

COMO A IA ESTÁ IMPULSIONANDO A EVOLUÇÃO DO PROJETO DE ESTAÇÕES DE TRABALHO MÓVEIS E COMPACTAS

Historicamente, a grande maioria das estações de trabalho móveis dependia de uma placa de vídeo dedicada (dGPU) para oferecer desempenho suficiente para renderização 3D, edição de vídeo e projetos em CAD/CAM. Embora essa tenha sido uma estratégia eficaz para alguns usuários, a energia adicional que essas placas consomem forçava os fabricantes de sistemas a trocar outros recursos que os usuários finais valorizam, incluindo a espessura do sistema, temperaturas operacionais mais baixas e maior duração da bateria.

Os realinhamentos estruturais de longo prazo no mercado de estações de trabalho estão prestes a tornar essas trocas ainda mais caras. Os modelos de IA e os aplicativos emergentes que dependem deles estão impondo demandas novas e exclusivas aos fabricantes de estações de trabalho. A adoção generalizada de vídeos 4K e 8K em cabines de edição, o uso crescente de modelagem de informação da construção (BIM) em engenharia e a mudança constante em direção à renderização fotorrealista também colocaram pressão sobre os requisitos da GPU. Essas pressões vão contra a preferência geral do consumidor por sistemas mais leves, mais rápidos e com mais eficiência energética.

O MERCADO DAS ESTAÇÕES DE TRABALHO EM EXPANSÃO

Espera-se que as vendas de estações de trabalho móveis cresçam nos próximos cinco anos, com a previsão da IDC de que um em cada 10 sistemas comerciais será uma estação de trabalho até 2030 "à medida que mais organizações reconhecerem seu valor para cargas de trabalho de missão crítica". De acordo com Jay Chou, gerente de pesquisa de Worldwide Client Device Trackers da IDC, "A exploração atual em torno da IA para o trabalho deve se traduzir em uma expansão de casos de uso, incluindo desenvolvimento de modelos, em diversos setores".

Embora os casos de uso de estações de trabalho possam estar aumentando, os espaços ocupados pelo sistema não estão necessariamente indo na mesma direção. A adoção de estações de trabalho móveis aumentou durante a pandemia à medida que as condições de trabalho mudaram. E embora as estações de trabalho de torre continuem essenciais para o setor, os notebooks e as estações de trabalho de desktop de formato pequeno (SFF) são cada vez mais importantes.

Os clientes de estações de trabalho empresariais estão procurando sistemas projetados com espaço ocupado mínimo e fluxo de dados máximo. Essa mudança simultânea no fator de forma e no caso de uso é um desafio de projeto e uma oportunidade de redefinir o que as empresas esperam de uma estação de trabalho móvel e compacta.

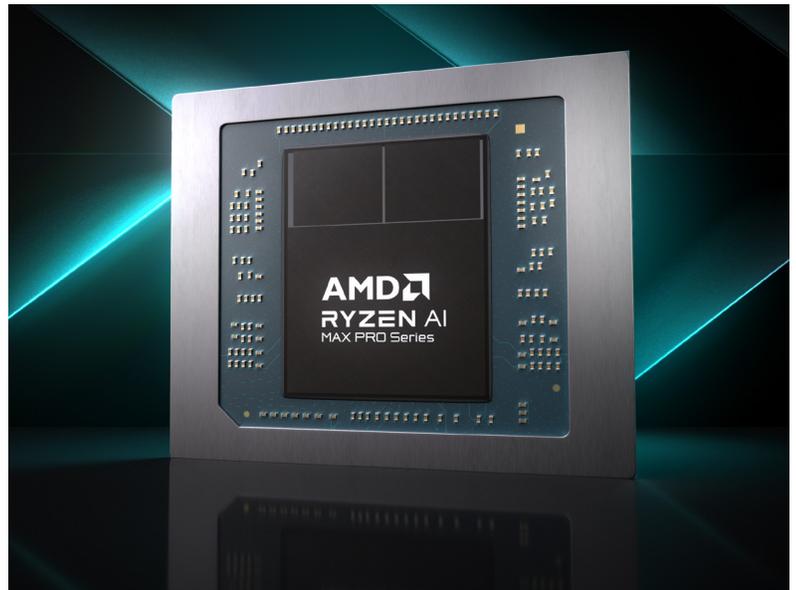
DESAFIANDO AS TROCAS QUE OS CLIENTES DE ESTAÇÕES DE TRABALHO SÃO FORÇADOS A ACEITAR

Muitas vezes, as estações de trabalho móveis e compactas são projetadas de forma a forçar os clientes a fazer escolhas binárias entre duas opções de produtos ou resultados desejáveis. As estações de trabalho de desktop são menos restritas, graças aos seus grandes volumes internos e aos sistemas de resfriamento agressivos.

Essa dicotomia força os usuários de estações de trabalho móveis e compactas a comprometer recursos e desempenho. É uma visão que está fora de sintonia com as tendências maiores em torno dos fatores de forma do sistema e requisitos computacionais aumentados da IA. Há uma nova maneira de atender às demandas de computação de hoje; e uma solução melhor para atendê-las.

APRESENTAMOS O AMD RYZEN™ AI MAX PRO:

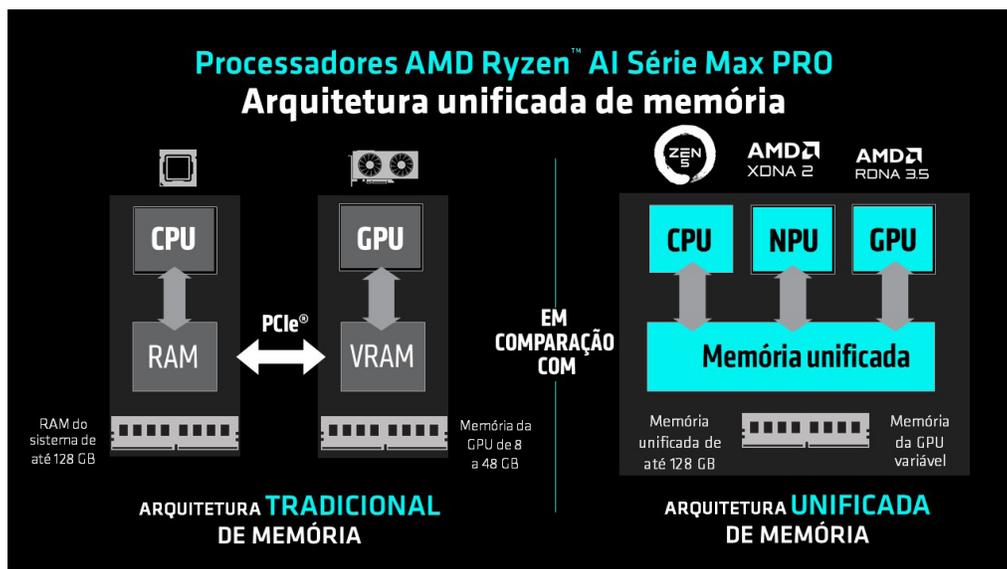
A série de processadores AMD Ryzen™ AI Max PRO representa um marco significativo para os sistemas x86 e o amplo mercado de estações de trabalho do Windows. Como os primeiros processadores x86 a combinar uma GPU integrada com desempenho de nível dedicado, núcleos de CPU de classe de desktop e uma unidade de processamento neural (NPU) em um único chip, eles são adaptados para clientes que desejam excelente desempenho, otimizações e certificações de aplicativos profissionais, e a capacidade de executar cargas de trabalho de IA que, de outra forma, seriam muito pesadas para a maioria das dGPUs encontradas em notebooks comerciais e desktops SFF comumente disponíveis hoje. Os processadores AMD Ryzen AI Série Max PRO foram desenvolvidos para lidar com projetos 3D complexos com vários aplicativos em execução em paralelo ou para explorar novas ideias usando modelos de linguagem grande local.



ELIMINANDO GARGALOS DE MEMÓRIA DE GPU

As placas de vídeo móveis dedicadas modernas geralmente oferecem entre 8 e 16 GB de RAM de vídeo dedicada (VRAM). Embora isso seja suficiente para muitos aplicativos de estação de trabalho, executar cargas de trabalho com grandes conjuntos de dados e modelos de IA localmente representa um desafio maior. Por exemplo, o Stable Diffusion 3.5 Large pode exceder até mesmo a capacidade de 16 GB de VRAM normalmente disponível em GPUs móveis de alta tecnologia. Os clientes devem usar versões menos complexas do modelo que *podem* se ajustar na memória limitada disponível ou ganhar tempo em serviços na nuvem.

Os processadores AMD Ryzen AI Série Max PRO resolvem esse problema compartilhando um pool comum de memória entre a CPU, GPU e NPU, conforme mostrado abaixo:



Esse tipo de compartilhamento é conhecido como arquitetura de memória unificada (UMA), em vez de uma arquitetura de memória distribuída tradicional (mostrada à esquerda), em que a CPU e a GPU têm seus próprios pools dedicados de memória.

Uma arquitetura de memória tradicional fornece à GPU uma grande oferta de largura de banda, mas muito menos memória total em comparação com a CPU. Isso também torna a comunicação entre a CPU e a GPU mais lenta. Conectar esses dois componentes juntos on-die, conforme mostrado à direita, permite que eles compartilhem um pool comum de memória. Isso é vantajoso, desde que o sistema ofereça largura de banda de memória suficiente. Os processadores AMD Ryzen AI Série Max PRO oferecem.

Enquanto a maioria dos outros processadores de estação de trabalho móvel x86 depende de dois canais de memória, o AMD Ryzen AI Série Max PRO usa quatro. O aumento resultante na largura de banda total do sistema é suficiente para alimentar a GPU, CPU e NPU simultaneamente. Os processadores AMD Ryzen AI Série Max PRO também incluem até 32 MB de cache MALL (Memory Attached Last Level) para ampliar a largura de banda da placa de vídeo e manter o núcleo da placa de vídeo dedicada alimentado.

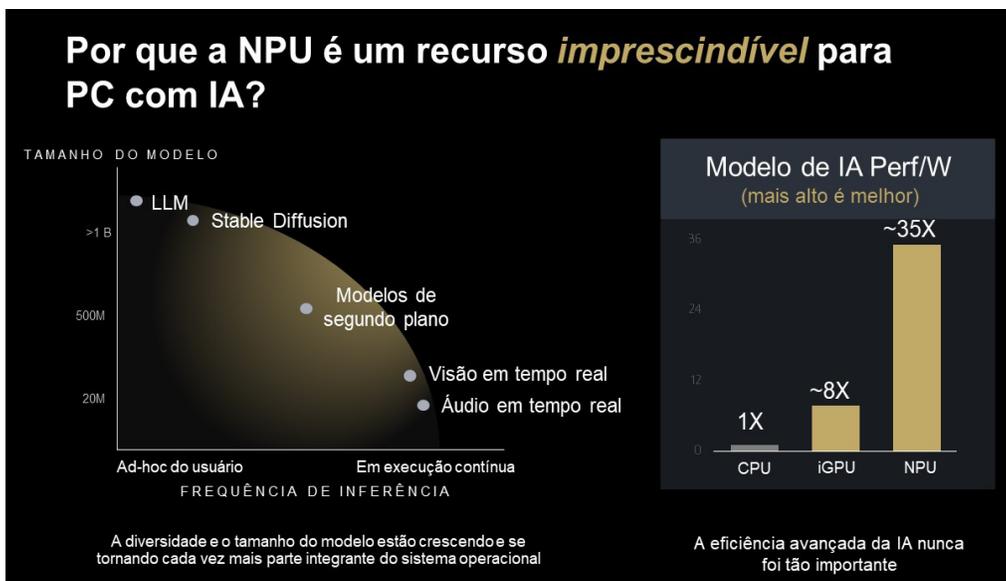
A AMD criou o Ryzen AI Série Max PRO dessa forma para aproveitar as vantagens de eficiência e o consumo de energia da integração em nível die, ao mesmo tempo em que oferece uma GPU potente e dedicada equivalente. Além disso, é possível dedicar até 96 GB dos 128 GB de memória total máxima disponível do sistema ao processamento gráfico, muito mais do que está disponível em qualquer dGPU hoje.

Esse enorme buffer de memória tem implicações particularmente interessantes para a IA. Como discutido anteriormente, muitas placas de vídeo comerciais e para o consumidor só podem carregar modelos mais antigos ou simplificados que foram ajustados para se encaixar nos buffers de quadro de 8-16 GB comumente disponíveis em sistemas de desktop móveis e mini. Os processadores AMD Ryzen AI Série Max PRO, por outro lado, podem oferecer memória GPU suficiente para executar modelos como Stable Diffusion 3.5 Large ou Llama 3.1 70B-Q4 localmente. As cargas de trabalho de inferência que, de outra forma, estariam fora de alcance em qualquer outro sistema, móvel ou desktop, podem ser executadas em um sistema AMD Ryzen AI Max PRO.

UMA NPU PARA CARGAS DE TRABALHO DE IA EMERGENTES

Até o momento, a maioria das cargas de trabalho de IA tem como alvo a CPU ou a GPU. As unidades de processamento neural, ou NPUs, são um tipo de processador emergente introduzido pela primeira vez no PC x86 pela AMD em 2023. O desempenho da NPU aumentou rapidamente, de 10 TOPS nos primeiros processadores AMD Ryzen PRO Série 7040 para 50 TOPS com os processadores Ryzen AI Max PRO disponíveis hoje.

Espera-se que o suporte ao software NPU cresça à medida que os PCs com IA se multipliquem e os ISVs se familiarizem mais com seus pontos fortes e capacidades. O potencial de eficiência que eles oferecem em comparação com CPUs ou GPUs integradas tradicionais os torna um alvo de otimização atrativo, conforme mostrado abaixo:



A decisão de incluir uma NPU no AMD Ryzen AI Série Max PRO reflete onde as cargas de trabalho de IA provavelmente serão executadas no futuro, enquanto os núcleos de CPU equivalentes a desktop e GPU de nível dedicado são projetados para os aplicativos convencionais e centrados em IA que as empresas estão usando hoje. À medida que os aplicativos são movidos para a NPU, eles desbloquearão a eficiência de energia adicional e liberarão a CPU e a GPU para se concentrarem em outras tarefas.

CONCLUSÃO

As empresas precisam de estações de trabalho que vão além da tradicional dicotomia móvel/torre que limita as cargas de trabalho e cenários que as estações de trabalho móveis e compactas poderiam atender de forma satisfatória. Os usuários de estações de trabalho, de desenvolvedores a criadores, provavelmente encontrarão IA tanto no nível do sistema operacional quanto do aplicativo à medida que as empresas fazem a transição para o Windows 11 e os ISVs integram inteligência artificial em ferramentas existentes de edição de fotos e vídeos, plataformas de gerenciamento de conteúdo, bases de conhecimento e pacotes empresariais.

Os processadores AMD Ryzen AI Série Max PRO são ideais para empresas e usuários finais que desejam trabalhar com grandes montagens de projetos, lidar com projetos mais complexos e acelerados por IA e desenvolver novos aplicativos baseados em LLM localmente. Os processadores AMD Ryzen AI Série Max PRO incluem poderosos gráficos certificados pelo ISV, oferecem suporte a recursos de segurança e confiabilidade das tecnologias AMD PRO e aproveitam as vantagens intrínsecas da integração para expandir os limites do que é possível em uma estação de trabalho móvel e compacta.

NOTAS DE RODAPÉ

1. [Veja que as remessas globais da IDC das estações de trabalho de PC diminuiram quase 9% em 2023, mas a recuperação era esperada à medida que vários motivadores de mercado se unissem em 2024, de acordo com a IDC, 13 de março de 2024](#)

ISENÇÕES DE RESPONSABILIDADE

As informações contidas neste documento são apenas para fins informativos e podem conter imprecisões técnicas, omissões e erros tipográficos. As informações aqui contidas estão sujeitas a alterações e podem ser imprecisas por vários motivos, incluindo, entre outras, alterações de produtos e roteiros, alterações de versões de componentes e placas-mãe, novos modelos e/ou versões de produtos, diferenças de produtos entre diferentes fabricantes, alterações de software, BIOS flashes, atualizações de firmware ou similares. Qualquer sistema de computador apresenta riscos de vulnerabilidades de segurança que não podem ser completamente evitados ou mitigados. A AMD não assume obrigação de atualizar ou de fazer quaisquer correções ou revisões a estas informações. No entanto, a AMD se reserva o direito de revisar essas informações e fazer alterações periodicamente no conteúdo, sem que a AMD seja obrigada a notificar alguém sobre tais revisões ou alterações.

AVISO DE COPYRIGHT

© 2025 Advanced Micro Devices, Inc. Todos os direitos reservados. AMD, o logotipo de seta AMD, Ryzen e combinações dos mesmos são marcas comerciais da Advanced Micro Devices, Inc.