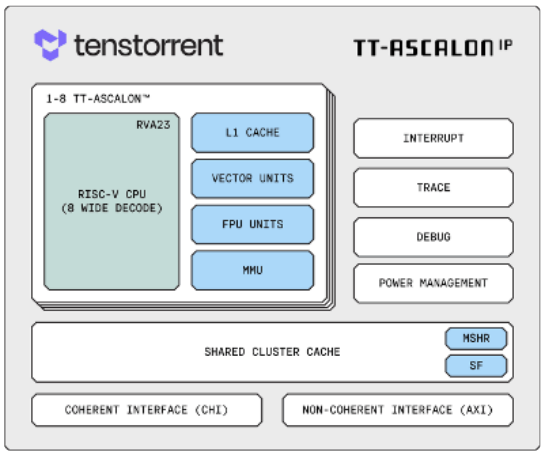
Ascalon CPU IP Cache Hierarchy Overview

Ascalon CPU IP Cache Hierarchy Overview:



Ascalon Cluster Block Diagram

The **Tenstorrent Ascalon CPU** features Shared Cluster Cache SCC memory designed for high performance and efficiency as shown in the diagram. It uses a 64 KB[a] instruction cache (I-Cache) and a 256 KB[a] data cache (D-Cache). The 3 load/store pipelines and two 256-bit RISC-V vector pipes optimize memory and data throughput for demanding tasks.

Instead of private L2 caches, Ascalon uses a 12 MB[a] shared cluster cache with 24-way associativity, allowing all cores to quickly access a common pool of data, reducing cache complexity and improving performance. There's no L3 cache, simplifying the design and keeping focus on the shared L2 cache for efficient memory access.

This design reduces latency, improves power efficiency, and simplifies cache management, making Ascalon well-suited for high-performance workloads while keeping power and area consumption low.

Tenstorrent Ascalon Cache Key Features:

- **Instruction Cache (I-Cache):**
  - 64 KB[a], 8-way set-associative
  - Fast access for instructions
- **Data Cache (D-Cache):**
  - 256 KB[a], for each core.

    - Provides quick access to data during processing
- **Shared Cluster Cache:**
    - 12 MB[a], 24-way set-associative
    - Shared across all cores in the cluster
- **No L3 Cache:**
    - Simplified memory hierarchy by eliminating L3 cache
    - Focus on maximizing the performance of the shared L2 cache

[a] Please note that the L1 and L2 cache sizes are under active optimizations as part of ongoing performance improvements. Final sizes may vary.