

SiFive Intelligence X280

The **SiFive® Intelligence™** X280 processor is a 64-bit RISC-V vector processor offering best-in-class high performance and efficiency targeting AI and ML workloads. The highly configurable multi-core, multi-cluster capable design has been optimized for the broadest range of applications requiring high-throughput, single-thread performance while under the tightest power and area constraints. With a specific emphasis on AI/ML compute, the X280 features the powerful combination of a RISC-V Vector 512-bit vector length along with the SiFive Intelligence Extensions, tightly integrated with an 8-stage dual-issue in-order scalar pipeline and decoupled vector ALU pipeline.

Building on the robust foundations of the comprehensive **SiFive® Essential™** product portfolio, the X280 enhances the high-end capability with a feature set designed to offer the utmost in flexibility to a designer, targeting applications requiring the most demanding vector computational capabilities.

The finely tuned combination of out-of-the-box features, high level of configurability and design scalability ensures designers can achieve the optimal balance of power, performance and area while achieving the fastest time to market.

RISC-V Vectors (RVV)

The RISC-V Vector (V) ISA standard extension enables processor cores based on the RISC-V instruction set architecture to process data arrays alongside traditional scalar operations, unifying vector and scalar capabilities into a single application processor. The X280 processor implements a 512-bit vector length architecture (VLEN), fully supporting the vector extension standard, with dynamic variable vector length operations. The vector ALU and load/store architecture data width (DLEN) is 256-bits.

RISC-V vectors offer:

- A single vector ISA (ratified at version 1.0 by RISC-V International in 2021), which greatly simplifies software development across the full range of vector processors
- A vector-length-agnostic architecture, which allows library and application code investment to be reused across multiple generations and broad ranges of processor implementations
- Dynamic (runtime) modification of vector parameters for the most efficient computation on a processor, enabling better targeting for market application computation needs
- Support for vector length multiplier (LMUL), the ability to concatenate multiple vector hardware computations for a single instruction, giving more efficient vector throughput with a smaller number of software instructions. X280 supports LMUL up to 8, thus a 4096-bit software vector length.
- Extensive range of vector data types and sizes, including BF16 / FP16 / FP32 / FP64, integer INT8 up to INT64 data types, and Q8.8 to Q15 fixed point, supporting a wide range of application requirements

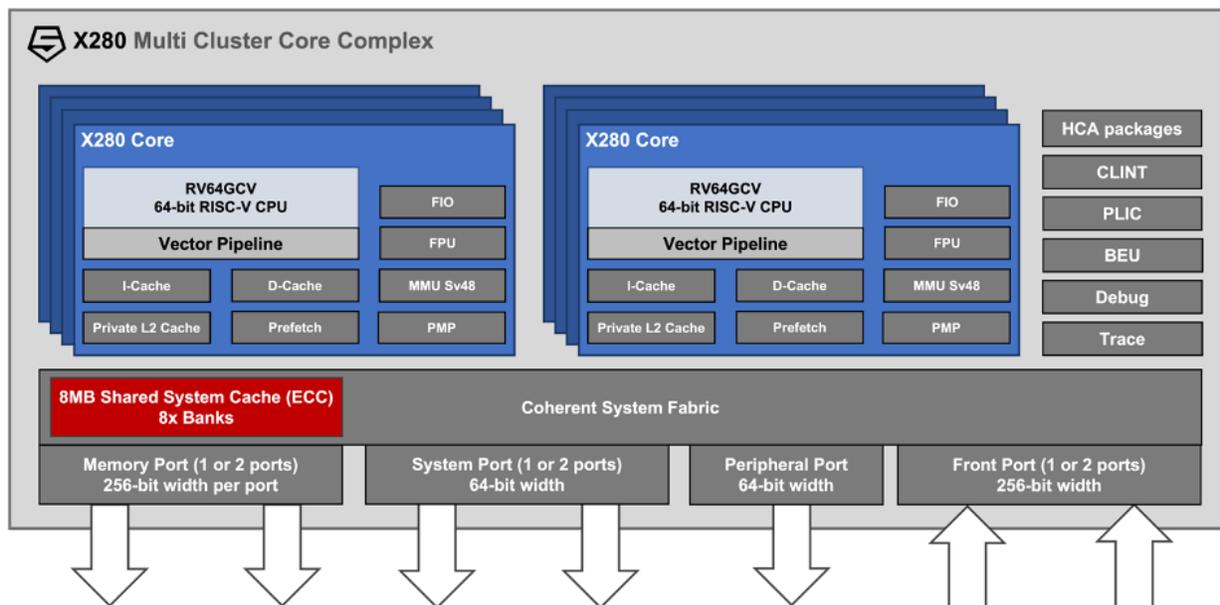
SiFive Intelligence Extensions

The SiFive Intelligence Extensions operations augment and extend the standard RISC-V Vector ISA and are specifically tuned for the acceleration of machine learning operations, extending the capability of the RISC-V Vector ISA Extension to greatly improve performance for specific AI/ML workloads.

The SiFive Intelligence Extensions enable the X280 processor to achieve a best-in-class 2.25 TFLOPs (bfloat16 MatMul) or 4.6 TOPS (INT8 MatMul)¹, supporting the broadest range of ML workloads and AI computation needs. The SiFive Intelligence acceleration gives a 5x improvement over that achievable using the out of the box RISC-V Vector ISA.²

Multi-Core, Multi-Cluster

The X280 processor can be configured up to 4 cores in a coherent multi-core cluster, and up to 2 coherent clusters, giving a maximum of 8-cores. In the multi-core, multi-cluster configuration, cache coherency is managed in the Shared System Cache (SSC) with a dedicated 1MB memory bank of up to 8MB. The Coherent System Fabric manages both coherency and a crossbar network to allow all cores full connectivity to the shared ports. A multi-cluster design may be configured to add a second port to enable more bandwidth to the multiple clusters.



¹ Both with a 2-cluster, quad core per cluster configuration, at 2.3GHz

² Measured using MobileNet v1 Inference with a batch size set to 1

SiFive Intelligence X280 Key Features

- 64-bit RISC-V ISA
- 8-stage dual-issue superscalar in-order pipeline for scalar computation
- SiFive Intelligence Extensions, which are custom instructions that accelerate AI/ML performance critical operations
- Multi-core, multi-cluster processor configuration options, with up to 8 cores
- Loosely coupled Vector Computation Pipeline, ALU implementing RISC-V Vectors extension specification 1.0
- INT8, INT16 & INT32, FP16, FP32 & FP64, and Q8.8 to Q15 fixed point data-types
 - Vector FP64 can be made optional for area and power-constrained markets
- 512-bit vector register length (VLEN)
 - Variable length operations, up to 512-bits of data per cycle, offering the Ideal balance of control logic and data parallel compute
- 256-bit Vector ALU and Load/Store architecture
- High performance vector memory subsystem
- Vector data stride L2 prefetcher unit
- Decoupled scalar and vector pipelines for optimum parallel execution of scalar and vector computation
- Memory parallelism provides cache miss tolerance
- Multi-layer caching support for optimum data movement
- Virtual memory support, with up to 48-bit addressing, with precise exceptions
- High performance, flexible connectivity to SoC peripherals

Memory System and Caches

The X280 memory system has been designed for scalability and flexibility, allowing the most suitable level of tuning for application workloads. A 32KB Level 1 Instruction Cache, a similarly sized Data Cache, both with 4-way cache associativity, alongside a private 256KB Level 2 cache offers the best selection of high performance while minimizing power and area. In multi-core systems a Level 3 cache can be configured to be either 1MB, 2MB or 4MB, with a multi-cluster option of 8MB.

For maximum flexibility in moving data in and out of the processor for general purpose I/O or other more demanding sub-system components, for example a Neural Network (NN) accelerator, there are several ports available within the bus matrix.

X280 PORTS

These ports provide connection interfaces to external system memories and peripherals; they are shared across all processors in a multi-core, multi-cluster configuration.

Memory Port 1 or 2 ports	Arm® AMBA® AXI4™ 256-bit	<ul style="list-style-type: none"> Interface with memory that offers the highest performance The only cacheable region of memory support accesses for data and instructions Supports up to 128 outstanding transfers per memory port
System Port 1 or 2 ports	AXI4 64-bit	<ul style="list-style-type: none"> Used typically for high-bandwidth, uncached memory or devices
Peripheral Port 1 port	AXI4 64-bit	<ul style="list-style-type: none"> Interface with lower speed peripherals Supports code execution supports the RISC-V standard Atomic (A) extension
Front Port 1 or 2 ports	AXI4 256-bit	<ul style="list-style-type: none"> External Initiators for accessing on Core Complex devices and ports Transactions through the Front Port are coherent with L1 Data Caches

X280 Local Ports

X280 local ports are dedicated per-core connection interfaces, running on the same clock as the core. They support single cycle pipelined data throughput.

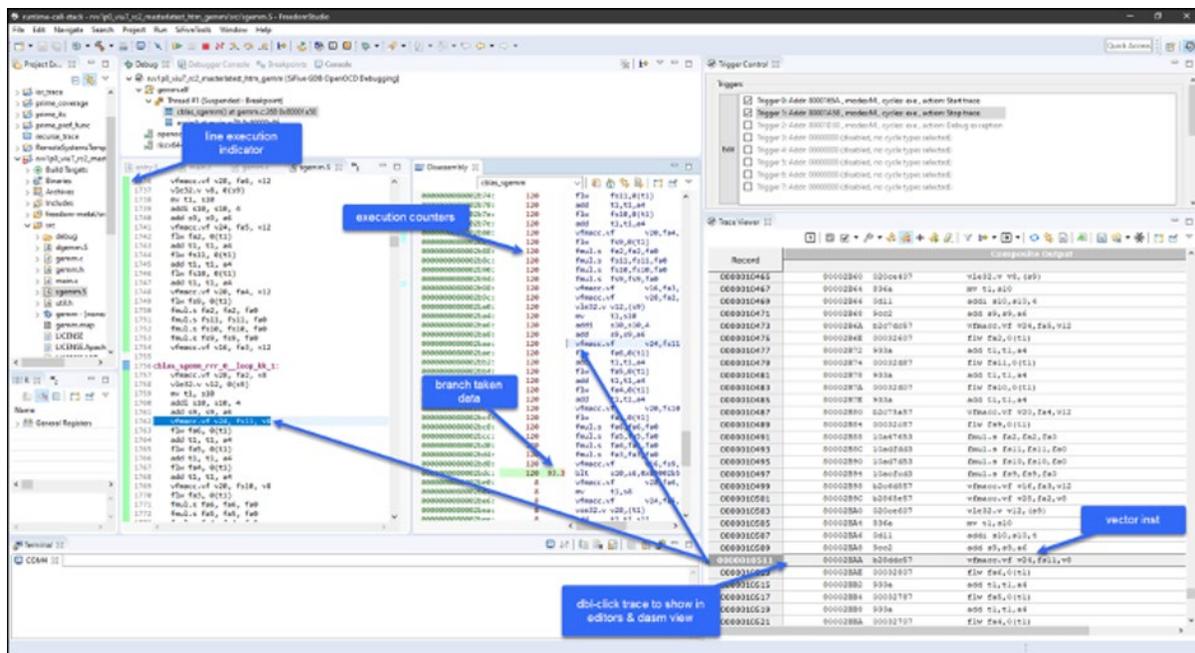
Data Local SRAM (DLS)	SRAM I/f 64b, 128b	<ul style="list-style-type: none"> Closely coupled memory, up to 8MB size, with ECC support Supports Data and Instruction accesses Option for 1 to 64 Memory Banks, offering increased bandwidth
Core Local Port (CLP)	AXI4, AHB-Lite 64b, 128b	<ul style="list-style-type: none"> Low latency direction connection to external peripherals / accelerators Supports code execution and RISC-V standard Atomic (A) extension
Core Local Front Port (CLFP)	AXI4, AHB-Lite 64b, 128b	<ul style="list-style-type: none"> External Initiators can access single processor DLS and CLP through CLFP Interface with high-speed data generator/consumer peripherals Accesses limited to the DLS and CLP of the same single core

Advanced Software Development Capabilities

SiFive Freedom Studio, built on top of the popular Eclipse IDE, is the fastest way to get started programming with your SiFive hardware. Freedom Studio is packaged with a pre-built tool suite, example projects, and includes comprehensive support for SiFive Insight Advanced Trace and Debug capabilities.

For SiFive Vector accelerated processors there is an Advanced LLVM C-Compiler, which performs auto-vectorization of C-code. This enables software developers to map their C-algorithms onto the SiFive Vector processors quickly and efficiently. The compiler is also capable of mapping code onto the SiFive Intelligence Extensions to take advantage of the higher AI performance of the X280.

Multiple execution targets are supported by Freedom Studio, including simulation models and a variety of FPGA platforms, in addition to fully featured silicon-based development boards. Freedom Studio is supported on Windows, macOS, and Linux host computers.



Comprehensive Linux Support

SiFive includes and supports Linux FPGA bitstreams for all the SiFive Intelligence, SiFive® Performance™, and SiFive Essential Application processors. The Linux FPGA platform, based on the Xilinx VCU118, provides a fast, hassle-free way to experience a complete Linux environment.

It takes just three simple steps to get started:

1. Download and flash the Xilinx VCU118 FPGA with the supplied bitstream for your chosen SiFive processor
2. Download and flash the SiFive Linux BSP onto an SD card
3. Boot Linux

TensorFlow Lite Support

TensorFlow Lite is an open-source deep learning framework that allows low-latency inference of on-device machine learning models with an optimized binary size and fast performance supported by scalar or RISC-V Vector hardware acceleration. The TensorFlow Lite software is a framework that enables the use of inference AI/ML models on the X280 for AI edge applications. The operators used within the TensorFlow Lite framework have been optimized for the most efficient computation on X280, giving the highest performance across Neural Network (NN) models ported onto X280. Multiple NN models have been ported onto X280 with TensorFlow Lite, including those optimized for Image Classification, Object Detection, Natural Language Processing, and Segmentation.

Broad Application Coverage

In addition to ML inferencing, the X280 is ideal for the broadest range of applications that require high-throughput, single-thread performance, while performing within specific power and area constraints:

- AI/ML Inference and Training
- Enterprise Switching/Routing/Storage, Smart NICs
- Edge Analytics, Big-Data Analytics
- Image processing, Object Detection, Recognition
- Autonomous Machines
- Edge Compute
- AR/VR/MR/XR