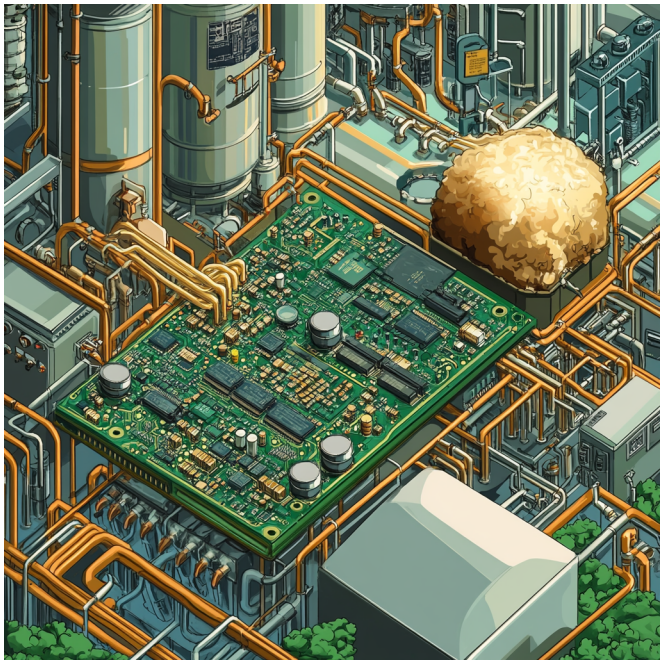
HXP HACK-4 AI1337 Device Specification

The HXP HACK-4 AI1337 is designed to fulfil the compute needs of the AI industry. The design is a significant extension to the existing X86 architecture to enable fast scratch operations.



Real picture of first sample (not vaporware)

High-Level Architecture

This section provides a high-level overview of the HXP HACK-4 and AI1337 architecture.

Processor Organization

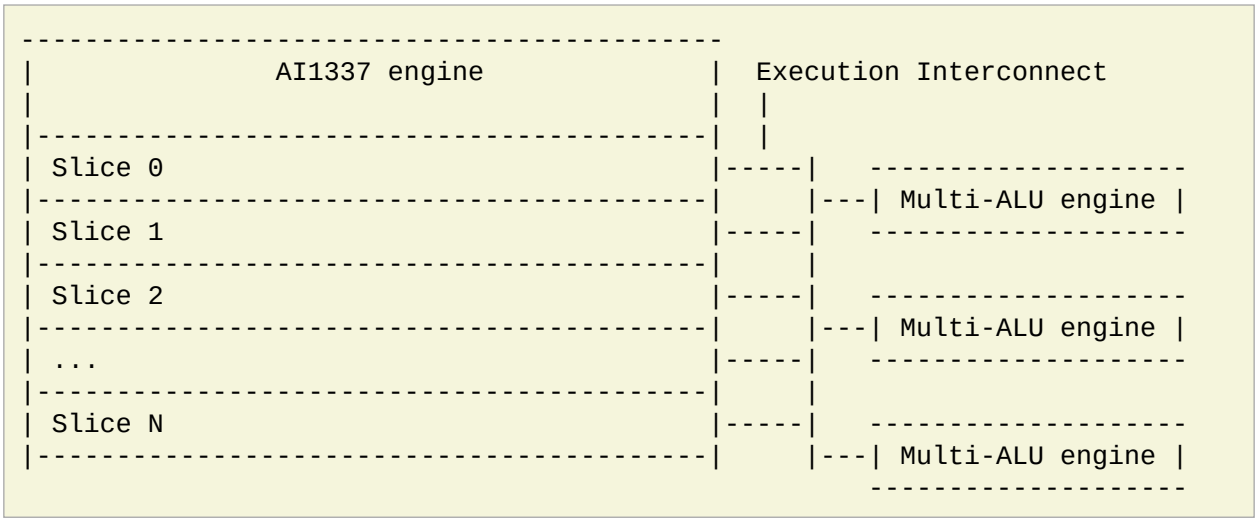
```
HXP HACK-4 application processor, optimized for scalar compute
AI1337 Engine, optimized for very-wide compute

            |------------|---------|-----------|
            |            |---------|           |
            | HXP HACK-4 |---------| AI1337 IP |
            |            |---------|           |
            |------------|---------|-----------|
                                 |
                                 |
                   PSCHORR interconnect

PSCHORR very-wide link
```

The HXP HACK-4 is the application processor responsible for boot and executing OS software. The AI1337 execution engine is on-die engine responsible for fast scratch operations.

AI1337 Engine Organization

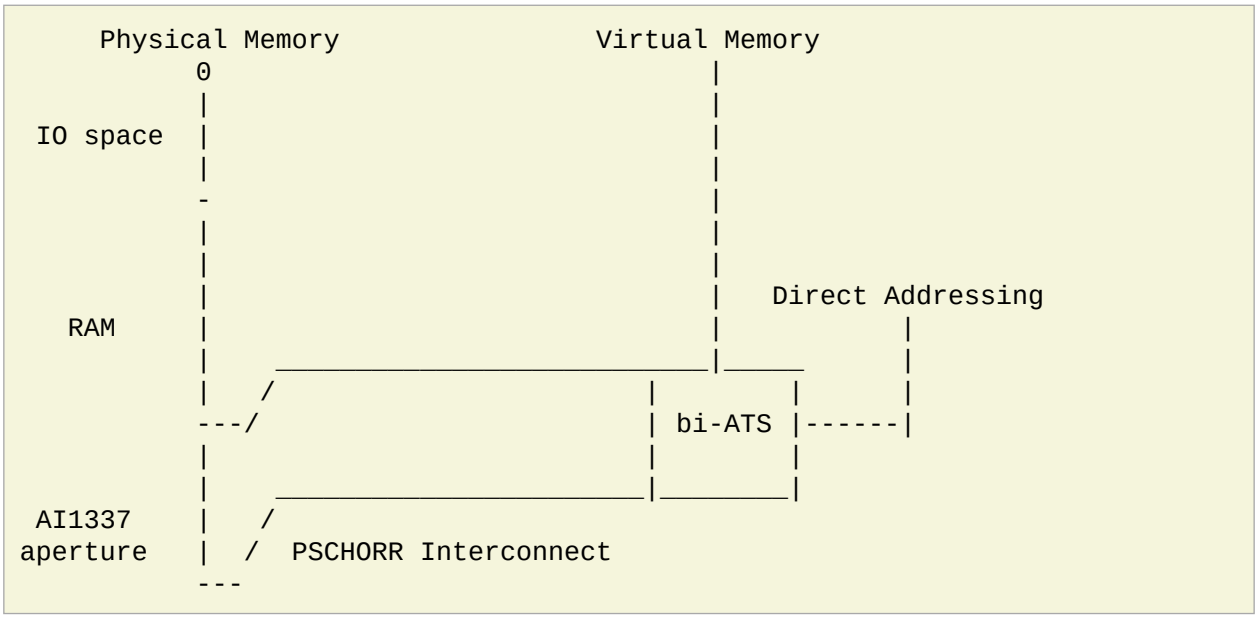


The AI1337 engine is organized as a vector of interconnected memory slices. Slices are interconnected via the 'execution interconnect' in an N-to-N fashion, and each cross-slice wide-link is connected to a series of multi-ALU engines that support fast addition, subtraction and multiplication.

PSCHORR Interconnect

The PSCHORR Interconnect connects the HACK-4 application processor and the AI1337 Engine using a multi-link organization for fast slice reads and writes.

The interconnect allows also for addressability of the scratch memory through an bi-ATS unit that supports bi-directional addressing of scratch and application processor memory.



ISA Contributions

This section describes the ISA contributions to the X86_64 ISA. The added instructions are responsible for updating scratch memory on the AI1337 engine and for submitting work to the AI1337 engine. The ISA also includes instructions for fast reconfiguration of the PSCHORR interconnect.

ISA

Opcode	Instruction	Description
0F 0A 83	MTS	Load RCX bytes from memory address (RSI) to slice (RBX) at slice offset (RDI)
0F 0A 84	STM	Read RCX bytes from slice (RBX) at slice offset (RDI) and write memory address (RSI)
0F 0A 85	FSCR	Clear all slices
0F 0A 86	SCRADD	Add the slices pointed by RDI and RSI, and store the result into slice pointed by RDX
0F 0A 87	SCRSUB	Subtract the slices pointed by RDI and RSI, and store the result into slice pointed by RDX
0F 0A 88	SCRMUL	Multiply the slices pointed by RDI and RSI, and store the result into slice pointed by RDX
0F 0A 89	SCRHLW (privileged)	Update scratch memory PSCHORR bi-ATS base VA
0F 0A 8A	SCRHLR	Read scratch memory PSCHORR bi-ATS base VA

System-Level Contributions

This section provides information on system-level specification and configuration, and it's primarily targeted towards kernel developers.

Specification

The AI1337 engine support is dictated by the existence of the 0x80000022 CPUID leaf. If the AI1337 CPUID leaf exists, the EAX, ECX, EDX and EBX registers provide the following information:

CPUID 0x80000022

Register	Bits	Information
EAX	0-31	Total scratch memory size
ECX	0-9	Maximum number of slices
ECX	10-31	Maximum slice size in bytes
EDX	0-31	Low 32 bits of the AI1337 Aperture
EBX	0-31	High 32 bits of the AI1337 Aperture

Configuration

The AI1337 engine is a multi-configurable engine that software can utilize for scaling up for high-computing workloads and scaling down for power-efficiency.

MSR

MSR	Identifier	Description
MSR_HACK4_SLICE_SIZE	0xC0000105	Read/Write slice size in the AI1337 engine

MSR_HACK4_NUM_SLICES	0xC0000106	Read/Write count of slices in the AI1337 engine