XUANTIE

Datasheet

XuanTie C950

Server-Oriented Flagship Processor

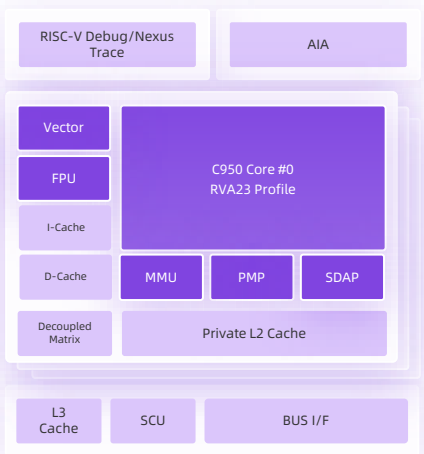
Overview

XuanTie C950 is a high-performance, 64-bit multi-core CPU IP. It features an out-of-order superscalar microarchitecture with 8-wide decode and incorporates advanced microarchitectural technologies such as TAGE/ITTAGE branch prediction, wide ALU execution pipelines, private L2 cache, and dynamic adaptation hardware prefetchers for high performance and robustness. It complies with the RISC-V RVA23 Profile and supports all optional extensions such as Vector Crypto, Zacas, and Zama16b. Additionally, XuanTie C950 supports the proprietary XuanTie AME (Attached Matrix Extension) ISA and enables seamless integration with the XuanTie TPE (Tensor Processing Engine) IP.

XuanTie C950 targets a wide range of high-performance applications, including cloud computing, edge computing, and AI computing.

Features

Feature	Description
Architecture	RVA23 Profile
Pipeline	Superscalar out-of-order microarchitecture with 8-wide decode
Floating Point	RISC-V F/D Extension
Vector	RISC-V Vector Extension v1.0 with Vector Crypto support
Matrix	XuanTie TPE coprocessor integration (AME v0.5)
Hypervisor	Suitable for Type #1 and Type #2 hypervisor
Cache system	Private L1 and L2 Cache Optional L3 shared cache
MMU	Sv57/Sv48/Sv39 with PA48
Bus Architecture	Direct Connect Mode: CHI.E/CHI.F Multi-Processor Mode: AXI4.0/ACE4.0
Security	CFI (Landing Pad, Shadow Stack)/Smmtt
QoS	CBQRI (Capacity and Bandwidth Controller QoS Register Interface) Ssqosid (Quality-of-Service Identifiers)
Interrupt	AIA (Advanced Interrupt Architecture) v1.0
Debug	RISC-V Debug Specification v1.0.
Trace	RISC-V Nexus Trace v1.0
RAS	RERI (RAS Error Record Interface)



Technical Highlights

Vector Processing

The dual 256-bit datapath vector engine (VLEN=256-bit), with full compliance to the RISC-V Vector Extension v1.0 and Vector Crypto Extension, covers a comprehensive set of data types from integer and floating-point to AI-oriented micro-scaling formats.

The wide vector engine significantly boosts data-parallel throughput for multimedia codec, signal processing, image processing, and scientific computing workloads. The Vector Crypto Extension provides hardware acceleration for cryptographic algorithms such as AES and SHA, enabling high-performance secure communication and storage encryption with minimal software overhead.

Key features:

- Supported floating-point element types: BF16 / FP16 / FP32 / FP64 (SEW=16/32/64)
- Supported integer element types: INT4 / INT8 / INT16 / INT32 / INT64 (SEW=8/16/32/64)
- XuanTie Vector Convert Instruction Extension: FP4 / FP8 / MX Scaling data format conversions

AI Acceleration

XuanTie C950 supports optional integration of the XuanTie TPE (Tensor Processing Engine), an AI co-processor implementing the XuanTie Matrix Extension architecture. The TPE supports a broad range of data types from FP16 down to INT4/FP8 and micro-scaling formats (MXFP8 / MXFP4 / RVFP4).

XuanTie TPE delivers powerful hardware acceleration for generative AI, large language model inference, and computer vision. Its extensive low-precision and micro-scaling data type support enables flexible model deployment, significantly improving throughput and energy efficiency while preserving inference accuracy. The on-chip tensor cache and local memory, coupled with the DMA engine, minimize data movement latency and maximize compute utilization. The matrix engine and vector engine can execute in parallel, accelerating the full inference operator pipeline including Softmax, SiLU, and GELU.

Key features:

- Support up to 8 TOPS per TPE
- Supported data types: INT4 / INT8 / FP8 / MXFP8 / MXFP4 / RVFP4 / FP16 / BF16

Memory Subsystem

The memory subsystem comprises a high-performance multi-level cache hierarchy with an ultra-low 4-cycle load-to-use L1 data cache latency, a private per-core L2 cache supporting large capacity configurations, and an MMU with multiple RISC-V virtual memory modes and two-stage address translation.

The low-latency L1 cache and large private L2 cache ensure sustained data supply at high core frequencies, minimizing pipeline stalls. Multiple virtual memory modes provide flexible support for OS and virtualization scenarios, and two-stage address translation natively serves hypervisor-based virtualization requirements.

Key features:

- L1 data cache: 64KB, 4-way set-associative, 4-cycle load-to-use latency
- L2 cache: Private per core, configurable up to 3MB
- MMU: Sv39 / Sv48 / Sv57 virtual memory addressing modes, two-stage address translation

Bus Interface

The flexible bus interconnect architecture supports both Direct Connect mode and Multi-Processor mode to accommodate diverse system integration requirements.

Direct Connect mode delivers maximum bandwidth and minimum latency through a point-to-point link, ideal for high-performance single-chip designs. Multi-Processor mode leverages the XL-300 interconnect to form clusters of up to 8 cores, meeting the demands of multi-core parallel computing in server, networking, and other high-throughput scenarios. Dual AXI/ACE protocol support further extends system integration flexibility.

Key features:

- Direct Connect mode: AMBA CHI.E / CHI.F protocol, 256-bit data width
- Multi-Processor mode: XL-300 interconnect, up to 8 cores per cluster, AMBA AXI4.0 or AMBA ACE4.0 bus ports

Interrupt Subsystem

The advanced interrupt architecture is compliant with the RISC-V AIA v1.0 standard. Each core embeds an IMSIC (Incoming MSI Controller) with support for multiple supervisor domains and physical interrupt files.

The AIA-compliant interrupt architecture provides efficient, scalable interrupt management for complex SoC designs with native MSI (Message Signaled Interrupt) support, reducing interrupt dispatch latency. The multi-domain design enables interrupt isolation for TEE (Trusted Execution Environment) and virtualization use cases.

Key features:

- RISC-V AIA v1.0 compliant
- Per-core IMSIC (Incoming MSI Controller)
- 2 supervisor domains, each with 5 physical interrupt files

Trace and Debug

Comprehensive debug and trace capabilities are provided in accordance with RISC-V standard debug and trace specifications, with multiple trigger types and an efficient trace data storage scheme.

Standardized debug interfaces and rich trigger types enable developers to quickly diagnose hardware and software issues, reducing development turnaround time. The Trace Data Scatter Memory Access scheme writes trace data directly to system memory without dedicated trace buffers, lowering hardware cost while supporting extended tracing sessions. Compatibility with mainstream commercial debug tools ensures seamless integration into existing development environments.

Key features:

- Debug: RISC-V Debug Specification v1.0
- Trace: RISC-V Nexus Trace v1.0, Trace Data Scatter Memory Access
- Supported tools: XuanTie debug tools and Trace32

Processor Configuration Options

Core-level configuration options:

Feature	Options
Direct Connect Mode	Yes/No
L2 Cache	256KB/512KB/1024KB/2048KB/3072KB
Core Memory ECC/Parity	Yes/No
Vector Unit	Yes/No
Vector Crypto	Yes/No
Coprocessor Interface	Yes/No
Tensor Processing Engine	Yes/No
Core Master Port	CHI.E/CHI.F

Interconnect (via XL-300) configuration options:

Feature	Options
Core Number	1/2/3/4/5/6/7/8
Snoop Filter Size	512KB/1MB/2MB/4MB/8MB/16MB
L3 Cache	0MB/1MB/2MB/3MB/4MB/8MB
Master Interface Protocol	AXI4.0/ACE4.0
Master Interface Data Width	128-bit/256-bit
Master Interface Number	1/2/3/4
LLP Interface Number	0/1/2/3
DCP Interface Number	0/1/2
DCP Interface Data Width	128-bit/256-bit
L3 Cache/Snoop Filter ECC	Yes/No

Software Ecosystems

- Compiler, assembler, linker, debugger and binary tools are contributed to GNU/LLVM and supported officially
- QEMU is contributed and supported officially
- Optimized runtime library for enhanced performance
- Integrated Development Environment (CDS)
- Multi-OS support