SILICON IP

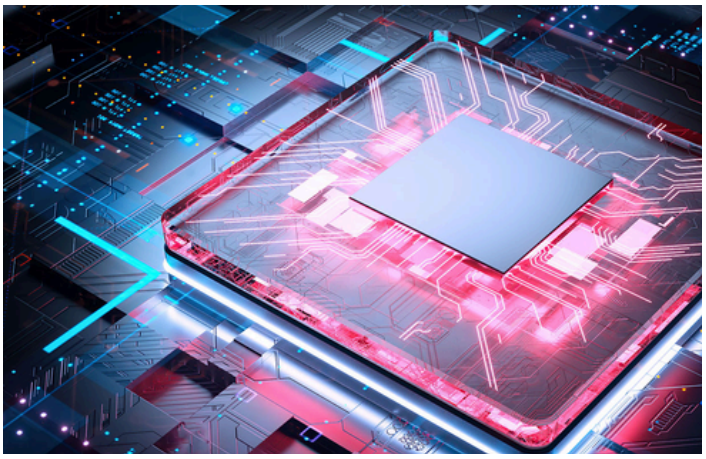
HIGH SPEED INTERFACE: CXL

High-speed interconnect standard that enables low-latency, coherent memory sharing between CPUs, accelerators, and memory devices, optimizing data-centric workloads in modern computing systems.

For sales, please contact info@xtremesilica.com

OVERVIEW

Compute Express Link (CXL) is a high-speed interconnect standard developed to enhance data center performance and efficiency by providing a unified and flexible connection between CPUs, GPUs, memory, and other accelerators. Introduced in 2019 by the CXL Consortium, CXL is designed to address the growing demands for high-bandwidth, low-latency communication in modern computing environments. It integrates features from PCI Express and provides additional capabilities such as cache coherence, which allows different types of processors to share and access data efficiently. CXL aims to improve scalability and performance in data centers, cloud computing, and high-performance computing (HPC) by enabling seamless and scalable memory and resource sharing across a variety of devices.



KEY FEATURES

High-Speed Interconnect

- CXL leverages the PCI Express (PCIe) infrastructure to provide high-speed data transfer rates, enhancing communication between CPUs, GPUs, memory, and accelerators.

Cache Coherence

- CXL supports cache coherence, allowing different processors to access and share data in memory efficiently. This reduces data duplication and ensures consistency across various computing units.

Unified Memory Space

- CXL enables the creation of a unified memory pool that can be shared across different types of processors and accelerators, improving resource utilization and flexibility.

Flexible Device Connectivity

- CXL provides a standardized interface for connecting a wide range of devices, including CPUs, GPUs, FPGAs, and memory modules, facilitating seamless integration and communication.

Low Latency

- CXL reduces data transfer latency, enhancing system responsiveness and performance, crucial for HPC and data-intensive applications.

Scalability

- CXL enables scalable architectures by allowing seamless communication between CPUs, memory, and accelerators. It helps data centers expand and adapt resources efficiently to handle growing workloads. This improves flexibility and resource utilization.

Backward Compatibility

- CXL is built on top of PCIe technology, ensuring compatibility with existing PCIe infrastructure while adding additional features for improved performance and resource management.

Advanced Resource Management

- CXL enables dynamic resource allocation and management, allowing systems to optimize performance based on current workloads and requirements.

Memory Expansion

- CXL supports memory expansion beyond traditional limitations, enabling systems to add and access more memory resources as needed.

Support for Emerging Technologies

- CXL is designed to support emerging technologies and evolving workloads, ensuring that it remains relevant and capable in the face of technological advancements.

CXL APPLICATIONS

Data Centers

- CXL enhances data transfer and resource sharing between CPUs, GPUs, and accelerators, improving performance for complex simulations, modeling, and large-scale data processing.
- Enables efficient resource allocation and scaling in cloud environments, optimizing the use of computing resources and improving overall performance and cost-effectiveness.

For sales, please contact info@xtremesilica.com

Memory Expansion and Management

- Facilitates the creation of large, shared memory pools that can be accessed by different types of processors and accelerators, improving memory utilization and flexibility.
- Supports advanced memory tiering strategies, allowing systems to efficiently manage and utilize different types of memory (e.g., DRAM, persistent memory) based on workload requirements.

Artificial Intelligence (AI) and Machine Learning (ML)

- Enhances communication between AI and ML accelerators (such as GPUs and FPGAs) and CPUs, improving data throughput and processing efficiency for training and inference tasks.
- Reduces latency and improves data sharing between various components in AI and ML systems, speeding up complex computations and data analysis.

Networking

- Supports high-speed data transfers and low-latency communication in networking equipment, improving performance and reliability in high-bandwidth network environments.

Storage Systems

- Improves the efficiency of storage systems by providing fast, coherent access to shared memory resources, optimizing data throughput and reducing latency for storage-intensive applications.

Enterprise Applications

- Enhances resource management and allocation in enterprise IT environments, enabling efficient use of computing resources for a wide range of applications and services.

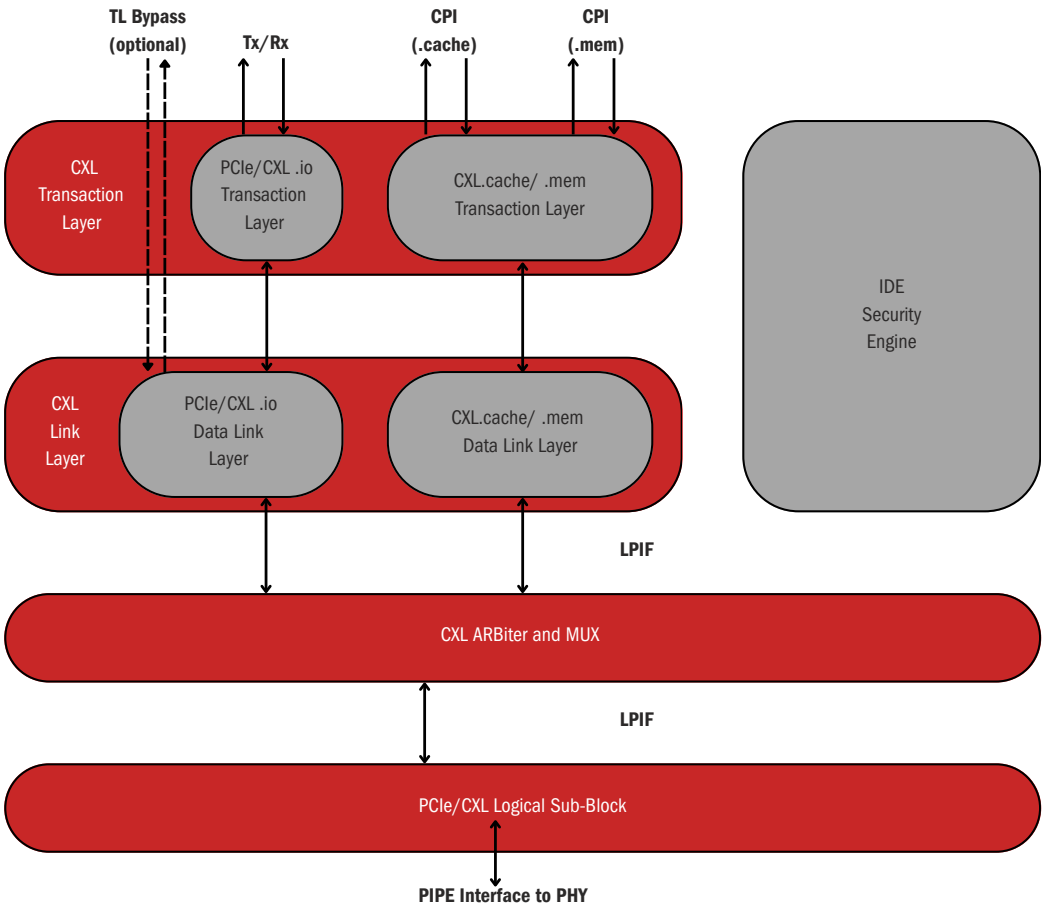
Scientific Research

- Facilitates high-speed communication and data sharing between various research components, improving the performance and scalability of scientific simulations and experiments.

Emerging Technologies

- Provides a scalable, flexible architecture that supports emerging technologies and new computing paradigms, ensuring long-term relevance with advancements.

CXL ARCHITECTURE



For sales, please contact info@xtremesilica.com

XtremeSilica Technologies Private Limited
494, 2nd Floor, CMH Road, Indiranagar,
Bengaluru, Karnataka 560038 India
www.xtremesilica.com
info@xtremesilica.com
+91 79932 79934