

CÓMO LA IA ESTÁ IMPULSANDO LA EVOLUCIÓN DEL DISEÑO DE WORKSTATIONS MÓVILES Y COMPACTAS

Históricamente, la gran mayoría de las workstations móviles ha dependido de una tarjeta gráfica independiente (dGPU) para proporcionar un rendimiento suficiente para el renderizado en 3D, la edición de videos y los proyectos CAD/CAM. Aunque esta ha sido una estrategia eficaz para algunos usuarios, la potencia adicional que consumen estas tarjetas ha obligado a los fabricantes de sistemas a sustituir otras características del valor para los usuarios finales, como el espesor del sistema, las temperaturas de funcionamiento más bajas y el aumento de la duración de la batería.

Las realineaciones estructurales a largo plazo en el mercado de las workstations están preparadas para hacer que estas concesiones sean aún más costosas. Los modelos de IA y las aplicaciones emergentes que dependen de ellos están planteando nuevas y exclusivas demandas a los fabricantes de workstations. La adopción generalizada de video 4K y 8K en las cabinas de edición, el uso creciente del modelado de información de construcción (BIM) en ingeniería y el cambio constante hacia el renderizado fotorrealista también han ejercido una presión ascendente sobre los requisitos de la GPU. Estas presiones son contrarias a la preferencia general de los consumidores por sistemas más ligeros, rápidos y eficientes en términos energéticos.

EL MERCADO DE WORKSTATIONS EN EXPANSIÓN

Se espera que las ventas de workstations móviles crezcan en los próximos cinco años, e IDC predice que uno de cada 10 sistemas comerciales será una workstation en 2030 “a medida que más organizaciones reconozcan su valor para las workstations esenciales”. Según Jay Chou, director de Investigación de Worldwide Client Device Trackers de IDC, “la exploración actual en torno a la IA para el trabajo debería traducirse en una expansión de los casos de uso, incluido el desarrollo de modelos, en numerosos sectores!”.

Si bien los casos de uso de las workstations pueden estar aumentando, los tamaños del sistema no necesariamente van en la misma dirección. La adopción de estaciones de trabajo móviles aumentó durante la pandemia a medida que las condiciones de trabajo cambiaban y, aunque las workstations en torre siguen siendo esenciales para el sector, las laptops y las workstations o de escritorio de formato pequeño (SFF) son cada vez más importantes.

Los clientes de workstations empresariales buscan sistemas diseñados con el mínimo espacio físico y el máximo flujo de datos. Este cambio simultáneo en el formato y el caso de uso es un reto de diseño y una oportunidad para redefinir lo que las empresas esperan de una workstation móvil y compacta.

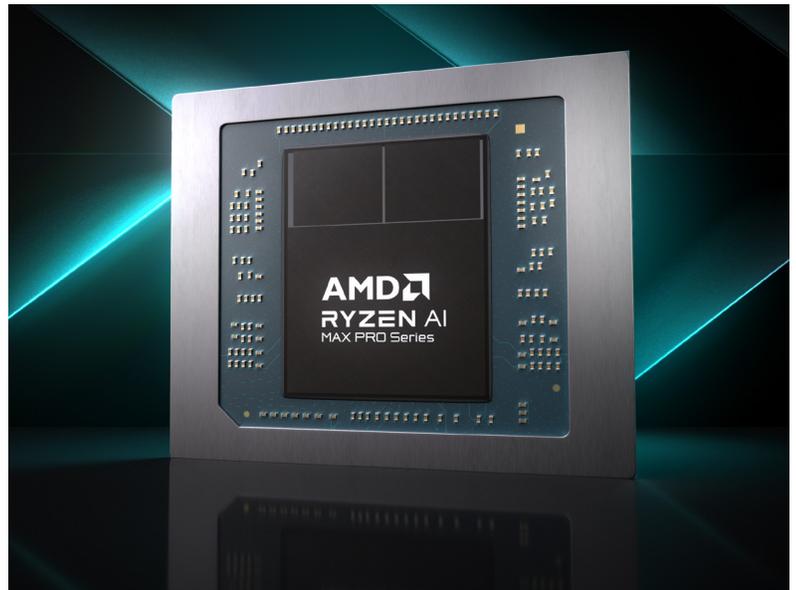
DESAFIAR LAS CONCESIONES QUE LOS CLIENTES DE LAS WORKSTATIONS SE VEN OBLIGADOS A ACEPTAR

Con demasiada frecuencia, las workstations móviles y compactas se diseñan de forma que los clientes se ven obligados a elegir entre dos opciones de producto o resultados deseables. Las workstation de escritorio son menos limitadas, gracias a sus grandes volúmenes internos y a sus potentes sistemas de refrigeración.

Esta dicotomía obliga a los usuarios de workstations móviles y compactas a comprometer las características y el rendimiento. Se trata de una visión que no está a la altura de las tendencias más importantes en torno a los factores de forma del sistema y los crecientes requisitos computacionales de la IA. Existe una nueva forma de satisfacer las demandas informáticas actuales y una mejor solución para satisfacerlas.

PRESENTAMOS AMD RYZEN™ AI MAX PRO:

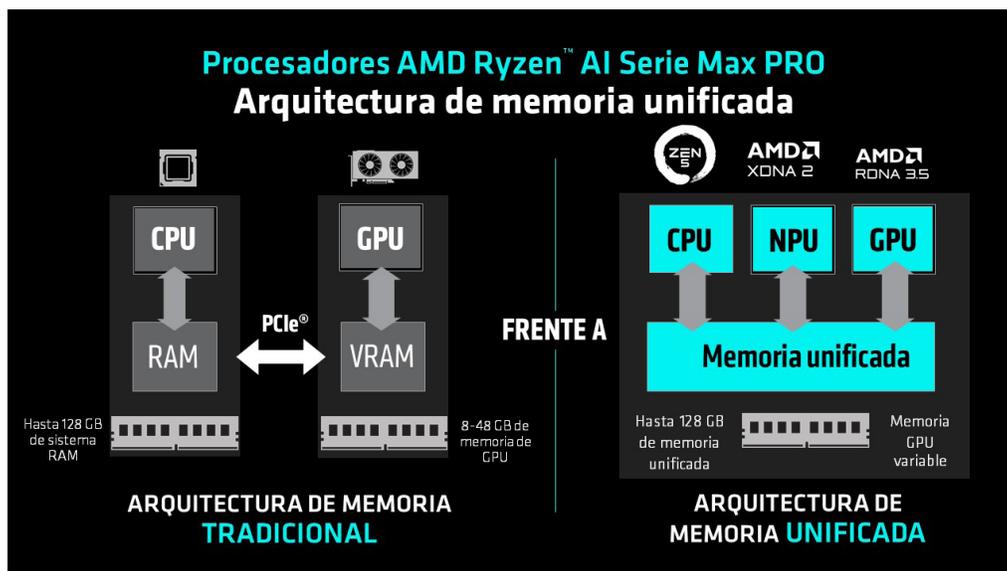
Los procesadores AMD Ryzen™ AI Serie Max PRO representan un hito significativo para los sistemas x86 y el mercado de las workstations Windows. Como los primeros procesadores x86 que combinan una GPU integrada con rendimiento de nivel independiente, núcleos de CPU de clase de escritorio y una unidad de procesamiento neuronal (NPU) en un solo chip, están diseñados para clientes que desean un rendimiento excelente, optimizaciones y certificaciones de aplicaciones profesionales y la capacidad de ejecutar cargas de trabajo de IA que de otro modo serían demasiado pesadas para la mayoría de las dGPU que se encuentran en los portátiles comerciales y los equipos de escritorio SFF que se envían comúnmente en la actualidad. Los procesadores AMD Ryzen AI Serie Max PRO se diseñaron para abordar proyectos 3D complejos con varias aplicaciones ejecutándose en paralelo o para explorar nuevas ideas mediante modelos lingüísticos locales de gran tamaño.



ELIMINACIÓN DE LOS CUELLOS DE BOTELLA DE LA MEMORIA GPU

Las tarjetas gráficas independientes y móviles modernas suelen ofrecer entre 8 y 16 GB de RAM de video (VRAM) dedicada. Aunque esto es suficiente para muchas aplicaciones de workstations, la ejecución local de cargas de trabajo con conjuntos de datos grandes y modelos de IA supone un mayor desafío. Por ejemplo, Stable Diffusion 3.5 Large puede superar incluso la capacidad de VRAM de 16 GB que normalmente está disponible en las GPU móviles de gama alta. Los clientes deben utilizar versiones menos complejas del modelo que *quepan* dentro de la memoria limitada disponible o comprar tiempo en servicios en la nube.

Los procesadores AMD Ryzen AI Serie Max PRO resuelven este problema compartiendo un grupo común de memoria entre la CPU, la GPU y la NPU, como se muestra a continuación:



Este tipo de uso compartido se conoce como arquitectura de memoria unificada (UMA), a diferencia de una arquitectura de memoria distribuida tradicional (mostrada a la izquierda), donde la CPU y la GPU tienen sus propios grupos de memoria dedicados.

Una arquitectura de memoria tradicional proporciona a la GPU una gran cantidad de ancho de banda, pero mucha menos memoria total en comparación con la CPU. También ralentiza la comunicación entre la CPU y la GPU. La conexión de estos dos componentes en el chip, como se muestra a la derecha, les permite compartir un conjunto común de memoria. Esto es ventajoso, siempre que el sistema ofrezca suficiente ancho de banda de memoria. Los procesadores AMD Ryzen AI Serie Max PRO pueden hacerlo.

Mientras que la mayoría de los procesadores de workstations móviles x86 dependen de dos canales de memoria, AMD Ryzen AI Serie Max PRO utiliza cuatro. El aumento resultante del ancho de banda total del sistema es suficiente para alimentar la GPU, la CPU y la NPU simultáneamente. Los procesadores AMD Ryzen AI Serie Max PRO incluyen además hasta 32 MB de caché MALL (Memory Attached Last Level, último nivel de memoria conectada) para amplificar el ancho de banda de los gráficos y mantener el núcleo de los gráficos de clase independiente.

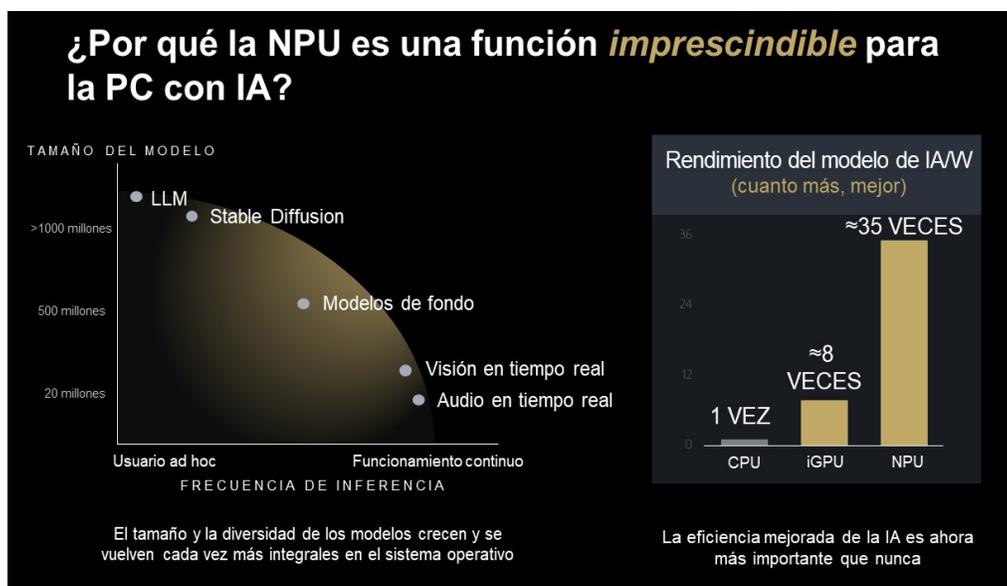
AMD ha creado Ryzen AI Serie Max PRO de esta forma para aprovechar las ventajas de consumo energético y eficiencia de la integración a nivel de chip, al tiempo que ofrece una potente GPU equivalente e independiente. Además, se pueden dedicar hasta 96 GB de la memoria total del sistema máxima disponible de 128 GB al procesamiento de gráficos, mucho más de lo que está disponible en cualquier dGPU actual.

Este enorme búfer de memoria tiene implicaciones especialmente interesantes para la IA. Como se ha comentado anteriormente, muchas tarjetas gráficas comerciales y de consumo solo pueden cargar modelos más antiguos o simplificados que se han recortado para ajustarse a los búferes de marco de 8 a 16 GB disponibles comúnmente en sistemas móviles y de escritorio mini. Los procesadores AMD Ryzen AI Serie Max PRO; por el contrario, pueden ofrecer suficiente memoria GPU para ejecutar modelos como Stable Diffusion 3.5 Large o Llama 3.1 70B-Q4 localmente. Cargas de trabajo de interferencia que, de otro modo, estarían fuera de alcance en cualquier otro sistema, móvil o de escritorio, pueden ejecutarse en un sistema AMD Ryzen AI Max PRO.

UNA NPU PARA CARGAS DE TRABAJO DE IA EMERGENTES

Hasta la fecha, la mayoría de las cargas de trabajo de IA se han dirigido a la CPU o a la GPU. Las unidades de procesamiento neuronal, o NPU, son un tipo de procesador emergente introducido por primera vez en la PC x86 por AMD en el 2023. El rendimiento de la NPU ha aumentado rápidamente, de 10 TOPS en los primeros procesadores AMD Ryzen Serie PRO 7040 a 50 TOPS con los procesadores Ryzen AI Max PRO disponibles hoy en día.

Se espera que la compatibilidad con el software de NPU crezca a medida que proliferen las PC con IA y los ISV se familiaricen más con sus puntos fuertes y capacidades. El potencial de eficiencia que ofrecen en comparación con las CPU o las GPU integradas tradicionales las convierten en un objetivo de optimización atractivo, como se muestra a continuación:



La decisión de incluir una NPU en AMD Ryzen AI Serie Max PRO refleja dónde es probable que se ejecuten las cargas de trabajo de IA en el futuro, mientras que la GPU de nivel independiente y los núcleos de CPU equivalentes de escritorio están diseñados para las aplicaciones convencionales y centradas en IA que las empresas utilizan hoy en día. A medida que las aplicaciones se trasladan a la NPU, ambas desbloquean un ahorro de energía adicional y liberan la CPU y la GPU para que se centren en otras tareas.

CONCLUSIÓN

Las empresas necesitan workstations que vayan más allá de la dicotomía tradicional móvil/torre, que limita las cargas de trabajo y las situaciones que las estaciones de trabajo móviles y compactas podrían abordar de forma plausible. Es probable que los usuarios de workstations, desde desarrolladores hasta creadores, encuentren IA tanto en el sistema operativo como en a nivel de aplicación, ya que las empresas realizan la transición a Windows 11 y los proveedores ISV integran la inteligencia artificial en las herramientas de edición de fotos y videos, las plataformas de gestión de contenido, las bases de conocimientos y los paquetes de oficina existentes.

Los procesadores AMD Ryzen AI Serie Max PRO son ideales para empresas y usuarios finales que desean trabajar con ensamblajes de proyectos más grandes, abordar proyectos más complejos y acelerados por IA y desarrollar nuevas aplicaciones basadas en LLM localmente. Los procesadores AMD Ryzen AI Serie MAX Pro incluyen potentes gráficos certificados por ISV, son compatibles con las funciones de seguridad y confiabilidad de tecnologías AMD PRO y aprovechan las ventajas intrínsecas de la integración para superar los límites de lo que es posible en una workstation móvil y compacta.

NOTAS AL PIE

1. Según IDC, los envíos globales de workstations para PC se redujeron casi un 9 % en el 2023, pero se espera una recuperación a medida que varios impulsores del mercado se fusionen en el 2024. 13 de marzo del 2024

DESCARGOS DE RESPONSABILIDAD

La información de este documento se proporciona solamente con fines informativos y puede que contenga imprecisiones técnicas, omisiones y errores de tipografía. Está sujeta a cambios y, como resultado, puede presentar imprecisiones por muchos motivos, incluidos, entre otros, cambios en el producto y el mapa de ruta, cambios en la versión de los componentes o la motherboard, lanzamientos de nuevos modelos o productos, diferencias de productos entre varios fabricantes, cambios de software, instalaciones de BIOS, actualizaciones de firmware o similares. Cualquier sistema informático tiene riesgos de vulnerabilidades de seguridad que no se pueden prevenir ni mitigar por completo. AMD no tiene obligación de actualizar o corregir o modificar esta información. No obstante, AMD se reserva el derecho de modificarla y de realizar cambios en su contenido de forma periódica sin obligación de notificar a ninguna persona al respecto.

AVISO DE COPYRIGHT

© 2025 Advanced Micro Devices, Inc. Todos los derechos reservados. AMD, el logotipo de la flecha de AMD, Ryzen y sus combinaciones son marcas comerciales de Advanced Micro Devices, Inc.