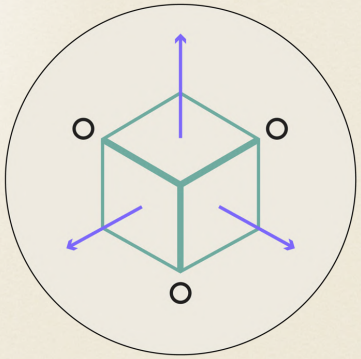
Chronical of AI Computation



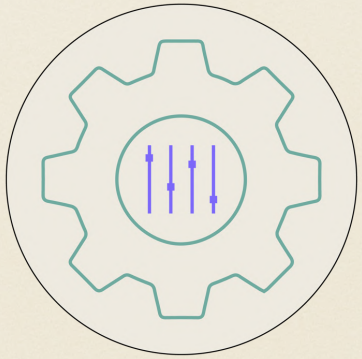
AI personalization requires ubiquitous AI computing

Scalable, open, portable, and composable hardware/software

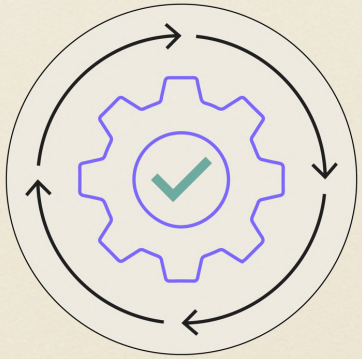
Benefits of Open Source Hardware



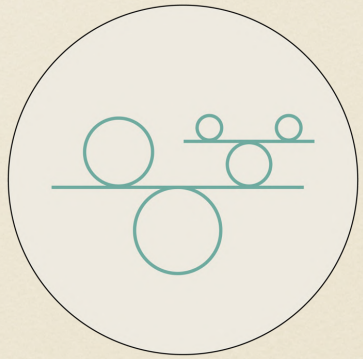
Scalable



Extensible



Efficient



Stable

Open Win with Price/Performance

All companies had proprietary Unix implementation;

1990-2010

All companies had proprietary ISA's and CPU's

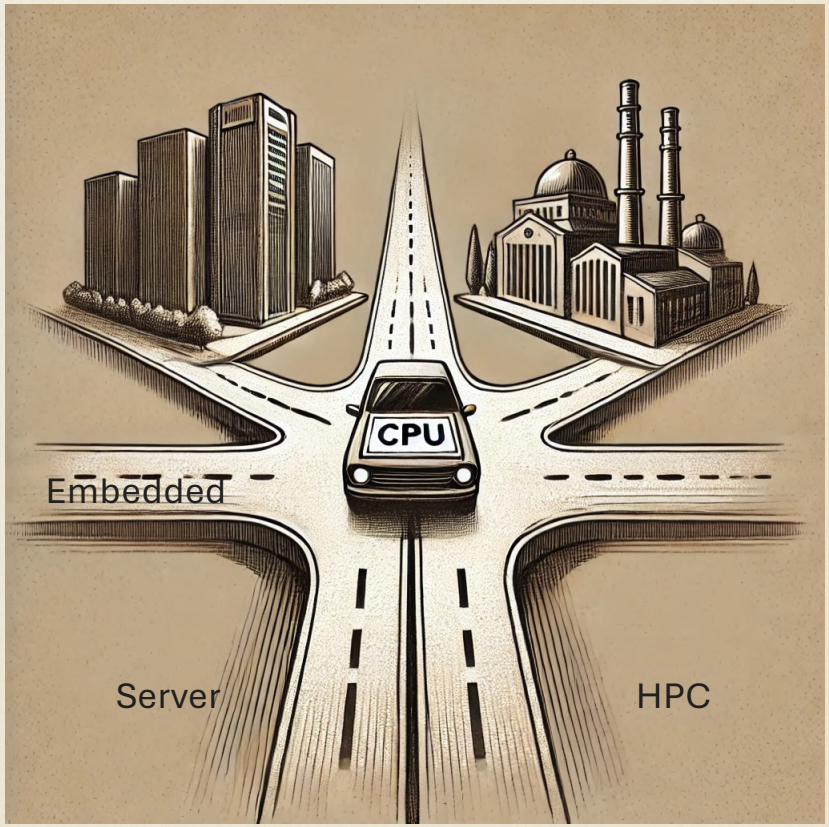
1990-2010

X86 won because it has 5x performance/price

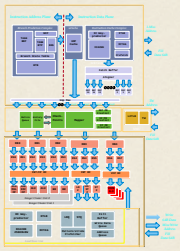
Tenstorrent builds the best performance/price chips (RISC-V+AI) to power Data Centers
Open + Best performance/price == Disruption

RISC-V in the Cross-road

- RISC-V's 15-year Journey
  - An academic project in UC Berkeley
  - Mainstream adoption in embedded and industrial applications
- Sustain momentum
  - High-performance computing implementation
- Key Requirements
  - Advanced O-o-O CPU design with high-bandwidth memory
  - Optimized compilers, firmware, and system software
  - Scalable heterogenous compute solution in AI/HPC/networking
- Without high-end investment, RISC-V confined to low-cost, niche markets



Foundational Compute IP's



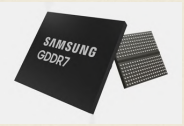
CPU

Tensix NEO

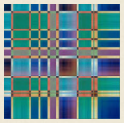
SOC Technology

PCIe

Networking



Memory



Fabric

Power/Thermal Management

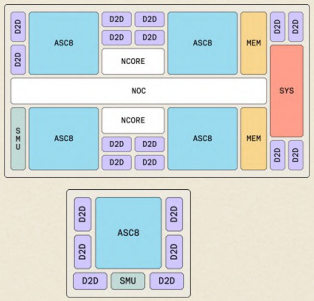
RAS

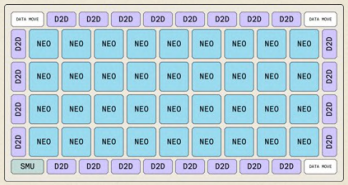
Security

Debug

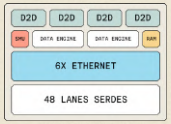
Performance Modeling

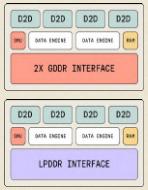
Chiplet Technology



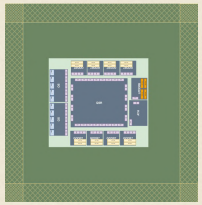


CPU, AI, IO, Memory chiplets





Applications



Data Center

Tenstorrent

Yayui

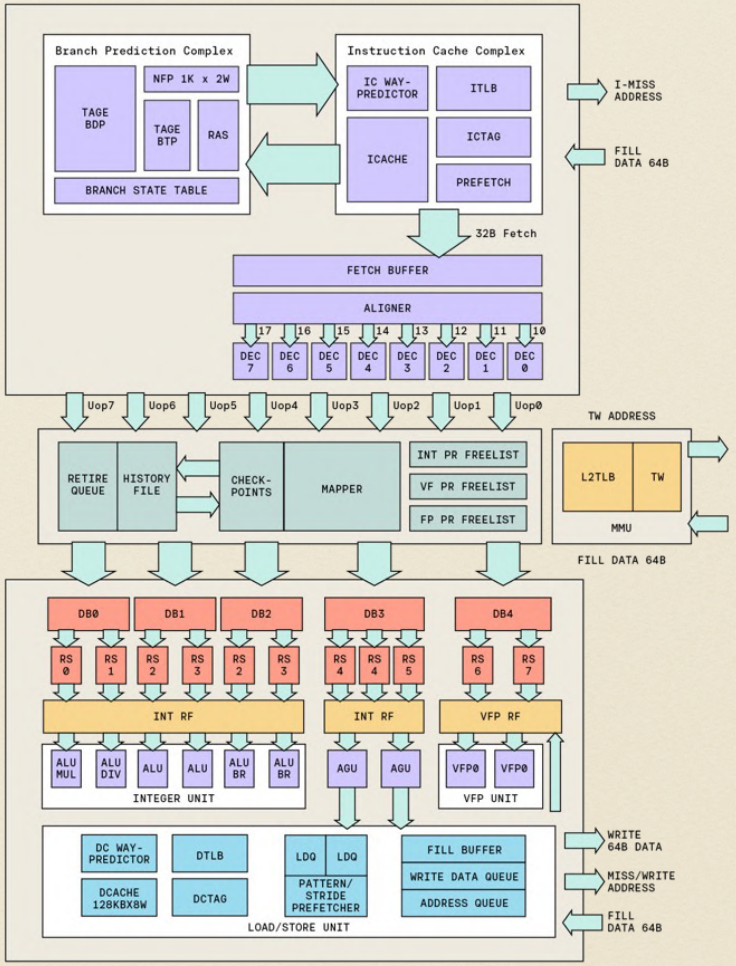
AIDC

RiscV CPU - Ascalon

- Disruptive high-performance RISC-V processor for AI and server
- Projected Zen5 performance in 2024

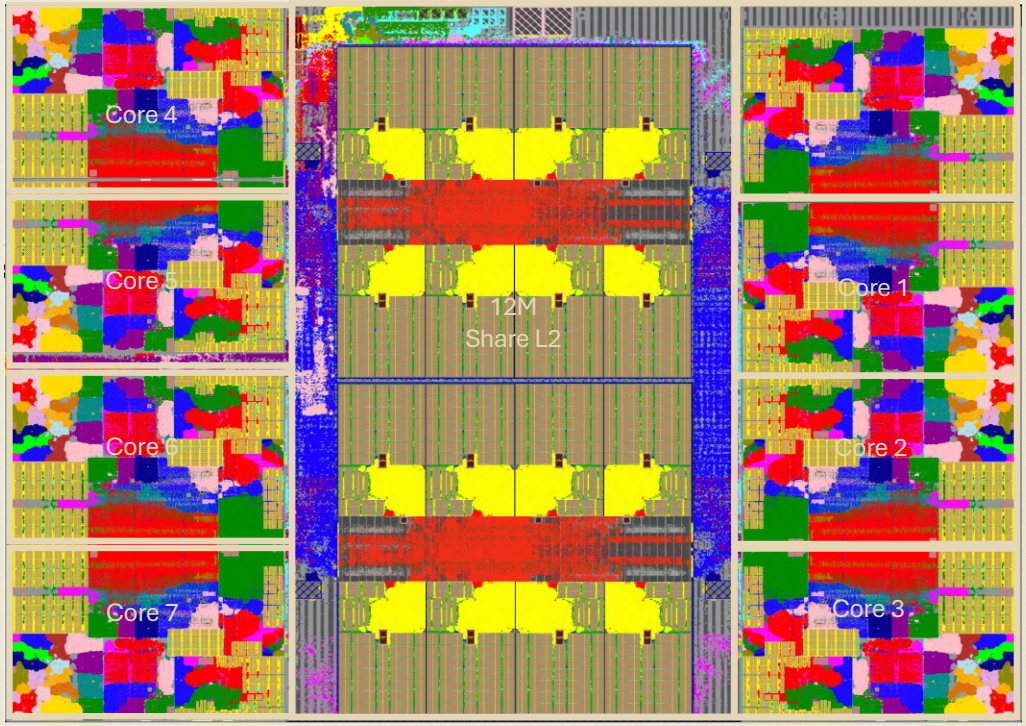
RVA-23

- Advanced branch predictions
- 8-wide decode
- 3 LD/ST with large load/store queues
- 6 ALU/2 BR
- 2 256-bit vector units
- 2 FPU units



Ascalon Cluster

- Athena Chiplet Q2/2025
  - Ascalon CPU cluster
  - PCIe Gen6
  - LPDDR5 memory controller
- Samsung SF4
- 18 SPEC2k6INT/GHz performance correlated

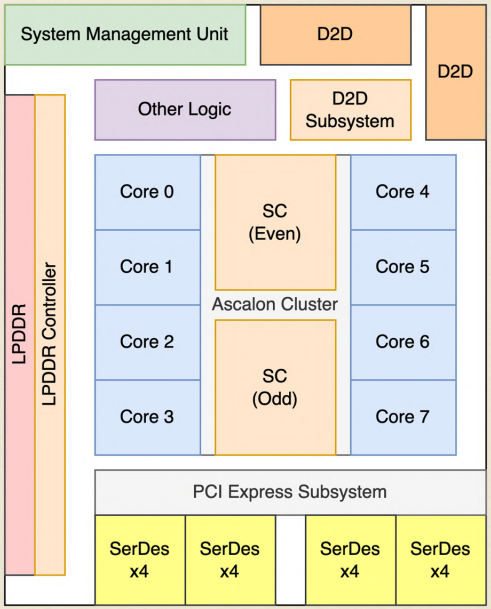


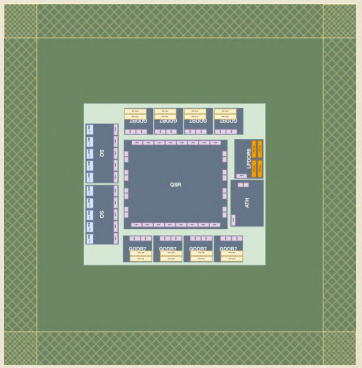
Ascalon Core

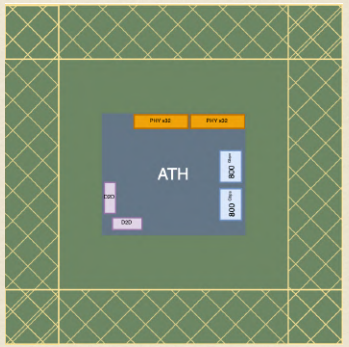
8 Ascalon Cluster
12M Shared L2

Athena Chiplet

- 8-core Ascalon CPU cluster
  - 35 SPECINT2K17
- PCIe Gen6x16, 128 GB/S TX/RX
- LPDDR5X 8533 68 GB/S
- Q4/2024 Tapeout
- Samsung SF4
- Tapeout

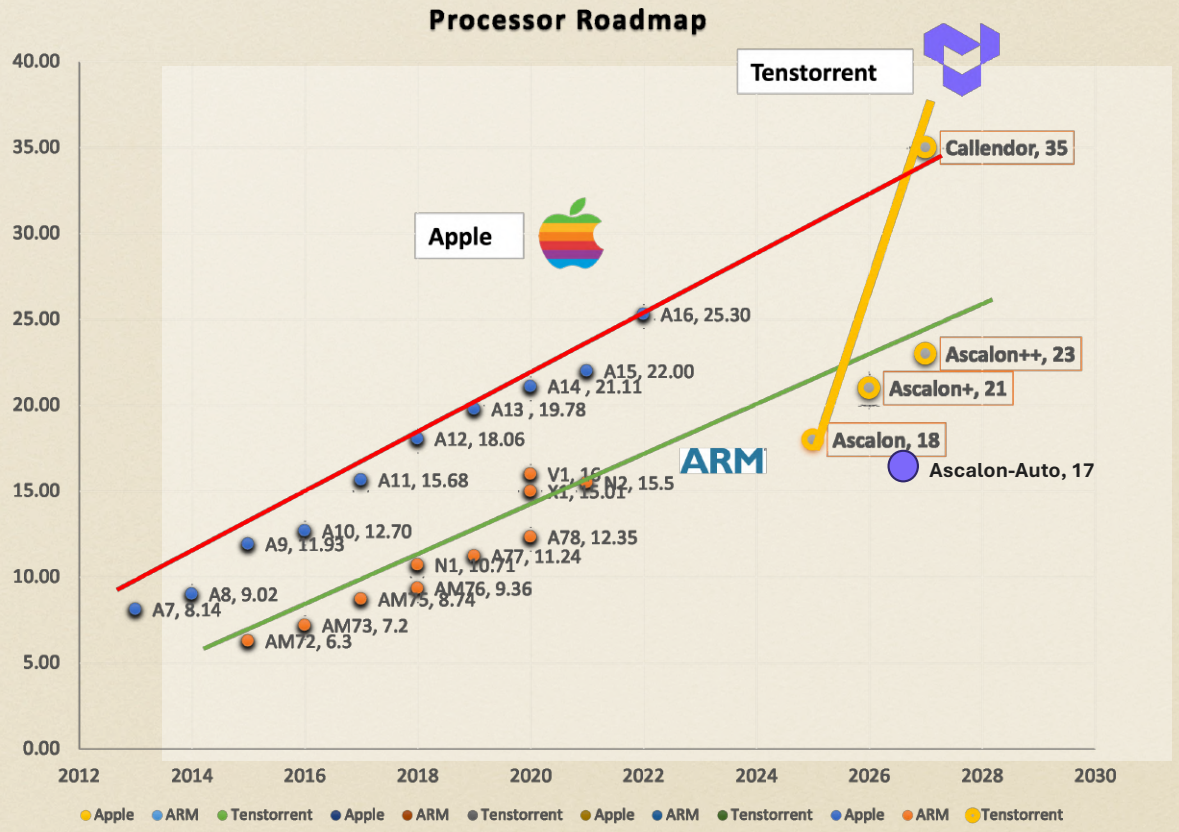






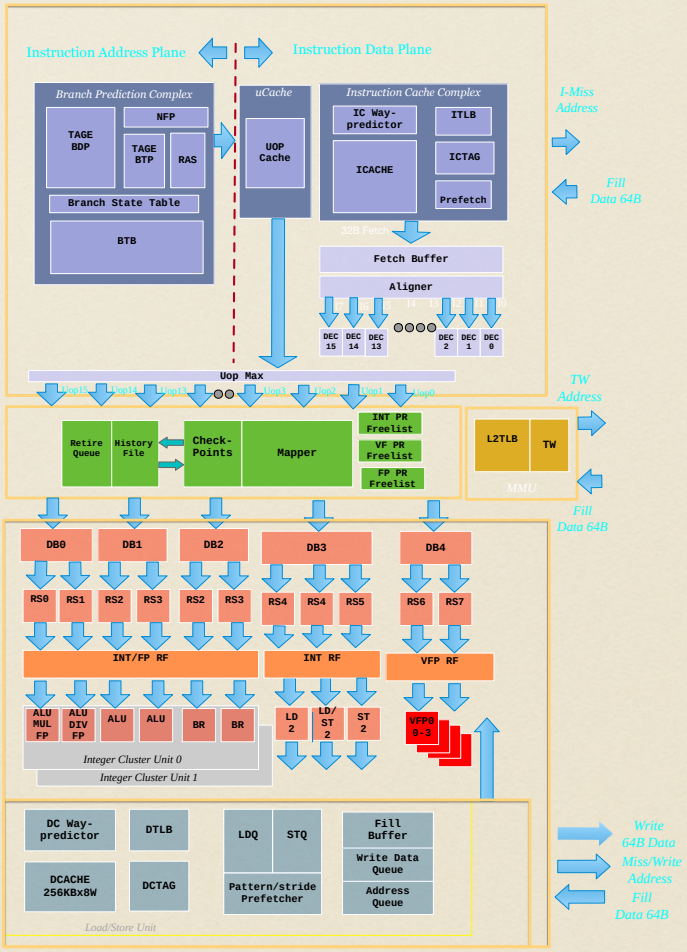
RISC-V CPU Roadmap

- Ascalon yearly 10% upgrade
  - Ascalon+ 20 SPEC2k6INT/GHz
  - Ascalon++ 22 SPEC2K6INT/GHz
- **Callandor 100%**
  - **35 SPECINT2K6/GHz**



Callandor O-o-O Superscalar Processor

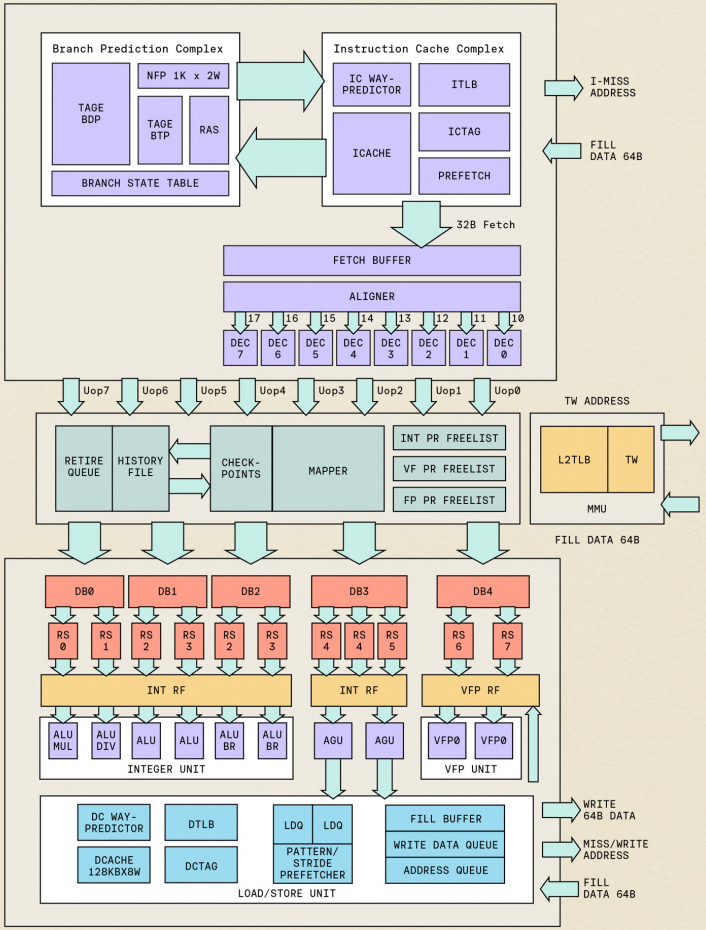
- At the forefront of CPU performance innovation in Q1/2027
- 35 SPEC2K6INT/GHz
- 3.5 SPEC2K17INT/GHz
- RVA25
  - Front-end
    - Advanced branch predictors: Two taken branches
    - Decouple Front-end
    - uOp cache
  - Mid-core
    - 16-wide decode
    - Register file sharing
    - 1K ROB
  - Execution uits
    - 6 wide LD/ST with large load/store queues
    - 8 ALU/4 BR
    - 4 256-bit vector units
    - 4 FPU units
    - Matrix engine



Ascalon-Auto IP

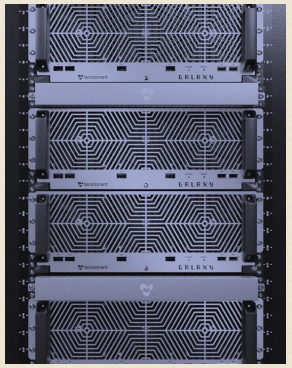
- ISO26262 Functional Safety Features
  - Dual Core Lock Stepped with Time Disparity
  - Coherent & Non-Coherent Bus Protection with CRC or ECC
  - EEC Protection for L1 & L2 Cache
  - AXI Interface Parity Protection
  - RAS & Fault Controller
  - Safety Bus for Debug
  - Software Test Library Support

- SPECInt2K17@2.25GHZ up to 30 per cluster
- Dhrystone MIPS : 12.38 DMIPS Rate/ MHZ



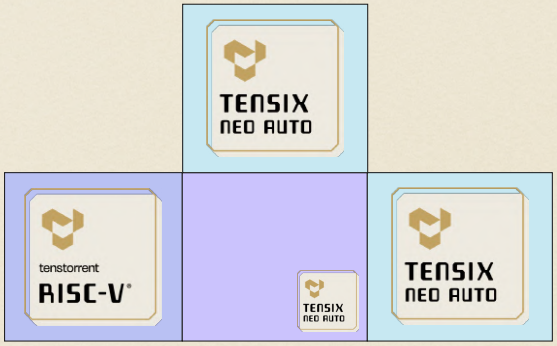
Auto IP potential user cases in ADAS/ADS

Level 2+/3 → Level 3/4 → Level 4/5 → Data Center



Dual SoCs

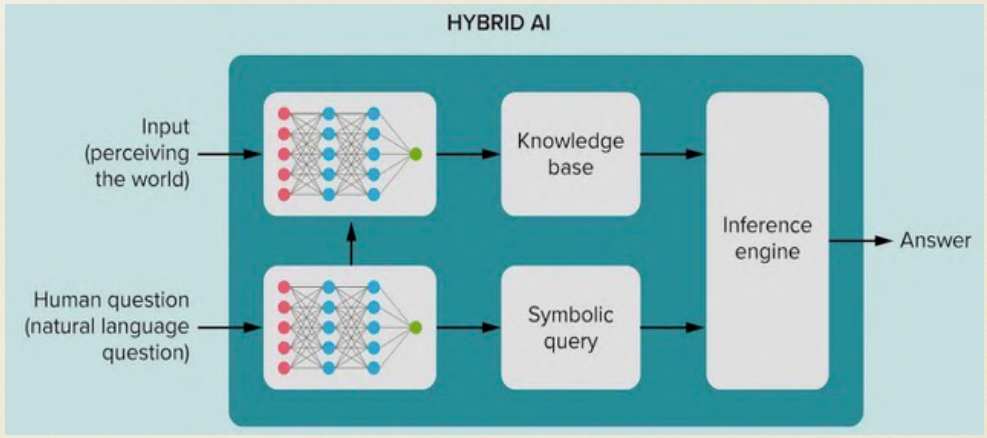
Dual SoCs + AI Co Processors



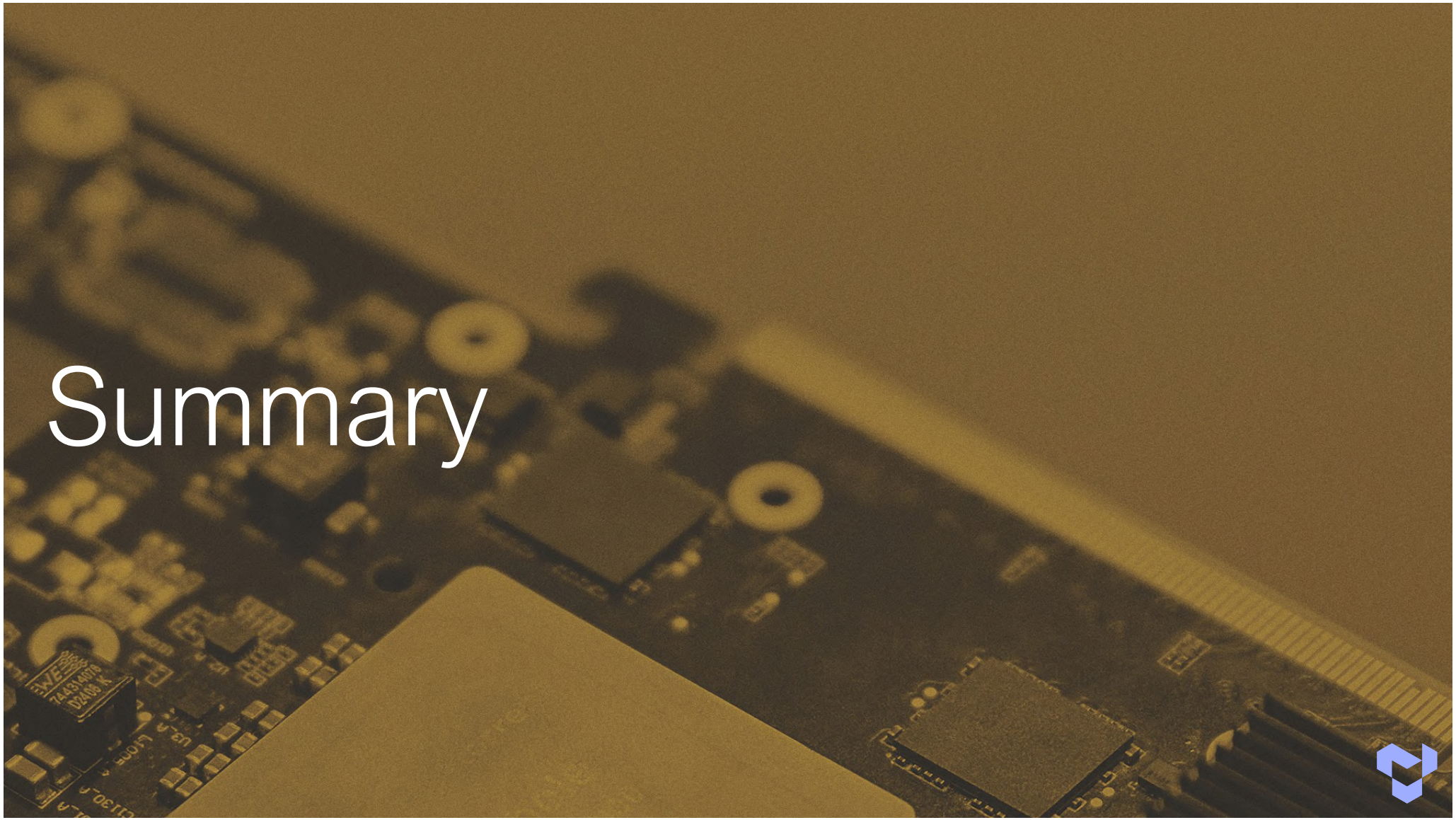
Central Compute Unit - Heterogenous

Neuro-symbolic AI Chiplet

- Heterogenous Compute with high-performance Ascalon
  - Neuro-symbolic AI
  - Optimize CPU/GPU data transfer latency

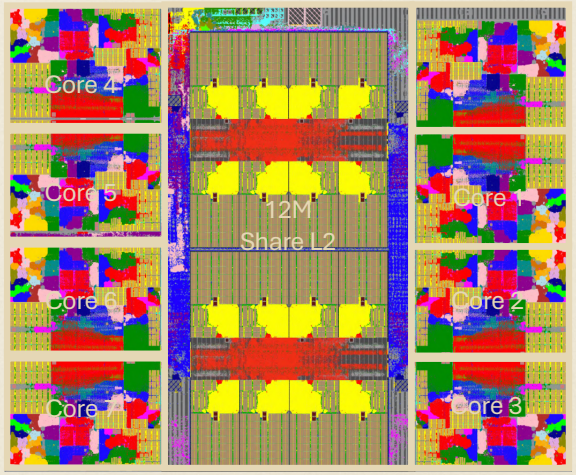


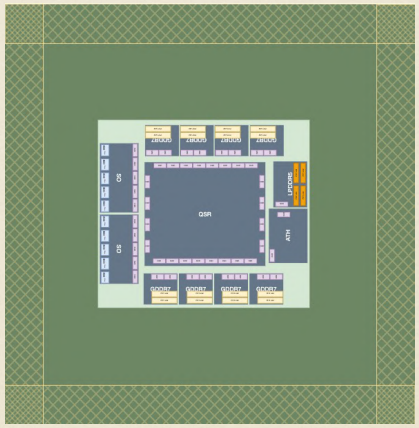
Summary

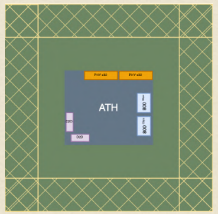


Summary

- AI personalization
  - Ubiquitous AI computing
- Efficient computers in all level of applications
  - Collaboration will be the key
- RISC-V provides scalable, open, portable, and composable customized hardware
- Portable computing software
- Meet future AI compute challenge with diverse fundamental technology portfolio
- But we need high-performance RISC-V implementations







We Are HIRING!!

- Tenstorrent just raised 700 M series D
- Open Japanese design center office in Apri !
  - 100 People capacity
  - Jim will be here. Welcome to talk to him.
- Work on AIDC RISC-V chiplet in Rapidus N2
- Cool leaders

Nakano
Mamoru

Yasuo
Ishii

Camille
Vuillaume