

# Meteor Lake



## Componentes de una CPU Moderna

Enfoque en la **NPU** (Unidad de Procesamiento Neuronal)

**New** Intel Arc graphics

**New** Media & Display standards

**New** Low power island E-cores

**First** Built-In NPU AI engine

**First** 32 performance hybrid architecture

**New** P-core & E-core microarchitectures

**First** on Intel 4 process technology

Intel's largest client SoC architectural shift in 40 years

# Introducción a las CPUs modernas

## CPU (Unidad Central de Procesamiento)

Núcleos de propósito general para tareas tradicionales del sistema operativo y aplicaciones.

## GPU (Unidad de Procesamiento Gráfico)

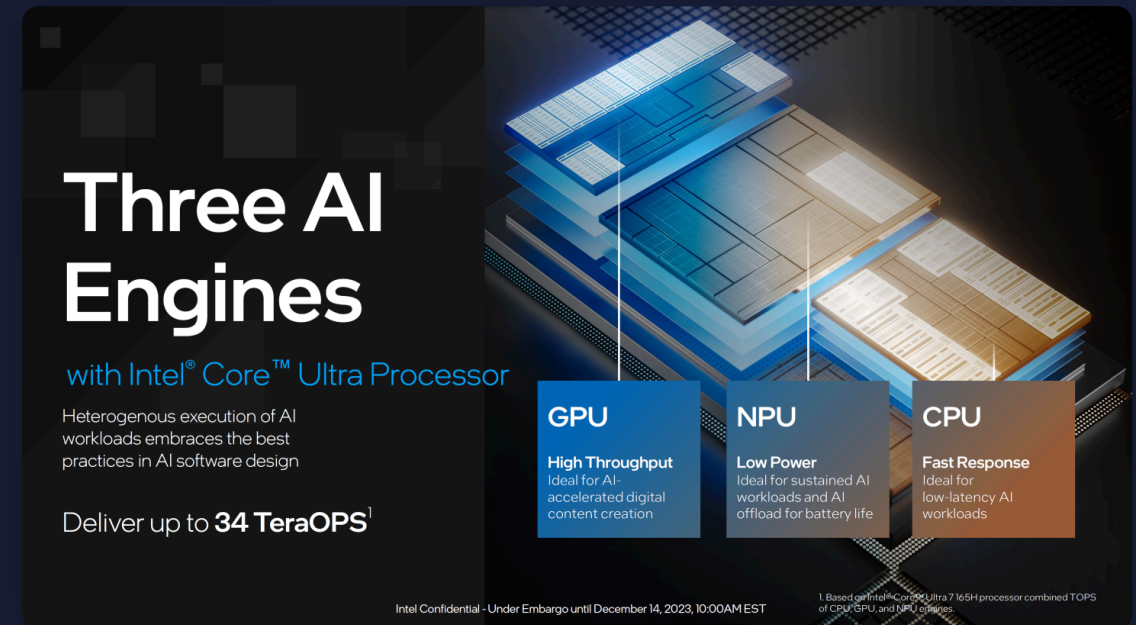
Procesamiento paralelo para gráficos, cálculos vectoriales y aceleración de IA.

## **NPU** (Unidad de Procesamiento Neuronal)

Hardware especializado para inferencia de IA con alta eficiencia energética.

### Evolución de las CPUs

- 2010s** CPUs con núcleos múltiples
- 2015+** Integración de GPU en el mismo die
- 2023+** Incorporación de **NPU** dedicada



**Three AI Engines**

with Intel® Core™ Ultra Processor

Heterogenous execution of AI workloads embraces the best practices in AI software design

Deliver up to **34 TeraOPS<sup>1</sup>**

GPU	NPU	CPU
<b>High Throughput</b> Ideal for AI-accelerated digital content creation	<b>Low Power</b> Ideal for sustained AI workloads and AI offload for battery life	<b>Fast Response</b> Ideal for low-latency AI workloads

Intel Confidential - Under Embargo until December 14, 2023, 10:00AM EST

<sup>1</sup> Based on Intel® Core™ Ultra 7 165H processor combined TOPS of CPU, GPU, and NPU engines

# Arquitectura de Intel Meteor Lake

## CPU Tile

- Fabricado por Intel en proceso **Intel 4**
- Núcleos P-Cores Redwood Cove
- Núcleos E-Cores Crestmont

## GPU Tile

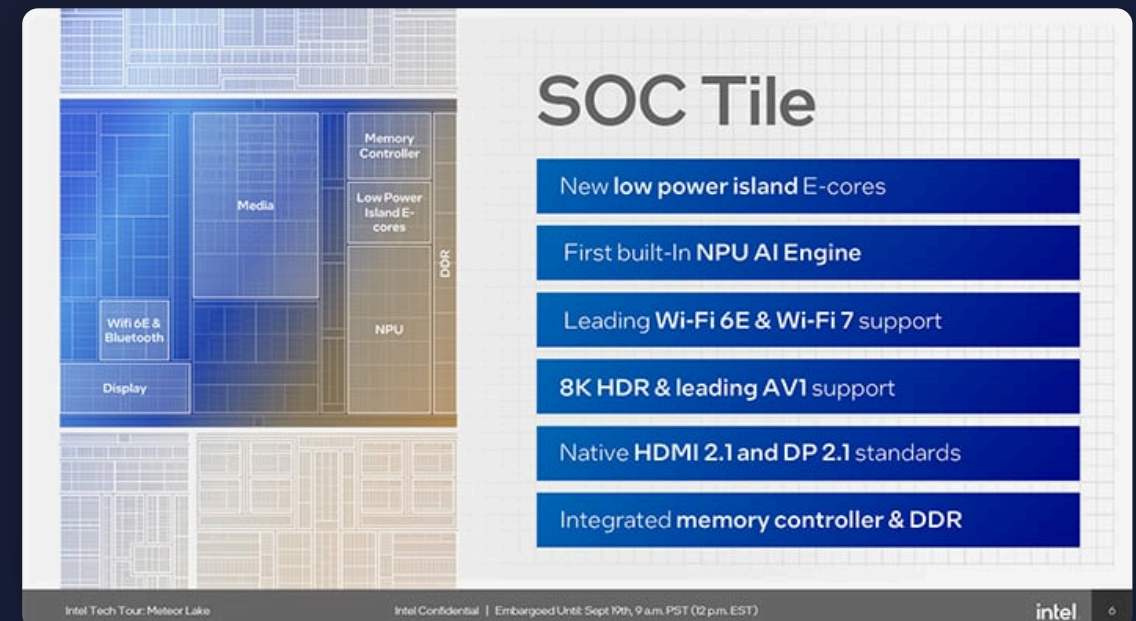
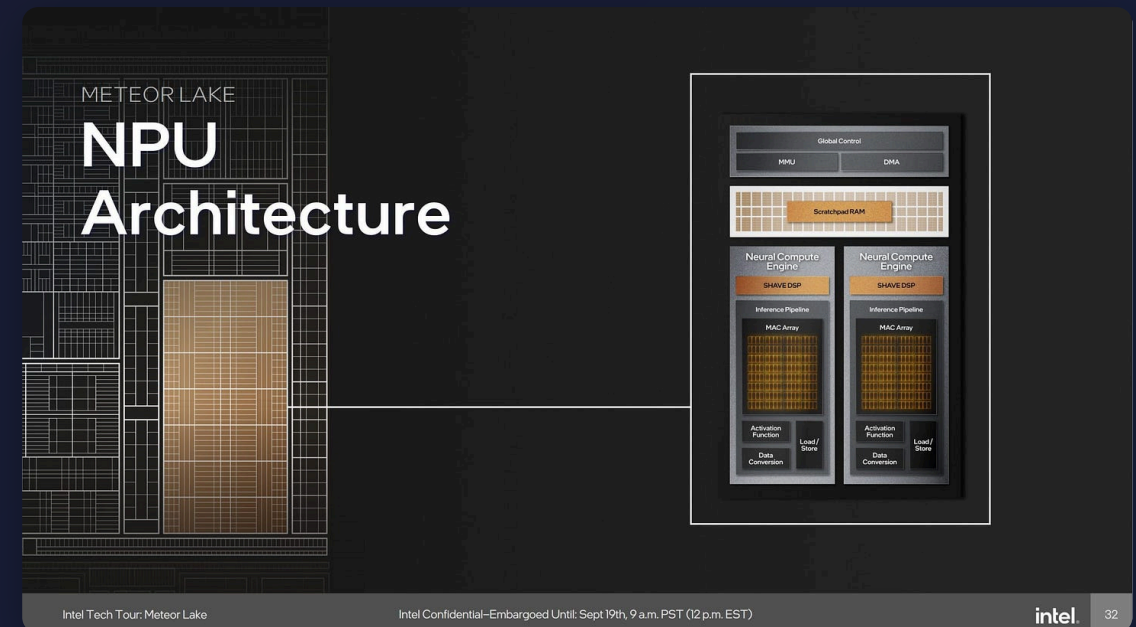
- Fabricado por TSMC en proceso **N5 (5nm)**
- Arquitectura Xe-LP
- Rendimiento duplicado vs generación anterior

## **NPU** en SoC Tile

- Fabricado por TSMC en proceso N6 (6nm)
- **NPU 3720** para cargas de IA
- 9.5 TOPS de rendimiento a 1.16 GHz
- Optimizada para eficiencia energética

### Características Clave

- ✓ Arquitectura de tiles desagregados
- ✓ Tres motores de IA: CPU, GPU y NPU
- ✓ Empaquetado 3D Foveros



# NPU de Intel: NPU 3720

## Arquitectura

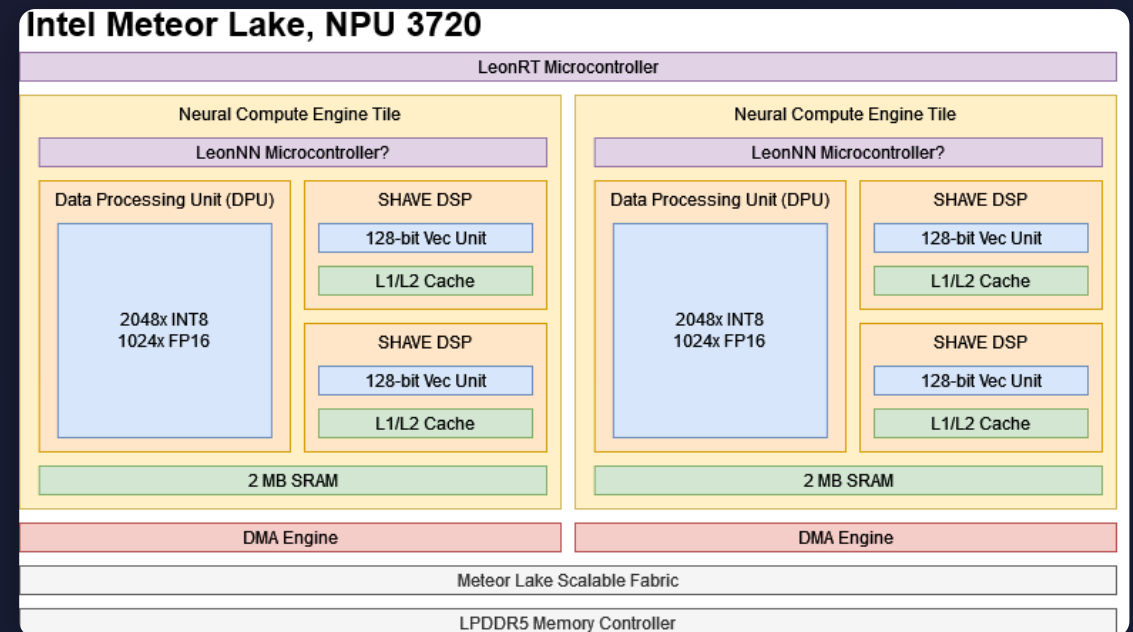
- Basada en diseños **Movidius** (adquirida en 2016)
- Matriz MAC dividida en dos tiles **Neural Compute Engine**
- Microcontroladores **LEON** de 32 bits (SPARC)
- DSPs **SHAVE** para operaciones especiales

## Rendimiento

- ⚡ **9.5 TOPS** a 1.16 GHz
- 📊 4096 operaciones INT8 MAC por ciclo
- 📈 3-4× más rápida que CPU en tareas de IA

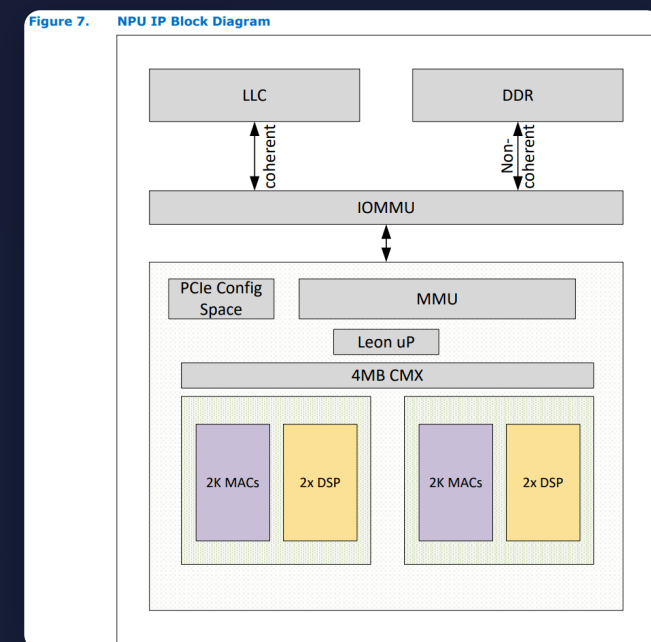
## Capacidades Destacadas

- 🌿 Alta eficiencia energética
- ⚡ Optimizada para cargas FP16
- 🔗 Integración con OpenVINO



## Especificaciones Técnicas

<b>2 MB</b>	<b>512</b>	<b>4.7</b>	<b>10 GB/s</b>
SRAM por NCE	MPEs	FP16 TFLOPS	Ancho de banda





# Arquitectura de AMD Ryzen AI

## 🧠 CPU (Zen 5)

- ✓ Hasta 12 núcleos en dos CCX
- ✓ 4 núcleos Zen 5 + 8 núcleos Zen 5c

## 🎮 GPU (RDNA 3.5)

- ✓ 8 WGP (16 Compute Units)
- ✓ 1024 stream processors

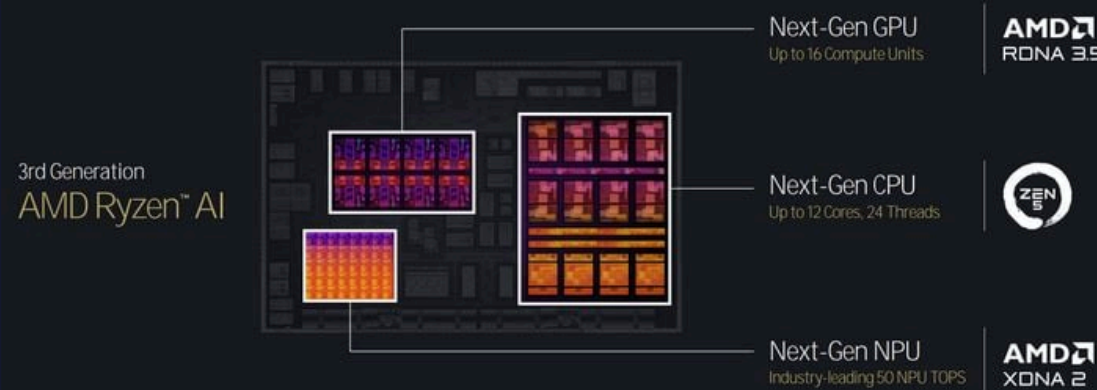
## 🧠 NPU(XDNA 2)

- ✓ Segunda generación XDNA
- ✓ **32 tiles** de motor de IA
- ✓ Memoria local de alta velocidad
- ✓ Interfaz con Infinity Fabric

### Características Clave

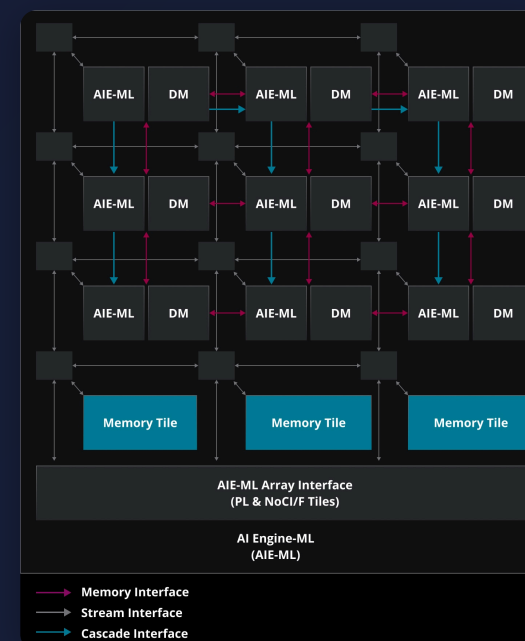
- ✓ Proceso TSMC N4P (4nm)
- ✓ Die más grande que generación anterior
- ✓ Soporte para memoria LPDDR5-7500

Providing the next level of NPU, CPU, and GPU architectures for next-gen AI PC experiences



### 📋 Especificaciones de la NPU XDNA 2

<b>50</b> TOPS de rendimiento	<b>32</b> Tiles de motor de IA
<b>2×</b> Más grande que XDNA 1	<b>4nm</b> Proceso de fabricación



# Comparación de NPUs: Intel vs AMD



## Intel NPU 3720

Arquitectura Basada en **Movidius**

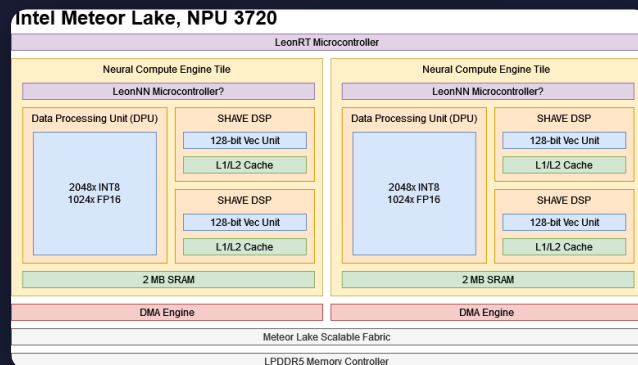
Rendimiento **9.5 TOPS** a 1.16 GHz

Unidades 2 tiles NCE con 512 MPEs

Memoria 2 MB SRAM por NCE

Ancho banda Menos de **10 GB/s**

Optimización Enfocada en **FP16**



## AMD XDNA 2

Arquitectura Segunda generación **XDNA**

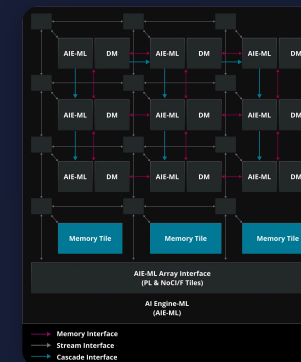
Rendimiento **50 TOPS** de rendimiento

Unidades **32 tiles** de motor de IA

Memoria Memoria local de alta velocidad

Interfaz Conexión con **Infinity Fabric**

Proceso Fabricado en **TSMC 4nm**



## ↔ Diferencias Clave



Intel: Mayor eficiencia energética



AMD: Mayor rendimiento bruto



Intel: Basada en adquisición de Movidius



AMD: Arquitectura XDNA propia