



本报记者 许子皓

AMD 董事长兼 CEO 苏姿丰在美国当地时间 10 月 10 日于旧金山举行的 Advancing AI 活动上带来了一整套 AI “撒手锏”，给整个业界带来了全新的 AI 震撼。其中，最大的惊喜正是 AMD 最新发布的旗舰 AI 芯片——AMD Instinct MI325X GPU（以下简称“MI325X”），苏姿丰也是再次信心满满地掏出了与老对手英伟达当下热销产品 H200 的对比图，她表示，MI325X 的内存容量是 H200 的 1.8 倍，内存带宽、FP16 和 FP8 峰值理论算力都达到了 H200 的 1.3 倍。

此外，苏姿丰还发布了第五代 EPYC 服务器 CPU，并称其为“面向云计算、企业级和 AI 的全球最好 CPU”，与老对手英特尔的第五代至强铂金 8592+ 处理器相比，这款 CPU 的 HPC（高性能计算）性能提高多达 3.9 倍，基于 CPU 的 AI 加速提高多达 3.8 倍。苏姿丰此次可谓是将满满的压力给到了英伟达和英特尔。

HBM3e 内存容量是 GPU 提升的关键？

当前，英伟达毫无疑问是数据中心 GPU 领域的霸主，TechInsights 公布的数据显示，2023 年全球数据中心 GPU 总出货量达到了 385 万颗，其中，英伟达以 98% 的市场份额稳居第一，AMD 虽位居第二，却仍然相差十分悬殊。但熟悉苏姿丰的人都知道，她受命于 AMD 危难之际，最不怕的就是打硬仗，此次带来的 MI325X 就是为了“硬刚”英伟达的 H200 而来。

据了解，MI325X 采用了与上一代 MI300X 相同的 CDNA 3 架构，最大的提升就是首次采用了业界目前最先进的 HBM3e，总容量达 256GB，基于 16-Hi 堆栈制程，内存带宽高达 6TB/s，单颗芯片总共拥有 1530 亿个晶体管。

专家告诉记者，HBM 内存提高，对于 GPU 来说，能够带来数据传输速度的提升、数据处理能力的增强、能效比的提高、空间

与延迟的节省、对更复杂应用场景的支持，以及系统稳定性与可靠性的提升等多方面的的好处。因此，MI325X 的 FP8 性能达到了 2.6PFLOP，FP16 性能也达到了 1.3PFLOP。

横向对比，英伟达 H200 采用的 HBM3e 容量为 141GB，带宽为 4.8TB/s，总晶体管数量约为 800 亿个，相比于 MI325X 肯定是不够看了，哪怕是英伟达最新推出的 B200，采用的 HBM3e 内存容量也仅为 192GB，与 MI325X 的 256GB HBM3e 内存容量仍存在差距。

虽然，两家对于 GPU 的设计各有不同，HBM3e 内存容量也并不是评判 GPU 性能的唯一标准，但对产品性能的上限有明显提升，这也是 AMD 直面英伟达的底气。苏姿丰很自豪地在发布会上强调：“MI325 服务器平台在运行 Llama 3.1 时，能提供比英伟达 H200 HGX 高出多达 40% 的性能。”

最新 CPU 性能已经赶超英特尔？

CPU 业务对于 AMD 来说是涅槃重生的开始，在 2018 年，其 EPYC 服务器的市占率仅为 2%，经过四代的升级，终于在 2024 年第一季度达到了 34%，从英特尔手里抢走了近 30% 的 CPU 服务器市场份额，也成了 AMD 的主要营收来源。

苏姿丰在发布今年第二季度财报时就表示：“我们在第二季度实现了强劲的营业额与收入增长，这得益于数据中心事业部创纪录的营业额。我们的 AI 业务继续加速攀升，在 Instinct 系列、EPYC 和 Ryzen 处理器的市场需求带动下，AMD 为下半年营业额的强劲增长已做好了充分准备。”

所以，AMD 此次的另一大撒手锏就是其第五代 EPYC 服务器 CPU，代号“Turin”。据介绍，Turin 拥有 1500 亿晶体管，采

用台积电 3/4nm 制程、全新“Zen 5”及“Zen 5c”核心兼容广泛部署的 SP5 平台，最多支持 192 核、384 个线程。当前，英特尔能效核版至强 6 虽然能在未来通过双芯封装做到 288 核心，但无法支持超线程技术，只有 288 个线程。因此，在核心数和线程数上，AMD 已经赶超了英特尔。

在技术全面升级的情况下，Turin 的性能得到了明显提升。AMD 表示，相比于上一代的英特尔 Xeon 服务器，Turin 在 SPEC CPU 测试中性能提升 2.7 倍，企业级性能最高提升 4.0 倍，HPC 性能最高提升 3.9 倍。

苏姿丰打了很形象的比方，她表示，如果用 Turin 服务器替代上一代的至强服务器，只需 131 台服务器就能达到原来 1000 台的性能水平，节省了 87% 的占地空间。同

MI325X 采用了与上一代 MI300X 相同的 CDNA 3 架构，最大的提升就是首次采用了业界目前最先进的 HBM3e。

根据 AMD 官方信息，MI325X 加速器目前有望在 2024 年第四季度投入生产，预计从 2025 年第一季度开始发货。

此外，根据当前“一年一迭代”的节奏，AMD 此次更新了其 AI 芯片的路线图，预计下一代 MI350 系列将在 2025 年上市，将采用 3nm 制程，并把 HBM3E 内存再提升至 288GB，新一代 CDNA 4 架构也将安排上，使得其推理性能比基于 CDNA 3 架构的加速器提高了 35 倍，该系列产品 AMD 计划在 2025 年下半年上市。

其中，MI350 系列的代表产品 MI355X 是 AMD 专门用来应对英伟达在 3 月刚刚发布的 Blackwell B200 的。据了解，MI355X 的 FP8 和 FP16 性能相比 MI325X 提升了 80%，FP16 峰值性能达到 2.3PFLOPS，FP8 峰值性能达到 4.6PFLOPS，FP6 和 FP4 峰值性能达到 9.2PFLOPS。

除了 GPU 和 CPU，苏姿丰此次还带来了最新的 DPU、AI 网卡以及 ROCm 6.2 生态系统，凑齐了“AI 全家桶”。

时，功耗可以节省最多 68%。

AMD 还优化了 Turin 在 AI 工作流程中的关键动作，包括数据预处理、内存复制、内核启动和任务协调等，这些优化使得 CPU 在处理 GPU 协调任务时更高效，比前代产品快了 28%。

除了 GPU 和 CPU，苏姿丰此次还带来了最新的 DPU、AI 网卡以及 ROCm 6.2 生态系统，凑齐了一整套“AI 全家桶”，给“双英”带来了十足的底气。但当天的 AMD 股价并没有买账，反而以大跌收场，日内跌幅一度扩大到约 5.3%，最终收跌 4%，创 9 月 3 日以来最大盘中收盘跌幅。可见，AMD 的新品虽然来势凶猛，可只有真正量产，经过客户检验，才能赢得市场的认可。

升级是一笔庞大的开支，但杨奕东始终认为，这是一笔划算的投资。“5G-A 网络真正实现了工厂内无线网络对有线网络的替换，让设备脱离笨重的固定线缆、实现‘即插即用’，极大地提高了产线变换的灵活性。5G-A 规模应用后，我们产线的换线时间已经缩短至 30 分钟以内，从产品上料到最后的成品下线最快仅需 15 分钟，整体生产效率提升约 45%，真正实现了‘质的提升’。”杨奕东表示。

聚焦 5G-A 在生产中发挥的关键作用，中国移动四川成都新部分公司副总经理欧一磊侃侃而谈：“相较于前代 5G 网

半导体行业 第三季度利润预增显著

本报讯 记者姬晓婷报道：近期，A 股市场三季报业绩预告密集披露，已有 115 家 A 股公司发布了 2024 年前三季度业绩预告，96 家公司业绩预喜（包括 63 家预增、3 家续盈、11 家扭亏、19 家略增），占比超过八成。其中半导体行业业绩增长显著，在前三季度预计归母净利润增幅位居前 10 位的公司中，半导体与半导体生产设备行业有 5 家，占据了“半壁江山”。

除了全志科技、晶合集成的净利润同比增幅目前暂时位居第二名和第三名之外，还有韦尔股份、思特威和瑞芯微，净利润同比增幅分别位列第六位、第七位和第九位，这三家公司预计归母净利润分别为 22.67 亿元至 24.67 亿元、2.52 亿元至 2.92 亿元和 3.4 亿元至 3.6 亿元，同比增幅分别为 515.35% 至 569.64%、485.41% 至 546.51% 和 339.75% 至 365.62%。

关于业绩增长原因，全志科技在业绩预告中表示，报告期内，公司把握下游市场需求回暖的机会，积极拓展各产品线业务，出货量提升使得营业收入同比增长约 50%，营业收入的增长带动了净利润的增长。晶合集成称，随着行业景气度逐渐回升，公司自今年 3 月起产能持续处于满载状态，并于今年 6 月起对部分产品代工价格进行调整，助益公司营业收入和产品毛利水平稳步提升。

韦尔股份在公告中称，报告期内，市场需求持续复苏，下游客户需求有所增长，伴随着公司在高端智能手机市场的产品导入及汽车市场自动驾驶应用的持续渗透，公司的营业收入和毛利率实现了显著增长；此外，为更好地应对产业波动的影响，公司积极推进产品结构优化及供应链结构优化，公司产品毛利率逐步恢复，整体业绩显著提升。

瑞芯微发布的业绩预告显示，依托公司在 AIoT 的持续增长，尤其是在汽车电子、工业、行业类、消费类市场的提升，叠加第三季度是行业传统旺季，公司在今年前三季度实现营业收入和净利润同比大幅增长。其中第三季度营业收入约 9.11 亿元，同比增长约 51.42%，环比增长约 29.18%，创历史单季度新高；第三季度净利润约 1.57 亿元至 1.77 亿元，同比增长约 199.39% 至 237.47%，环比增长约 36.57% 至 53.95%，均实现突破性增长。

中信证券研究分析指出，结合当前半导体行业销售额、咨询机构的预测情况，以及国内外半导体厂商的表现及展望，半导体周期在前期云端算力及存储涨价的因素拉动下显著修复，目前处于温和复苏状态，预计后续主要的成长动力为云端算力的景气度持续和端侧 AI 的爆发。

记者注意到，重视研发投入、促进芯片高阶化转型是上述半导体预喜企业的共性。

全志科技表示，报告期内，公司为满足客户持续增长的产品及服务需求，加大在芯片新产品开发及智能车载、扫地机器人等新兴应用领域方案的研发投入，研发费用同比增长约 10%。

晶合集成表示，公司高度重视研发体系建设，持续增加研发投入。目前 55nm 中高阶单芯片及堆栈式 CIS 芯片工艺平台已大批量生产，40nm 高压 OLED 芯片工艺平台已实现小批量生产，28nm 逻辑芯片通过功能性验证，28nm OLED 驱动芯片预计将于 2025 年上半年批量量产。公司将加强与战略客户的合作，加快推进 OLED 产品的量产和 CIS 等高阶产品的开发。

第四季度 NAND Flash 合约价 将下调 3% 至 8%

本报讯 TrendForce 集邦咨询最新数据显示，NAND Flash 产品受 2024 年下半年旺季不旺的影响，晶圆合约价于第三季度率先下跌，预期第四季度跌幅将扩大至 10% 以上。模组产品部分，除了 Enterprise SSD 因订单动能支撑，有望于第四季度小涨 0% 至 5% 外，PC SSD 及 UFS 因买家的终端产品销售不如预期，采购策略更加保守。TrendForce 集邦咨询预估，第四季度 NAND Flash 产品整体合约价将出现季减 3% 至 8% 的情况。

从需求角度分析，即使厂商积极推出 AI PC，但由于通胀和 AI 实用性不足等因素，未出现明显换机潮。供给部分，多家原厂的稼动率于第三季度恢复满载，加上其他供应商推动制程升级，产能小幅增加。然而，除了服务器端需求稳定，消费市场的疲软则难以支撑涨价。现货和渠道市场价格与 OEM 合约价的差距扩大，也导致原厂调价受阻。据此，TrendForce 集邦咨询预估 PC SSD 合约价第四季度将季减 5% 至 10%。

因部分企业级客户延迟购置 AI 服务器，第四季度来自服务器 OEM 的订单量明显下调，加上 CSP 采购高峰已过，整体采购容量较第三季度下滑。此外，智能手机和笔记本电脑客户因采取去库存策略，NAND Flash 订单保守，但在原厂持续增产的情况下，导致供过于求。

TrendForce 集邦咨询表示，由于 Enterprise SSD 订单动能及单价优于其他 NAND Flash 产品，供应商积极抢单并提高出货量，这种策略将抑制价格增长。因

此，预计第四季度 Enterprise SSD 合约价将大幅收敛，仅季增 0% 至 5%。

此外，由于带动需求的智能手机第三季度市况未好转，加上手机厂商积极去 eMMC 库存和技术性抵制涨价，因此 eMMC 未出现明显交易量。第四季度部分品牌推出新机，以及 iPhone 16 系列和华为三折叠机发布，看似为 eMMC 市场注入新动能，但买方为避免再有库存过高压力，将采取更谨慎的备货策略。TrendForce 集邦咨询表示，经过第三季度买卖双方价格僵持，原厂库存增加，模组厂和现货市场货源充足，议价天平倾向买方，预估第四季度合约价将季减 8% 至 13%。

UFS 主要应用在高端和旗舰智能手机，市况与 eMMC 相同。整体经济增长趋缓导致换新机的频率从不到两年延长至三年，以及尚未有智能手机的杀手级应用出现，预估第四季度需求不会明显改变。UFS 产品目前有原厂和模组厂竞争供货，因市场需求减弱，原厂为避免堆积库存和达成业绩目标，第四季度必须在价格上让步，预估合约价将季减 8% 至 13%。

2024 年以来，零售端的 Client SSD、闪存卡和 U 盘需求低迷，欧美地区返校季和节庆效应不佳，第四季度 NAND Flash 晶圆需求恐进一步减少。

TrendForce 集邦咨询指出，在模组厂库存过高和部分原厂削价竞争的态势下，第四季度 NAND Flash 晶圆合约价将出现较大幅度衰退，预估季减 10% 至 15%。（文 编）

（上接第 1 版）基于此，工业和信息化部明确，要强化 5G-A、6G 等关键技术突破，强化人工智能、量子通信等前沿技术攻关，强化创新体系和生态构建，使信息通信领域原创性、颠覆性技术创新成果竞相涌现。

记者观察到，5G-A 在制造场景中的应用已经初见萌芽。

“未来工厂的发展方向是逻辑离散化、调度集中化、网络柔性化。”面向未来制造业的发展方向，上海勃傲自动化系统有限公司总经理何流提出上述预测。这一发展趋势在如今的 3C 行业中得到了充分体现。作为补充，微网优联 5G 工厂解决方案负责人杨奕东分享了他基于 3C 行业现状的思考：“随着网络技术不断更迭、应用行业不断深化、市场需求不断细分，网通行业的市场需求由单一大规模的刚性生产向小批量、定制化的柔性生产转变已成大势所趋。”

面临着这样的市场发展趋势，轻量化、低时延、高可靠，并有多种自适应工业总线

协议加持的 5G-A 网络成为了制造业在时代发展下的必然选择。而作为一家以 Wi-Fi 技术深度开发为主营业务，以网关、模组等 3C 产品为核心产品的企业，微网优联对相关网络技术的应用表现出了极高的敏感度。“早在成都厂区的土建阶段，微网优联就按 5G+ 工业互联网的标准进行了提前规划，直接建设 5G 智慧工厂。今年是 5G-A 商用元年，我们紧随技术发展步伐，对 RedCap 数据网关与制造设备进行全面升级，现已实现 5G-A 网络在厂房内的规模应用。”杨奕东说道。

诚然，5G-A 基站与网关设备的更新

5G-A 智慧工厂落地开花

升级是一笔庞大的开支，但杨奕东始终认为，这是一笔划算的投资。“5G-A 网络真正实现了工厂内无线网络对有线网络的替换，让设备脱离笨重的固定线缆、实现‘即插即用’，极大地提高了产线变换的灵活性。5G-A 规模应用后，我们产线的换线时间已经缩短至 30 分钟以内，从产品上料到最后的成品下线最快仅需 15 分钟，整体生产效率提升约 45%，真正实现了‘质的提升’。”杨奕东表示。

聚焦 5G-A 在生产中发挥的关键作用，中国移动四川成都新部分公司副总经理欧一磊侃侃而谈：“相较于前代 5G 网

背后的默默努力。纵观全局，在推动 5G-A 深入工业制造业的征程上，电信运营商可谓不遗余力、硕果累累：2023 年 2 月，中国联通联合中兴通讯发布了 5G-A 确定性工业互联网系列技术创新实践成果；今年 5 月，中国电信在紫江新材料公司完成了上海首个无源物联试点验证；今年 7 月，中国移动研究院联合江苏移动、立讯精密和华为公司，在立讯精密工业生产核心环节完成业界首例端到端“双频 5G-A 确定性网络”试点；今年 10 月，基于 5G-A 网络、面向汽车喷涂焊装产线的工位样机纷纷登场……

3C 行业只是 5G-A 网络走向千行百业的一个落脚点。在大会中国移动展台前，一位现场工作人员向记者表示：“在 5G-A 网络的加持下，未来数智工厂的控制架构一定是矩阵式、模块化、柔性化的，工业控制设备永远在线、工艺改进和流程对接变得更加灵活……种种技术构想，都将借由 5G-A 网络成为现实。”