

超威半导体 (AMD)

算力龙头产品矩阵完善，数据中心助推成长

买入 (首次)

2023年08月10日

证券分析师 张良卫

执业证书: S0600516070001

021-60199793

zhanglw@dwzq.com.cn

证券分析师 卞学清

执业证书: S0600523070002

bianxq@dwzq.com.cn

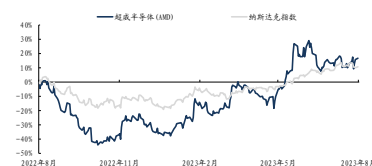
| 盈利预测与估值 | 2022A | 2023E | 2024E | 2025E |
|---------------------|--------|--------|--------|--------|
| 营业总收入 (百万美元) | 23,601 | 23,105 | 32,179 | 35,840 |
| 同比 | 44% | -2% | 39% | 11% |
| Non-GAAP 净利润 (百万美元) | 5504 | 4869 | 7822 | 9226 |
| 同比 | 60% | -12% | 61% | 18% |
| 每股收益 (美元/股) | 3.40 | 3.04 | 4.95 | 5.92 |
| PE (Non-GAAP) | 32 | 37 | 23 | 19 |

关键词: #新产品、新技术、新客户

投资要点

- **数据中心潜力大，AMD 拥有丰富产品组合。**云计算和人工智能推动数据中心发展，云计算 2017 年到 2022 年年均复合增长率为 22%，增长迅速；AI 相关产业 2022 年到 2026 年年均复合增长率将达到 27%。两者都对数据中心和算力有着海量需求，AMD 提前布局，有着最丰富完善的数据中心产品组合，形成了“CPU+GPU+FPGA+DPU”的产品矩阵，迎接 AI、数据中心浪潮。
- **Instinct MI 300X 为大模型而生，AI 生态有望改善，AMD2024 年业绩将迎来爆发。**Instinct MI 300X 集成了 192GB 的 HBM3，5.2TB/s 存储带宽，内存密度是 NVIDIA H100 的 2.4 倍，内存带宽是 H100 的 1.6 倍，一张 MI300X 就能进行有着 400 亿参数的大语言模型的推理。AMD 对应的 ROCm 开放平台也在微软、Meta、Hugging Face 等厂商扶持下不断完善，AMD 也将 ROCm 平台逐渐开放给消费级显卡完善生态，此外，AMD 还收购了 Pensando，获得了 DPU 技术来解决集群问题。所以硬件领先和软件生态的不断完善下，AMD 业绩有望因为 AI 加速显卡迎来爆发。
- **消费级显卡和客户端业务营收随着 PC 市场回暖复苏，消费级显卡或因 AI 生态改善获益。**AMD 消费端业务在 2022 年第二季度后营收大幅下降，受到了因为疫情后 PC 市场库存高位的影响。现今 PC 市场库存问题在 2023 年第三季度后明显好转，AMD 消费端业务营收将恢复。同时，随着 AMD 开放 ROCm 平台至消费级显卡，消费级显卡生产力将逐渐提高，消费级显卡市场占有率将逐渐提高。
- **盈利预测与投资评级：**我们预计公司 2023-2025 年 NON-GAAP 净利润分别为 48.7 亿、78.2 亿、92.3 亿美元，对应现价（8 月 9 日）PE 分别为 37 倍、23 倍、19 倍，考虑到 Instinct MI 300X 加速器和消费级显卡重新带动增长预期，首次覆盖予以“买入”评级。
- **风险提示：**AMD 软件团队规模不足、业务和份额受到英特尔挑战、计算卡集群技术瓶颈、数据中心市场发展受阻

股价走势



市场数据

| | |
|--------------|--------------|
| 收盘价(美元) | 110.47 |
| 一年最低/最高价 | 129.19/55.94 |
| 流通美股市值(百万美元) | 178483 |

基础数据

| | |
|-----------|-------|
| 资产负债率(%) | 19.0% |
| 总股本(百万股) | 1611 |
| 流通股本(百万股) | 1611 |

内容目录

| | |
|--|-----------|
| 1. AMD 历久弥新的半导体芯片设计公司 | 4 |
| 1.1. 半导体行业巨头，四大业务线覆盖面广 | 4 |
| 1.2. 管理层产品布局清晰明确，市值超过英特尔 | 4 |
| 1.3. AMD 历史复盘：几经调整，迎接 AI、数据中心浪潮 | 5 |
| 2. 数据中心需求持续增长，AMD 丰富产品组合潜力较大 | 7 |
| 2.1. 云计算和人工智能推动数据中心发展 | 7 |
| 2.2. EPYC 霄龙处理器：保持竞争力，扩大市场份额 | 8 |
| 2.3. Instinct MI AI 加速器：生态、集群问题有望解决，AMD 业绩迎来爆发 | 10 |
| 2.4. 自适应 SOC 和 AI 引擎：数据中心业务的重要补充 | 13 |
| 2.5. Pensando DPU：补全数据中心最后一块拼图 | 14 |
| 3. 游戏业务：半定制业务短期走弱，游戏显卡有望增长 | 15 |
| 3.1. 半定制业务：AMD 独树一帜，营收短期下降 | 15 |
| 3.2. 游戏显卡业务：营收和市占率不断提升 | 17 |
| 4. 客户端业务：Zen 架构为消费 CPU 基石，远期增速相对平稳 | 20 |
| 4.1. Zen 架构和 3D V-Cache 技术得到市场认可 | 20 |
| 4.2. 客户端营收随 PC 市场恢复，但增长遭遇瓶颈 | 23 |
| 5. 嵌入式业务：覆盖多行业，营收稳定增长 | 23 |
| 5.1. 收购 Xilinx 拓展 AMD 业务范围，边缘侧需求不断增长 | 23 |
| 5.2. AMD 嵌入式业务是信息技术行业的重要参与者 | 24 |
| 6. 盈利预测与投资评级 | 25 |
| 7. 风险提示 | 26 |

图表目录

| | |
|---|----|
| 图 1: AMD 产品线 | 4 |
| 图 2: AMD、Intel 和 NVIDIA 股价走势 | 5 |
| 图 3: AMD 第一个山寨 CPU——Am9080 | 5 |
| 图 4: “推土机”架构 | 6 |
| 图 5: 全球云计算市场规模 (亿美元) | 8 |
| 图 6: 2017-2023E 全球数据中心市场规模 (亿美元) | 8 |
| 图 7: EPYC 霄龙处理器迭代历史 | 9 |
| 图 8: Rome 系列霄龙处理器 | 9 |
| 图 9: EPYC 9654 和 Xeon 8490H 性能对比 | 10 |
| 图 10: AMD 和 Intel 服务器 CPU 份额 | 10 |
| 图 11: CDNA 架构内存模式迭代 | 10 |
| 图 12: Instinct MI300A 性能提升预测 | 10 |
| 图 13: Instinct MI300X 内存性能 | 11 |
| 图 14: 不同大模型推理端所需 GPU 数量 (2023 年, 个) | 11 |
| 图 15: ROCm 平台转译 CUDA 代码 | 12 |
| 图 16: ROCm 和 CUDA 工具链对应图 | 12 |
| 图 17: Xilinx 芯片 | 13 |
| 图 18: VITIS AI 开发环境 | 13 |
| 图 19: AMD AI 产品覆盖规划图 | 14 |
| 图 20: Pensando 产品路线图 | 15 |
| 图 21: Rockstar 游戏发布速度 | 16 |
| 图 22: 3DMark API 开销测试 | 16 |
| 图 23: 全球主机市场规模 (亿美元) | 17 |
| 图 24: PS4 系列历年销量 (百万台) | 17 |
| 图 25: AMD 显卡产品演进 | 18 |
| 图 26: AMD 桌面级独立显卡市占率 | 18 |
| 图 27: AMD 和 NVIDIA 消费级显卡生产力对比 | 19 |
| 图 28: AMD 消费级显卡营收 (百万美元) | 19 |
| 图 29: NVIDIA 和 AMD 消费级显卡性价比对比 | 20 |
| 图 30: AMD Zen 架构路线图 | 21 |
| 图 31: AMD 和 Intel 桌面端 CPU 份额 | 21 |
| 图 32: AMD 和 Intel 移动端 CPU 份额 | 22 |
| 图 33: 3D V-Cache 技术 | 22 |
| 图 34: AMD 客户端业务营收 (百万美元) | 23 |
| 图 35: 全球 FPGA 市场规模 (亿美元) | 23 |
| 图 36: Xilinx 汽车平台 | 24 |
| 图 37: Vivado 设计软件 | 25 |
| 表 1: 公司各业务营收预测 (百万美元) | 25 |
| 表 2: 公司营收利润预测摘要 | 25 |
| 表 3: 可比公司估值 | 26 |

1. AMD 历久弥新的半导体芯片设计公司

1.1. 半导体行业巨头，四大业务线覆盖面广

美国 AMD 半导体公司专门为计算机、通信和消费电子行业设计和制造各种创新的微处理器，公司成立于 1969 年。AMD 的主要业务有客户端业务、游戏业务、数据中心业务和嵌入式业务，主要产品有 CPU 处理器，独立显卡，半定制 SOC，加速计算卡，嵌入式处理器等等。

公司在 2022 年重构了业务组合，将所有产品拆分为了**数据中心业务、游戏业务、客户端业务和嵌入式业务**。数据中心业务中，包含了 EPYC 霄龙服务器处理器，主要面向云计算、企业和高性能计算，还包含了 Instinct MI GPU 加速器，面向高性能计算和人工智能，除此以外，数据中心业务还包括了 Xilinx 中的 AI 部分以及 Pensando 的 DPU 产品；在游戏业务中，包含了为游戏机厂商制作半定制 SOC 的半定制业务，还有消费级 Radeon 系列显卡和为工作站推出的 Radeon Pro 系列显卡；在客户端业务中，包括了面向个人台式机和笔记本的锐龙、速龙处理器，以及为工作站推出的 Threadripper PRO 处理器和锐龙 Pro 处理器；在嵌入式业务中，包含了锐龙和霄龙嵌入式处理器，以及 Xilinx 的 ALVEO、VERSAL 和 ZYNQ 系列产品，主要面向医疗保健，工业，机器人，汽车，计算机视觉等领域用户。

图1: AMD 产品线



数据来源：公司公告，东吴证券研究所

1.2. 管理层产品布局清晰明确，市值超过英特尔

AMD 现任 CEO 为苏姿丰 (Lisa Su)。苏姿丰在 1986 至 1994 在 MIT 完成了本科，硕士和博士的学习，于 1994 年加入德州仪器作为技术专员，1995 年加入 IBM 担任半导体研发中心副总裁，于 2007 年加入飞思卡尔半导体。苏姿丰在 2012 年加入 AMD，从 2014 年至今担任 AMD 首席执行官。

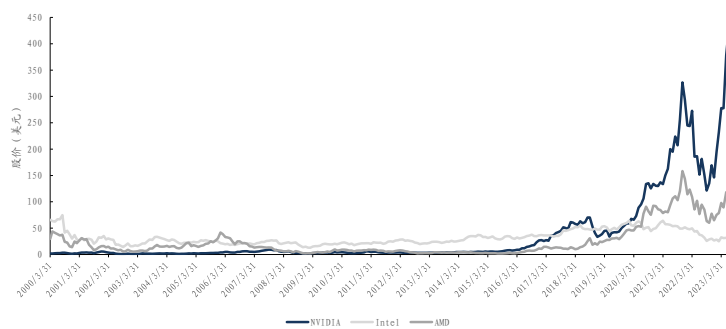
2014 年就任前公司连续三个年度亏损，AMD 简化了研发流程。AMD 还建立了新

的产品周期，即每年推出新一代 GPU、每 1.5 年推出新 CPU 内核。其次，加深与客户的战略关系。AMD 近年和索尼、微软的合作就是绝佳例证。最后，简化业务使其与新策略和重点保持一致。AMD 进行了自上而下的组织结构微调，提高了高管沟通的透明度和频率，并鼓励培养根据包容性的企业文化。

2016 年，发布 Zen 架构，Zen 架构有着质的飞跃，AMD 原本预计其 IPC 性能上比起 Excavator 挖掘机架构可提高 40%，实际上最终以 52% 的提升幅度超越了预期目标。ZEN 架构处理器的出现，一下子打破了英特尔的垄断地位，市值也超过了英特尔。

2022 年，AMD 提出“AI 优先”。2020 年收购 Xilinx（2022 年完成），2022 年收购 Pensando，完成了“GPU+CPU+FPGA+DPU”的产品组合布局，瞄准不断增长的数据中心市场。公司股价走势紧跟 AI 龙头 NVIDIA。

图2：AMD、Intel 和 NVIDIA 股价走势



数据来源：Bloomberg，东吴证券研究所

1.3. AMD 历史复盘：几经调整，迎接 AI、数据中心浪潮

1969~1987 年，AMD 公司成立：初创团队都是来自于仙童半导体。公司成立之初采取的是第二供应商策略，AMD 先后是是仙童半导体、国家半导体等公司的第二替代供货商，在获得原厂芯片设计方案授权的基础上自行制造并销售芯片。在 1976 年 Intel 推出整合指令集代码的 CPU 之后，AMD 和 Intel 进行了合作，成为了 Intel 的第二替代供应商。

图3：AMD 第一个山寨 CPU——Am9080



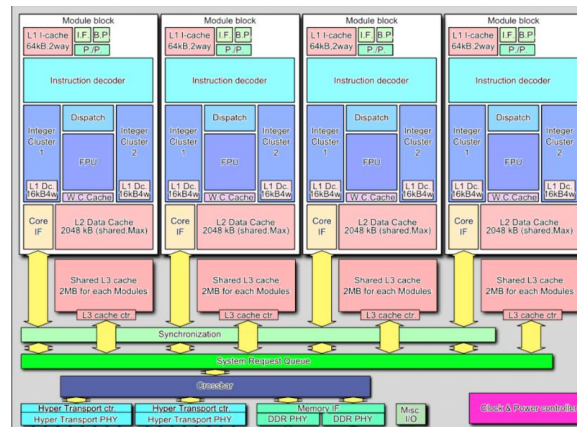
数据来源：公司公告，半导体行业观察，东吴证券研究所

1980~2006 年，AMD 自主创新，迅速发展：在 1980 年后，AMD 不再满足于第二替代供应商的身份，开始自主研发。在 1987 年 Intel 停止授权 386 之后，AMD 营收大

幅下滑，但是 AMD 还是在 1991 年推出了对标了 80386 的 Am386 系列，Am386 系列芯片当年收入达到 1.5 亿美元，销量达到 950 万枚。在此之后，AMD 坚定了自主创新的方向，在 1995 年，AMD 的营收达到了 25 亿美元，利润超过 2 亿美元。1997 年，AMD 推出了 K6 架构的处理器，在 2003 年，AMD 推出了基于 K8 架构的速龙处理器，两者都取得了成功。

2007~2016 年，AMD 收购 ATI，衰退的十年：2006 年，AMD 以 54 亿美元收购了 ATI，成为了第一家可以同时设计高性能 CPU 和 GPU 的半导体公司。**收购 ATI 成为 AMD 走下坡的转折点**，由于高估了 ATI 的商誉估值，AMD 2006-2010 三个财年的营业利润都处在亏损状态。在这期间，Intel 推出了酷睿系列 CPU，取得了成功。而 AMD 在此期间推出了“推土机”架构，这个技术后来被证明是一个失败的设计，所以在这 10 年中，AMD 只能靠低价来维持营收，股价不断走低。

图4：“推土机”架构



数据来源：公司公告，半导体行业观察，东吴证券研究所

2017~2021 年，Zen 架构助 AMD 涅槃重生，游戏业务另辟蹊径：在 2017 年，AMD 发布了 Zen 架构，该架构的 IPC 相比“打桩机”架构提升了 40%，同时，AMD 还选择和 TSMC 合作，享受到了先进制程的红利，其服务器霄龙处理器取得了较大成功。同时，AMD 由于它能够同时设计 CPU 和 GPU 的优势，它为游戏主机推出了高性价比的半定制方案，开拓了业务范围。在 2020 年 AMD 的市值也再次超过了英特尔。

2021 年~至今，AI、数据中心优先，其它业务齐头并进：数据中心产业有着广阔的发展前景。伴随 5G、人工智能、云服务等等业务的快速发展，催生出了大量的数据中心的需求。随着未来人工智能大模型等信息技术的不断发展，对于数据中心的需求也会不断增长。AMD 除了有着的服务器处理器之外，还有为 HPC 和 AI 构建的 Instinct GPU 加速器、收购 Xilinx 获得的 FPGA 和自适应 SoC，以及通过收购 Pensando 获得的 DPU 技术。AMD 有着业内最广泛的数据中心产品组合。

2. 数据中心需求持续增长，AMD 丰富产品组合潜力较大

2.1. 云计算和人工智能推动数据中心发展

#1) 云计算不断增长，AMD 服务器 CPU 为重要参与者

云计算是一种资源和服务共享的方式，在云计算平台，硬件资源、网络、应用环境都可以被随时随地、按需按量地分配调用。云计算是当前数字时代的核心技术，发挥着重要作用，可实现各类复杂的业务的协调发展。

根据中国信通院数据，全球云计算市场规模增长迅速，初步统计 2022 年全球云计算市场规模为 4053 亿美元，2017 年到 2022 年市场规模年均复合增长率为 22%。由于云计算的需求不断增长，云数据中心规模未来也会不断扩大。在云数据中心所使用的服务器 CPU 主要来自于 Intel 和 AMD，2023 年第一季度，AMD 的服务器 CPU 市场占比为 18%，Intel 占比为 82%。

#2) 人工智能迅速发展，数据中心 AI 芯片前景广阔

根据 IDC 报告，包括包括软件、硬件、以 AI 为中心的系统服务在内，AI 相关产业规模支出在 2022 年达到 1212 亿美元，IDC 预测 2023 年将达到 1540 亿美元，同比增长 26.9%。

随着越来越多的产品融合 AI 服务，到 2026 年 AI 相关产业规模支出将超过 3000 亿美元，2022-2026 年的复合年增长率 (CAGR) 将达到 27%。据 Frost&Sullivan 数据显示，2021 年中国 AI+金融服务的市场规模约为 23 亿元人民币，从 2017 年到 2021 年的复合年增长率为 104.3%，未来其市场规模预计将在 2026 年达到约 107 亿元人民币。同时，AI 解决方案在医疗领域的市场规模从 2017 年的约 4 亿人民币增加到 2021 年的 24 亿人民币左右，CAGR 为 54.1%。

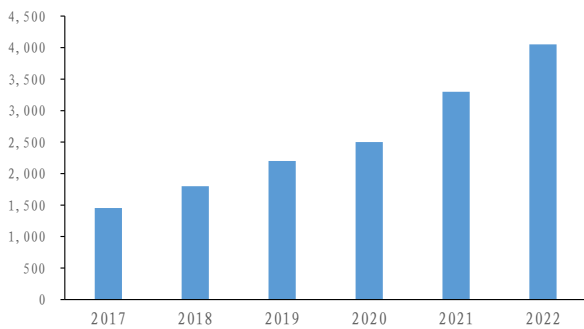
人工智能需要巨量的计算能力。数据中心必须提供强大的计算能力和存储资源，人工智能才能实时地处理大量数据集并进行训练和推理。通过 GPU 和 TPU 等专用硬件，数据中心可以加速复杂的计算，支持人工智能应用程序和工作负载。TrendForce 数据显示，2022 年，配备通用 GPU (GPGPU) 的 AI 服务器仅占全球服务器年出货量的 1%。预计从 2022 年到 2026 年，人工智能服务器的出货量将以 10.8% 的 CAGR 增长。Reportlinker 在其发布的 2023 人工智能芯片报告中提出，全球 AI 芯片市场将从 2022 年的 156.5 亿美元增长到 2023 年的 232.9 亿美元，复合年增长率 (CAGR) 为 48.8%。预计 2027 年人工智能芯片将增长到 888.5 亿美元，2023-2027 年 CAGR 为 39.8%。现在，AI 芯片市场的主要参与者包括 NVIDIA、Intel、AMD、Alphabet、Mediatek、Qualcomm、NXP 等。但在数据中心市场，竞争的焦点主要集中在 NVIDIA、Intel 和 AMD 之间。

#3) 数据中心为 AMD 首要战略重点，具有丰富产品组合

伴随 5G、人工智能、云服务等业务的快速发展，催生出了大量的数据中心的需求。根据中国信通院的数据，数据中心的规模从 2017 年的 465.5 亿美元增长到 2022 年的 746.5 亿美元，过去 5 年的年均复合增长率为 9.9%。随着未来人工智能大模型等信息技术的不断发展，对于数据中心的需求也会加速增长。

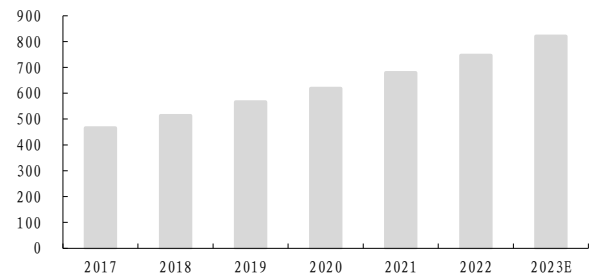
数据中心是 AMD 的首要战略重点。AMD 除了有着领先的 Epyc 处理器之外，还提供了非常完整的产品组合，包括为 HPC 和 AI 构建的 Instinct GPU 加速器、收购 Xilinx 获得的领先 FPGA 和自适应 SoC，以及通过收购 Pensando 获得的领先 DPU。AMD 有着业内最广泛的数据中心产品组合。

图5：全球云计算市场规模（亿美元）



数据来源：中国信通院，东吴证券研究所

图6：2017-2023E 全球数据中心市场规模（亿美元）



数据来源：中国信通院，东吴证券研究所

2.2. EPYC 霄龙处理器：保持竞争力，扩大市场份额

#1) EPYC 霄龙处理器发展历史回顾

AMD 在 2017 年发布了第一代霄龙处理器。第一代 EPYC 处理器代号为“那不勒斯”，还支持 128 条 PCIe 3.0 通道，具有 8 个内存通道。AMD 在 2019 年推出了代号为“罗马”的第二代霄龙处理器，其最高支持 64 核 128 线程，128 条 PCIe 4.0 通道，8 个内存通道；在 2021 年推出了代号为“米兰”的第三代霄龙处理器，其最高支持 64 核 128 线程，128 条 PCIe 4.0 通道，8 个内存通道；在 2022 年推出了代号为“热那亚”的第四代霄龙处理器，其最高支持 128 核 256 线程，128 条 PCIe 5.0 通道，12 个内存通道。同时，AMD 分别在 2022 年和 2023 年对“米兰”和“热那亚”系列处理器加入了 3D V-Cache 技术。

#2) EPYC 霄龙处理器技术性价比领先，份额不断扩大

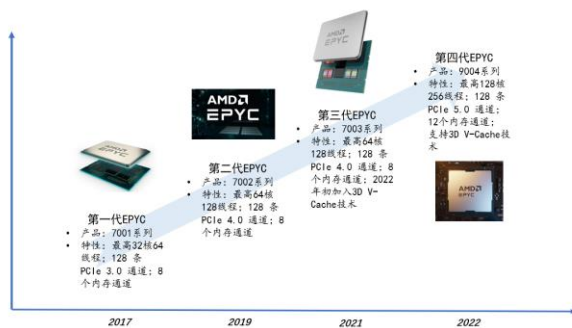
过去几十年来，半导体行业一直按照摩尔定律的规律发展，凭借着芯片制造工艺的迭代，使得每 18 个月芯片性能提升一倍。但是当工艺演进到 5nm，3nm 节点，提升晶体管密度越来越难，同时由于集成度过高，功耗密度越来越大，供电和散热也面临着很大挑战。Chiplet 技术是摩尔定律逐渐放缓情况下，持续提高集成度和芯片算力的重要途径。

使用 Chiplet 技术，可以将大型单片芯片划分为多个相同或者不同的小芯片，这些小芯片可以使用相同或者不同的工艺节点制造，再通过跨芯片互联和封装技术进行封装级别集成，**降低成本的同时获得更高的集成度。**

同时，由于数据中心应用端的算力需求仍在不断增加，chiplet 式的设计也有利于堆算力，AMD 最先进的霄龙处理器实现了 128 个核心 256 个线程。所以，**对于现代数据中心，Chiplet 设计满足了云计算的多核需求，成本更加低，能耗比更加优秀。**

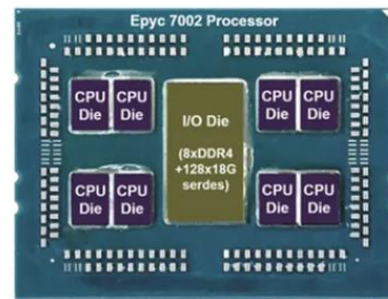
AMD 在 Chiplet 技术上有先发优势，AMD 在 2019 年推出的 Zen2 架构中，就采用了 Chiplet 小芯片设计，使用 8 块 CPU 芯片实现 64 核，是当时英特尔性能最佳处理器的两倍。Rome 设计增加了第九个小芯片，它集中了所有的 DRAM 和 I/O 电路。

图7: EPYC 霄龙处理器迭代历史



数据来源：公司官网，东吴证券研究所

图8: Rome 系列霄龙处理器



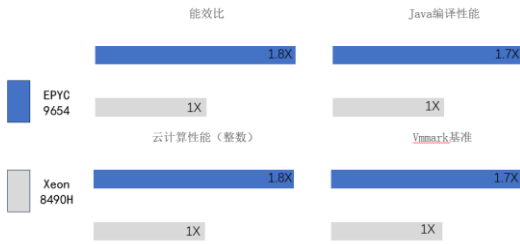
数据来源：公司公告，东吴证券研究所

Intel 在 23 年推出的最新一代至强处理器中也首次使用了 Chiplet，最高支持 60 核/128 线程，包括未来的 14 代酷睿处理器也会使用 Chiplet 技术。但是 AMD 的 EPYC 处理器受益于 AMD 的技术积累和 Zen 架构的特点，**在每瓦性能、核心/线程数量、运营成本都占据了优势。**AMD 发布的 EPYC 9754 可以支持 128 核/256 线程。这意味着第四代 Epyc 的核心密度优势可以使云服务提供商能够**支持超过两倍的服务器实例数量。**

在 AMD 数据中心和 AI 首映式中，AMD 对比了 EPYC 霄龙处理器最新的 EPYC 9654 和 Intel 至强处理器中最新的 Xeon 8490H，AMD 的处理器性能领先。AMD 的处理器能效比比英特尔强 80%，Java 编译性能要强 70%，云计算性能（整数）要强 80%，Vmmark 基准要强 70%。AMD CEO 称，AMD 的 Epyc 处理器在前 10 名最快的超级计算机中的占据了 5 台，包括 Frontier，这是第一台使用惠普企业硬件构建的百亿亿次计算计算机。

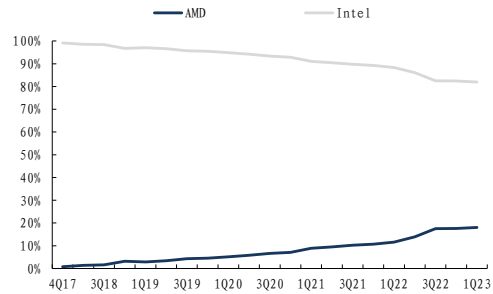
霄龙处理器性能和运营成本的独特优势得到业界的广泛认可。微软 Azure、AWS、谷歌云、HPE 和其它厂商都将 EPYC 处理器运用在了机密计算、通用工作负载、内存密集型工作负载、视觉工作负载等场景。**AMD 的服务器 CPU 份额也不断提高**，从 2017 年第四季度的 0.8%提升到了 2023 年的第一季度的 18%。预计 2024 年份额达到 20%，2027 年份额达到 25%。

图9: EPYC 9654 和 Xeon 8490H 性能对比



数据来源: AMD 数据中心和 AI 首映式, 东吴证券研究所

图10: AMD 和 Intel 服务器 CPU 份额



数据来源: Mercury Research, 东吴证券研究所

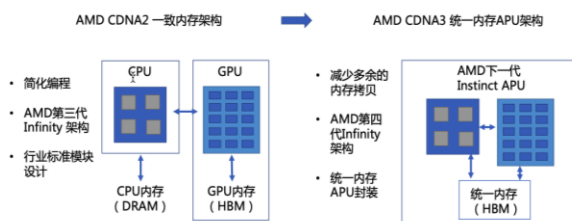
2.3. Instinct MI AI 加速器: 生态、集群问题有望解决, AMD 业绩迎来爆发

#1) Instinct MI 300A: 面向高性能运算 HPC

MI300A 是面向 HPC 产品定位, 因此产品形态是 CPU+GPU/APU 合封 Chiplet 的方式(24 个 Zen4 core 及其 I/O, 128G HBM3, 封装了 13 个 Chiplet); 并改造了底层 IF 高速互连和 UMA 内存架构。首批采购方也是美国国家超算 (EI Capitan Exascale)。

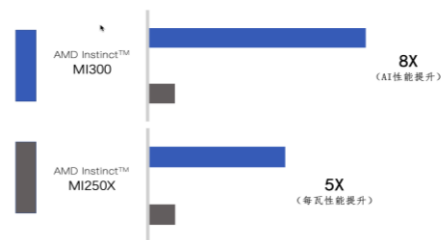
在 Instinct MI 300A 中, 使用了新一代的 APU 架构。在 Instinct MI300A 中, 它允许两种处理器类型共享高速、低延迟的统一内存空间。这将使得在 CPU 和 GPU 核心之间快速、轻松地传递数据成为可能, 让每个处理器类型分别处理它们最擅长的计算方面。此外, 它还将显著简化高性能计算 (HPC) 编程, 因为它让两种处理器类型直接访问相同的内存池, 而不仅仅是一个隐藏物理差异的统一虚拟内存空间的副本, 而是一个真正共享和物理统一的内存空间。AMD 宣称, instinct MI300A 对比 instinct MI250X 在 AI 性能上有着八倍的提升, 每瓦的 AI 性能上有着 5 倍的性能提升。

图11: CDNA 架构内存模式迭代



数据来源: 公司官网, CES2023, 东吴证券研究所

图12: Instinct MI300A 性能提升预测



数据来源: 公司官网, CES2023, 东吴证券研究所

#2) Instinct MI 300X: 面向大模型人工智能

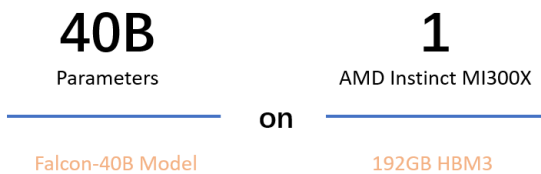
MI300X 是 AI/DL 加速器产品, 是 AI 客户的第一选择。4 个 SoC die 全部选择放 GPU, 而 8 个 HBM3 的槽位也从 16GB 的规格升级到了 24GB 的规格——集成了 192GB 的 HBM3, 5.2TB/s 存储带宽, 896GB/s 的 Infinity Fabric 互连带宽。而 NVIDIA 的

GH200 实际上是 256 组 CPU 480GB LPDDR5X+256 组 GPU 96GH HBM3, 低速 LPDDR5X 是延迟瓶颈。在市场需求发生大语言模型这种重大变化情况下, AMD 的产品比 NVIDIA 更加符合市场需求。

Instinct MI300X 有着领先的计算和内存性能。MI300X 的内存密度是 H100 的 2.4 倍, 内存带宽是 H100 的 1.6 倍。更多的 CDNA3 GPU 核心也会比 Instinct MI300A 有着更强的 AI 计算性能。

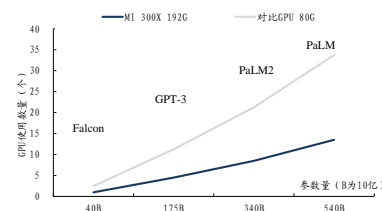
Instinct MI300X 为大模型而生。生成式 AI 和大预言模型对于加速卡的计算性能和存储性能要求很高, MI300X 的计算性能和内存性能符合未来 AI 推理对于硬件的要求。在 AI 模型参数越来越庞大的当下, 一张 MI300X 就能进行有着 400 亿参数的大语言模型的推理。对于有着更多参数的大模型的推理, Instinct MI300X 的优势会更加明显。

图13: Instinct MI300X 内存性能



数据来源: AMD 数据中心和 AI 首映会, 东吴证券研究所

图14: 不同大模型推理端所需 GPU 数量 (2023 年, 个)



数据来源: AMD 数据中心和 AI 首映会, 东吴证券研究所; 横轴单位为 10 亿

#3) ROCm 平台: 工具链完善, 兼容 CUDA, 支持主流平台

ROCm 平台的开发工具链已经相对完善。编程模型和 API 有 HIP、OpenCL 和 OpenMP; 编译及工具链有 ROCmCC、ROCgdb、HIPify 和 ROCm Profiling Tools; 支持数学库有 rocBLAS、rocFFT、rocSOLVER、rocSPARSE 和 rocWMMA; 支持并行算法库有 Parallel STL; 支持通信库有 RCCL; 支持深度学习库有 MIOpen、MIGraphX 和 MIVisionX; 开发工具有 ROCm Data Center Tools、rocm-smi、ROCm Profiling Tools 和 ROCmDebugger。

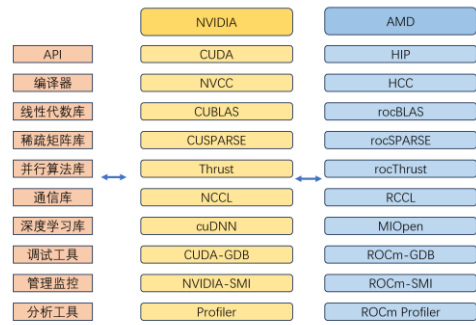
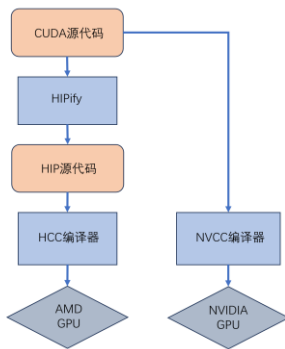
ROCm 的 HIP 对应 CUDA API, 只需要替换源码中的 CUDA 为 HIP 就可以完全移植, ROCm 工具链中的 HIPify 可以实现将 CUDA 原生代码转化为 HIP 原生 c++代码 ROCm 的工具链基本能够覆盖 CUDA, 工具加上编译模型基本能够做到全兼容。

AMD 对于 TensorFlow、Pytorch 这样的主流深度学习框架进行了专门的优化。所以在 pytorch 上使用 ROCm 非常容易。在 2023 年第一季度公司宣布其 ROCm 系统融入 PyTorch 2.0 框架, 目前 TensorFlow 和 Caffe 深度学习框架也已加入第五代 ROCm。在文章《TorchBench: Benchmarking PyTorch with High API Surface Coverage》中, 在 Pytorch 框架中分别使用 A100 和 MX210 进行性能测试, 有些模型在 A100 上的表现更加好, 有些模型在 MX210 上的表现更加好。总体上 MX210 和 A100 差距不大, 对于

经过优化的主流框架，ROCm 是完全可用的。

图15: ROCm 平台转译 CUDA 代码

图16: ROCm 和 CUDA 工具链对应图



数据来源: 公司官网, 东吴证券研究所

数据来源: 公司官网, 东吴证券研究所

#4) AI 加速器: 存在不足, 有望改善业绩迎来爆发

ROCm 平台起步晚, 应用场景覆盖窄。由于 ROCm 平台起步晚, 对于 GPU 加速库的支持没有 NVIDIA CUDA 全面, ROCm 相比 CUDA 缺失了标准数学函数库、随机数生成库、图像和视频相关库等等。现在 ROCm 主要的应用场景为 HPC 计算, AMD 为 Oak Ridge 国家实验室设计的 Frontier 超级计算机是世界上速度最快的 HPC 计算机。而 CUDA 由于多年的发展, 有着丰富的加速库支持, 基本应用场景能够覆盖全场景, 构成了软硬件结合的完整生态体系。

此外, ROCm 的社区不完善, 对消费级显卡支持差。ROCm 平台的适配性没有那么好, 对于个人用户, 如果不是在主流平台上运行, 使用的环境场景没有做特定的适配, 配置 AMD 的 HIP 平台还是相当费时费力的, 编译报错也难以处理。而对于 ROCm 而言, 整体的社区资源没有那么多, AMD 的技术支持也不可能做到全面的覆盖。ROCm 平台只支持 Instinct 系列 GPU 的部分 SKUs, 消费级上只支持少量 Radeon RX 6000 系、7000 系显卡, 以及 Radeon R9 Fury。

但是, 软件生态有望在 AMD 大力投入和其它厂商积极扶持下有望彻底改善。微软和其他云服务商大力扶持促进 ROCm 生态成型。微软使用 AMD Instinct™ MI200 加速器来支持大规模的 AI 训练工作负载。其 Azure 被打造成首个部署 AMD Instinct MI200 加速器集群用于大规模 AI 训练的公共云; Pytorch 也和 AMD 一块作为合作伙伴构建 ROCm 的软件堆栈。在 AI 开发中, 神经网络工作负载从一个平台移动到另一个平台很困难。Pytorch 正在帮助 AMD ROCm 简化这个过程; Hugging Face 也和 AMD 建立了合作, Hugging Face 将针对 AMD 平台优化所有模型, 从 InstinctGPU 开始, 再之后是 AMD 的其它产品。Hugging Face 还将一些最受欢迎的库, 通过不断的测试和调整使之更加 AMD 硬件。Hugging Face 认为, 由于内存容量和带宽优势, AMD 能够为数据中心的大语言模型提供动力。AMD 也在逐渐完善 ROCm 生态, 将更多用户引入 ROCm。据

AMD GPU 业务负责人，AMD 在将 ROCm 平台拓展到使用 RDNA 架构的消费端的 Radeon GPU。虽然目前仅有部分 SKU 支持 Windows 系统，但主流 Radeon 显卡用户可以开始试用过去仅专业显卡才能使用的 AMD ROCm (5.6.0 Alpha)，这一趋势会逐渐完善 ROCm 社区生态。

此外，对于 AI 加速器。随着如今大模型需求的参数越来越多，需要将 AI 加速器大规模部署，但 AMD 缺乏足够的大规模部署 AI 加速器的案例。不过 AMD 收购的 Pensando 公司带来的交换机、DPU 相关技术有望帮助 AMD 提升大规模集群的部署的能力。

当下，MI300X 更加符合当下大语言模型需求。同时期 NVIDIA 发布的 GH200 实际上是 256 组 CPU 480GB LPDDR5X+256 组 GPU 96GH HBM3，低速 LPDDR5X 是延迟瓶颈。而 MI300X 则是堆料在了大语言模型更加侧重的显存和带宽，集成了 192GB 的 HBM3，5.2TB/s 存储带宽，896GB/s 的 Infinity Fabric 互连带宽。如果能够解决软件生态和集群的问题，再加上 AMD 产品性价比优势，MI300X 会更加受到客户青睐。

AMD 会在 2023 年第四季度生产和出货 Instinct MI300 系列，初期用户主要是超算中心。随着产能增加，AMD 会将 Instinct MI300 系列进一步推广给 AI 用户。AMD 在 2024 年的业绩有望依靠 Instinct MI AI 加速器迎来爆发。

2.4. 自适应 SOC 和 AI 引擎：数据中心业务的重要补充

#1) 自适应 SOC 和 Vitis 平台拓展 AMD 计算业务

自适应 SOC 拓展了 AMD 计算业务。2020 年，AMD 宣布对 Xilinx 进行收购，赛灵思的自适应加速平台归入了 AMD 的版图。其中 AI Core 系列提供强大的 AI 推断和无线加速能力，其 AI 引擎的计算性能超过当前服务器级 CPU 的 100 倍；AI Edge 系列对于功率和热度受限的边缘应用，提供了超过领先 GPU 的 4 倍 AI 性能/瓦特。

Vitis 平台也是 ROCm 平台生态重要补充。其中，对于 ROCm 平台有重要影响力的工具是 Vitis AI 开发环境，它是一个专门的开发环境，用于在 Xilinx 嵌入式平台、Alveo 加速卡或云端 FPGA 实例上加速 AI 推断。Vitis AI 开发环境不仅支持业界领先的深度学习框架，如 Tensorflow 和 Caffe，而且还提供全面的 API 进行剪枝、量化、优化和编译训练过的网络，从而可以为部署的应用提高 AI 推断性能。

图17: Xilinx 芯片



数据来源：公司官网，东吴证券研究所

图18: VITIS AI 开发环境



数据来源：公司官网，东吴证券研究所

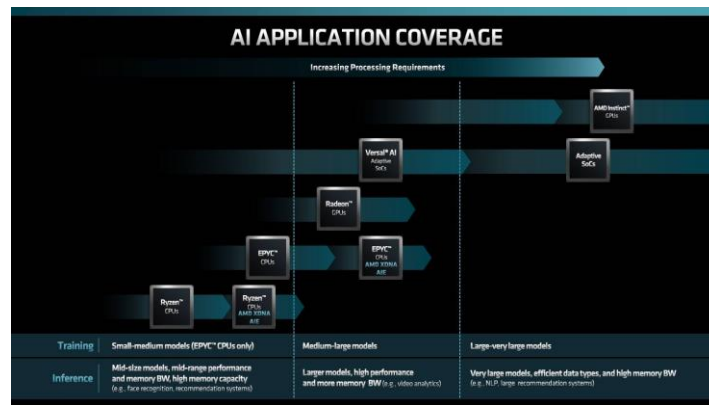
#2) AI 引擎整合 AI 平台，覆盖 AI 应用全场景

AMD 使用 Xilinx IP 来构建 AI 引擎，引入到其它硬件中。在 AMD 新命名的"自适应架构"构建模块 XDNA 中，AI 引擎将被纳入其中，并将在未来的产品中应用于公司的多个领域。AMD 已经将 AI 引擎整合到了代号为 Phoenix Point 的移动端芯片，也就是 CES2023 上发布的 Ryzen 7040 系列 CPU。与 Intel 处理器上的 GNA 相比，XDNA 提供了四个并行的 AI 处理流，用于处理多任务 AI 工作负载，而几乎不需要使用 x86 核心或 RDNA3 CUs 进行其处理堆栈。AMD 将利用 XDNA 引擎处理各种 AI 加速任务，包括图像处理。

Xilinx AI Engine 会促进 AMD 在加速卡领域的定制化服务大幅领先英伟达，协助云厂商在特定算法模块上进行训练，进一步降本增效。这种技术和硬件的融合可以帮助 AMD 未来在约 1500 亿美元的云计算、边缘计算和智能设备市场机遇中占据更大份额。

将自适应 SOC、AI 引擎、EPYC 处理器和 Instinct 加速卡组合，可以覆盖全场景覆盖 AI 数据计算需求。针对 AI 应用需求进行如下定位：Ryzen 和 Epyc CPU，包括搭载 AI 引擎的 Ryzen CPU，将覆盖用于训练和推理小型到中型模型；搭载 AI 引擎的 Epyc CPU、Radeon GPU 和 Versal 芯片将覆盖用于训练和推理中型到大型模型；Instinct GPU 和 Xilinx 的自适应芯片将覆盖用于训练和推理超大型模型。

图19: AMD AI 产品覆盖规划图



数据来源：公司官网，东吴证券研究所

2.5. Pensando DPU: 补全数据中心最后一块拼图

#1) DPU 使现代数据中心降低成本，提高性能

现代数据中心需求复杂。数据中心经历了 3 个演变，第一个发生在 2000 年前后，此时数据中心的模式是客户机/服务器的模式，然后在 2010 年演变成了云和虚拟化的模式，最后在 2020 年演变成了云和边缘混合的模式。在这种云和边缘混合的数据中心模式中，面临了很多挑战，第一是虚拟化的负载，第二是大规模网络的复杂性，第三是资源管理的挑战，第四是安全性上的问题。

在当下的数据中，服务器的负载能够从传统的 30-50%提高至 80-90%；这能为企业带来大量资源节省。但与此同时，由于一台物理机往往运行着 N 台虚拟机或容器，CPU 也需要花费更多资源来应对业务 APP 的各类 IO 需求。根据 AWS 对旗下数据中心的分析统计，CPU 平均会花 30%的计算能力来满足底层虚拟化和各类网络 IO 所带来的开销。而伴随云数据中心容器化趋势的推进，业务的细粒度还在快速增加，而这又会进一步增加 CPU 的负荷。所以，DPU 被设计来卸载这些数据管理和网络处理任务，让 CPU 能够更高效地执行其主要的计算任务，提高数据中心的整体效率，提高计算资源的利用率，降低数据中心的总体成本，同时提高服务的性能和可靠性。

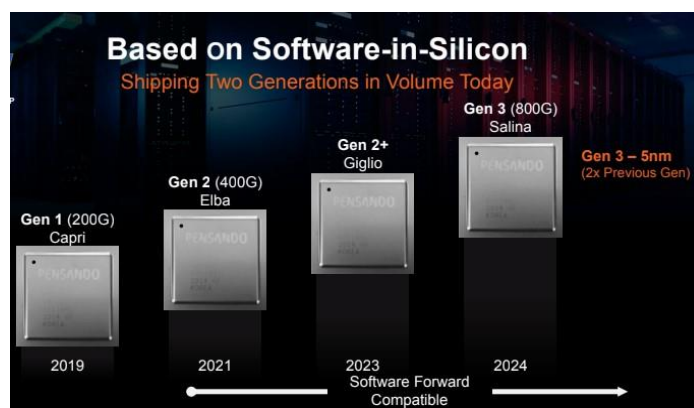
#2) 收购 Pensando 补全 AMD 数据中心拼图

2022 年 6 月，AMD 宣布 19 亿美元收购 Pensando。Pensando 的加入协助 AMD 为他们的 CPU、GPU、FPGA 和自适应 SoC 组合增加了一个领先的分布式服务平台，协助建立性能最优、安全、灵活、总拥有成本最低的前沿数据中心。

Pensando 创新能力强。 Pensando 不使用现成的 FPGA，而是拥有专门为这种功能设计的自有芯片。网络路径主要基于 P4 可编程流水线，并且 Pensando 在将 P4 作为边缘网络范例方面投入了大量资源。其产品已经在云和企业客户中大规模部署，包括高盛、IBM 云、微软 Azure 和甲骨文云。

Pensando 保持竞争力。 现在最新的产品为第二代 Elba，它和第一代的一个区别是从 HBM 切换到 DDR4/5，因为 DDR 的部署更加灵活，且 DDR 的成本也比 HBM 低，但是速度也会更慢。AMD 计划于 2023 年发布 Giglio DPU，它是 Elba 的成本优化更新版本。在 2024 年，Pensando 将推出 Salina。它被设计为一款采用 5 纳米工艺制造的产品，具备 800G 速度的能力。

图20: Pensando 产品路线图



数据来源：公司官网，东吴证券研究所

3. 游戏业务：半定制业务短期走弱，游戏显卡有望增长

3.1. 半定制业务：AMD 独树一帜，营收短期下降

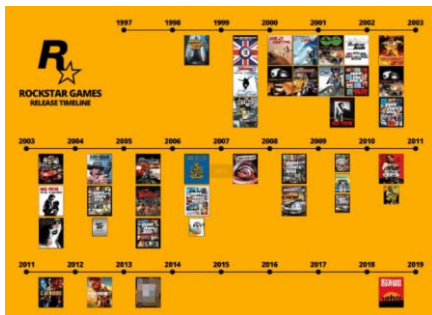
#1) AMD 在主机半定制业务上占据优势地位

如今 AAA 级电子游戏对于图形技术、游戏规模越来越高，开发成本越来越高，开发时长越来越长。其中，R 星的《GTA5》开发总成本为 2.7 亿美元，开发时间为 5 年，《荒野大镖客 2》的开发成本为 8 亿美元，时间为八年；索尼的《地平线：西之绝境》成本为 2.12 亿美元，时长为 5 年，《最后生还者：第二部》成本为 2.2 亿美元，时长超过 70 个月。Rockstar Game 是知名游戏开发商，其游戏开发速度随着游戏图形水平的提高不断变慢，2011 年前，每年发布数个作品，2011~2014 年，每年只发布 1 个作品，此后，2014 年至今，只发布了一部新作。

同时为了延长一个电子游戏的寿命和拓宽该游戏的营收，那么就需要考虑不同平台之间和跨代平台之间的兼容性，根据《Global Games Market Report》分析，截至 2021 年，PC 游戏在游戏市场中占比 20%，而主机游戏占比为 28%，PC 游戏是一个不容忽视的市场，而 PC 采用的都是复杂指令集的处理器的。所以一个趋势就是不同平台之间都采用 x86 的 CPU，微软和索尼的这一代主机同样是采用了 x86 的 CPU。同样，掌机为了兼容 PC 游戏生态，也会采用 x86 的 CPU，Valve 的 steamdeck 掌机和华硕的 ROG Ally 掌机都采用了 AMD 的定制方案。

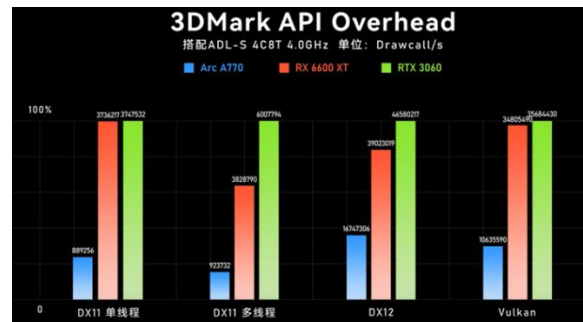
英伟达有高性能的 GPU，截至 2022 年占据独立显卡市场 80% 以上的份额。但是英伟达没有能力设计 x86 的 CPU。Intel 也开发了高性能 GPU，在 2022 年发布了独立显卡 A750 和 A770，但是英特尔显卡的驱动堪忧，在测试驱动和 API 在极限情况下沟通效率的时候，和相同规格的 AMD 和 NVIDIA 显卡有着明显的差距。AMD 是唯一能够同时提供成熟性价比高的 x86 高性能 CPU 和高性能 GPU 的厂商。

图21: Rockstar 游戏发布速度



数据来源：公司公告，东吴证券研究所

图22: 3DMark API 开销测试



数据来源：极客湾 Geekerwan，东吴证券研究所

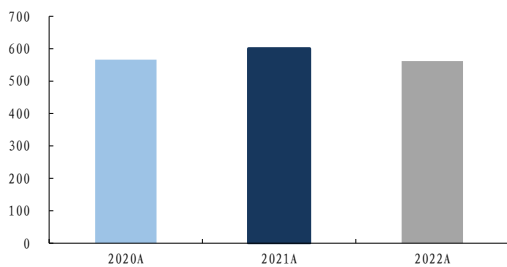
#2) 主机游戏市场增长缓慢，半定制业务短期营收下降

当下主机市场增长缓慢，新产品数量减少是主要原因。2022 年全球新上线主机游戏数量为 416 款，较 2021 年减少 311 款，降幅为 42.8%。受到资金获取难度加剧、人力研发成本上升、区域性局势不稳定等因素影响，2022 年新上线主机游戏数量大幅度减少，

主机游戏企业，特别是抗风险能力较差的中小企业开始出现生存困难，产品供给侧压力增大。游戏主机市场预计未来几年增幅不大。

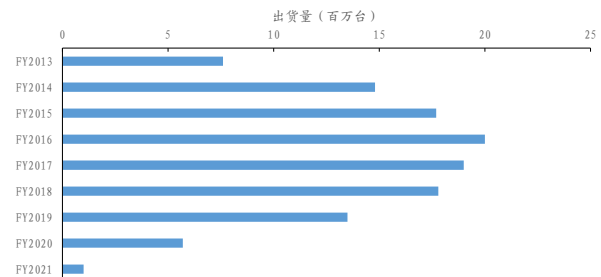
游戏机的销量通常在第四年达到峰值，而平台持有者在此之后往往会重新谈判零部件定价。根据 ps4 的数据，其销量在发布后的第四年 2016 年达到峰值。而当代的游戏主机发布于 2020 年，所以半定制业务预计 2023 年达到一个峰值，24 年开始收入逐渐减小。按照惯例，在当代主机发布 3~4 年后，会迎来一个性能上的迭代，预计将继续使用 AMD 方案，所以收入降幅不会很大。

图23: 全球主机市场规模(亿美元)



数据来源: Ampere Games, 东吴证券研究所

图24: PS4 系列历年销量(百万台)



数据来源: 公司公告, 东吴证券研究所

#3) AIGC 发展促进游戏行业重新发展, 半定制业务营收长期增长

游戏进入次时代之后, 游戏制作成本和时间的急剧上升, 导致的游戏数量减少, 这是影响游戏行业快速发展的重要阻力。但是 AIGC 技术出现, 会降低游戏开发成本, 改变游戏行业现状, 促进游戏行业未来不断增长。许多游戏厂商都推出了基于 AIGC 的开发工具: 腾讯 AILab 发布自研 3D 游戏场景自动生成解决方案, 通过 AIGC 技术, 帮助开发者在极短时间内打造出高拟真、多样化的虚拟城市场景; 育碧公开内部 AI 工具 Ghostwriter, 可自动生成 NPC 脚本、编写游戏剧情、创建任务内容; Roblox 首次推出两款 AIGC 游戏创建工具 Code Assist 和 Material Generator, 用户可根据简短提示, 生成代码片段和游戏内物件纹理; Unity 通过 AIGC 技术, 实现将文字输入转换为游戏资产创造。未来游戏行业的增长也会促进消费者对于游戏机的需求。

3.2. 游戏显卡业务: 营收和市占率不断提升

#1) 显卡技术不断迭代

2019 年, AMD 推出了基于 RDNA 架构的 RX 5000 系列 GPU, 并采用台积电 7nm 工艺, RDNA 架构较先前使用长达七年的 GCN 架构技术大幅提升, 每瓦性能比较上一代提升了 50%。2022 年, AMD 推出 RadeonRX 7000 系列, 采用突破性的 RDNA3 架构和小芯片设计。AMD RDNA3 架构的小芯片设计结合了 5nm 和 6nm 工艺节点, 突破性的架构使得每瓦性能比 RDNA2 架构高出 54%。

对标 NVIDIA 的 DLSS 技术，AMD 提出了 FSR 技术。FSR 代表 FidelityFX 超分辨率，是 AMD 的 FidelityFX 图像工具包的最新成员之一。它是一种升级技术，旨在在不损失图像质量和图形细节的情况下提高游戏性能。如果使用 4K 等高分辨率玩游戏，会导致明显的帧速率下降，如果游戏再使用要求苛刻的效果，例如光线追踪，则游戏流畅度会进一步下降。而 FSR 可以将输入分辨率降低到 1080p，然后将输出放大到接近 4K，它的工作原理是降低游戏的渲染质量，然后分析图像，检测边缘，并使用空间放大算法以更高的目标分辨率重建它们。

图25: AMD 显卡产品演进

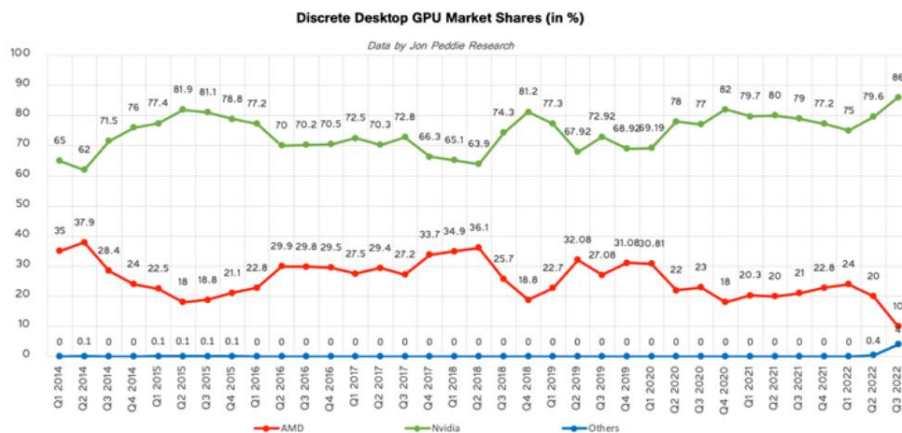
| | | | | | |
|-------|----------------|---------------|----------------|----------------|----------------|
| GPU产品 | Radeon HD 5000 | Radeon RX 500 | Radeon RX 5000 | Radeon RX 6000 | Radeon RX 7000 |
| 核心架构 | TeraScale | GCN | RDNA | RDNA2 | RDBA3 |
| 发布年份 | 2009 | 2016 | 2019 | 2020 | 2022 |

数据来源：公司公告，东吴证券研究所

#2) 加密货币和生产力欠缺拉低占有率，营收受二手显卡和 PC 市场库存影响

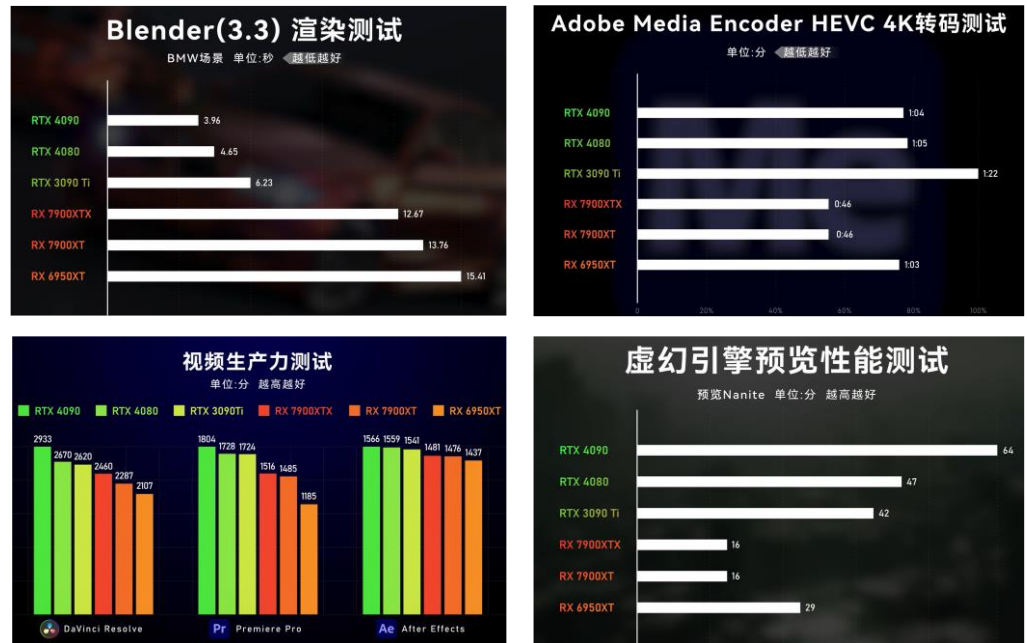
AMD 游戏显卡业务的市占率在 2022 年 Q1 之后持续走低。第一个原因是受到加密货币的影响。数字货币的“挖矿”业务对 GPU 业务产生了较大影响，NVIDIA 显卡在“挖矿”效率上优于 AMD 显卡，在 2020 年 Q4 到 2022 年 Q3 “矿潮”期间，AMD 显卡的市场因为加密货币的影响受到了严重挤压。第二个原因是生产力性能差，不被有生产力需求的业务用户或者小型工作室选择。大量生产力软件都可以使用 NVIDIA 的 CUDA 加速来提高效率，如果使用 AMD 来进行兼容，配置会相对繁琐且效能低。在其它的一些生产力场景中，AMD 显卡的表现也比 NVIDIA 显卡差。

图26: AMD 桌面级独立显卡市占率



数据来源：Jon Peddie Research，东吴证券研究所

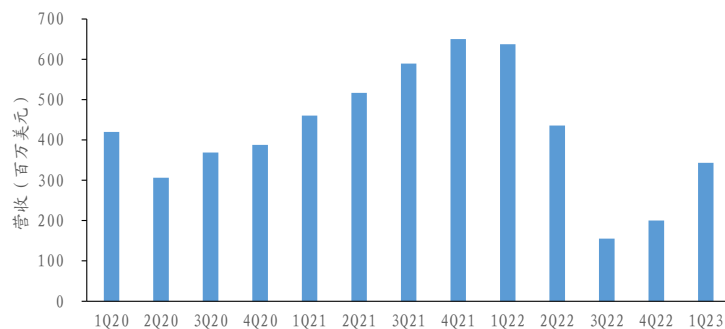
图27: AMD 和 NVIDIA 消费级显卡生产力对比



数据来源: 极客湾 Geekerwan, 东吴证券研究所

AMD 游戏业务营收在 2020 年第二季度后不断提升, 但是在 2022 年第二季度后营收大幅下降。最首要的原因是受到了 PC 市场库存高位的影响, 在 2020~2021 年疫情期间, 居家办公和娱乐需求激增, 笔电 PC 等消费性市场享受阶段红利后, 需求出现滑落, 导致终端库存在 22 年下半年处于库存高位。AMD 采取了降低下游库存压力的策略, 减少了向下游出货, 导致了营收下跌。其次是受到了二手显卡冲击, 2022 年 9 月 15 日, 以太坊正式转向 PoS 机制, 告别大规模矿机“挖矿”时代, 用于“挖矿”的显卡在这段时间前后大量流入市场。

图28: AMD 消费级显卡营收 (百万美元)



数据来源: 公司公告, 东吴证券研究所

#3) PC 市场回暖和生产力提高助游戏显卡营收和市占率不断恢复

以太坊转 PoS 之后, 消费级显卡市场恢复正常。2022 年 9 月 15 日, 以太坊正式转

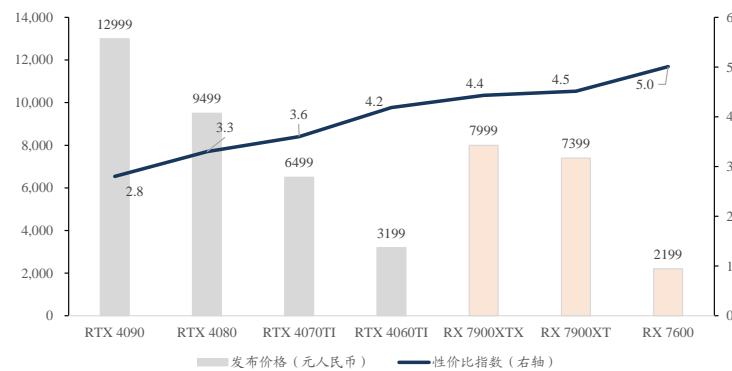
向 PoS 机制，告别大规模矿机“挖矿”时代。以太坊是“挖矿”收益最高的币种，其它任何一个币种都不能通过游戏显卡获得相比较的收益，所以**消费级显卡市场在二手矿卡影响结束后会恢复正常。**

PC 市场恢复，库存恢复正常。品牌厂宏碁表示，2023 年第二季度会是 PC 市场库存的转折点，本季下半起已明显订单涌进，除教育标案外，商用、消费及电竞等领域都有较大量的订单；华硕表示，库存去化已经见效，2023 年第二季度动能优于第一季，可预期下半年的传统季节性需求浮现，但全年市场出货量仍会低于去年。到 2024 年，老旧的设备将开始需要更新换代，**届时 PC 市场将恢复正常。**

AMD 发力提高消费级显卡生产力，ROCm 开放给消费级显卡。 AMD 在 2016 年推出了 ROCm，这是一个开放式软件平台，作为英伟达 CUDA 平台的一个开源替代。**2023 年 4 月份，ROCm 首次来到 Windows 系统，如今支持的消费级显卡有 RX 6900 XT、RX 6800 XT、Radeon RX 7900 XTX 24GB。AMD 承诺，未来会支持更多的消费级显卡使用 ROCm 平台。AMD 的消费级显卡生产力提高之后，会受到更多业务消费者的青睐。**

AMD 消费级显卡具有性价比优势。将 3Dmark 中的图形测试分除以显卡售价，可以发现 AMD 的显卡在图形性能的性价比整体优于 NVIDIA 显卡。

图29: NVIDIA 和 AMD 消费级显卡性价比对比



数据来源：公司公告，3Dmark，东吴证券研究所

4. 客户端业务：Zen 架构为消费 CPU 基石，远期增速相对平稳

4.1. Zen 架构和 3D V-Cache 技术得到市场认可

#1) Zen 架构不断调整更新，路线明确

2017 年，AMD 正式推出了 Zen 架构。IPC 比上代挖掘机提升了超过 40%。Zen 架构的最小 CPU Complex (CCX) 内有四个 x86 核心，每个核心都有独立的 L1 与 L2 缓存，共享 8MB L3 缓存，每个核心都可以选择性的附加 SMT 超线程，另外 CCX 内部的核心是可以单独关闭的。2019 年，AMD 推出了 Zen2 架构。AMD 第一次在消费级产品

中采用了 Chiplet 设计。整个处理器上有个单独的中央 I/O die，Zen 2 不需要像上代那样为每个 die 添加内存控制器和单独的 I/O，核心所在 die 的尺寸也就相对比较小，L3 cache 得以做大。在 2020 年底发布的 Zen 3 架构上，AMD 保留并改进了小芯片设计。改进的一个主要目标是每个 CCD 中的 CCX 结构。在 Zen 2 中，后者包含两个 CCX，每个都有 4 个内核和 16MB 的 L3 缓存。对于 Zen 3，设计师保留了相同的 8 核整体 CCD 结构，但将它们全部组合成一个 CCX。现在，I/O 芯片必须管理的唯一 L3 缓存事务是在单独的 CCD 之间，从而大大改善了每个 CCD 内的数据流。AMD 在 2022 年末发布了 Zen 4 架构。前端部分变化较大，这里包括指令缓存、分支预测、解码器、指令缓存、微指令队列等模块。执行引擎部分变化较小，存储队列维持 64 个不变，二级缓存 DTLB(数据页表缓冲)从 2K 条目增大了多达 50% 至 3K 条目，另外还减少了数据缓存端口的冲突几率。同时 AMD 会在 2024 年前推出 Zen 5 架构。

图30: AMD Zen 架构路线图

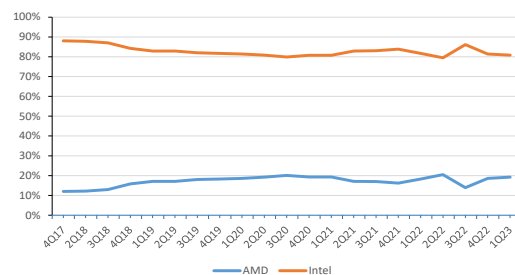


数据来源: 公司官网, 东吴证券研究所

#2) Zen 架构助力 AMD 客户端份额提升

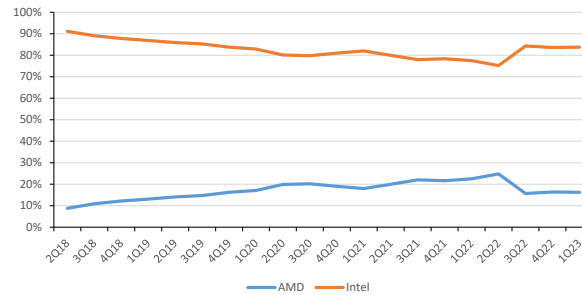
Zen 架构发布后，AMD 客户端业务份额不断增长。客户端业务包括了桌面端消费级 CPU 和移动端消费级 CPU。从 18 年后，AMD 客户端业务份额持续增长。但是在 22 年 Q3 受到了影响，因为 Intel 发布的大小核异构 CPU 在消费端营销上取得了领先。实际上同代产品性能上没有拉开很大差距。

图31: AMD 和 Intel 桌面端 CPU 份额



数据来源: Mercury Research, 东吴证券研究所

图32: AMD 和 Intel 移动端 CPU 份额



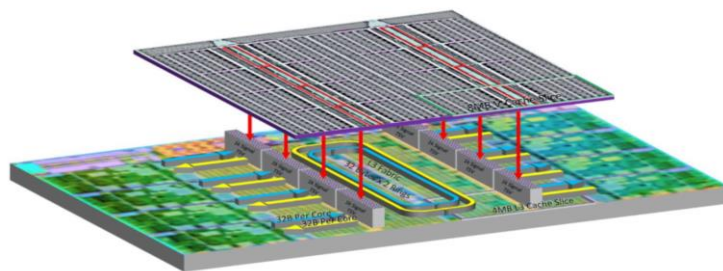
数据来源: Mercury Research, 东吴证券研究所

#3) 3D V-Cache 技术得到市场认可

3D V-Cache 技术最早在 Zen3 架构中推出。在缓存层次结构中, Zen3 架构最大的变化是共享 L3 缓存, 容量为 32 MB, 最多可支持 8 个核心。核心计算复合体以前被划分为 2x 16 MB。从理论上讲, 现在每个内核都可以使用完整的 32 MB 三级缓存。但 L3 缓存的生产也发生了变化。AMD 也留了一个“后门”, 这在实践中还没有发挥作用。那就是 L3 高速缓存可通过额外的 SRAM 芯片从 32 MB 扩展到 96 MB。尺寸仅为 36 mm² 的缓存芯片放置在 CCD 的 L3 区域, 并通过铜焊相互连接。两个芯片的界面之间的纯粘合就足够了。不需要焊接。现有缓存阵列和附加缓存之间的连接是通过 TSV 实现的。AMD 在缓存块之间提供两行 TSV 连接。对于每个 8 MB 3D V-Cache Slide, AMD 提供 1,024 个连接。因此, CCX 和 3D V-Cache 之间有 8,192 个连接。TSV 接口提供每片超过 2 TBit/s 的带宽。L3 高速缓存的环形总线在两个方向上也实现了超过 2 TB/s 的速度, 因此能够为内核提供最大的 L3 高速缓存带宽。

由于 3D V-Cache 的性能提升, 采用该技术的锐龙 5800X3D 受到市场热捧。AMD 发布的 5800X3D 中运用了 3d V-Cache 技术, 其优秀的游戏性能和性价比受到了消费者的青睐。即使在 Intel 推出了全新一代的 CPU 以及 AMD 推出了非 3d V-Cache 版本的新一代 CPU, 5800X3D 的销量依然领先, 根据德国在线零售商 Mindfactory 的数据, 5800X3D 自从发布以来一直是销量榜的榜首。

图33: 3D V-Cache 技术



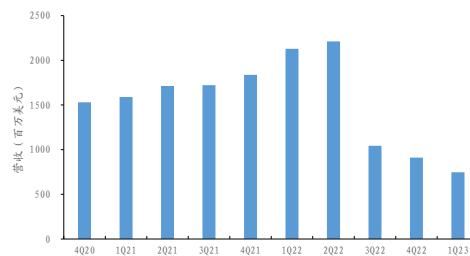
数据来源: 公司官网, 东吴证券研究所

4.2. 客户端营收随 PC 市场恢复，但增长遭遇瓶颈

PC 市场正在回暖。在 2020~2021 年疫情期间，居家办公和娱乐需求激增，笔电 PC 等消费性市场享受阶段红利后，需求出现滑落，导致终端库存在 22 年下半年处于库存高位。AMD 采取了降低下游库存压力的策略，减少了向下游出货，导致了营收下跌。但是到 23 年 2 季度，库存问题已经大大改善。到 2024 年，老旧的设备将开始需要更新换代。随着 PC 市场需求回暖，AMD 的客户端业务营收也会恢复正常。

AMD 客户端产品暂时失去性价比优势。AMD 推出 Zen4 架构的 7000 系 CPU 时，同时也更新了桌面端 CPU 平台至 AM5 平台，且 7000 系 CPU 只支持 DDR5 内存。过去 AMD 有着便宜的 AM4 平台，配合性价比很高的 CPU，很容易受到消费者青睐。但如今三者的组合失去了性价比优势。所以 AMD 客户端业务市占率短期之内提升空间较小。

图34: AMD 客户端业务营收 (百万美元)



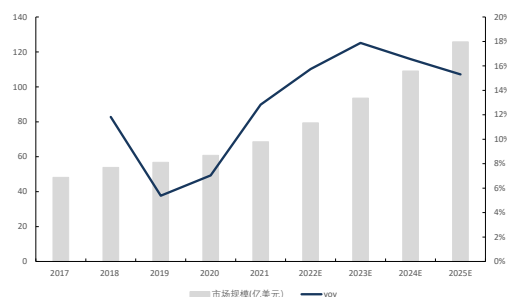
数据来源: 公司公告, 东吴证券研究所

5. 嵌入式业务: 覆盖多行业, 营收稳定增长

5.1. 收购 Xilinx 拓展 AMD 业务范围, 边缘侧需求不断增长

2020 年, AMD 宣布收购 Xilinx, 并且在 2022 年完成了收购。有了 Xilinx 的参与, AMD 极大拓宽了业务范围。如今 AMD 的嵌入式业务在医疗保健, 工业, 机器人, 汽车, 计算机视觉, 军工, 航空电子, 高速网络, 高性能计算, 通信领域都是重要参与者。

图35: 全球 FPGA 市场规模 (亿美元)



数据来源: Frost&Sullivan, 东吴证券研究所

同时, 边缘侧, FPGA 产品需求不断增长。根据 Frost&Sullivan 统计和预测, 全球 FPGA 市场规模在 2021 年达到了 68.6 亿美元, 2025 年有望达到 125.8 亿美元, CAGR 为 16%, 相关业务需求强劲。

5.2. AMD 嵌入式业务是信息技术行业的重要参与者

#1) 汽车智能化的新玩家

随着汽车行业创新步伐的不断加快，人们对高性能计算和图形技术的需求也在不断增加。智能车机主要有两条路线，第一种是 ARM 架构芯片+安卓系统，第二种是 X86 架构芯片+Linux 系统,AMD 属于后者。AMD 利用自身优势迎接高性能智能座舱潮流，除开和特斯拉展开智能座舱合作外，还和 ECARX 合作，结合 AMD 的锐龙嵌入式 V2000 处理器和 Radeon RX 6000 系列 GPU 开发了沉浸式数字驾驶舱计算平台，添加更多的功能到数字驾驶舱中。

图36: Xilinx 汽车平台



数据来源：公司公告，东吴证券研究所

AMD 还参与了高级驾驶辅助。Xilinx 通过与 Seeing Machines 合作，为 Seeing Machines 的 Fovio 芯片提供了其车规级芯片的半定制版本。斯巴鲁就采用了 Xilinx 的 FPGA 方案。同时，赛灵思汽车（XA）平台在为自动驾驶模块提供动力方面发挥着关键作用。XA 平台实现了高速数据聚合、预处理及分配（DAPD）并计算加速。该平台不仅可优化处理越来越多的复杂安全关键型应用，而且还可满足传感器和域控制器之间的计算时延、性能、电源效率和功能安全性需求。

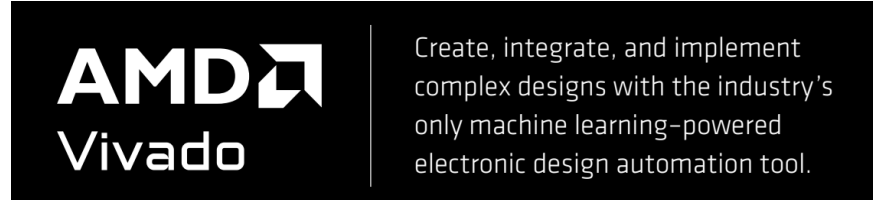
#2) Versal Premium VP1902 加速半导体设计

Versal Premium VP1902 自适应 SoC 是世界上最大的自适应 SoC 和 FPGA。该 FPGA 简化日益复杂的半导体设计的验证，范围从 Raspberry Pi 中的小型 SoC 到 AMD Instinct MI300 等大型下一代加速器。

Versal Premium VP1902 加速半导体设计。与 VU19PFPGA 相比，该自适应 SoC 有着 2 倍可编程逻辑密度和 2 倍聚合 I/O 带宽，随着容量的增加和 I/O 的降低，芯片制造商可以以更高的速率运行其仿真设计，可以缩短设计周期；人工智能工作负载正在推动芯片制造的复杂性不断增加，需要下一代解决方案来开发未来的芯片。调试对于投产前验证和并行软件开发至关重要。VP1902 自适应 SoC 利用 Versal 架构，与上一代 VU19PFPGA 相比，调试速度提高了 8 倍。AMD Vivado 机器学习设计套件为客户提

供了全面的开发平台，支持在 VP1902 自适应 SoC 上进行更高效开发的新功能包括自动设计收敛辅助、交互式设计调整、远程多用户实时调试和增强的后端编译，使最终用户能够更快地迭代 IC 设计，有着先进的设计软件，生态完善。

图37: Vivado 设计软件



数据来源：公司公告，东吴证券研究所

6. 盈利预测与投资评级

核心假设：

- 1) AI 计算软件生态和计算卡大规模集群技术进步符合预期，2024 年数据中心业务爆发，对应数据中心业务 23-25E 营收增速分别为-1.5%/116.6%/19.4%。
- 2) PC 市场回暖，消费端业务产品营收增长。客户端业务，我们预期 23-25E 营收增速分别为-27.6%/50.2%/4.6%，游戏业务，23-25E 营收增速预期 1.5%/-8.3%/2.5%。
- 3) 边缘侧产品需求保持增长，对应嵌入式业务 23-25E 营收增速分别为 30.2%/3.1%/10.4%。

表1: 公司各业务营收预测（百万美元）

| | 2022A | 2023E | 2024E | 2025E |
|---------------|---------|--------|--------|--------|
| 总营收 (百万美元) | 23,601 | 23,105 | 32,179 | 35,840 |
| yoy | 43.6% | -2.1% | 39.3% | 11.4% |
| 客户端业务 | | | | |
| 营收 (百万美元) | 6,294 | 4,554 | 6,839 | 7,154 |
| yoy | -8.3% | -27.6% | 50.2% | 4.6% |
| 营收占比 | 26.7% | 19.7% | 21.3% | 20.0% |
| 数据中心业务 | | | | |
| 营收 (百万美元) | 6,175 | 6,083 | 13,173 | 15,730 |
| yoy | 61.0% | -1.5% | 116.6% | 19.4% |
| 营收占比 | 26.2% | 26.3% | 40.9% | 43.9% |
| 游戏业务 | | | | |
| 营收 (百万美元) | 6,437 | 6,535 | 5,992 | 6,140 |
| yoy | 21.5% | 1.5% | -8.3% | 2.5% |
| 营收占比 | 27.3% | 28.3% | 18.6% | 17.1% |
| 嵌入式业务 | | | | |
| 营收 (百万美元) | 4,600 | 5,990 | 6,175 | 6,816 |
| yoy | 1267.7% | 30.2% | 3.1% | 10.4% |
| 营收占比 | 19.5% | 25.9% | 19.2% | 19.0% |

数据来源：公司年报，东吴证券研究所

表2: 公司营收利润预测摘要

| | 2022A | 2023E | 2024E | 2025E |
|---------------------|--------|--------|--------|--------|
| 总营收 (百万美元) | 23,601 | 23,105 | 32,179 | 35,840 |
| 总营收yoy | 43.6% | -2.1% | 39.3% | 11.4% |
| 营业成本 (百万美元) | 11,357 | 11,333 | 14,780 | 15,940 |
| 毛利润 (百万美元) | 12,244 | 11,772 | 17,399 | 19,899 |
| 毛利率 | 51.9% | 50.9% | 54.1% | 55.5% |
| Non-GAAP 净利润 (百万美元) | 5,504 | 4,869 | 7,822 | 9,226 |
| Non-GAAP 净利率 | 23.3% | 21.1% | 24.3% | 25.7% |

数据来源：公司年报，东吴证券研究所

盈利预测: 预计 2023 年第三季度 PC 市场开始恢复回暖, 但全年 PC 市场仍然处于较低水平, 加上嵌入式业务受到通信行业需求下降影响, 2023 年营收出现下滑; 2024、2025 年度因 PC 市场恢复和 Instinct MI 300X 加速卡出货而迎来业绩爆发。预计 2023-2025 年营收增速分别为 -2%、39%、11%, 由于数据中心产品高毛利的特点, 毛利率 (Non-GAAP) 将为 51%、54%、56%, 对应净利率 (Non-GAAP) 分别为 21%、24%、26%。

可比公司: GPU 和 CPU 都呈寡头垄断格局, 考虑到业务类似性, 美股的英伟达 (NVIDIA, 可比 GPU 业务) 和英特尔 (INTEL, 可比 CPU 业务) 作为可比公司, 2023-2024 年平均的 NON-GAAP PE 分别为 62 倍、32 倍。

我们预计公司 2023-2025 年 NON-GAAP 净利润分别为 48.7 亿、78.2 亿、92.3 亿美元, 对应现价 (8 月 9 日) PE 分别为 37 倍、23 倍、19 倍, 考虑到 Instinct MI 300X 加速器和消费级显卡重新带动增长预期, 首次覆盖予以“买入”评级。

表3: 可比公司估值

| 股票代码 | 公司名称 | 总市值 (亿美元) | NON-GAAP净利润 (亿美元) | | | PE (NON-GAAP) | | |
|---------|--------|--------------|-------------------|-------|-------|---------------|-------|-------|
| | | | 2022A | 2023E | 2024E | 2022A | 2023E | 2024E |
| NVDA. 0 | 英伟达 | 10511 | 84 | 188 | 236 | 126 | 56 | 44 |
| INTC. 0 | 英特尔 | 1436 | -72 | 21 | 71 | N. A. | 68 | 20 |
| | 可比公司均值 | | | | | N. A. | 62 | 32 |
| AMD. 0 | 超威半导体 | 1784. 83 | 55 | 49 | 78 | 32 | 37 | 23 |

注: 收盘价更新至 2023 年 8 月 9 日, 美元兑人民币汇率为 7.2。英伟达盈利预测来自东吴, 英特尔盈利预测来自 Bloomberg。

7. 风险提示

1) AMD 软件团队规模不足。 AMD 的产品驱动会出现优化问题, 显卡 7900xtx 发布初期, 其在部分软件游戏上表现不如上一代产品 6950xt; AMD 游戏者开发大会 GDC2023 上, AMD 发布了 FSR3.0, 这项技术在 FSR 2 的基础上新增了游戏插帧技术, 使得游戏性能提升一倍。AMD 承诺很快推出这个技术, 但是直到 2023 年 7 月, AMD 并没有再给出这项技术的相关信息; AMD 的 7040 系列在更换 5nm 和 Zen4 后, 在笔记本相对 Intel 的 Raptor Lake 竞争力充分, 并且有了 AI 引擎的加入, 在移动端 CPU 上有望大大提高市场份额。但是出现了 LPDDR5 内存会出现睡眠睡死的 bug, LPDDR5 的产品上市时间严重拖延, 错失市场份额。

2) 业务和份额受到英特尔挑战。 AMD 历年的年度营收都不到 Intel 的 1/10。而且 Intel 的盈利来源也远不止 CPU 市场; Intel 近些年的财务状况始终十分健康。在财力雄厚, 重新拥有技术方面的领先优势或许也并不算难事; 英特尔制定推动了很多半导体行业的标准, 仍然是半导体行业的领导者; 英特尔也重新入局了独立 GPU, 英特尔在显卡技术上的储备并不差, 可能凭借其技术储备和研发能力挤压 AMD 独立显卡的份额。

3) Instinct MI300X 集群技术瓶颈。 Instinct MI300X 非常符合当下大语言模型

的需求，但是集群技术还存在瓶颈。AMD 并未明确说明 MI300X Platform 标准机型的机内和机间 Pod 的互连拓扑，仅仅谈到了更高速率的 Infinity Fabric / 架构，以及 8 颗 MI300X 互连为一组的标准。如果未解决 IMP 机内铜缆互连以及 Rack-to-Rack 机间 RDMA 的硬问题，这就十分不利于 MI300X in AIDC 场景的生产级部署。

4) 数据中心市场发展受阻。人工智能发展可能会带来伦理问题，可能会遭到政府和相关监管部门的限制；同时生成式人工智能的生产内容未来也会受到监管和限制；人工智能发展需要大量数据，这些数据同样会带来隐私的问题，数据的匮乏也会导致人工智能发展受阻。

超威半导体三大财务预测表

| 资产负债表 (百万美元) | | | | | 利润表 (百万美元) | | | | |
|----------------|--------|--------|--------|--------|---------------------|--------|--------|--------|--------|
| | 2022A | 2023E | 2024E | 2025E | | 2022A | 2023E | 2024E | 2025E |
| 流动资产 | 15,019 | 15,771 | 18,659 | 22,255 | 营业总收入 | 23,601 | 23,105 | 32,179 | 35,840 |
| 现金及现金等价物 | 4,835 | 3,863 | 4,949 | 7,507 | 营业成本 | 11,357 | 11,333 | 14,780 | 15,940 |
| 应收账款及票据 | 4,126 | 4,183 | 5,103 | 5,684 | 销售费用 | 1,906 | 1,843 | 2,403 | 3,016 |
| 存货 | 3,771 | 4,166 | 5,049 | 5,505 | 管理费用 | 0 | 0 | 0 | 0 |
| 其他流动资产 | 2,287 | 3,558 | 3,558 | 3,558 | 研发费用 | 5,005 | 5,772 | 6,787 | 7,537 |
| 非流动资产 | 52,561 | 51,949 | 51,713 | 51,477 | 其他费用 | 0 | 0 | 0 | 0 |
| 固定资产 | 1,513 | 1,518 | 1,342 | 1,166 | 经营利润 | 5,333 | 4,156 | 8,210 | 9,346 |
| 商誉及无形资产 | 0 | 0 | 0 | 0 | 利息收入 | 80 | -72 | 40 | 40 |
| 长期投资 | 0 | 0 | 0 | 0 | 利息支出 | 0 | 0 | 0 | 0 |
| 其他长期投资 | 0 | 0 | 0 | 0 | 其他收益 | -4,069 | -3,788 | -3,800 | -3,800 |
| 其他非流动资产 | 51,048 | 50,431 | 50,371 | 50,311 | 利润总额 | 1,184 | 440 | 4,370 | 5,506 |
| 资产总计 | 67,580 | 67,720 | 70,372 | 73,732 | 所得税 | -122 | 87 | 568 | 716 |
| 流动负债 | 6,369 | 7,041 | 7,658 | 7,995 | 净利润 | 1,320 | 355 | 3,802 | 5,506 |
| 短期借款 | 0 | 0 | 0 | 0 | 少数股东损益 | 0 | 0 | 0 | 0 |
| 应付账款及票据 | 0 | 0 | 0 | 0 | GAAP 净利润 | 1,320 | 355 | 3,802 | 5,506 |
| 其他 | 6,369 | 7,041 | 7,658 | 7,995 | EBIT | 1,522 | 196 | 3,274 | 4,830 |
| 非流动负债 | 6,461 | 6,363 | 6,363 | 6,363 | EBITDA | 5,959 | 4,792 | 8,846 | 9,982 |
| 长期借款 | 2,467 | 2,467 | 2,467 | 2,467 | Non-GAAP 净利润 | 5,504 | 4,869 | 7,822 | 9,226 |
| 其他 | 3,994 | 3,896 | 3,896 | 3,896 | | | | | |
| 负债合计 | 12,830 | 13,404 | 14,021 | 14,358 | | | | | |
| 股本 | 0 | 0 | 0 | 0 | 主要财务比率 | 2022A | 2023E | 2024E | 2025E |
| 少数股东权益 | 0 | 0 | 0 | 0 | NON-GAAP EPS (美元) | 3.40 | 3.04 | 4.95 | 5.92 |
| 归属母公司股东权益 | 54,750 | 54,316 | 56,351 | 59,374 | 每股净资产 (美元) | 33.99 | 33.93 | 35.67 | 38.08 |
| 负债和股东权益 | 67,580 | 67,720 | 70,372 | 73,732 | 发行在外股份 (百万股) | 1,611 | 1,601 | 1,580 | 1,559 |
| | | | | | ROIC (%) | 3% | 0% | 4% | 6% |
| | | | | | ROE (%) | 2% | 1% | 7% | 9% |
| | | | | | 毛利率 (%) | 52% | 51% | 54% | 56% |
| | | | | | NON-GAAP 净利率 (%) | 23% | 21% | 24% | 26% |
| | | | | | 资产负债率 (%) | 19% | 20% | 20% | 19% |
| | | | | | 收入增长率 (%) | 44% | -2% | 39% | 11% |
| | | | | | NON-GAAP 净利润增速 (%) | 60% | -12% | 61% | 18% |
| | | | | | P/E (NON-GAAP) | 32 | 37 | 23 | 19 |
| | | | | | P/B | 3.4 | 3.5 | 3.3 | 3.2 |
| | | | | | EV/EBITDA | 31 | 39 | 21 | 18 |

数据来源:Wind, 东吴证券研究所, 全文如无特殊注明, 相关数据的货币单位均为美元, 美元: 人民币汇率为2023年8月9日的7.2, 预测均为东吴证券研究所预测。

免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准,已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司(以下简称“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。在任何情况下,本报告中的信息或所表述的意见并不构成对任何人的投资建议,本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

在法律许可的情况下,东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易,还可能为这些公司提供投资银行服务或其他服务。

市场有风险,投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息,本公司力求但不保证这些信息的准确性和完整性,也不保证文中观点或陈述不会发生任何变更,在不同时期,本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有,未经书面许可,任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的,应当注明出处为东吴证券研究所,并注明本报告发布人和发布日期,提示使用本报告的风险,且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的,应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期(A 股市场基准为沪深 300 指数,香港市场基准为恒生指数,美国市场基准为标普 500 指数,新三板基准指数为三板成指(针对协议转让标的)或三板做市指数(针对做市转让标的)),具体如下:

公司投资评级:

- 买入: 预期未来 6 个月个股涨跌幅相对基准在 15%以上;
- 增持: 预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间;
- 中性: 预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间;
- 减持: 预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间;
- 卖出: 预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级:

- 增持: 预期未来 6 个月内,行业指数相对强于基准 5%以上;
- 中性: 预期未来 6 个月内,行业指数相对基准-5%与 5%;
- 减持: 预期未来 6 个月内,行业指数相对弱于基准 5%以上。

我们在此提醒您,不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系,表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况,如具体投资目的、财务状况以及特定需求等,并完整理解和使用本报告内容,不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
 苏州工业园区星阳街 5 号
 邮政编码: 215021
 传真: (0512) 62938527
 公司网址: <http://www.dwzq.com.cn>