中泰证券 ZHONGTAI SECURITIES

证券研究报告

报告日期：2024年6月11日

【中泰电子】Computex 2024系列—AMD主题演讲：CPU+GPU+UA互联厂商

分析师：

王芳 S0740521120002

杨旭 S0740521120001

李雪峰 S0740522080004

中泰证券研究所

专业 | 领先 | 深度 | 诚信

AMD ComputeX 2024主题演讲

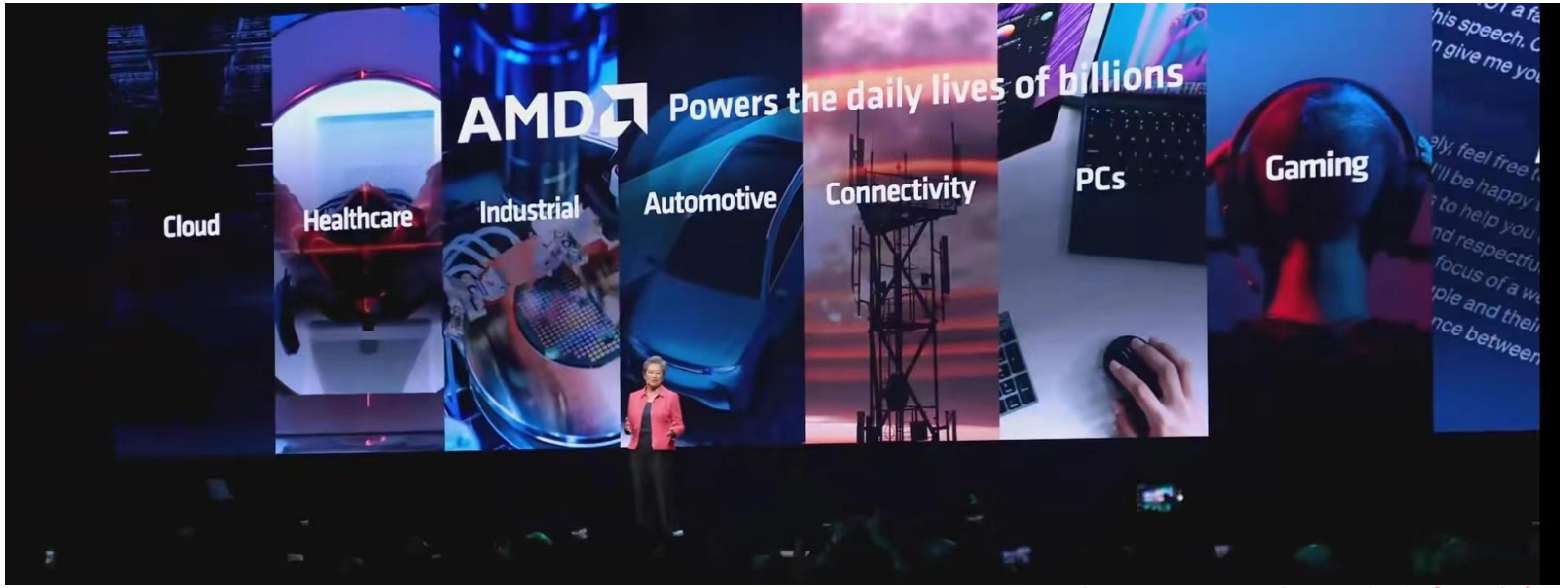
■ **6月3日，AMD CEO苏姿丰在台北ComputeX 2024大会上详细展示了其在CPU、GPU及UA互联等方面的最新产品：**

- **Zen 5：**展示被苏姿丰称之为“迄今为止性能最高、能效最高的处理器核心”——全新Zen 5核心。
- **第二代XDNA NPU架构：**XDNA NPU 2引入了全新的Block FP16（BF16）浮点精度，其AI引擎性能是第二代AMD锐龙AI的三倍，是目前唯一可提供50 TOPS的AI处理性能的产品。
- **消费级PC处理器：**推出全新Ryzen 9000 CPU，全球速度最快的消费级PC处理器，采用Zen5核心，并引入AM5平台，支持最新的I/O和内存技术。首发有4款产品均于今年7月出售。
- **AI PC处理器：**AMD下一代超薄和高端笔记本电脑的处理器“Strix Point”，苏姿丰称之为“面向下一代AI PC/Copilot+PC的世界一流处理器”。该系列产品展示的有Ryzen AI 9 HX370。
- **全新Versal AI Edge Gen 2系列：**提供首个集成预处理、推理和后处理的单芯片自适应解决方案。
- **AI GPU：**缩短产品更新时间，计划今年将推出速度更快、内存更大的MI325X；25年推出新的cDNA4架构的MI350系列；26年，将推出带有全新cDNA架构的MI400系列。
- **UA-Link：**展示AMD在推动高性能AI网络基础设施系统的发展方面也取得了重大进展：计划将于今年推出UA-Link 1.0标准。

来源：AMD官网，中泰证券研究所

AMD ComputeX 2024主题演讲

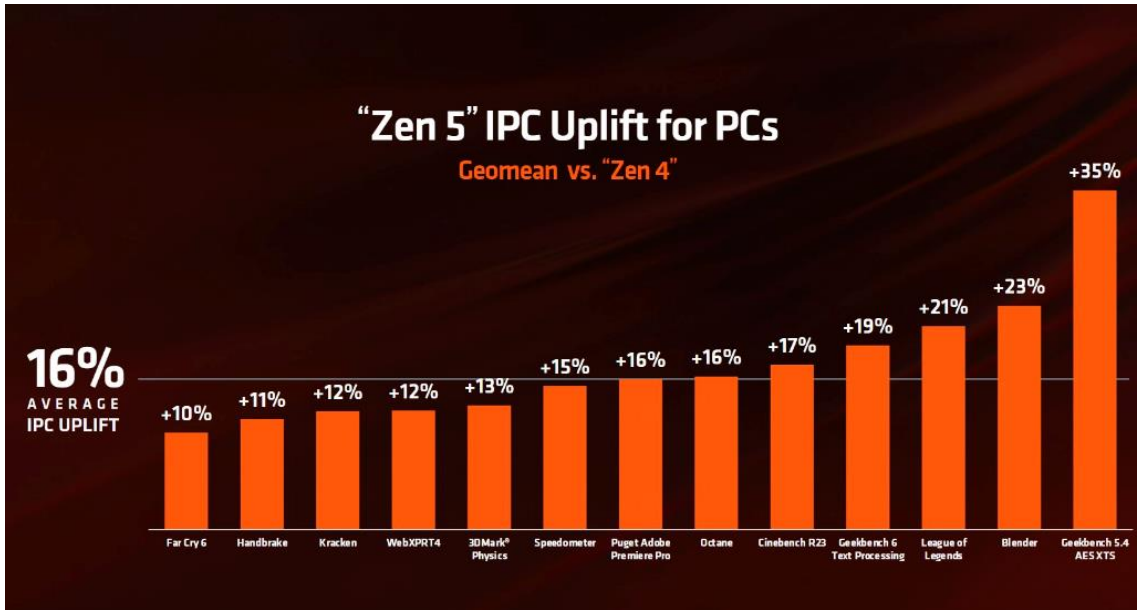
■ **AMD一直以来致力于推动高性能和自适应计算的发展，从云端和企业数据中心到5G网络，再到医疗保健、工业、汽车、个人PC、游戏和AI，AMD技术无处不在。**目前AI是AMD的首要任务，因为AI几乎改变了每一个企业，重塑了计算市场的每个部分。在这个行业，AMD具有独特优势，可以为定义人工智能计算时代端到端基础设施提供动力。现在，为了提供这些领先的AI解决方案，AMD专注于三个优先事项：首先，AMD为AI训练和推理提供了广泛的高性能和节能计算引擎，包括CPU、GPU和NPU；其次，AMD致力于创建一个开放、成熟且对开发人员友好的生态系统，确保所有领先的AI框架、库和模型都能在AMD硬件上完全应用；第三，AMD致力于合作创新，与全球最大的云服务提供商、软件公司和AI公司合作，共同将最好的人工智能解决方案推向市场。



来源：AMD官网，中泰证券研究所

桌面处理器—Ryzen 9000系列

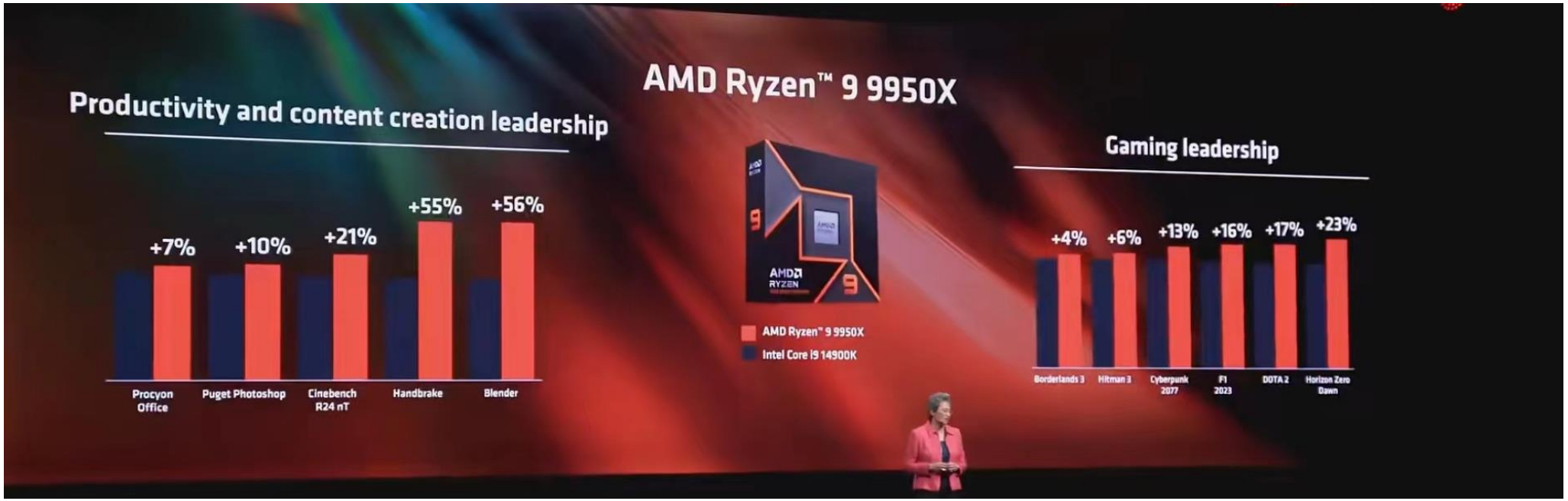
- **AMD全新Ryzen 9000 CPU是全球速度最快的消费级PC处理器，**采用全新的Zen 5核心并引入AM5平台，支持最新的I/O和内存技术，包括PCIE 5和DDR5。
- **Zen 5是是高性能 CPU 的下一大步，是一种全新的设计，性能极高，而且非常节能。**从超级计算机到数据中心和个人PC，Zen 5随处可见。Zen 5采用了新的并行双管道前端，它的作用是提高分支预测准确性并减少延迟，还能使我们能够在每个时钟周期提供更高的性能。此外，Zen 5具有更宽的CPU引擎和指令窗口，可以并行运行更多指令，以实现领先的计算吞吐量和效率。与Zen 4相比，它获得了双倍的指令带宽、双倍的缓存和浮点单元之间的数据带宽以及双倍的AI性能，具有完整的AVX-512吞吐量，所有这些都汇集在Ryzen 9000系列产品中。与Zen 4相比，Zen 5在广泛应用程序基准测试和游戏中高出平均16%的IPC，尤其在GeekBench5上大幅提升。



来源：AMD官网，中泰证券研究所

桌面处理器—Ryzen 9000系列

- **顶配的Ryzen 9 9950X：**拥有16个Zen5核心，32个线程，高达5.7GHz的提升，80兆字节的缓存，TDP为170瓦，这是世界上最快的消费级CPU。
  - 与酷睿i9-14900K相比，锐龙9 9950X在多项基准、实际应用以及游戏测试里面都有领先，在Cinebench 2024的多线程测试中就比酷睿i9-14900K快21%之多，而在Handbrake和Blender测试里面领先幅度更是超过50%，在游戏测试中，则有4%到23%的优势。

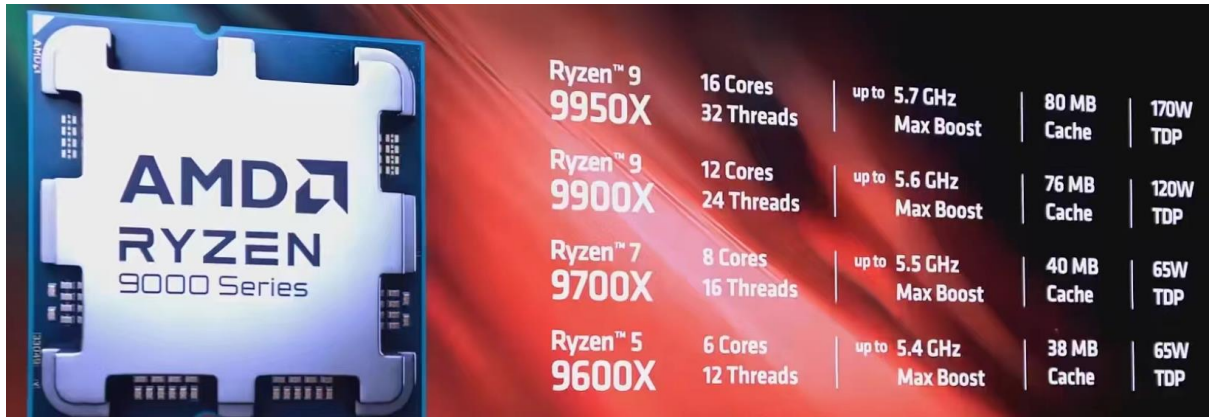


来源：AMD官网，中泰证券研究所

锐龙系列产品矩阵更新

■ **如今Ryzen能帮助用户可以拥有一个跨多代产品升级的基础设施。**最初的Ryzen平台 Socket AM4于2016年推出，现在在平台Socket AM4的基础上，AMD有11个不同产品系列的145款CPU和APU，下个月我们将会推出几款Ryzen 5000 CPU。对于Socket AM5也采取同样的策略，目前计划支持到2027年及以后。此外， Ryzen 9000处理器将于今年7月销售，首发产品如下：

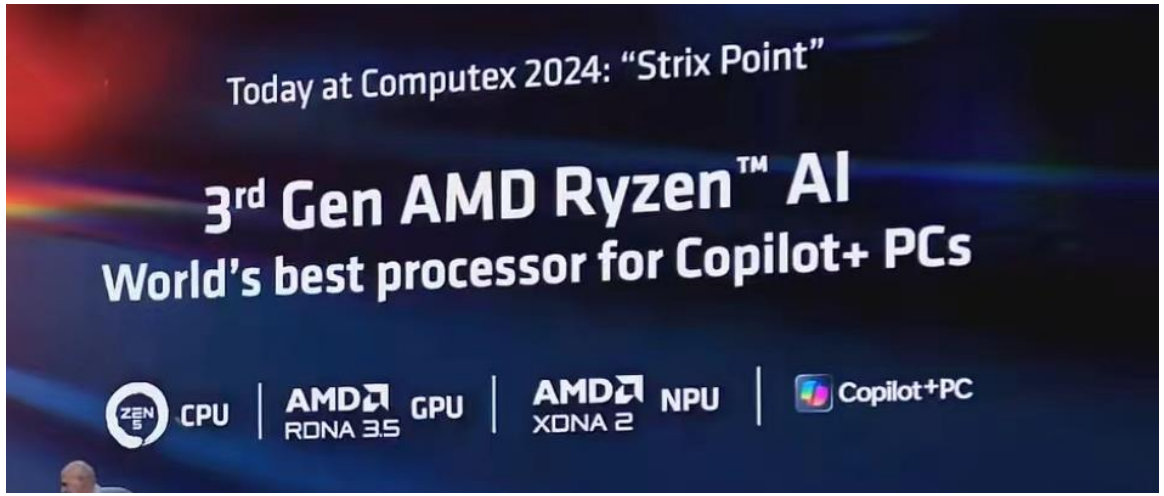
- 锐龙9 9950X，16核32线程，最高频率5.7GHz，80MB Cache，170W TDP；
- 锐龙9 9900X，12核24线程，最高频率5.6GHz， 76MB Cache ，120W TDP；
- 锐龙7 9700X，8核16线程，最高频率5.5GHz， 40MB Cache ，65W TDP；
- 锐龙5 9600X，6核12线程，最高频率5.4GHz， 38MB Cache ，65W TDP。



来源：AMD官网，中泰证券研究所

AI PC—锐龙AI 300系列

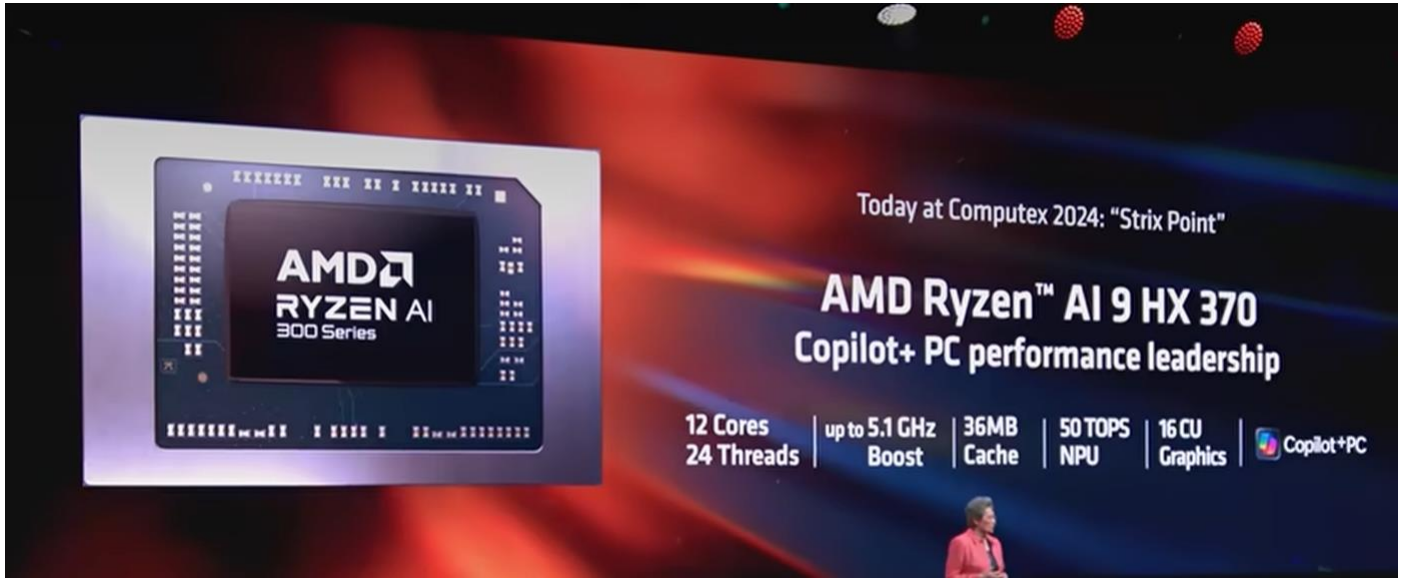
- **AI PC：**自去年AMD推出第一代Ryzen AI以来，AMD实际上一直在引领AI PC的浪潮。如今AI PC正彻底改变了我们与PC交互的方式，它能够实现更智能化、个性化的体验，使PC成为我们日常生活中更重要的一部分。为了更好的实现这些，我们需要更好的AI硬件：AMD第三代Ryzen AI系列显著提高了计算和AI性能，并为Copilot PC设定了新标准。
- Strix Point是AMD下一代超薄和高端笔记本电脑的处理器，它采用了全新的Zen 5 CPU架构，GPU内核升级为RDNA3.5架构，NPU是全新的XDNA 2 架构，新的NPU提供行业领先的50 TOPS计算能力，能够在极低功耗下实现新的人工智能体验，号称是“面向下一代AI PC/Copilot+ PC的世界一流处理器”。



来源：AMD官网，中泰证券研究所

AI PC—锐龙AI 300系列

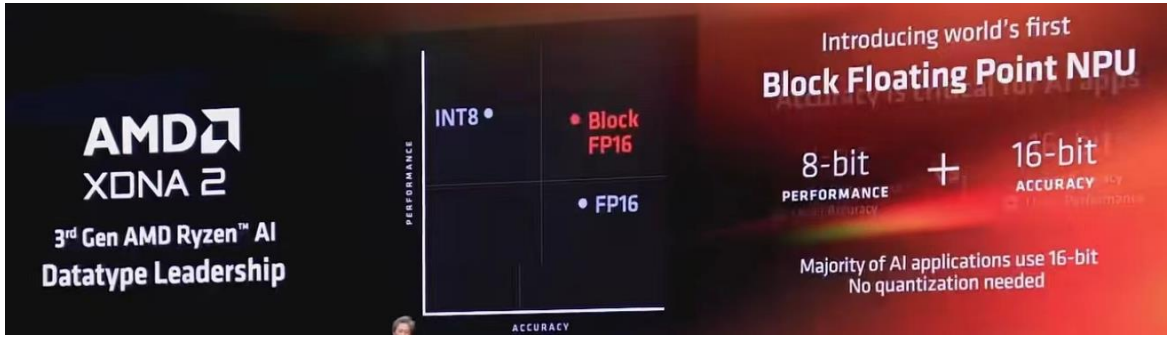
- Ryzen AI 9 HX370作为旗舰产品，CPU部分拥有12核24线程，36MB 缓存，最高主频5.1GHz。GPU部分不但升级了架构，CU单元数量增至16个，NPU算力则提升到了50TOPS。

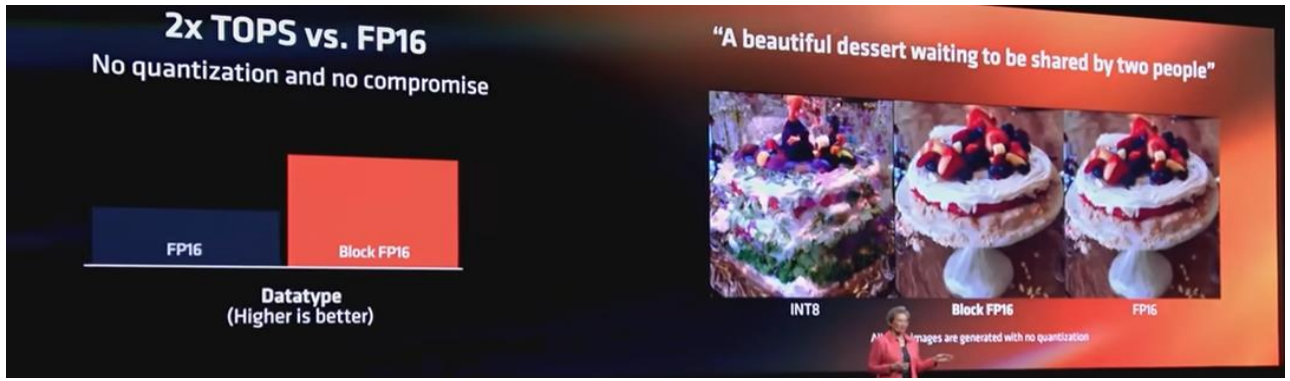


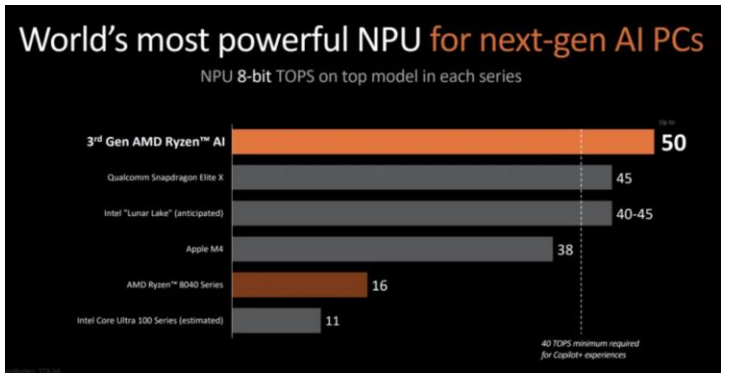
来源：AMD官网，中泰证券研究所

AI PC—锐龙AI 300系列

■ **第二代XDNA NPU架构的提升：**与AMD第一代XDNA NPU架构对比，第二代XDNA NPU架构具有32个AI计算引擎模块；本地内存模块，增加至8个，可以更好地配合更大规模的本地调度、运算；多任务处理性能高了一倍；在运行Gen AI工作负载时，其能效高出两倍，计算能力提升5倍。此外，XDNA2首发引入了全新的Block FP16 （BF16）浮点精度，AI引擎性能是第二代 AMD 锐龙 AI 的三倍，是目前唯一可提供 50 TOPS 的AI 处理性能的产品。它在CPU、GPU上已经很常见，而在NPU上还是第一次。传统的FP8浮点格式性能高而精度不足，FP16浮点格式精度高而性能略逊，而将二者融合起来的BF16可以在精度、性能上达到较好的平衡，灵活性也更高。并且，如今大多数AI应用都采用了16位精度，因此有了BF16，不再需要量化为8位精度，减少了转换步骤，提高了执行效率。







来源：AMD官网，中泰证券研究所

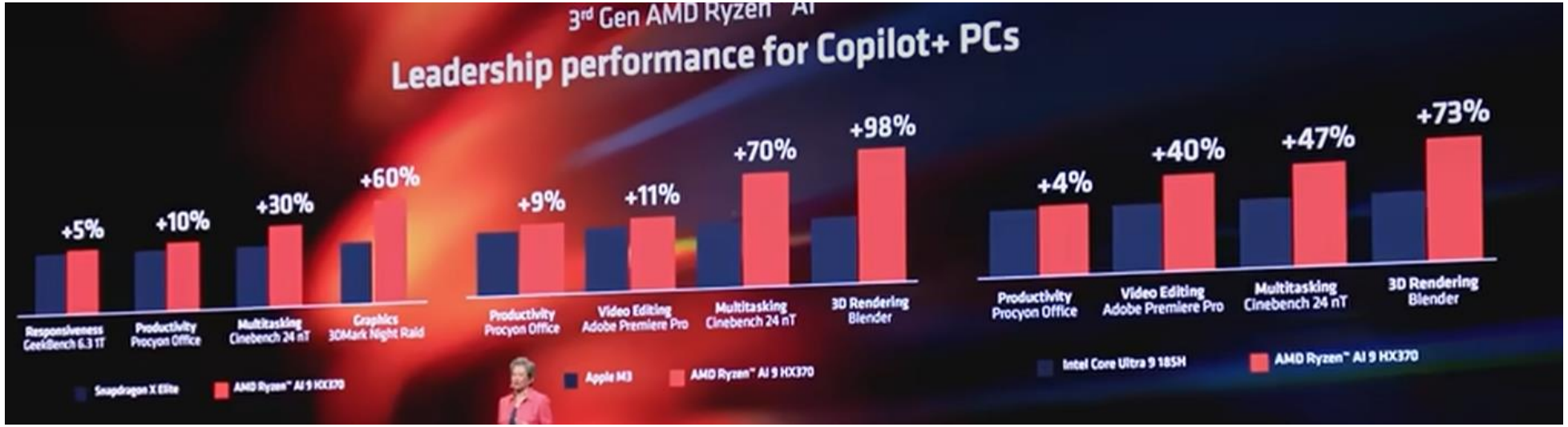
AI PC—锐龙AI 300系列

- AMD现在的客户有：微软、Adobe、索尼、zoom等用户为了加速采用支持AI的PC应用程序，到今年年底这些用户有望超过150家。

来源：AMD官网，中泰证券研究所

AI PC—锐龙AI 300系列

■ **苏姿丰认为Strix是市场上最好的笔记本CPU：**将Ryzen AI 300系列与所有最新的x86和ARM CPU进行比较时，无论是单线程响应、生产力应用、内容创作还是多任务处理方面，第三代Ryzen AI处理器都提供了更高的性能，而且绝大多数情况下都是以两位数的百分比领先竞品。



来源：AMD官网，中泰证券研究所

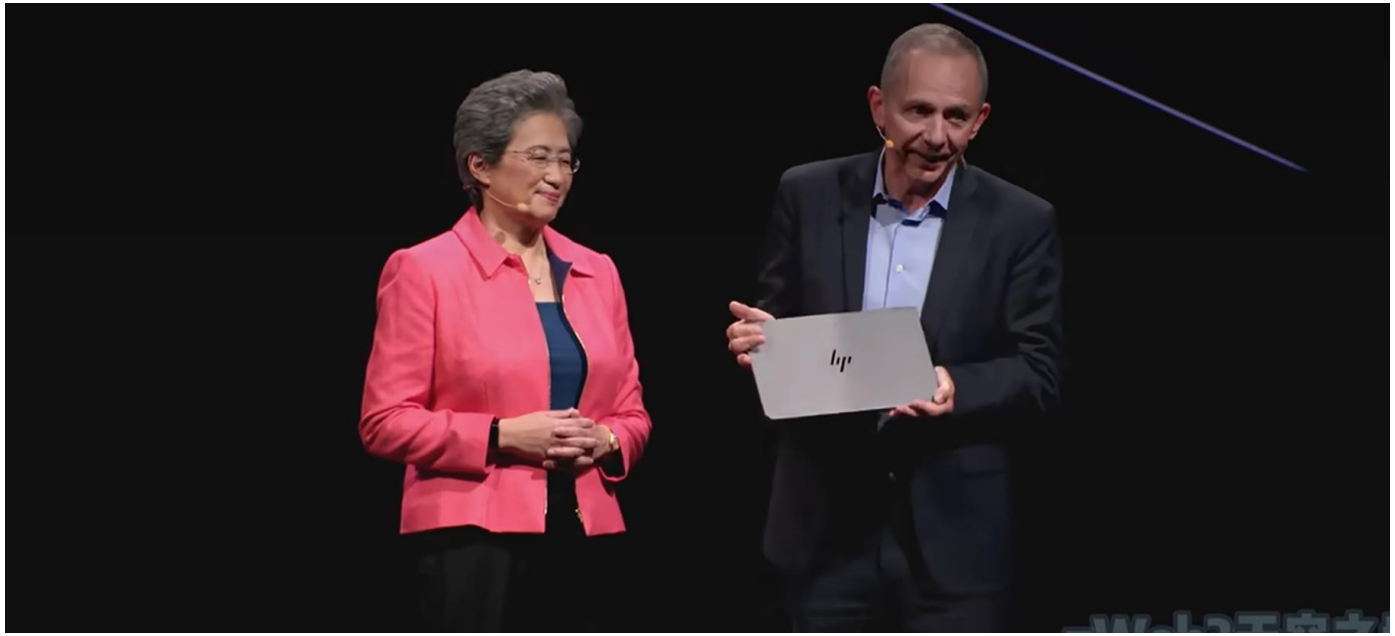
AI PC各厂商进展更新

- **微软公司副总裁：**宣布了一个为人工智能设计的新型个人电脑，即Copilot Plus个人电脑。为了挖掘人工智能在个人电脑上的全部潜力，重新设计了整个系统，从芯片到Windows的每一层。这是最快、最智能的个人电脑，很高兴能与AMD合作推出基于Strix的Copilot Plus个人电脑。
- Copilot Plus个人电脑与几年前的传统个人电脑相比，性能提升了20倍，人工智能工作负载的效率提高了100倍。为了实现这一目标，每台Copilot Plus个人电脑都需要至少40 TOPS的NPU。而Strix的NPU提供了50 TOPS，非常强大。

来源：AMD官网，中泰证券研究所

AI PC各厂商进展更新

- **惠普公司总裁兼首席执行官Enrique Lores：**即将推出的新一代人工智能个人电脑，这是整合最新Ryzen AI 300系列的产品，这将是第一个在设备中集成50 TOPS性能的产品。



来源：AMD官网，中泰证券研究所

AI PC各厂商进展更新

- **联想智能设备集团总裁Luca Rossi：**今年晚些时候，将推出搭载第三代Ryzen AI处理器的联想AI笔记本电脑，面向消费者的是Yoga系列，面向商用的是Thinkpad品牌，面向中小企业的是Thinkbook系列。无论创意人士、企业专业人士还是创业者，联想都将拥有完美的Copilot Plus笔记本电脑，搭载第三代Ryzen AI，提供业界领先的50 TOPS性能。此外还将推出独家联想AI体验，其中之一是Creator Zone，这是其与AMD共同生成并经过微调AI模型后的独家联想软件，专为创意人士提供工具和功能，提升创造力和生产力。

来源：AMD官网，中泰证券研究所

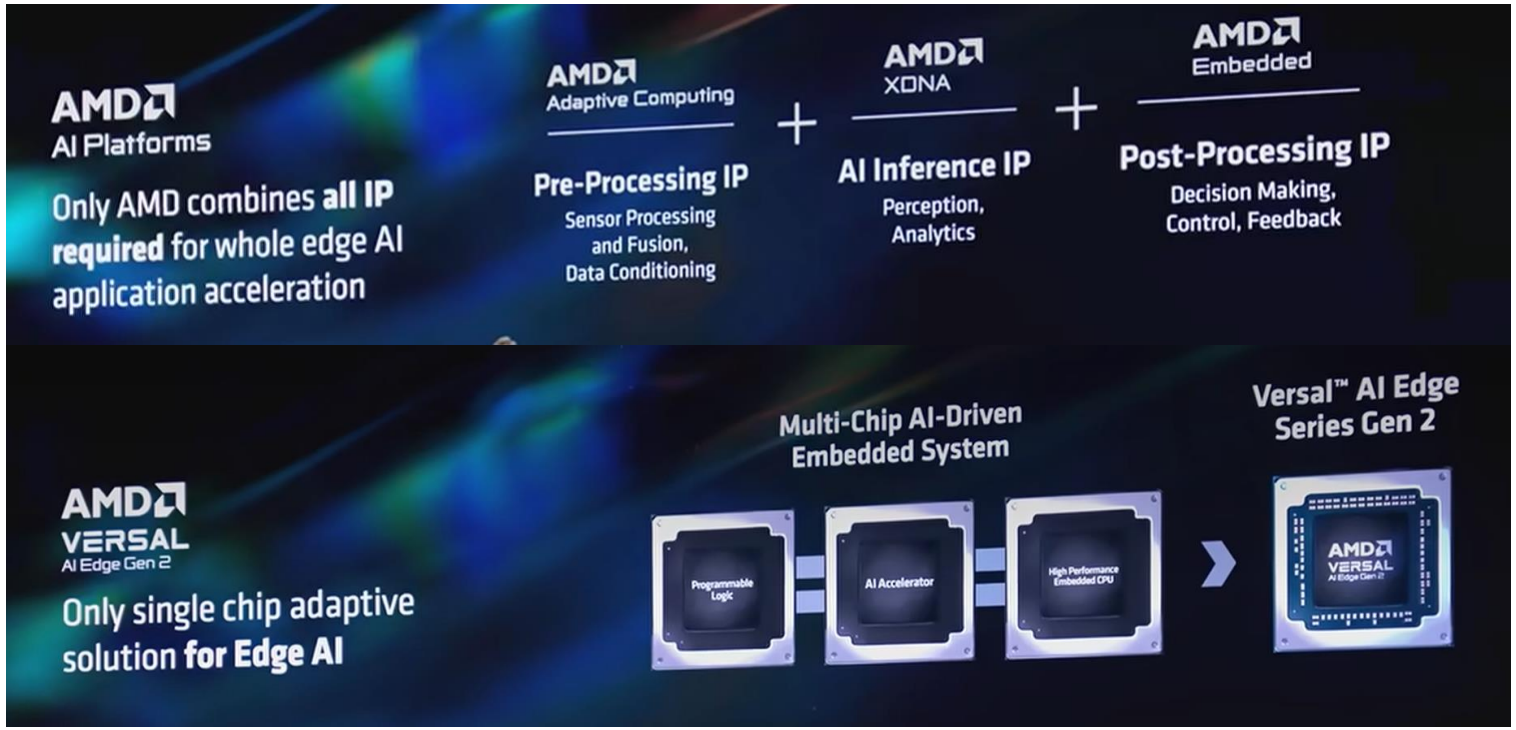
AI PC各厂商进展更新

■ **华硕董事长Jonney Shih：**2024年6月3日下午4点，华硕将推出一系列前沿的人工智能个人电脑，包括全新的Zenbook、ProArt、VivoBook、Asus TUF和LG笔记本电脑，这些产品都搭载了第三代AMD Ryzen AI处理器。新的产品配备了全球最强大的NPU，拥有50 TOPS性能，以及领先的Zen5架构，在计算和人工智能性能方面引领行业。这些产品预计将在7月上市。

来源：AMD官网，中泰证券研究所

AMD 边缘端Versal AI Edge Gen 2

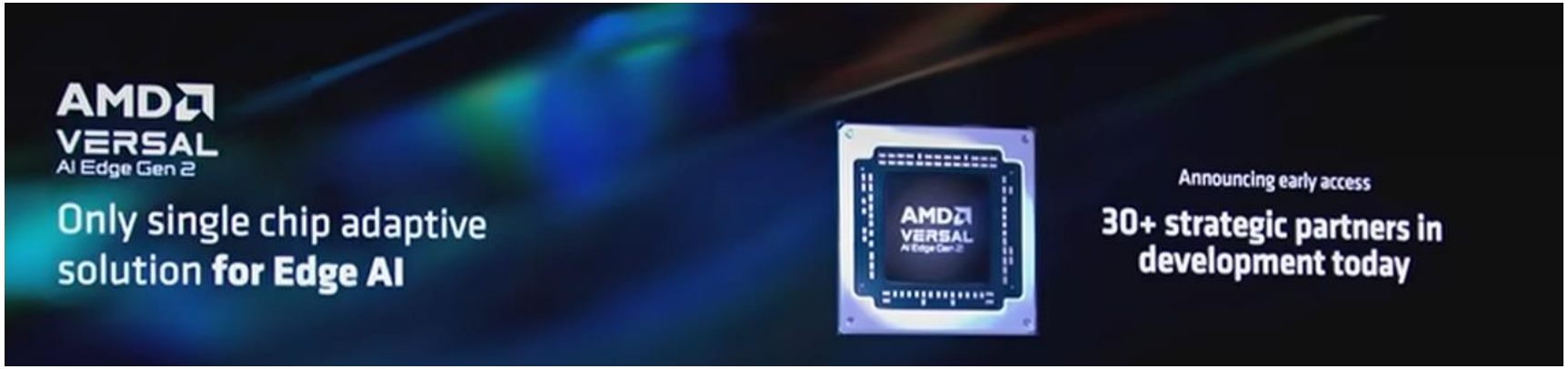
- **全新Versal AI Edge Gen 2系列能同时解决需要三个独立芯片解决的问题。**现在，AI在边缘设备上的应用是一个难题。它需要设备内部完成预处理、推理和后处理的能力。而只有AMD拥有加速端到端边缘AI所需的所有组件。公司结合了用于预处理的自适应计算引擎、传感器和其他数据，用于推理的AI引擎，以及用于后处理决策的高性能嵌入式计算核心，当前，要实现这些功能通常需要三个独立的芯片。而AMD的全新Versal AI Edge Gen 2系列，将所有这些领先的计算能力整合在一起，创造了首个集成了预处理、推理和后处理的单芯片自适应解决方案。



来源：AMD官网，中泰证券研究所

AMD 边缘端Versal AI Edge Gen 2

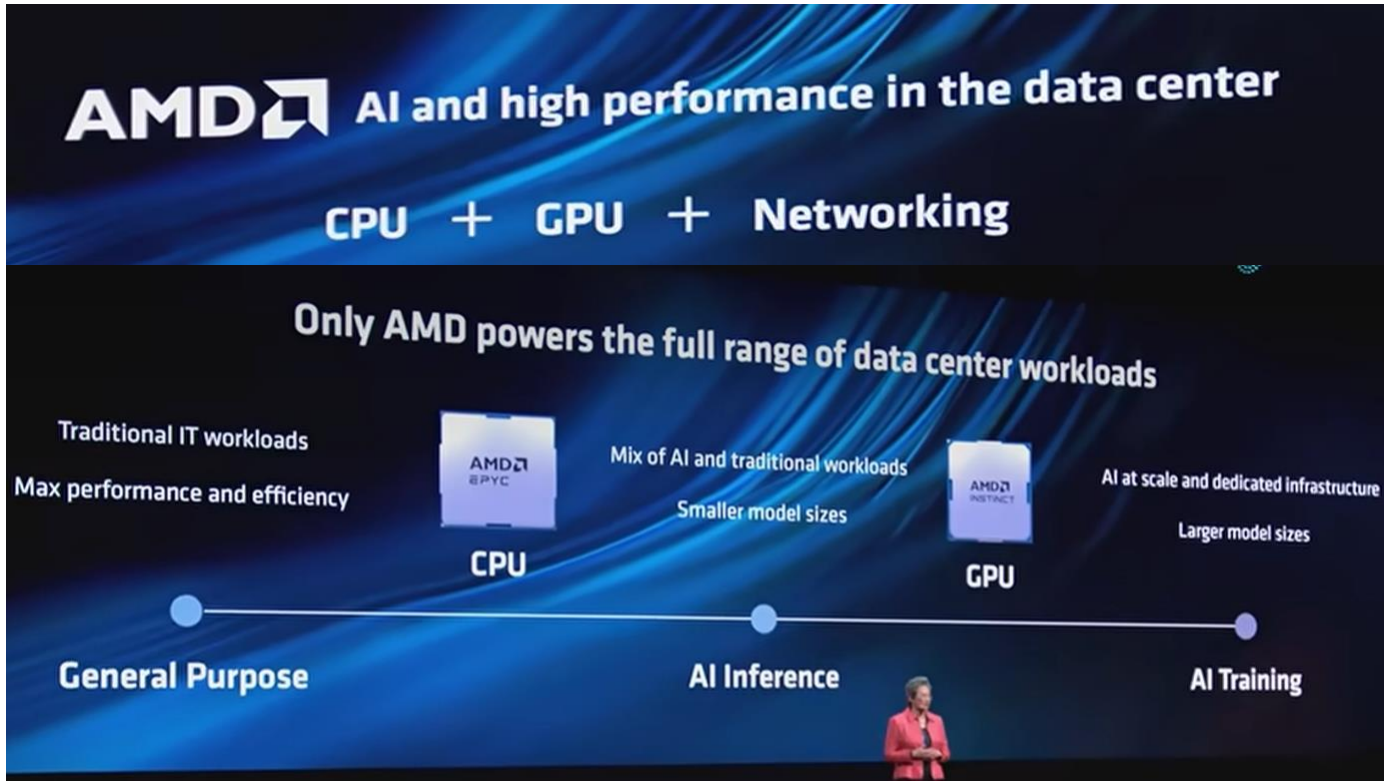
- AMD宣布了其下一代Versal平台的早期访问。30多家战略合作伙伴已经在开发通过全新单芯片Versal解决方案驱动的边缘AI设备，公司表示看到了推动边缘AI的机会以及利用新技术扩大嵌入式市场领导地位的重大机会。



来源：AMD官网，中泰证券研究所

云计算CPU产品—AMD Epic

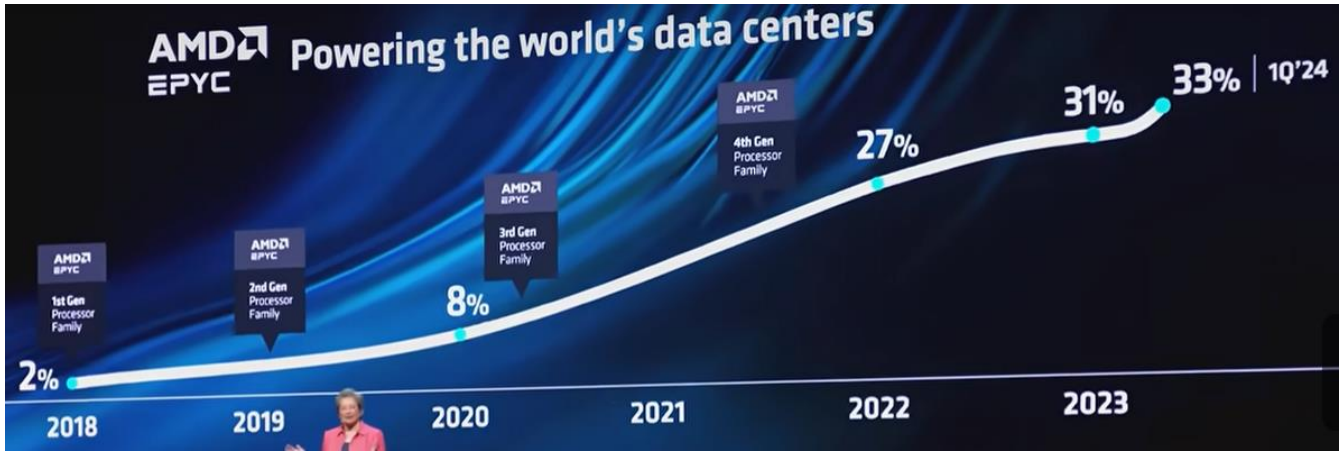
■ **数据中心：AMD称自己是唯一能提供全套数据中心的工作负载厂商。**目前AMD已经建立了业界最广泛的高性能CPU、GPU和网络产品的组合。现在的数据中心实际上运行许多不同的工作负载，包括传统的IT应用程序，小型企业级大型语言模型（LLMs）以及大规模的AI应用程序，这需要为每种工作负载提供不同的计算引擎。而只有AMD拥有全面的高性能CPU和GPU产品组合来满足所有的工作负载。例如：AMD的EPYC处理器，它在通用和混合推理AI工作负载中提供领先性能；业界领先的Instinct GPU，它是为了加速大规模AI应用程序而构建的。

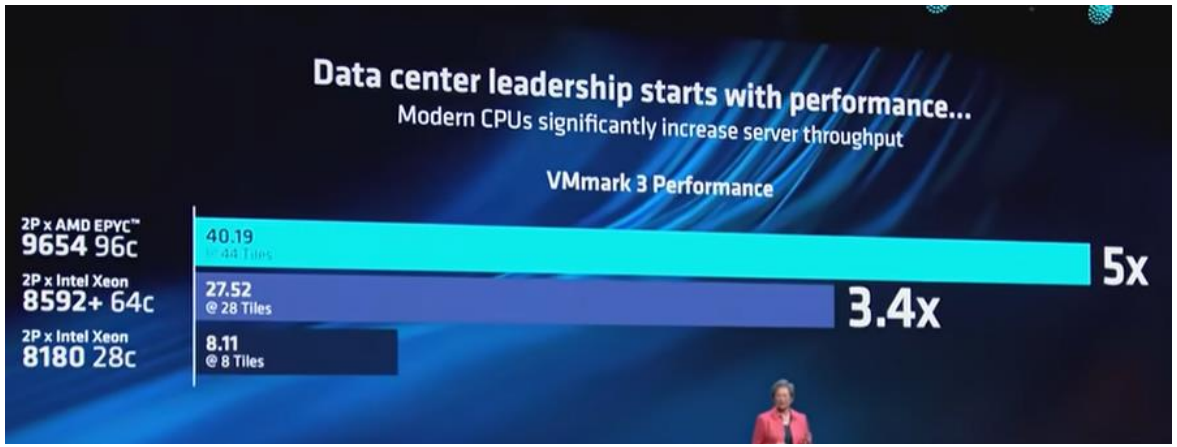


来源：AMD官网，中泰证券研究所

云计算CPU产品—AMD Epic

- **CPU市场：**EPYC作为AMD2017年推出的产品，随着每一代的更新，成为了云计算的首选处理器，每天全球数十亿人使用由EPYC提供的云服务。目前该市场份额占比33%，预计还会增长。
- 在最新一代服务器CPU的虚拟化性能方面，在新技术的加持下，EPYC的性能比传统处理器高出5倍，甚至和现在行业内竞争对手最好的处理器相比，AMD的性能也快了一半。





来源：AMD官网，中泰证券研究所

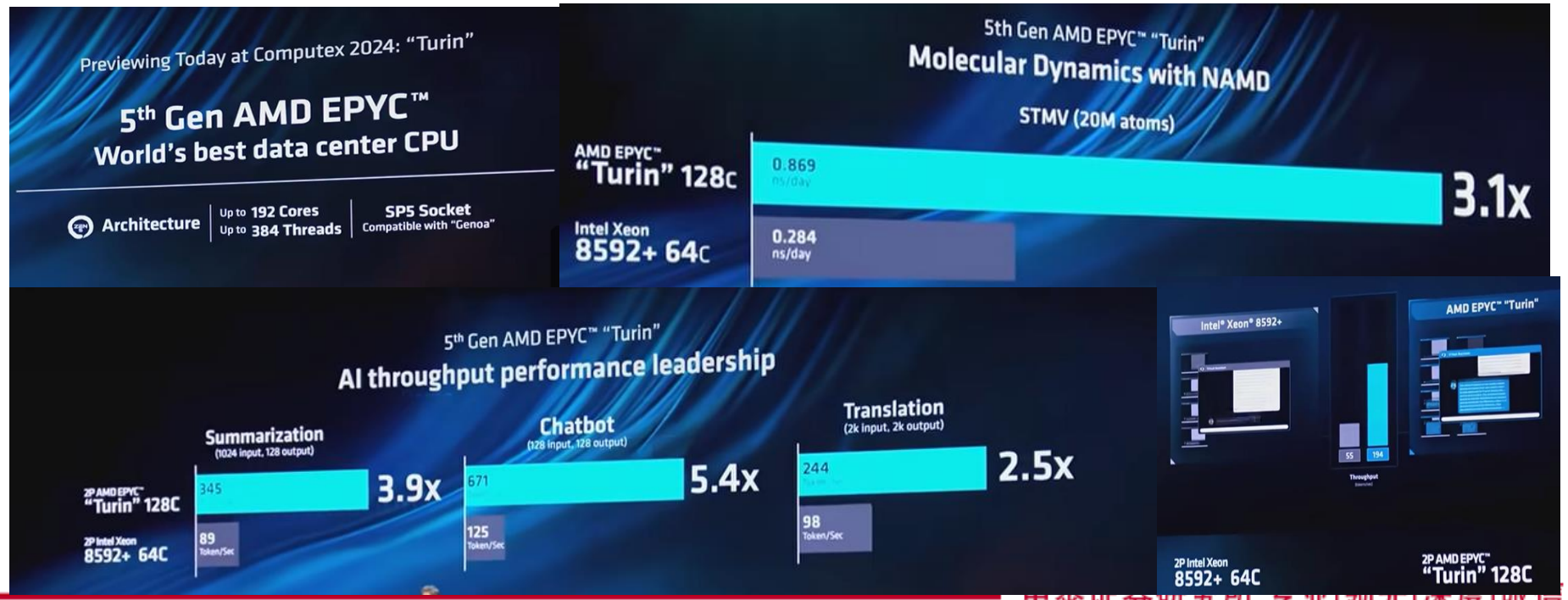
云计算CPU产品—AMD Epic

- **第四代EPYC的优点：**如今许多企业都在寻求对其通用计算基础设施进行现代化改造，并添加新的AI功能，而且通常是在相同的空间内。通过使用第四代EPYC更新他们的数据中心，可以实现这些目标。实际上，你可以用一台服务器替换五台旧服务器，从而减少80%的机架空间并减少65%的能源能耗。
- 现在，许多企业客户也希望在不添加GPU的情况下运行通用和AI工作负载的组合。而EPYC是最佳选择。EPYC在运行行业标准TPCx-AI基准测试时速度提高了1.7倍。

来源：AMD官网，中泰证券研究所

云计算CPU产品—AMD Epic

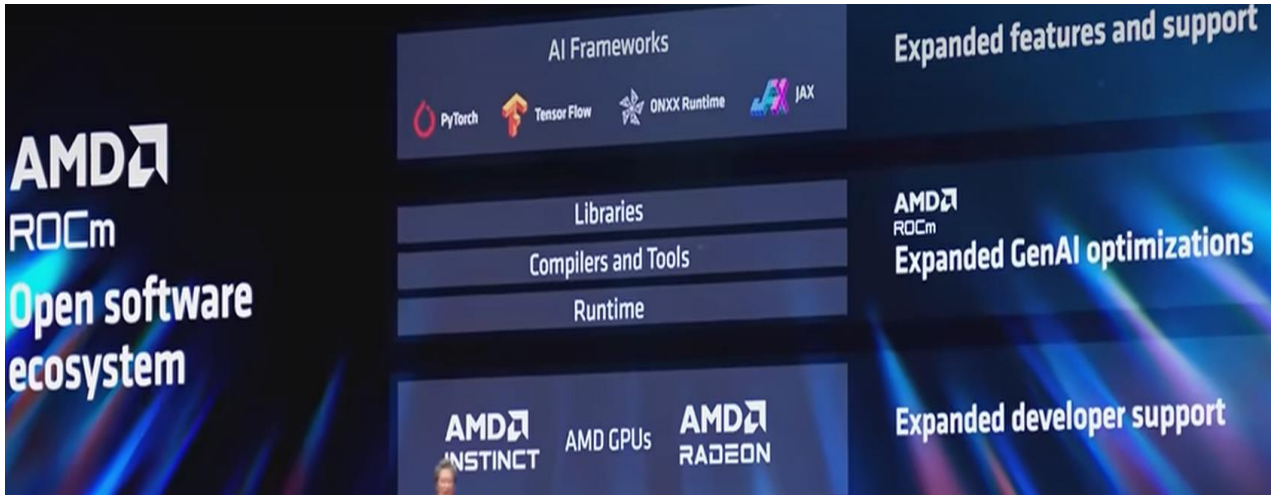
- **AMD下半年将推出的第五代EPYC处理器，代号Turin。**Turin具有192个内核和384个线程，并采用3纳米和6纳米工艺技术，内置了13种不同的芯片。它支持所有最新的内存和I/O。Turin将扩大EPYC在通用和高性能计算工作负载方面的领导地位。计划今年下半年推出Turin。
  - 在模拟2000万个原子模型时，128核版本的Turin比竞争对手的最佳版本快三倍以上。
  - Turin在运行较小的大型语言模型时也表现出色，具有出色的AI推理性能。例如：两台服务器都加载多个Llama2实例，并要求每个助手总结和上传文档。Turin服务器在相同时间内增加了两倍的会话数量，同时响应用户请求的速度明显快于竞品。而另一台服务器达到最大会话数量时，它很快停止并且基本上无法再支持延迟要求。而Turin继续扩展并提供每秒近四倍Token的持续吞吐量。这意味着当用户使用Turin时，需要更少的硬件来完成相同的工作。
  - 此外，在其他方面也有提升，翻译大型文档时的性能提高了2.5倍，运行支持聊天机器人时的性能提高了5倍以上等。

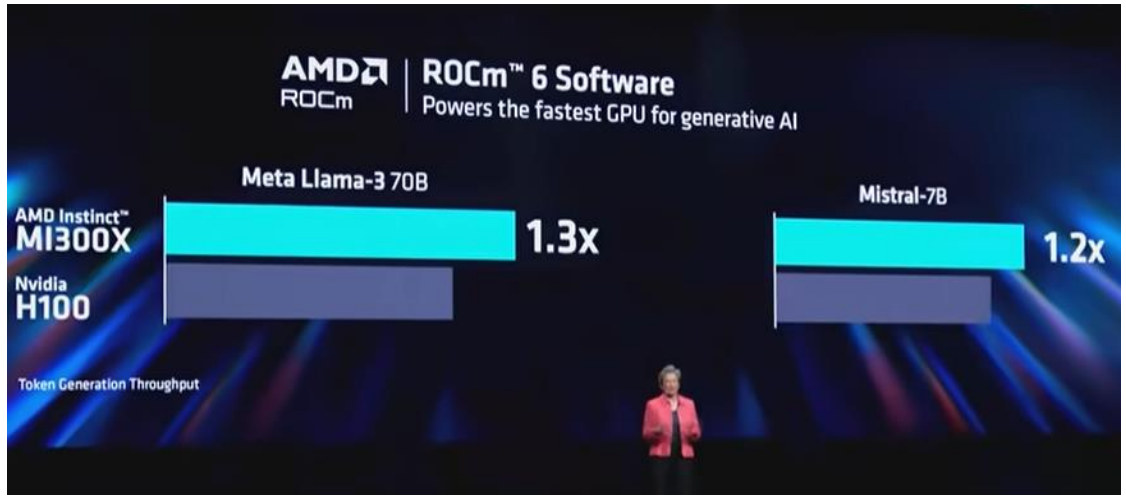


来源：AMD官网，中泰证券研究所

AI加速芯片—MI300系列产品矩阵

- **GPU以及Instinct加速器：**AMD于去年12月推出MI300，它迅速成为AMD历史上增长最快的产品，Microsoft、Meta和Oracle都采用了MI300。去年，AMD在ROCm软件堆栈上取得了巨大进展，与堆栈每一层的开源社区密切合作，同时添加了新功能，使客户能够非常轻松地在其软件环境中部署AMD Instinct。在过去的六个月中，AMD增加了对更多AMD AI硬件和操作系统的支持，公司集成了VLLM等开源库和JAXX等框架，采用了最先进的注意力算法的支持，改进了计算和通信库，所有这些为MI300新一代AI性能的显著提升做出了贡献。
- **MI300X比同行业的竞争对手有更好的推理性能：**与H100相比，公司在Meta最新的Llama370B模型上的性能提高了1.3倍，在Mistral的7B模型上的性能提高了1.2倍。

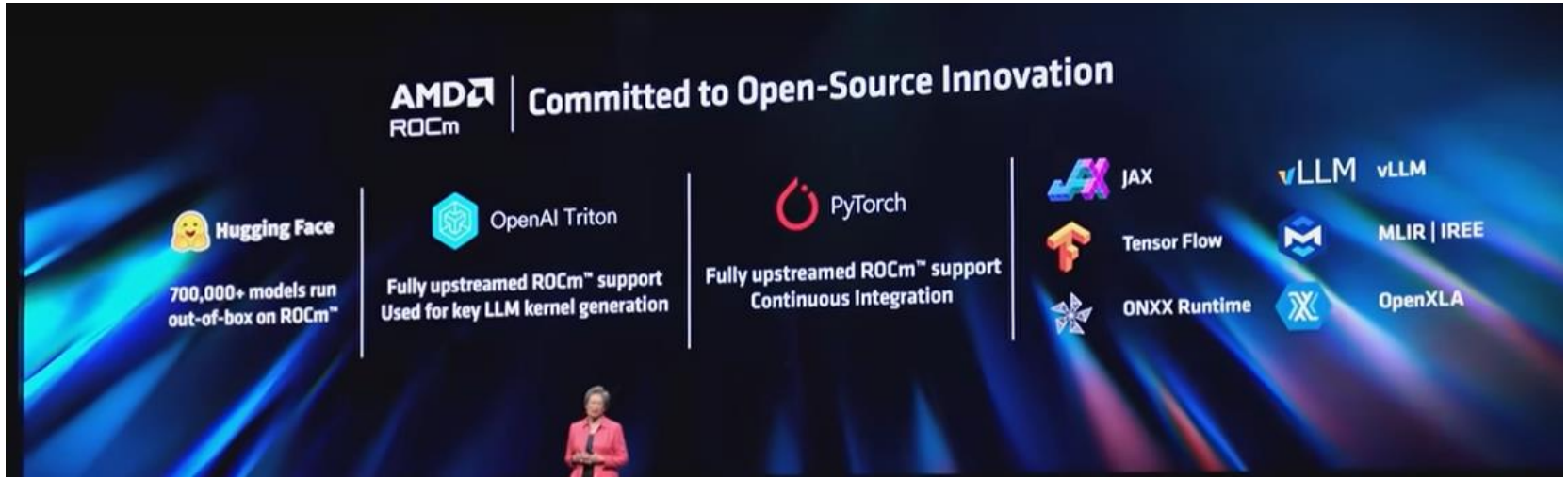




来源：AMD官网，中泰证券研究所

AI加速芯片—MI300系列产品矩阵

- **AMD扩大开源AI社区的合作。**目前超过700,000个Hugging Face模型使用MI300x上的ROCKm“开箱即用”。
- **此外，AMD在工作上和合作伙伴也取得了进展。**如AMD与OPEN AI的密切合作确保了对Triton的MI300X的全面支持，可以快速开发高性能的LLM内核。公司将会继续取得出色的进展，将对AMD AI硬件的支持添加到PyTorch、TensorFlow和JAX等领先框架中。现在公司正在与领先的AI开发人员密切合作，以优化MI300x模型。



来源：AMD官网，中泰证券研究所

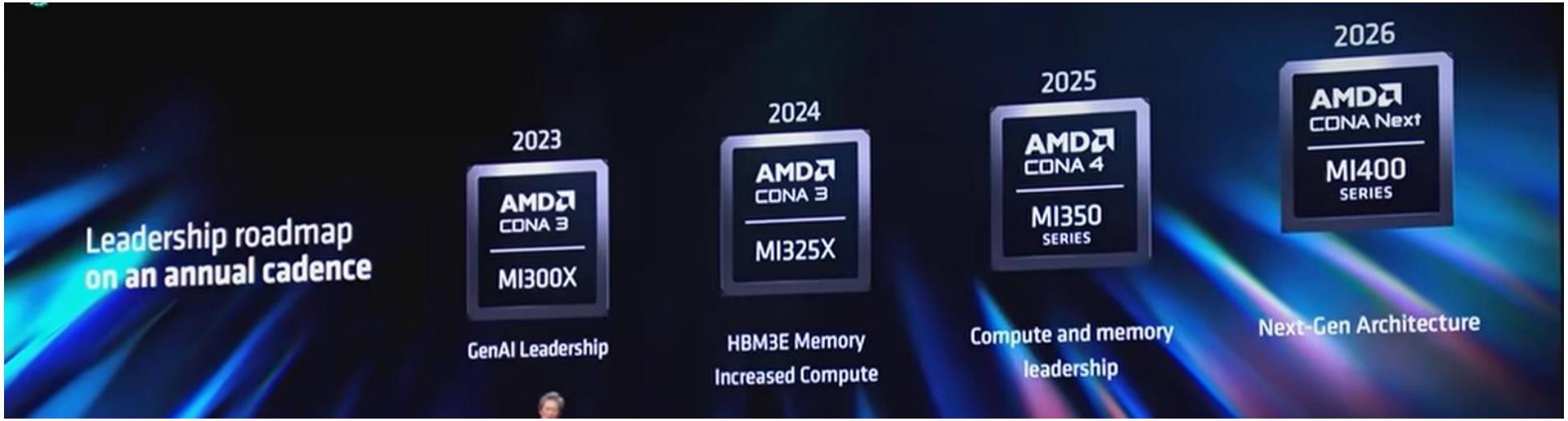
AI加速芯片—MI300系列产品矩阵

- **Stability AI的CTO和联合首席执行官：**预计6月12日，发布Stable Diffusion 3（SD3）中型模型供所有人下载。这个SD3中型模型SD3的优化版本，实现了前所未有的视觉质量，社区将能够根据自己的特定需求进行改进，帮助AMD发现生成式AI的下一个前沿。它在MI300上运行得非常快，也足够紧凑，而且可以在刚刚发布的Ryzen AI笔记本电脑上运行。
  - MI300通常是帮助SD3更快、更有效地训练更大模型的首要因素。例如：API中有一个创造性的放大特性，它基本上可以拿一个不到1兆像素的旧照片和旧图像，放大分辨率并同时提高质量。

来源：AMD官网，中泰证券研究所

AI加速芯片—MI300系列产品矩阵

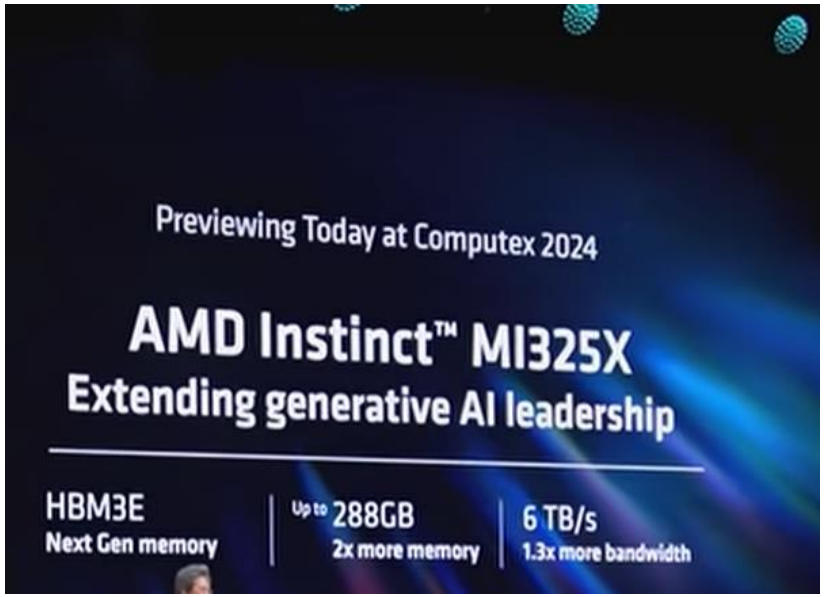
■ **产品更新时间缩短：**苏姿丰称，23年AMD推出推理性能、内存大小和计算能力处于业界领先水平的MI300X；今年晚些时候，计划推出速度更快、内存更大的MI325X；25年推出新的cDNA4架构的MI350系列，MI325和350系列都将采用与MI300相同的行业标准通用基板OCP服务器设计，这意味AMD的客户可以非常快速地采用这项新技术。26年，将推出带有全新cDNA架构的MI400系列。



来源：AMD官网，中泰证券研究所

AI加速芯片—MI300系列产品矩阵

- 从MI325开始，MI325扩展了AMD在生成AI方面的领导地位，拥有高达288GB的超快HBM3E内存和每秒6TB的内存带宽，它使用与M1300相同的基础设施，这使客户可以轻松过渡。
- 与竞争对手相比，MI325提供两倍的内存、1.3倍的内存带宽和1.3倍的峰值计算性能。基于这种性能，一台配备八个MI325加速器的服务器可以运行高达1万亿参数的高级模型，这是H200服务器支持的两倍大小。

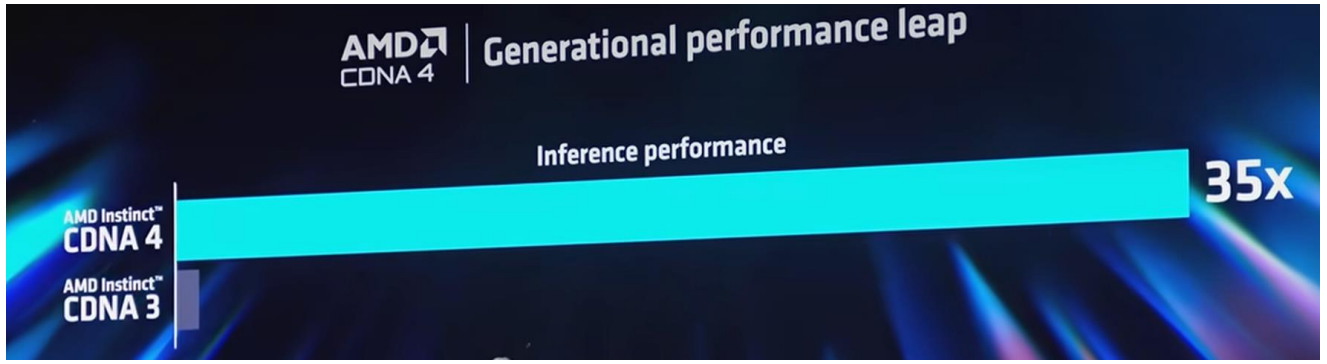


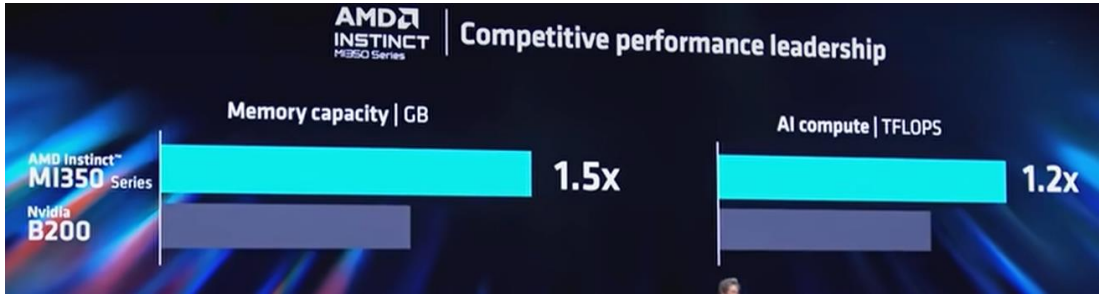
	MI325X	Advantage vs. H200
Memory	288 GB HBM3E	2x
Memory Bandwidth	6 TB/s	1.3x
Peak Theoretical FP16	1.3 PF	1.3x
Peak Theoretical FP8	2.6 PF	1.3x
Model Size	1 trillion	2x

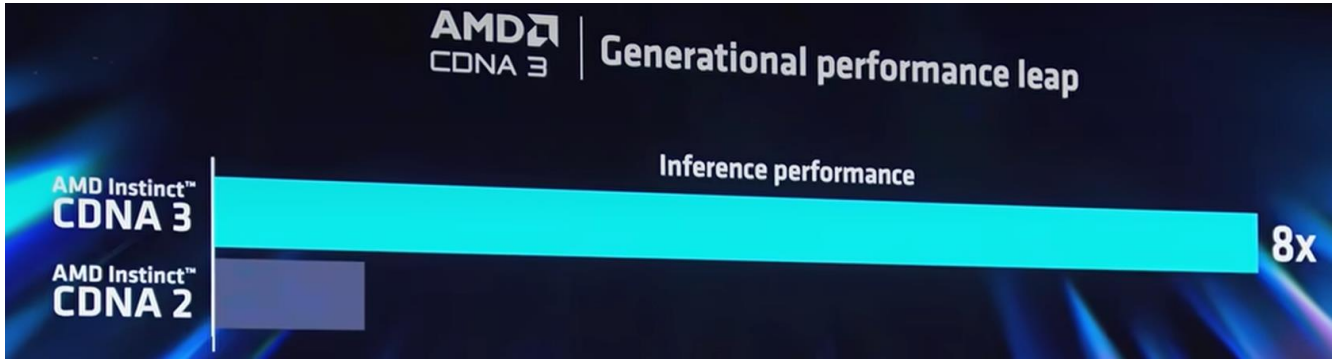
来源：AMD官网，中泰证券研究所

AI加速芯片—MI300系列产品矩阵

- 2025年，AMD将推出cDNA4架构，这将实现其历史上AI性能最大的代际飞跃。MI350系列将采用先进的3纳米工艺技术制造，支持FP4和FP6数据类型，并将再次采用与MI300和MI325相同的基础架构。当AMD推出cDNA3时，AI性能比上一代高出八倍。而借助cDNA4，有望实现35倍的提升。MI350系列与B200 Instinct进行比较时，它支持高达1.5倍的内存，并提供1.2倍的整体性能。







来源：AMD官网，中泰证券研究所

AI加速芯片—MI300系列产品矩阵

- **AMD在推动高性能AI网络基础设施系统的发展方面也取得了重大进展。**AI网络结构需要支持快速切换速率和极低延迟，并且必须可扩展以连接数千个加速器节点。AMD相信AI网络的未来必须是开放的，开放才能让业内每个人都能共同创新并推动最佳解决方案。对于推理和训练而言，扩展数百个加速器的性能实际上至关重要，将机架或吊舱中的GPU与速度极快、高度弹性的互连连接起来，以便它们可以作为单个计算节点运行最大的模型并实现最快的响应。
- 上周宣布计划开发一种高性能结构的开放标准，可以正式连接数百个加速器，称之为UA-Link，它是一种优化的负载存储结构，旨在以高数据速率运行，并利用AMD成熟的Infinity Fabric技术。我们确实相信UA-Link将成为扩展所有类型加速器（不仅是GPU）的最佳解决方案，并且成为专有选项的绝佳替代方案，UA-Link 1.0标准，AMD将在今年晚些时候推出，多家供应商已经开发出支特UA-Link的芯片。

来源：AMD官网，中泰证券研究所

投资建议

- AI依然是2024年最强赛道，重点关注AI算力及端侧产业链机遇：
- （1）AI PC：华勤技术、联想集团、奥海科技、飞荣达、春秋电子等；
- （2）苹果AI链：立讯精密，鹏鼎控股，水晶光电，歌尔股份，长盈精密，长电科技，领益智造，赛腾股份等；
- （3）算力链：沪电股份、深南电路、工业富联、寒武纪、通富微电、香农芯创等。

风险提示

- 行业需求不及预期的风险：若包括手机、PC、可穿戴等终端产品需求不及预期，则产业链相关公司的业绩增长可能不及预期。
- 大陆厂商技术进步不及预期、中美贸易摩擦加剧、研报使用的信息更新不及时的风险、报告中各行业相关业绩增速测算未剔除负值影响，计算结果存在与实际情况偏差的风险、行业数据或因存在主观筛选导致与行业实际情况存在偏差风险。

重要声明

- **中泰证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。**
- 本报告基于本公司及其研究人员认为可信的公开资料或实地调研资料，反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。本公司力求但不保证这些信息的准确性和完整性，且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，可能会随时调整。本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。
- 市场有风险，投资需谨慎。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。
- 投资者应注意，在法律允许的情况下，本公司及其本公司的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。本公司及其本公司的关联机构或个人可能在本报告公开发布之前已经使用或了解其中的信息。
-