



Xuantie-910 : Innovating Cloud and Edge Computing by RISC-V



Yu Pu



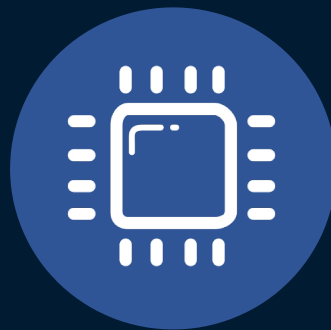
Open Source, Building the Chip Ecosystem in the New AIoT Era

**Chen Chen, Xiaoyan Xiang, Chang Liu, Yunhai Shang, Ren Guo, Dongqi Liu,
Yimin Lu, Ziyi Hao, Jiahui Luo, Zhijian Chen, Chunqiang Li,
Yu Pu, Jianyi Meng, Xiaolang Yan, Yuan Xie and Xiaoning Qi**

Building the Chip Infrastructure of the AIoT Era



Infrastructure Provider of the AIoT Era



MCU



Security



Intelligent Computing

...



Industry Control



Memory Control

AliOS

Domain Specific SoC Platforms (IPs from Partners)

RISC-V Compatible Processor

Domain Specific Architecture

Xuantie - The Evolving Processor Architecture



First RISC-V
Processor
with HW TEE

902



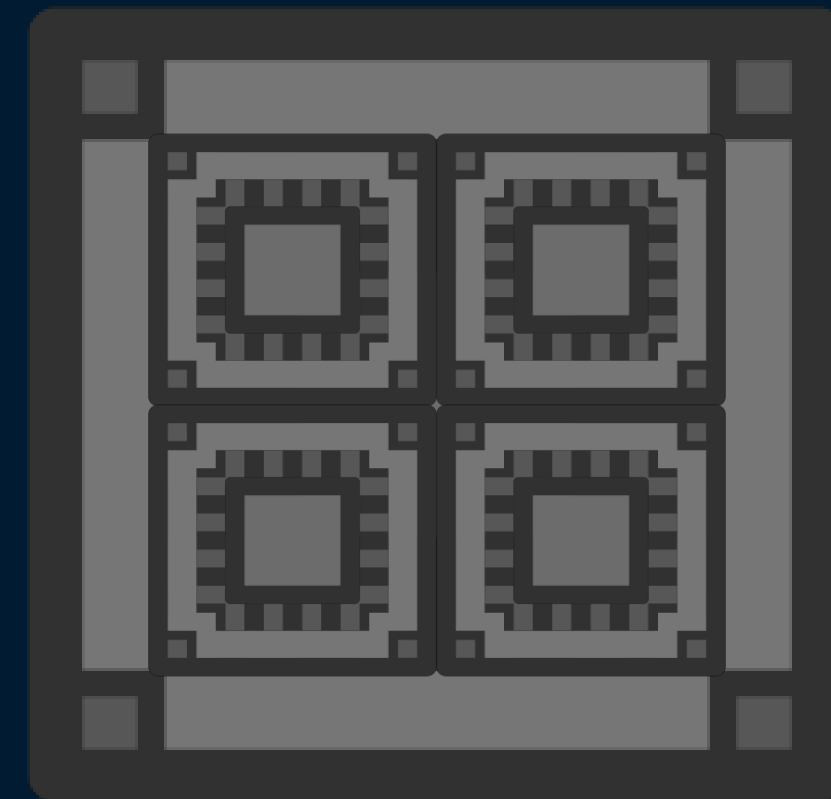
In progress

9xx

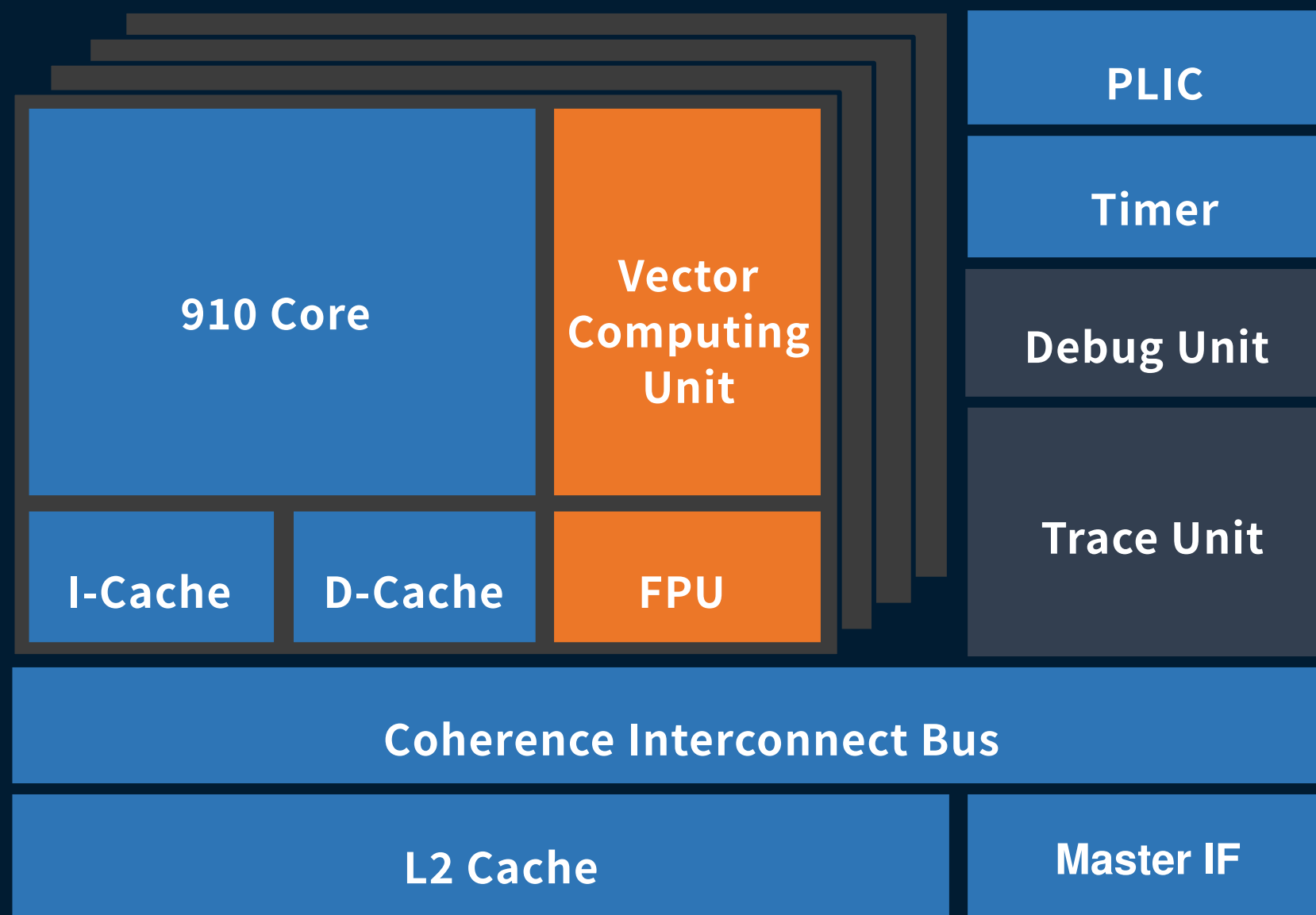


Ultra High
Performance
Processor with AI
Acceleration Engine

910



Ultra High Performance Architecture – Xuantie910

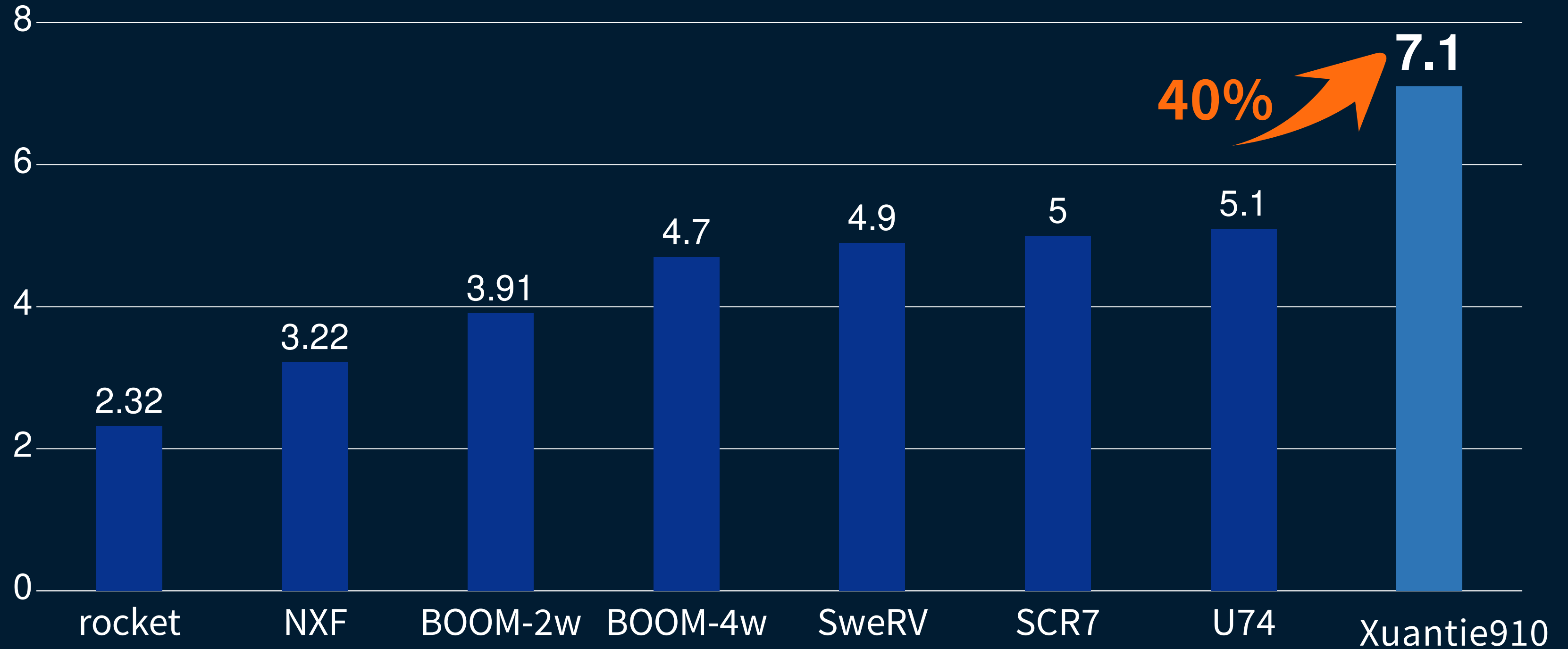


- RISC-V RV64GCV
- Cluster Based Multi-core Architecture
- 1/2/4 Cores per Cluster
- 32KB/64KB L1 D\$; 32KB/64KB L1 I\$
- 64-bit , 12-stage 、 Out-of-Order
- 3-decode , 8-issue
- Dual Issue Out-of-Order Memory Access
- High Performance Hybrid Branch Processing
- Multi-mode Dynamic Data Prefetch
- Vector Engine for AI Acceleration
- AI, Edge servers, Industrial control , ADAS

Remarkable Performance



CoreMark/MHz



Data source:
<http://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-151.pdf>
https://content.riscv.org/wp-content/uploads/2019/04/RISC-V_SweRV_Roadshow-.pdf
https://content.riscv.org/wp-content/uploads/2019/06/17.00-syntacore_zurich_ws.pdf
<https://www.sifive.com/cores/u74-mc>

Compatible with RISC-V Specification

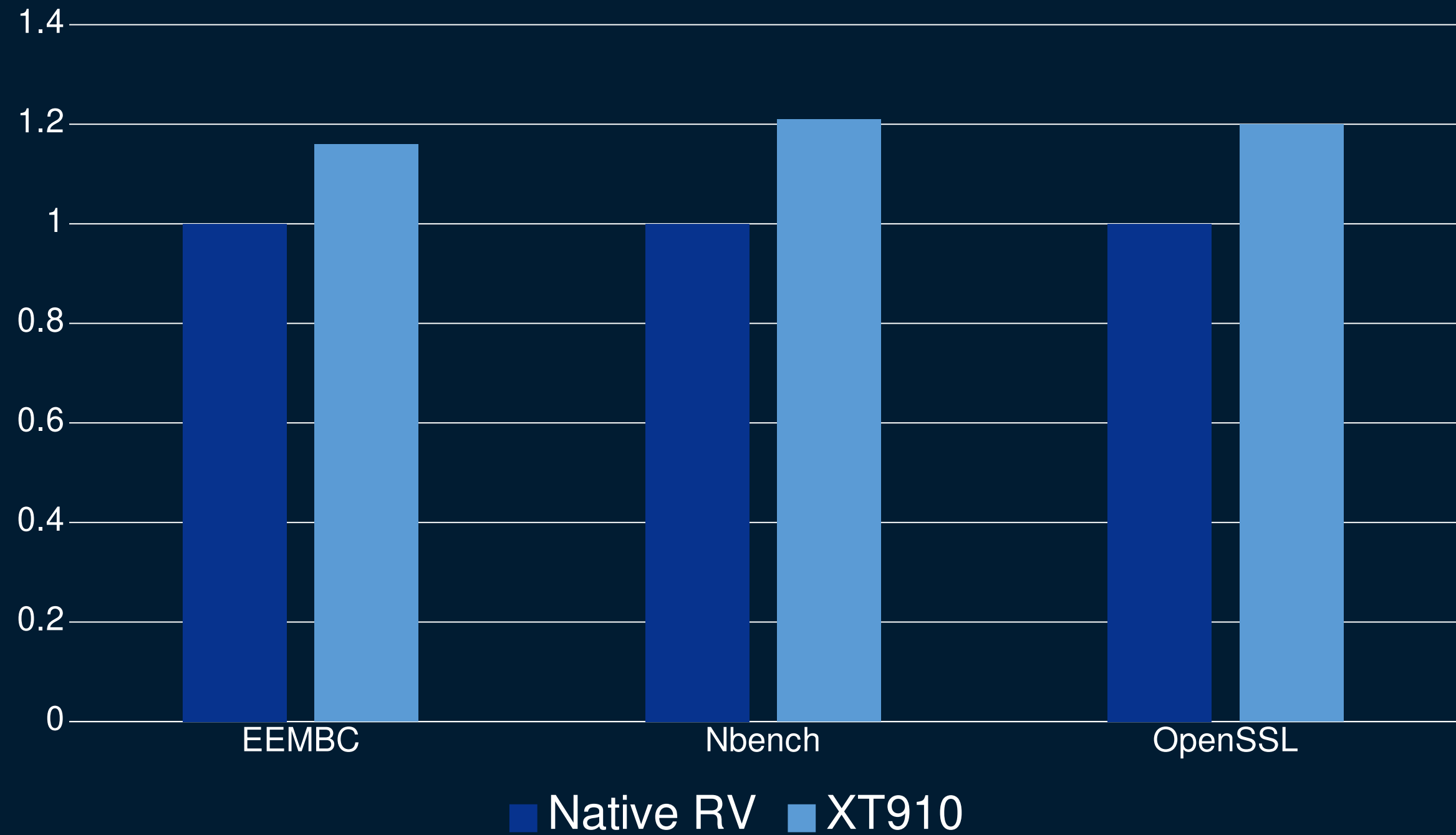


ISA	RV64GCV
Vector	RISC-V 0.7.1 Vector Extension FP16/32/64, INT8/16/32/64
Privilege Mode	Machine + Supervisor + User
Memory Management	Sv39 MMU + 8/16 PMP
Interrupt Controller	Clint + PLIC

Extended Enhancement - RISC-V Turbo

RISC-V Turbo

- Computing
 - Bit operation
 - Memory access
 - Multi-core synchronization
-
- Memory management
 - Cache、TLB



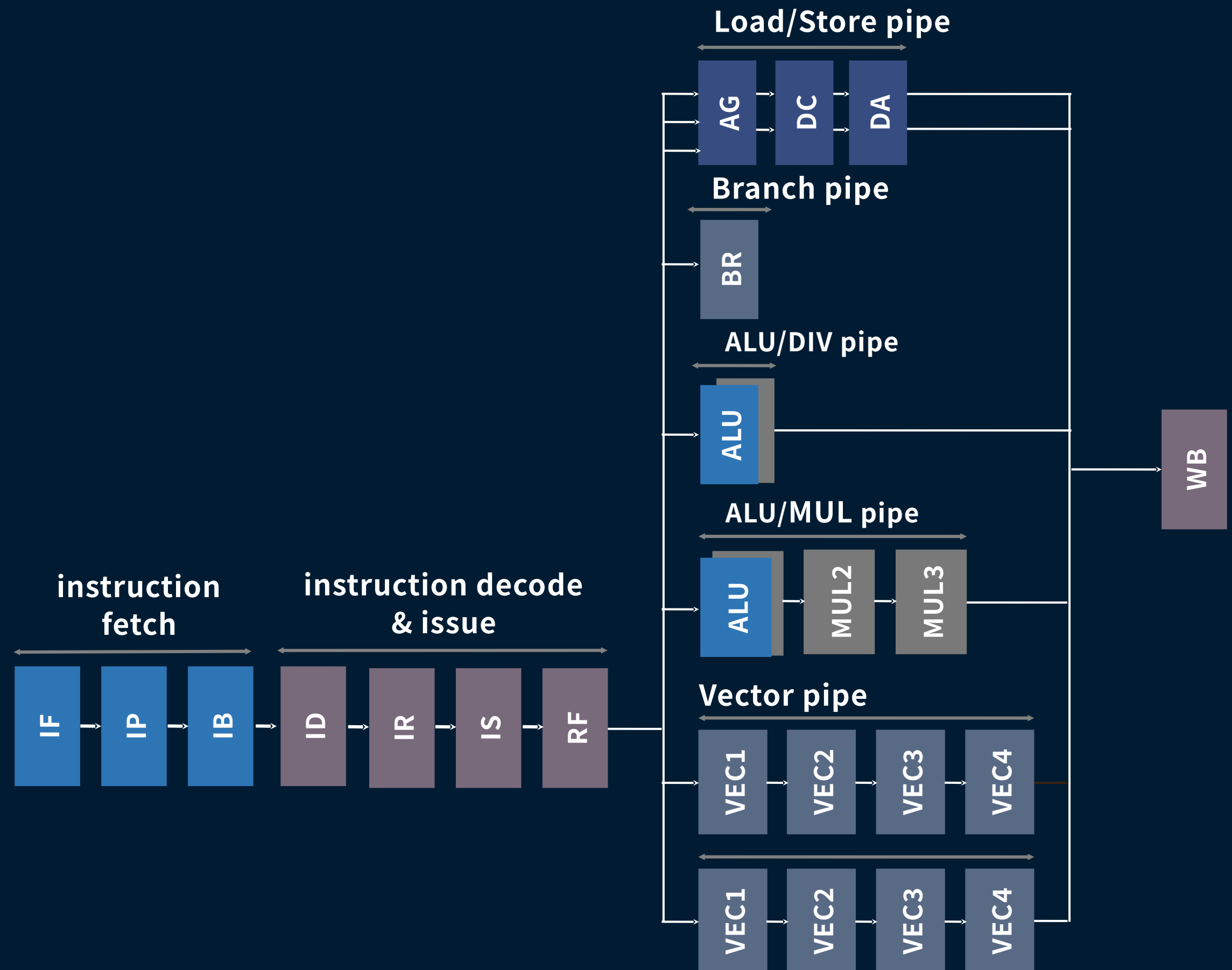
Deep Superscalar Out-of-Order Pipeline

- **Front-End**

- Fetch 8 Instructions/cycle
- Decode 3 Instructions/cycle
- Issue 8 Instructions/cycle

- **Back-End**

- Out-of-Order Memory Access
- Dedicated Branch Processing
- Out-of-Order Vector Computing



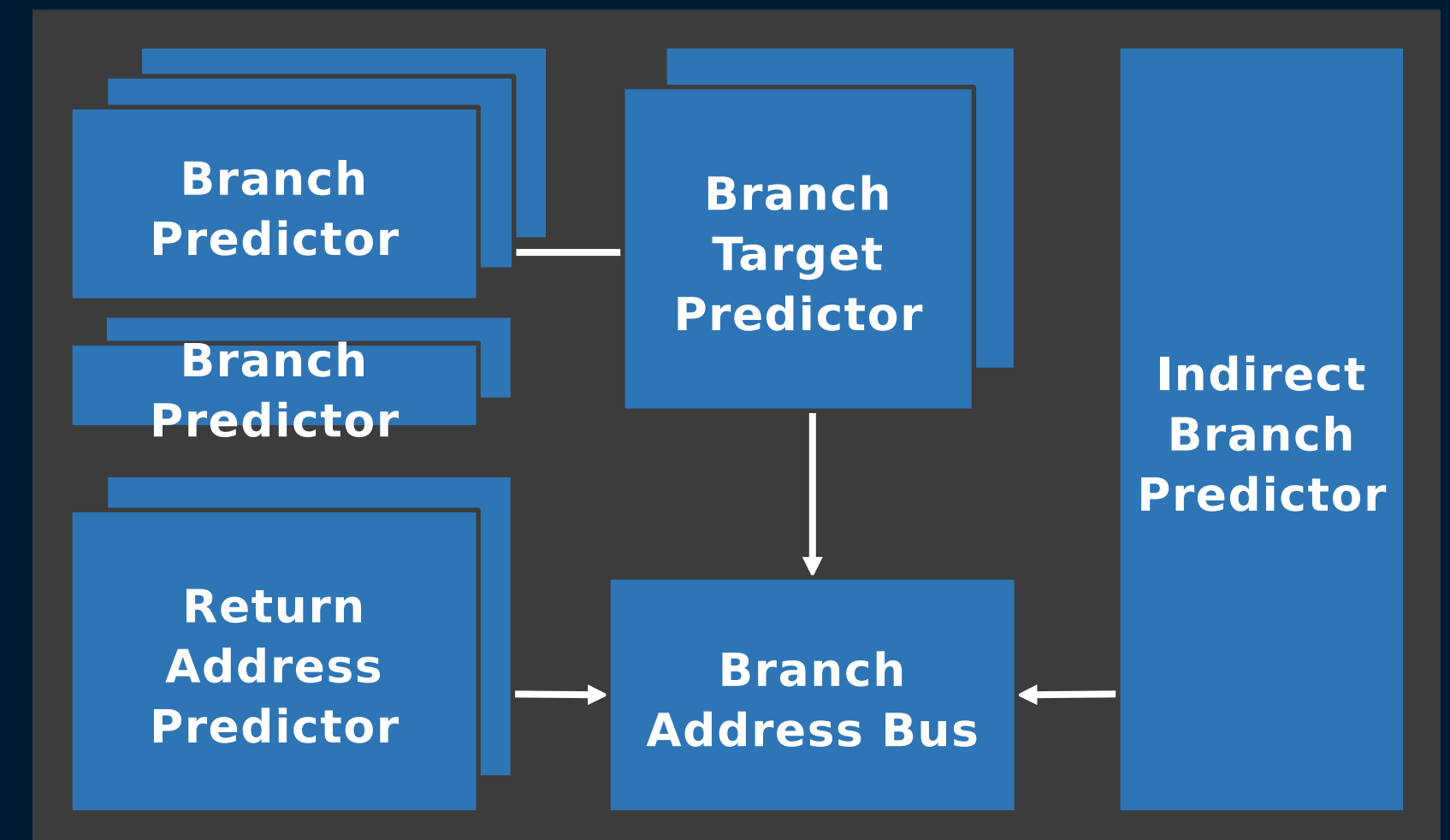
Instruction Fetch Unit with Hybrid Prediction

- **Hybrid Multi-mode Branch Prediction**

- Branch Direction Prediction
- Branch Target Prediction
- Return Address Prediction
- Indirect Branch Prediction

- **High-bandwidth Parallel Fetch**

- 128-bit Fetch
- Up to 8 Instructions Packaged in Parallel
- Instruction Cache Way Prediction
- Loop Acceleration



Dual Issue Out-of-Order Load Store Unit



- **Out-of-Order Dual Issue**

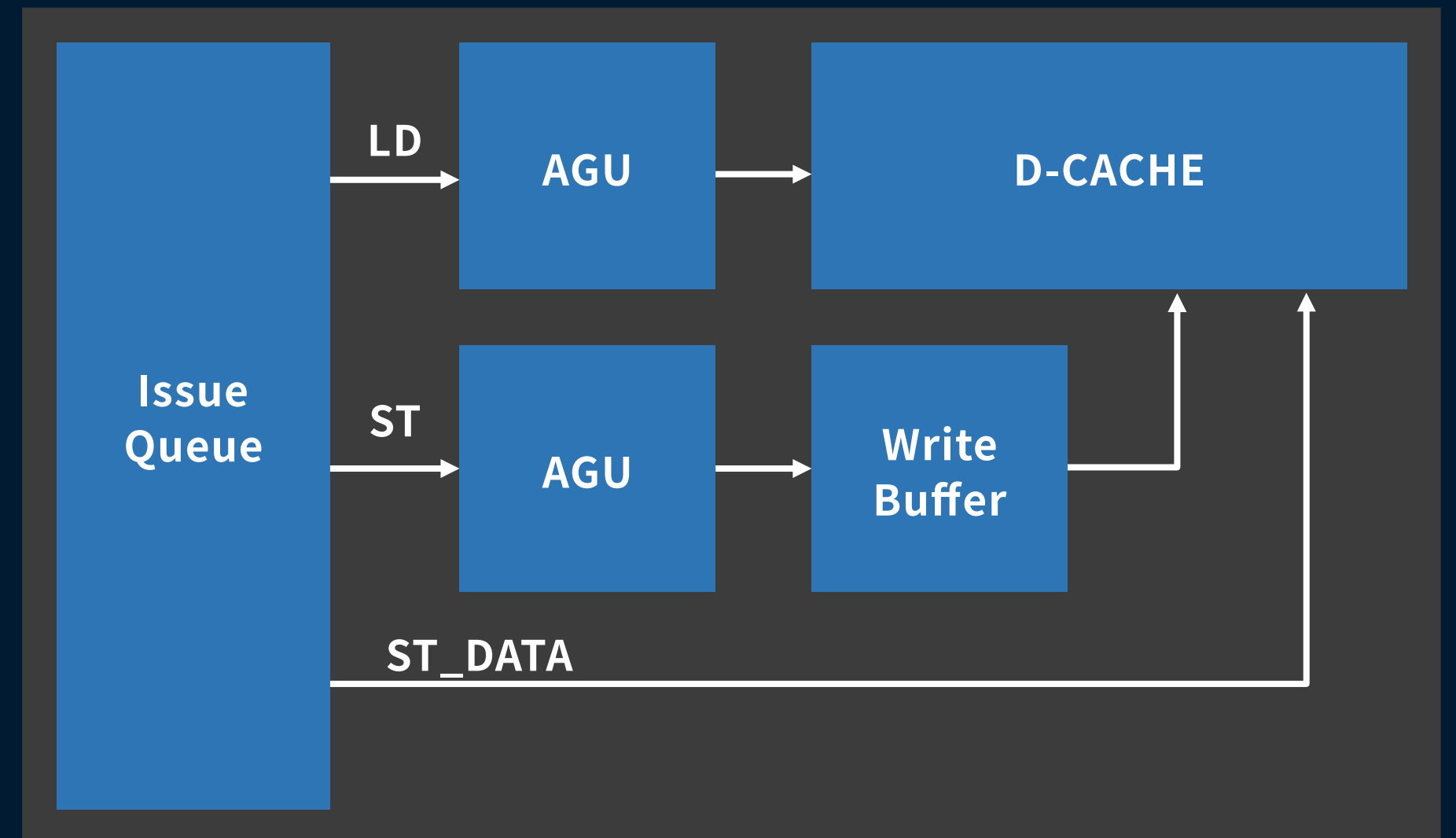
- Load/store Address Pipeline
- Independent Store Data Pipeline
- Speculation Fail Prediction

- **Load/store Fast Complete**

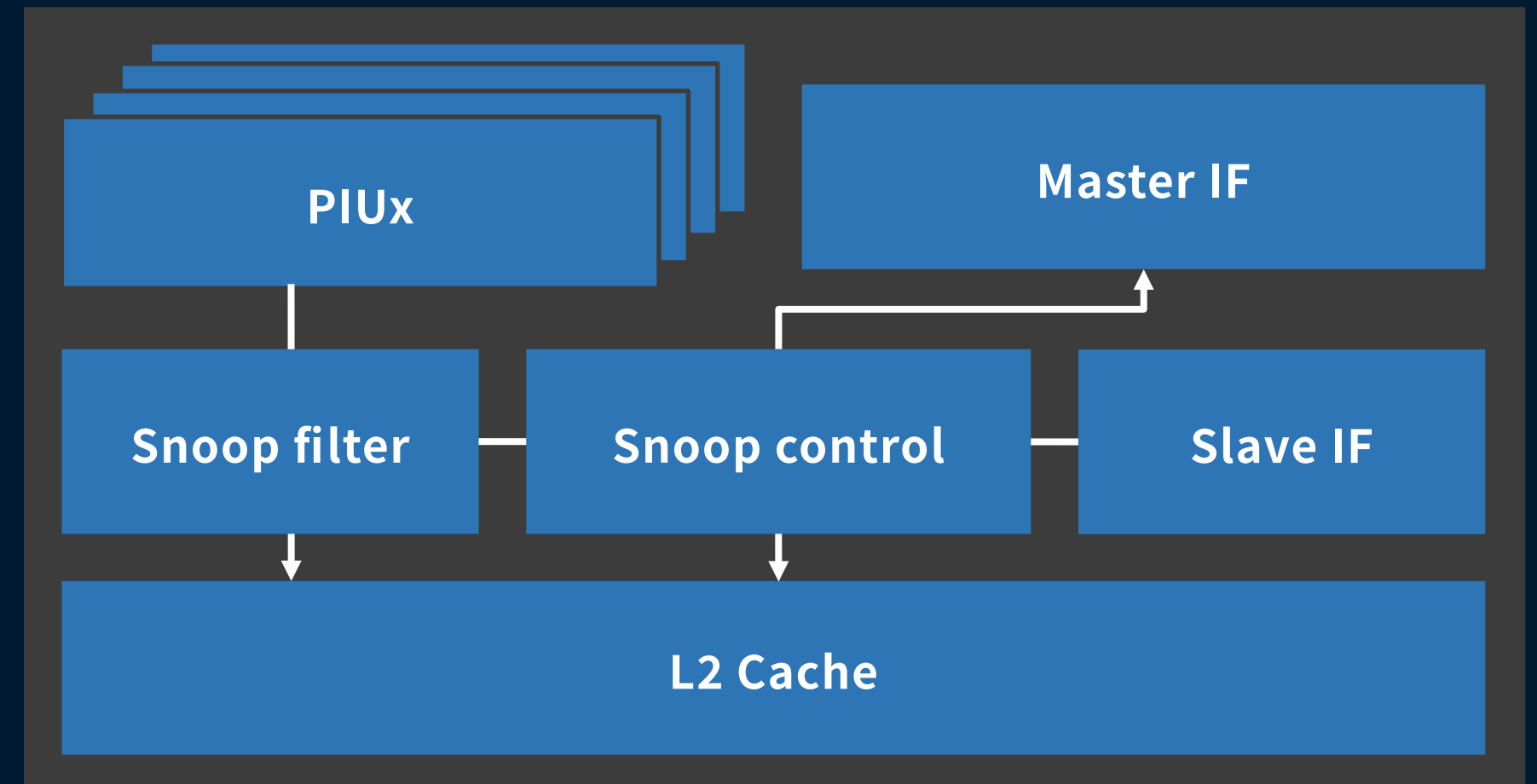
- 3 cycle Load-to-Use
- 1 cycle Store Execution

- **Powerful Prefetching Capabilities**

- Multi-mode and Multi-stream
- Both Virtual and Physical Address Prefetching
- Configurable Prefetch Capacity



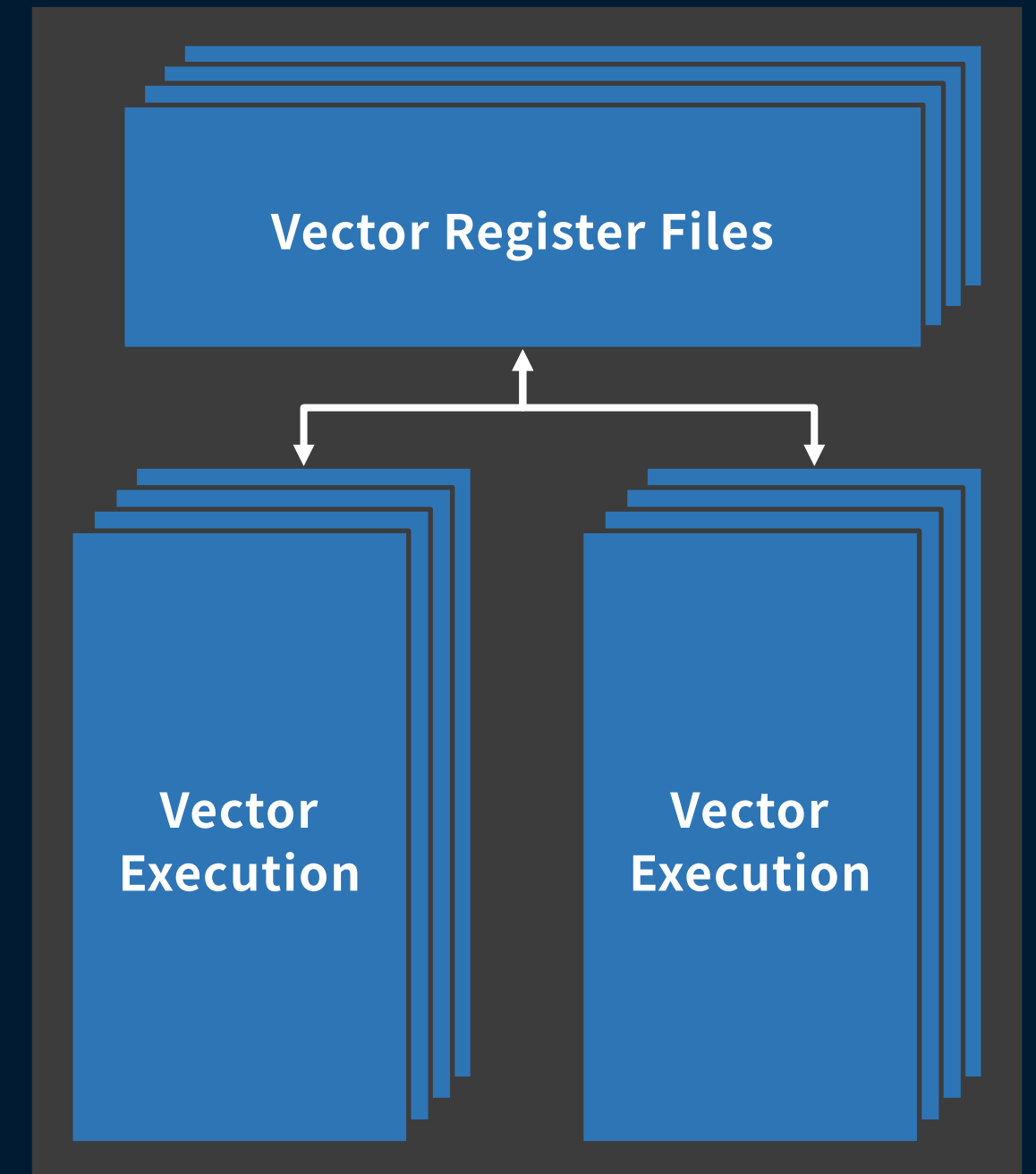
- Decoupled Processor Interface Units (PIUx)
- MOESI Coherence Protocol
- Directory-based Architecture
- Snoop Filter Supported
- Configurable L2 Cache , up to 8MB
- ECC Supported



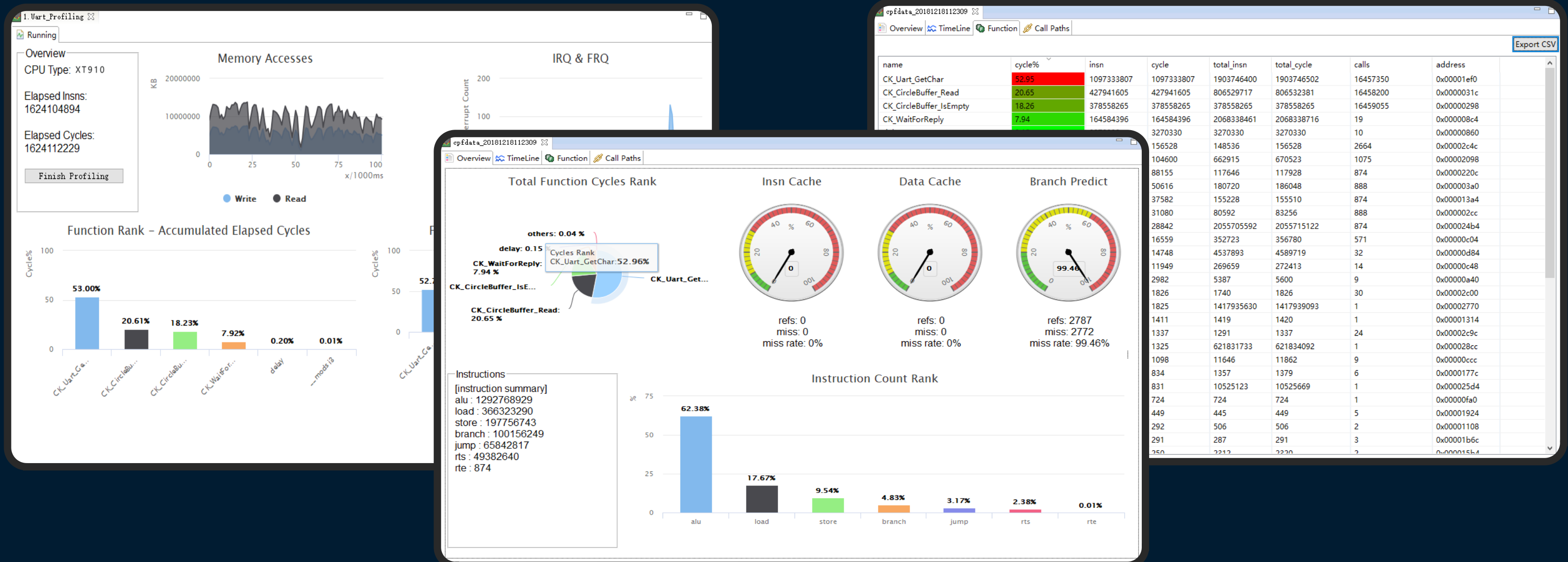
AI Optimized Vector Computing Engine



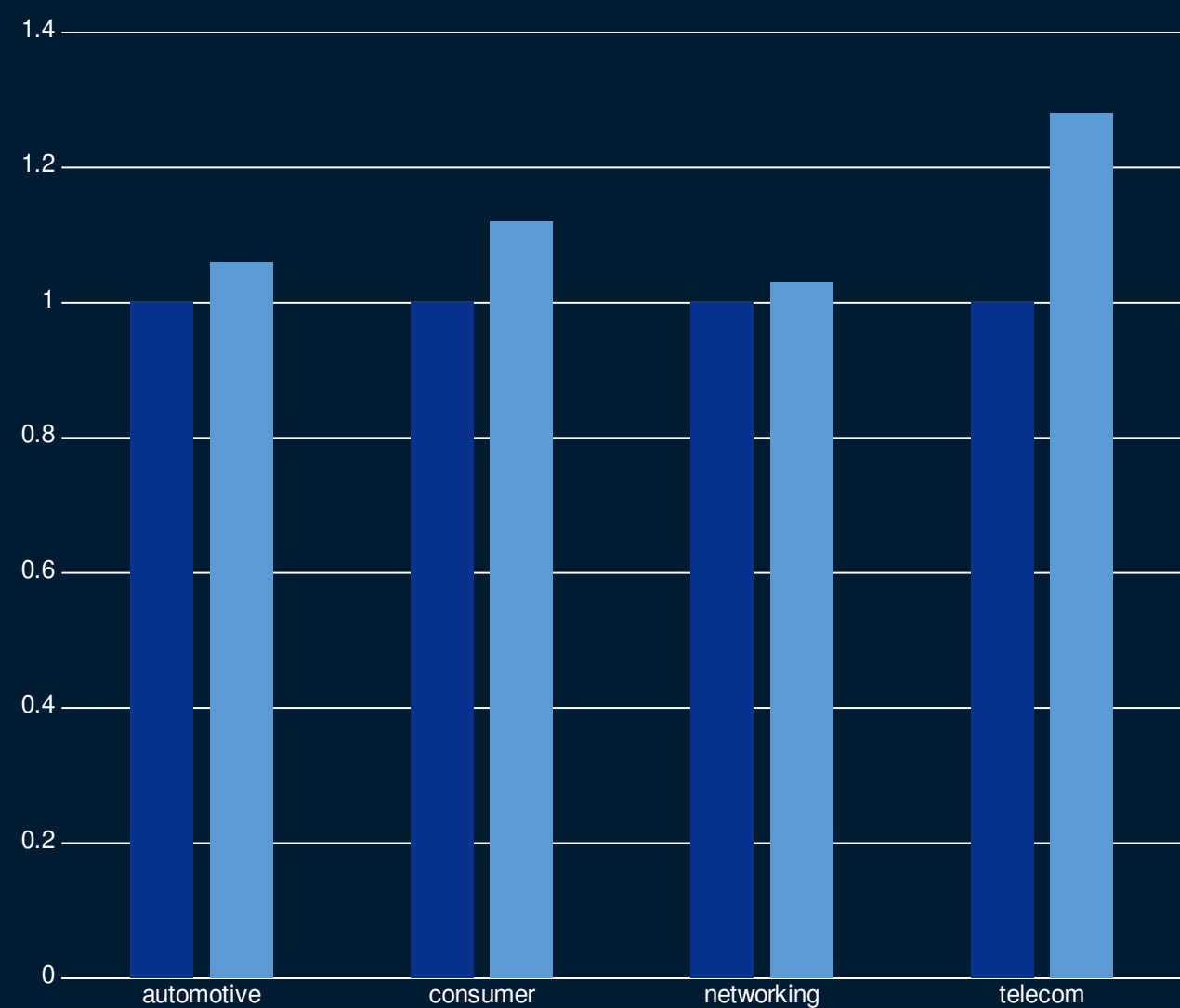
- **Compatible with RISC-V 0.7.1 Vector Extension**
- **Supports FP16/32/64, INT8/16/32/64**
- **256-Bit Operation Width, VL = 128 and 2 pipelines**
- **Two 128-Bit Vector ALU Ops per Cycle**
- **One 128-Bit Vector Load and One 128-bit Vector Store per Cycle**
- **Direct Access to L1\$ on Vector Load and Vector Store**
- **Dual-issue Out-of-Order Vector Execution Pipeline**
- **More than 300GFLOPS of FP16 Computing Power per Cluster**
(32 FLOPS/core/cycle x 2.5 GHz x 4 Cores)
- **Half of FP16 computing power when widening to FP32**



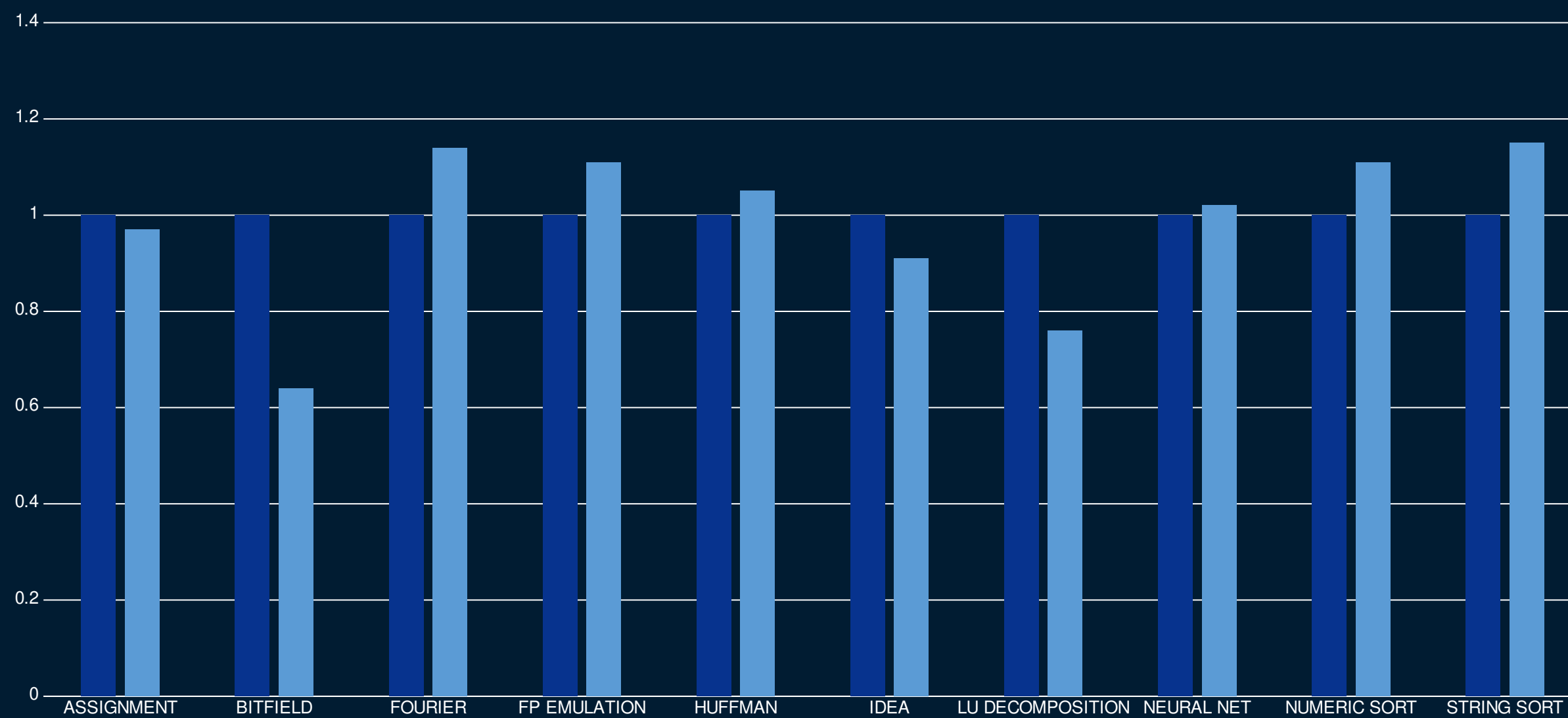
Efficient Profiling Engine



Experimental Results



EEMBC



nBench

Target Chip Info:

- CPU Type is C910V, Endian=Little, Vlen=128.
- DCache size is 64K.
- ICache size is 64K.
- Target is 2 cores.
- MMU has 1024 JTLB items.
- MGU zone size is 4KB.
- MGU zone num is 8.
- MISA: (RV64IACDFMVX, Imp M-mode, S-mode, U-mode)

GDB connect command for CPU 0:

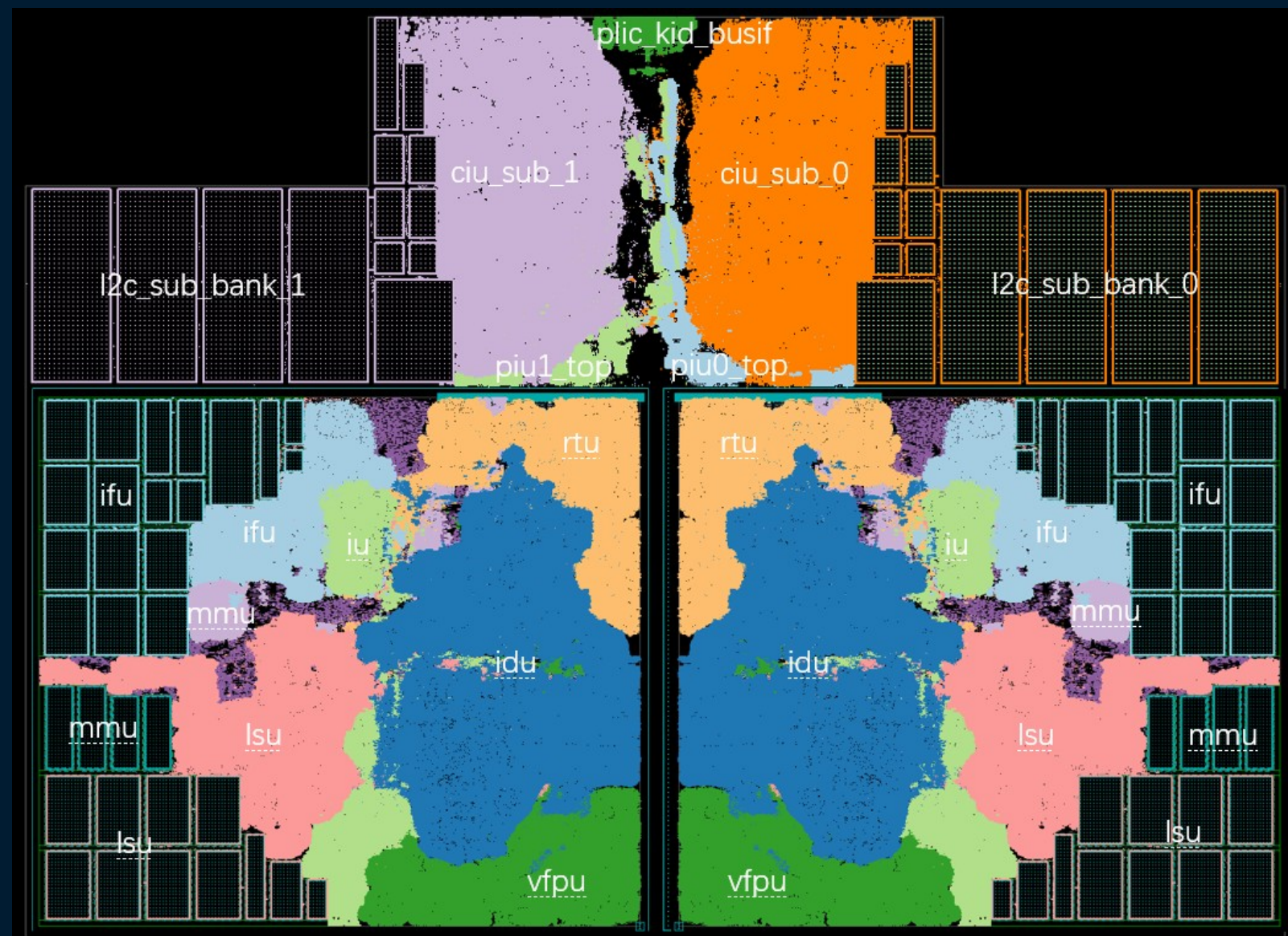
- target remote 172.16.150.172:1025
- target remote 192.168.5.1:1025
- target remote 192.168.126.1:1025

功能演示

图片上传
选择文件: cat0.jpg
上传图片 清除结果

位置	名称	置信度
B-05L, T-04 #275, #551	猫	99%

ASIC Implementation



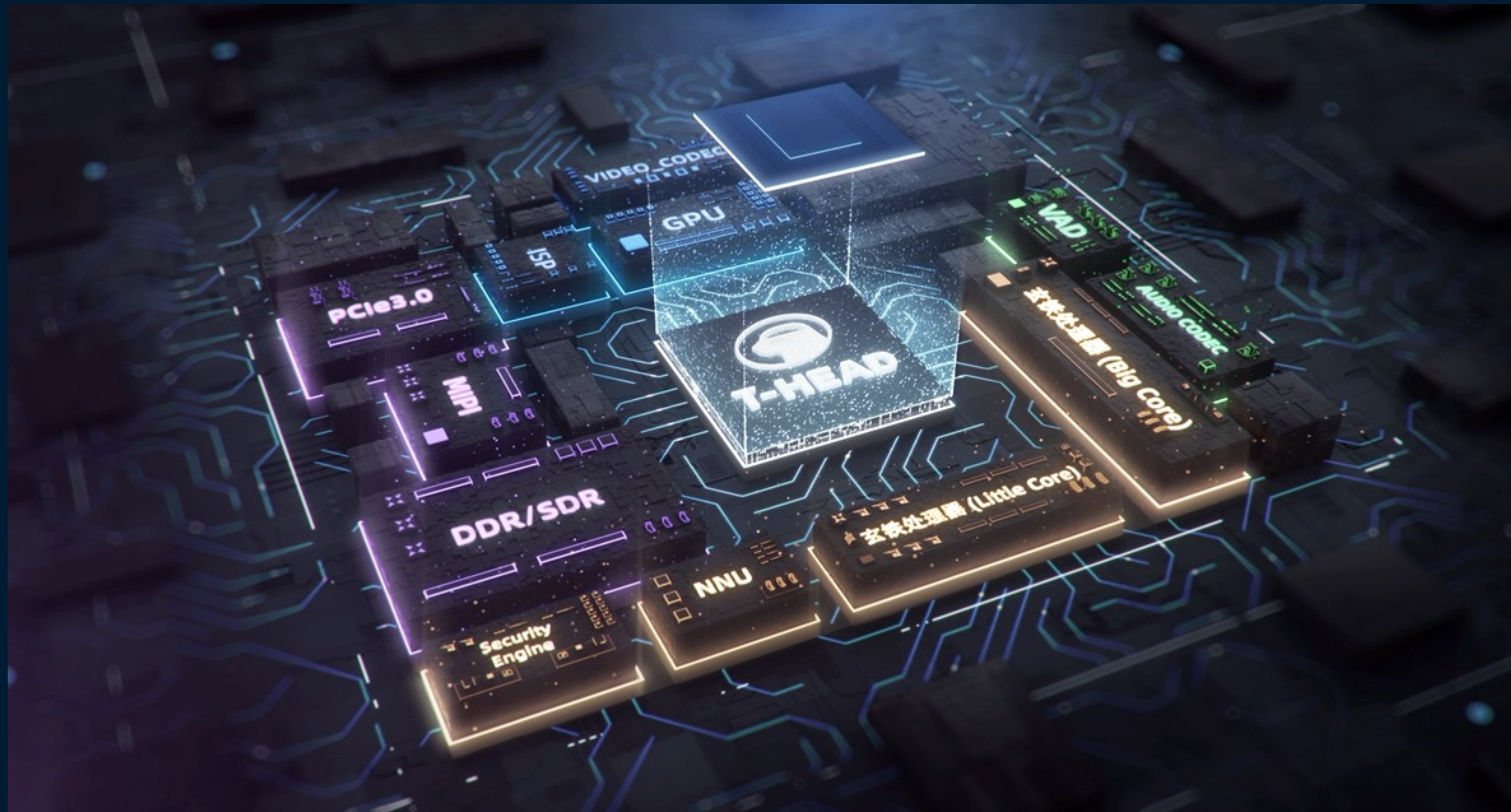
Process Technology	TSMC 12nm FinFET
Operating Frequency	2.0 GHz^a ~ 2.5 GHz^b
Area per Core (excl. L2\$)	0.6 mm² (without VEC) 0.8 mm² (with VEC)

a. LVT 6T-turbo STD cell, 0.8V VDD, TT 85°C
b. 30% ULVT STD cell, 1.0V VDD, TT 85°C

Wujian SoC Platform with Xuantie



Enabling chip differentiation competition



50%

Reducing Chip Design Time by 50%

50%

Saving up to 50% on Design Cost

- **Ultra High Performance Superscalar Processor**
- **RISC-V Compatible plus RISC-V Turbo Technology**
- **Dual issue Out-of-Order Memory Subsystem**
- **AI Vector Acceleration Engine**

