

???????

1. NVIDIA Jetson Orin

2. HAILO [Hailo-8](#) 15 [domain-specific-dataflow-processing](#)

- 5W 10 token TPS Llama2-7B Stable Diffusion 2.1 Hailo-10 Hailo-10 40 Hailo-10 5 Hailo-10 2 Core Ultra NPU Hailo-10 NPU Intel

3. [1684X](#) 17.6T INT8 LPDDR4x 68.3GB/s 16GB 17W

4. [https://d-robotics.cc](#)

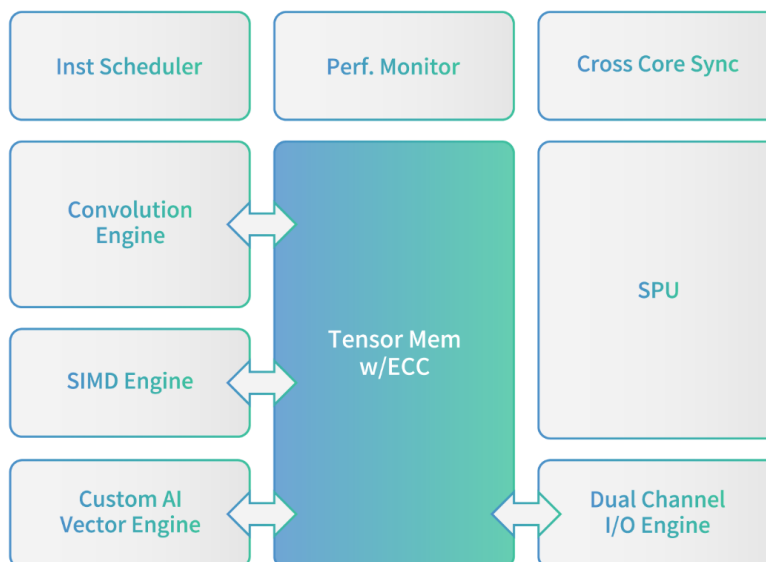
5. [https://www.listenai.com/products/chips/csk6](#)

6. AMD Versal (SoC) AI

Versal AI Edge Prime

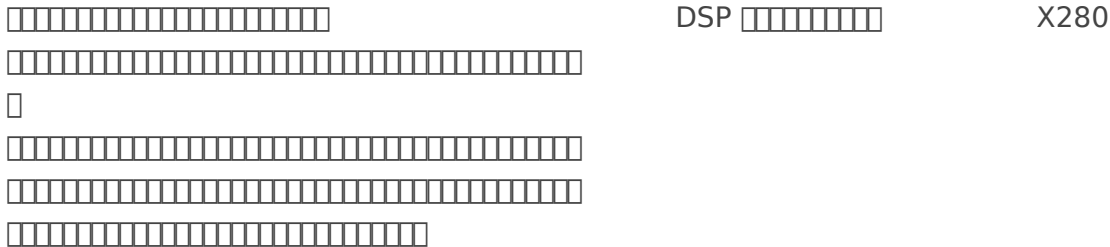
8. [Sophgo SG2380](#)

9. [https://www.novauto.com.cn/](#)



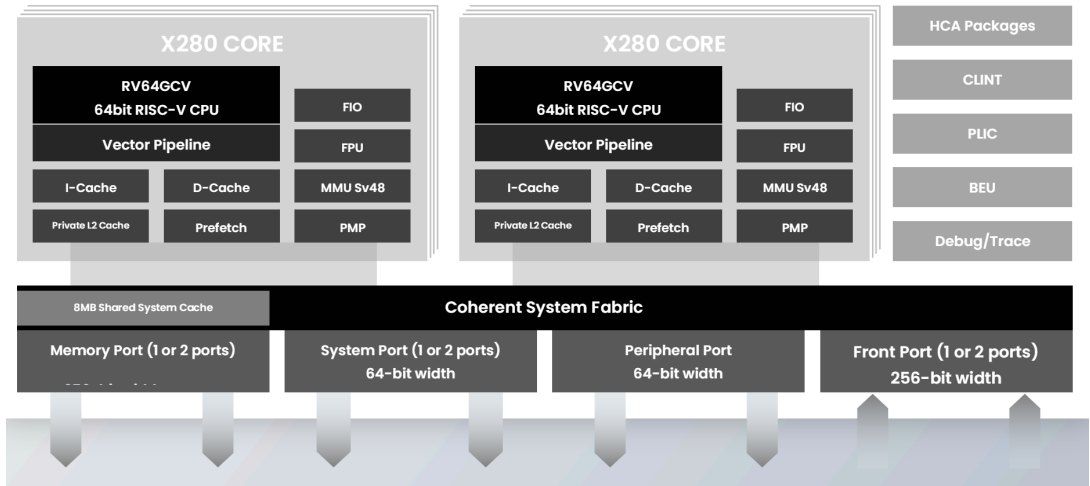
10. [sifive-intelligence-x280](#)

- the Vector Coprocessor Interface Extension (VCIX) RISC-V Vector ISA SiFive Intelligence Extensions



- RVV 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.

### X280 Series Multi-Cluster Core Complex



### 11. AI DeepEdge10 NNP400T



DeepEdge10Max

一款面向边缘大模型推理的高性能SOC芯片

采用D2D的chiplet片上互联技术，实现单芯片算力的灵活可扩展，满足人工智能大模型边缘推理对算力的需求；可广泛用于边缘CV大模型、NLP大模型的私有化部署场景。

通用处理器内核：

- 自研神经网络处理器 NNP400T；
- 提供48 TOPS (INT8)/24 TOPS (INT16)/8 TFLOPS (FP16) 的算力；
- 支持常用深度学习网络，如：CNN, RNN, Transformer, GNN 等；

[下载更多参数 +](#)

DeepEye1000 DeepEdge10C DeepEdge10 DeepEdge10Max

12. Meta MTIA 256MB 1.3GHz v1 128MB
- 800GHz 8x8 (PE) PE
- MTIA v1 3.5 7
- PE Meta PE
- SRAM 3.5 LPDDR5



### First Gen MTIA

**Technology**  
TSMC 7nm

**Frequency**  
800MHz

**Instances**  
1.12B gates, 65M flops

**Area**  
19.34mm x 19.1mm, 373mm<sup>2</sup>

**Package**  
43mm x 43mm

**Voltage**  
0.67V logic, 0.75V memory

**TDP**  
25W

**Host Connection**  
8x PCIe Gen4 (16 GB/s)

**GEMM TOPS**  
102.4 TFLOPS/s (INT8)  
51.2 TFLOPS/s (FP16/BF16)

**SIMD TOPS**  
Vector core:  
3.2 TFLOPS/s (INT8),  
1.6 TFLOPS/s (FP16/BF16),  
0.8 TFLOPS/s (FP32)  
SIMD:  
3.2 TFLOPS/s (INT8/FP16/BF16),  
1.6 TFLOPS/s (FP32)

**Memory Capacity**  
Local memory: 128 KB per PE  
On-chip memory: 128 MB  
Off-chip LPDDR5: 64 GB

**Memory Bandwidth**  
Local memory: 400 GB/s per PE  
On-chip memory: 800 GB/s  
Off-chip LPDDR5: 176 GB/s

### Next Gen MTIA

**Technology**  
TSMC 5nm

**Frequency**  
1.35GHz

**Instances**  
2.35B gates, 103M flops

**Area**  
25.6mm x 16.4mm, 421mm<sup>2</sup>

**Package**  
50mm x 40mm

**Voltage**  
0.85V

**TDP**  
90W

**Host Connection**  
8x PCIe Gen5 (32 GB/s)

**GEMM TOPS**  
708 TFLOPS/s (INT8) (sparsity)  
354 TFLOPS/s (INT8)  
354 TFLOPS/s (FP16/BF16) (sparsity)  
177 TFLOPS/s (FP16/BF16)

**SIMD TOPS**  
Vector core:  
11.06 TFLOPS/s (INT8),  
5.53 TFLOPS/s (FP16/BF16),  
2.76 TFLOPS/s (FP32)  
SIMD:  
5.53 TFLOPS/s (INT8/FP16/BF16),  
2.76 TFLOPS/s (FP32)

**Memory Capacity**  
Local memory: 384 KB per PE  
On-chip memory: 256 MB  
Off-chip LPDDR5: 128 GB

**Memory Bandwidth**  
Local memory: 1 TB/s per PE  
On-chip memory: 2.7 TB/s  
Off-chip LPDDR5: 204.8 GB/s

Revision #1

Created 11 January 2025 09:46:28 by Colin

Updated 12 January 2025 06:39:52 by Colin