



【中泰电子】AI系列报告5 AMD：发布MI300，指引Chiplet等AI芯片新方向

中泰电子王芳团队

分析师：王芳

执业证书编号：S0740521120002

分析师：杨旭

执业证书编号：S0740521120001

分析师：游凡

执业证书编号：S0740522120002

分析师：李雪峰

执业证书编号：S0740522080004

研究助理：张琮

研究助理：刘博文

中泰证券研究所
专业 | 领先 | 深度 | 诚信

■ AMD MI300 VS 英伟达GH200:

- 1) 在制程上，MI300属台积电5nm，相较MI200系列的6nm实现了跃迁，并与英伟达Grace Hopper的4nm制程（属台积电5nm体系）看齐。
- 2) 封装技术，AMD MI300使用台积电SoIC（3D）和CoWoS（2.5D）两种封装技术，而Nvidia GH200仅使用CoWoS（2.5D）封装技术。3D Chiplet封装技术可以提升封装密度，并在降低能耗的基础上进一步提高性能。

■ Chiplet+存算一体+异构计算，构建了AMD MI300强大的AI竞争力:

- **Chiplet:** MI300利用3D封装技术将CPU和加速计算单元集成在一起，增强了集成度。
- **存算一体:** 在MI300采用的CDNA3架构中，CPU和GPU共享一块“统一存储”HBM（Unified Memory），缩短了计算和存储芯片的“沟通距离”，提升了CPU与GPU间的连接速度。
- **异构计算:** MI300是市场上首款“CPU+GPU+内存”一体化产品，能更好满足如今AI大模型的算力需求。

■ 看好Chiplet产业链:

- 封测：通富微电/长电科技；减薄：华海清科；ABF：兴森科技；设备/第三方测试：长川科技/华峰测控/伟测科技；IP：芯原股份；EDA：华大九天/概伦电子等；

■ 看好AMD产业链:

- 封测：通富微电；
- PCB：沪电股份/奥士康；
- 设计：芯原股份；
- 服务器：工业富联；

- 风险提示：Chiplet相关技术路径尚未定型，存在技术路径被颠覆的风险。AMD AI相关产品市场销售不及预期。竞争加剧对行业的盈利性造成影响。

目 录

一、逐鹿AI：AMD发布最强AI芯片MI300系列

二、Chiplet产业链及投资建议

三、AMD产业链及投资建议

1.1.1 紧随Nvidia，AMD加入超算竞赛

- 北京时间6月14日AMD举行新品发布会，推出重磅产品MI300系列，包括MI300A/MI300X两颗芯片：1) MI300 A是一款针对AI和HPC (高性能计算)的APU，采用CDNA 3架构的GPU和24个Zen 4CPU内核，配置128GB的HBM 3内存，在13个小芯片中集成**1460**亿个晶体管，多于英伟达H100的**800**亿个；2) MI300 X是一款GPU专用的AI芯片，配置192GB的HBM 3内存、5.2TB/s的存储带宽，针对大语言模型的优化版，在12个5nm的小芯片中集成1530个晶体管。MI300X提供的HBM密度最高是Nvidia AI芯片H100的2.4倍，其HBM带宽最高是H100的1.6倍，单个MI300X可以运行一个参数多达800亿的模型；3) AMD还发布了AMD Instinct平台，它配置了八个MI300 X，能提供总计1.5TB的HBM3内存，采用行业标准OCP设计。
- 根据AMD预期，随AI发展，模型规模扩大，算力需求将不断增长，数据中心人工智能加速器的潜在市场总额将从2023年的300亿美元增长到2027年的1500亿美元以上，复合年增长率超过50%。

图表： Nvidia与AMD主要产品参数比较

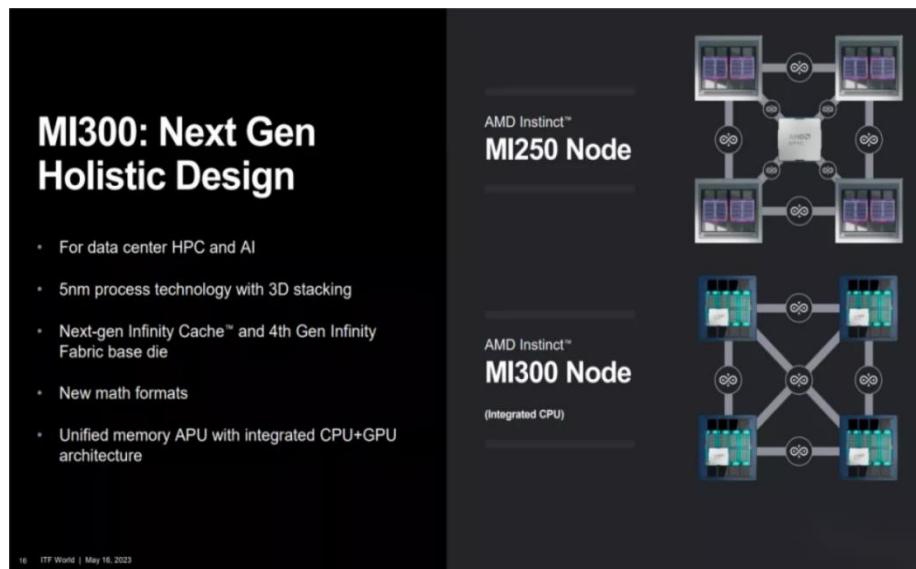
产品名称	Nvidia		AMD	
	A100	H100	MI300A	MI300X
发布时间	2020.06	2022.03	2023.06	2023.06
峰值算力（TFLOPS）	FP16: 312 624 FP32: 19.5 FP64: 19.5	FP8: 3,026 3,958 7,916 FP16: 1,513 1,979 3,958 FP32: 51 67 134 FP64: 51 64 134		
工艺制程	7nm	4nm	5nm和6nm	5nm和6nm
晶体管数量（亿）	540	800	1460	1530
内存容量	80GB(HBM2e)	PCIe: 80 GB(HBM2e) SXM: 80 GB(HBM3) NVL: 188GB(HBM3)	128GB(HBM3)	192GB(HBM3)
封装技术	CoWoS	CoWoS	3D Chiplet	3D Chiplet
互联带宽	600 GB/s NVLink for 2 GPUs 64GB/s PCIe Gen4	600 900 900 GB/s NVLink 125 GB/s PCIe Gen5	-	5.2TB/s的带宽和896GB/s的 Infinity Fabric带宽

来源：AMD官网、CES 2023、中泰证券研究所

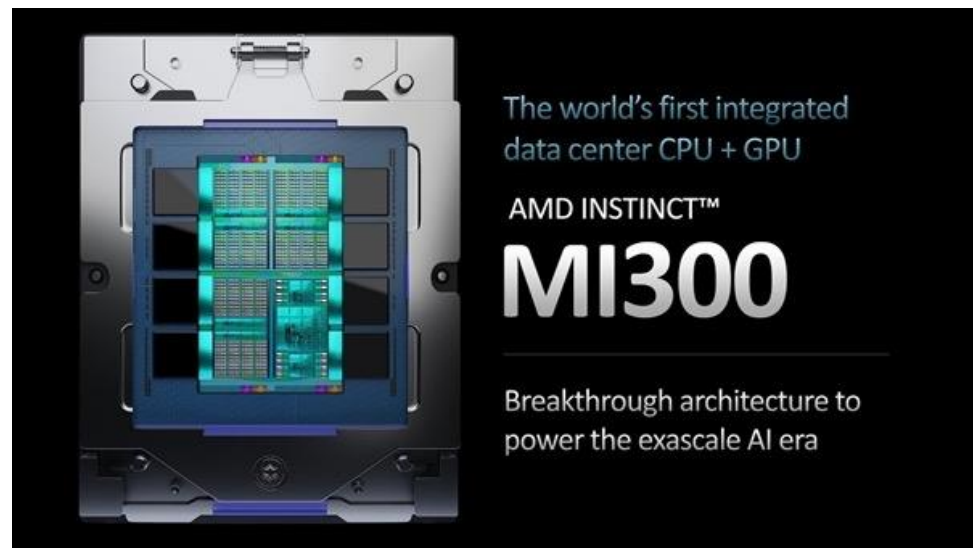
1.1.1 紧随Nvidia，AMD加入超算竞赛

- **Instinct MI300简化系统结构，大幅提升AI性能。**MI300采用3D堆叠技术和Chiplet设计，配备了9个5nm制程的芯片组，置于4个6nm制程的芯片组之上。MI300属台积电5nm，相较MI200系列的6nm实现了跃迁。AMD 上一代Instinct MI250 芯片包含独立CPU和GPU，中间依靠EPYC CPU来协调工作负载。相比之下，Instinct MI300包含一个内置的24核第四代 EPYC Genoa 处理器，从而减少了一个独立CPU的存在，简化了系统拓扑结构。由此实现了一个四个元素完全连接的all-to-all拓扑，使所有处理器可以直接相互通信，而无需另一个CPU或GPU作为数据中继，从而减少了延迟和不确定性。MI300的AI性能是上一代产品MI250X的8倍，每瓦性能则是它的5倍，实现了颠覆式升级。
- **AMD MI300使用台积电SoIC（3D）和CoWoS（2.5D）两种封装技术。**MI300采用的3D Chiplet封装技术将13个chiplets 分布在两层：上层为9个5nm制程的chiplets，而下层为4个6nm制程的chiplets，芯片两侧排列8个HBM3。3D Chiplet封装技术可以提升封装密度，并在降低能耗的基础上进一步提高性能。

图表： MI300与MI250结构比较



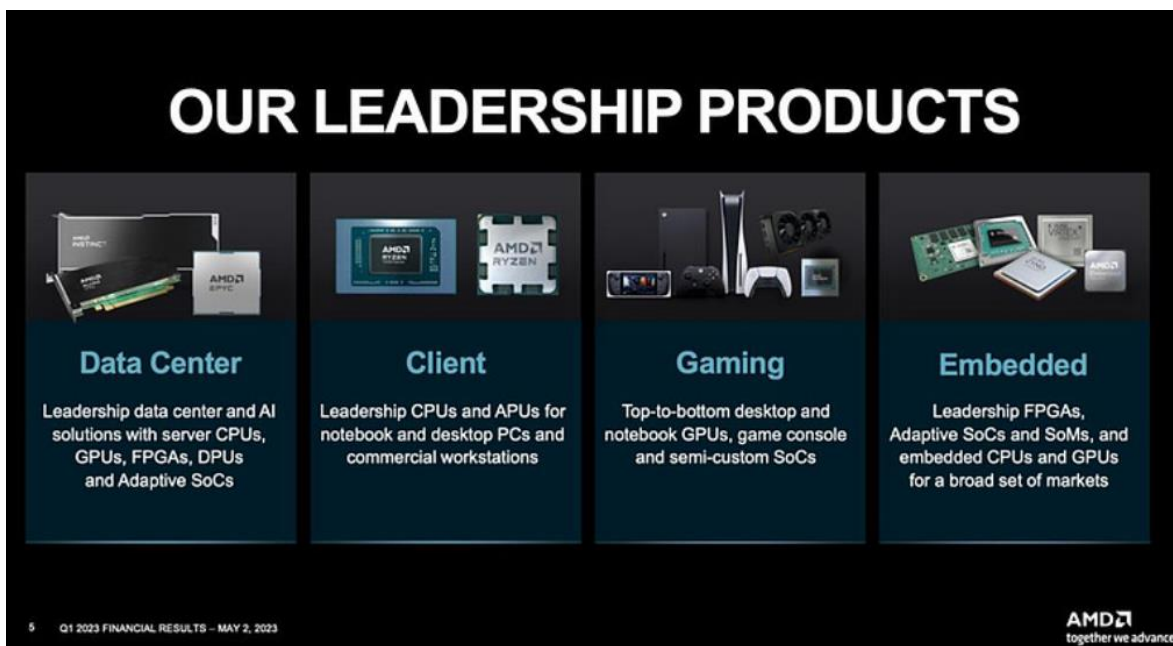
MI300产品示意图



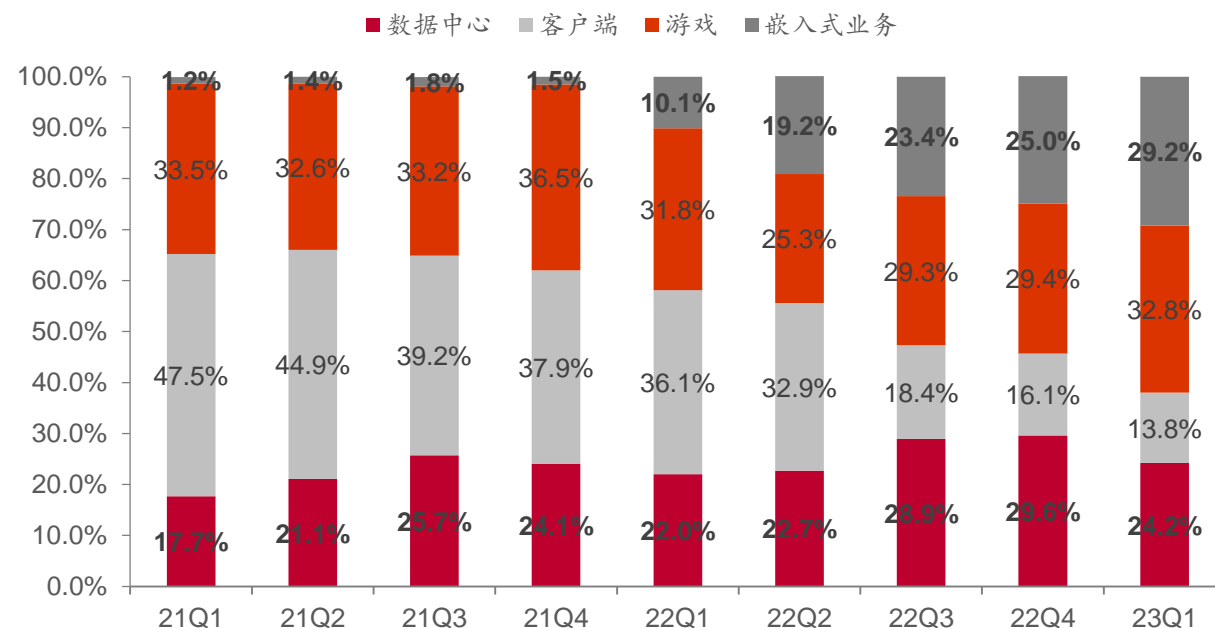
1.1.2 AMD与英伟达比较：AMD产品矩阵

- **AMD产品矩阵包括CPU、GPU、FPGA等**，业务线主要分为客户端（台式机、笔记本电脑处理器（CPU）和芯片组）、数据中心、游戏业务、嵌入式业务（包含收购赛灵思后的FPGA业务）四个板块。23Q1 AMD营收53.53亿美元，同比下降9%，其中1）客户端23Q1营收7.39亿美元，同比跌幅达65.2%，主要受一季度行业景气度下降影响。同期英特尔CCG业务（客户端电脑组）也受到了影响，23Q1营收下滑38%。2）嵌入式业务23Q1营收15.62亿美元，同比上涨163%，大幅增长受益于AMD在2022年2月14日完成对赛灵思的收购，23Q1已完成并表。3）数据中心23Q1营收12.95亿美元，同比基本持平。4）游戏业务23Q1营收17.57亿美元，同比下滑6%。游戏业务主要包括独立GPU、半定制SoC产品和开发服务，例如为Sony PlayStation 5等游戏机提供芯片。目前，AMD超过50%的营收由嵌入式和数据中心两个to B业务贡献。

图表：AMD产品矩阵



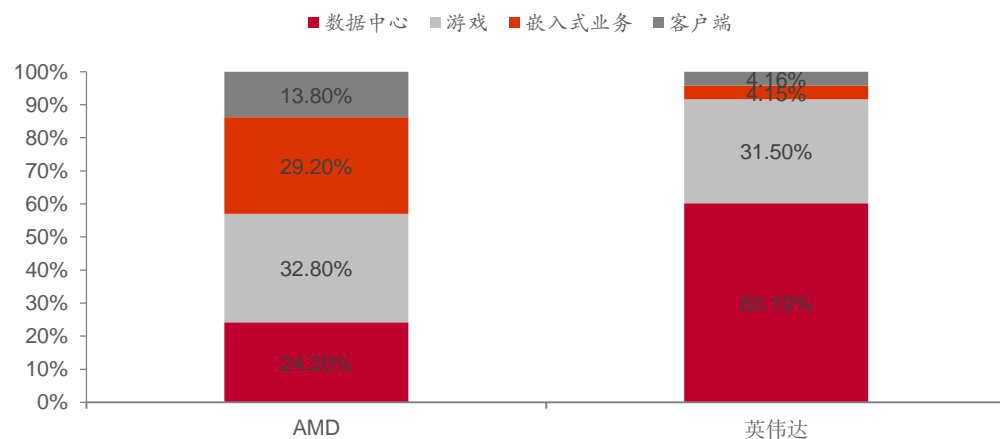
图表：AMD季度分业务营收占比



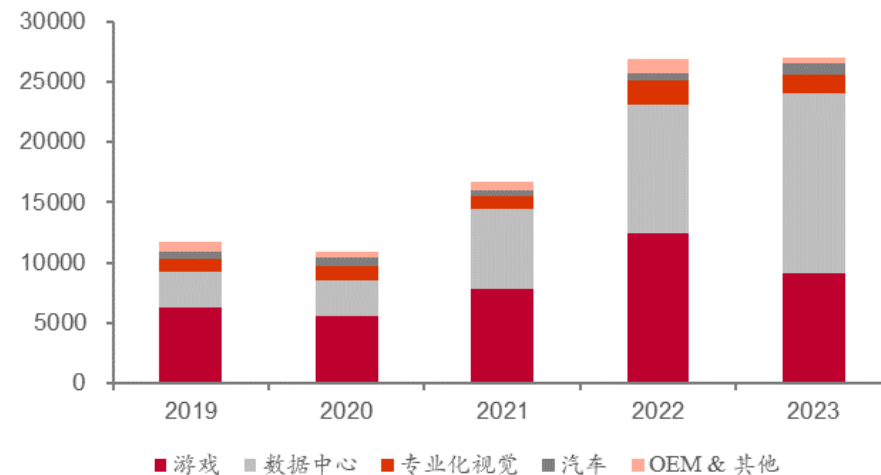
1.1.2 AMD与英伟达比较：英伟达产品矩阵

- 英伟达实行“CPU+GPU+DPU”三芯布局。旨在全面提升竞争力，满足云计算、人工智能及机器学习等高端应用领域的需求。CPU的加入使英伟达能够更好地应对各种计算任务，尤其是那些需要快速逻辑判断和高度并行处理能力的应用。而DPU则针对数据中心和网络设备的需求，具有高效处理数据包和协议的能力，为英伟达的产品线增添了新的价值。
- 随人工智能发展，数据中心业务收入增速最高，逐步成为英伟达最大营收占比，23年占比56%。从下游应用来看，英伟达产品主要集中于游戏、专业可视化、数据中心以及自动驾驶领域：1) 游戏市场：英伟达提供的产品包括PC游戏的GeForce RTX和GeForce GTX，用于游戏和流媒体的SHIELD设备，用于云端游戏的GeForce NOW，以及用于专门控制台游戏设备的平台和开发服务；2) 专业可视化市场：英伟达除了加速GPU计算解决方案，同时也为汽车、娱乐、建筑工程、石油和天然气、医疗等行业引入新的解决方案；3) 数据中心市场：英伟达使用NVlink技术将多个GPU结合在一起，加速神经网络训练和推理。同时开发出DGX超级计算机，进行科学计算、深度学习和机器学习；4) 自动驾驶市场：英伟达Drive作为一个人工智能汽车平台，涵盖了从交通拥堵到机器人出租车自动驾驶的所有领域。2018年有超过370家自动驾驶汽车公司开始使用Drive，共同开发自动驾驶的人工智能系统。

图表：AMD与英伟达分业务营收占比对比



图表：英伟达2019-2023财年分业务收入（百万美

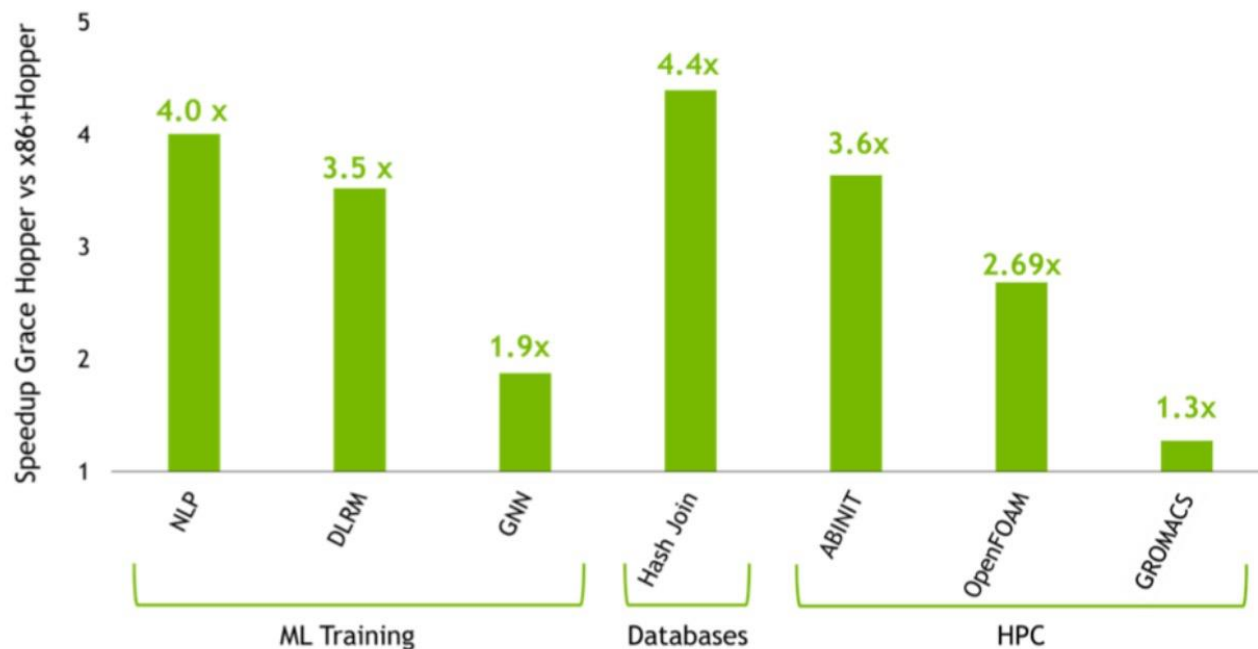


1.1.3 AMD与Nvidia争锋：MI300 vs GH200

■ 芯片架构相似，而CPU采用架构不同

- Nvidia将Grace CPU和Hopper H100 GPU融合到GH200超级芯片上，采用的是CPU+GPU相结合的架构；AMD MI300同时集成24个Zen4架构的CPU和CDNA3架构的GPU单元(具体数量未公开)，是AMD首款CPU+GPU架构的数据中心产品，两者芯片架构相似。
- Nvidia GH200中Grace CPU基于ARM架构，而AMD MI300中CPU基于x86架构，两种CPU架构在AI应用中难分优劣：Arm架构能耗较低，能为能耗不断提升的AI芯片减轻负担；而x86架构旨在通过最少的指令完成计算任务，性能较高，可以进一步赋能AI应用。Nvidia GH200与AMD MI300采用不同CPU架构都是基于将自身芯片性能最优化的考虑。

图表：Nvidia Grace Hopper与x86+Hopper终端应用性能模拟



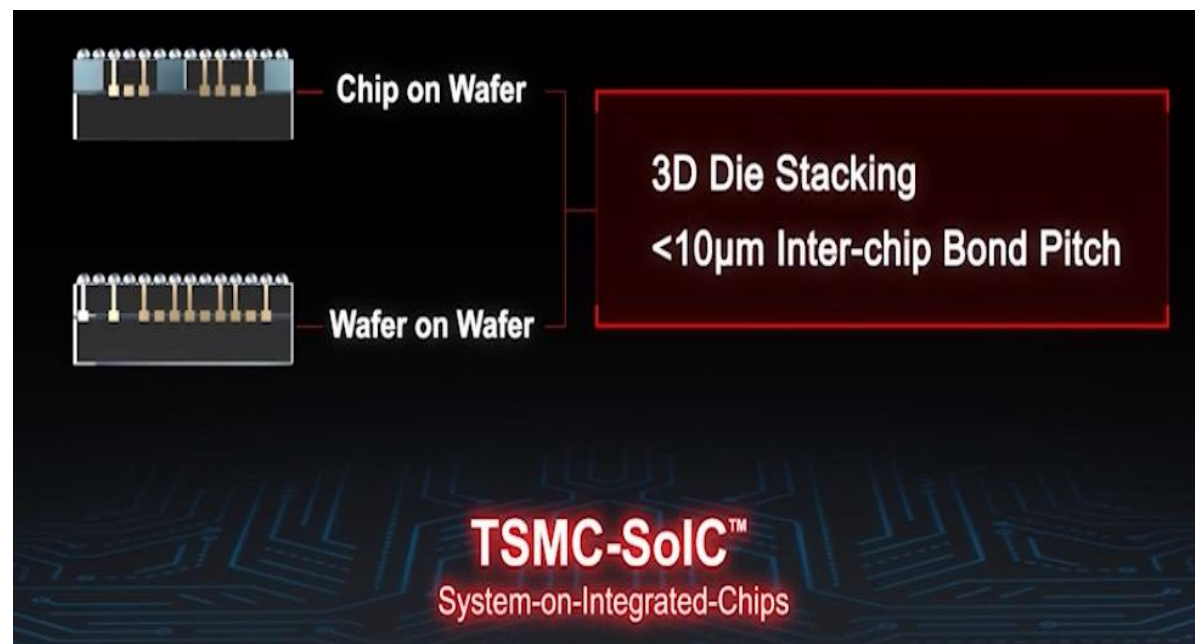
1.1.3 AMD与Nvidia争锋：MI300 vs GH200

- **AMD MI300制程技术逼近Nvidia GH200。**MI300采用3D堆叠技术和Chiplet设计，配备了9个基于5nm制程的芯片组，置于4个基于6nm制程的芯片组之上。因此在制程上，MI300属台积电5nm，相较MI200系列的6nm实现了跃迁，并与英伟达Grace Hopper的4nm制程（属台积电5nm体系）看齐。
- **AMD封装技术略微领先Nvidia。**AMD MI300使用台积电SoIC（3D）和CoWoS（2.5D）两种封装技术，而Nvidia GH200仅使用CoWoS（2.5D）封装技术。MI300采用的3D Chiplet封装技术将13个chiplets分布在两层：上层为9个基于5nm制程的chiplets，而下层为4个基于6nm制程的chiplets，芯片两侧排列8个HBM3。GH200则是通过CoWoS 2.5D工艺将Grace CPU、Hopper GPU、96GB的HBM3和512GB的LPDDR5X集成在同一封装中。3D Chiplet封装技术可以提升封装密度，并在降低能耗的基础上进一步提高性能。

图表：台积电CoWoS封装图



图表：台积电SoIC封装技术



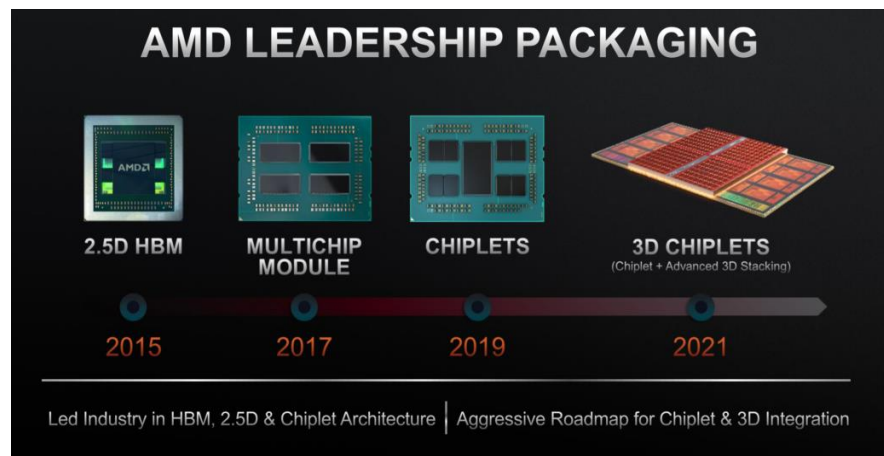


1.2.1 MI300背后：Chiplet+存算一体+异构计算

■ Chiplet+存算一体+异构计算，构建了AMD MI300强大的AI竞争力：

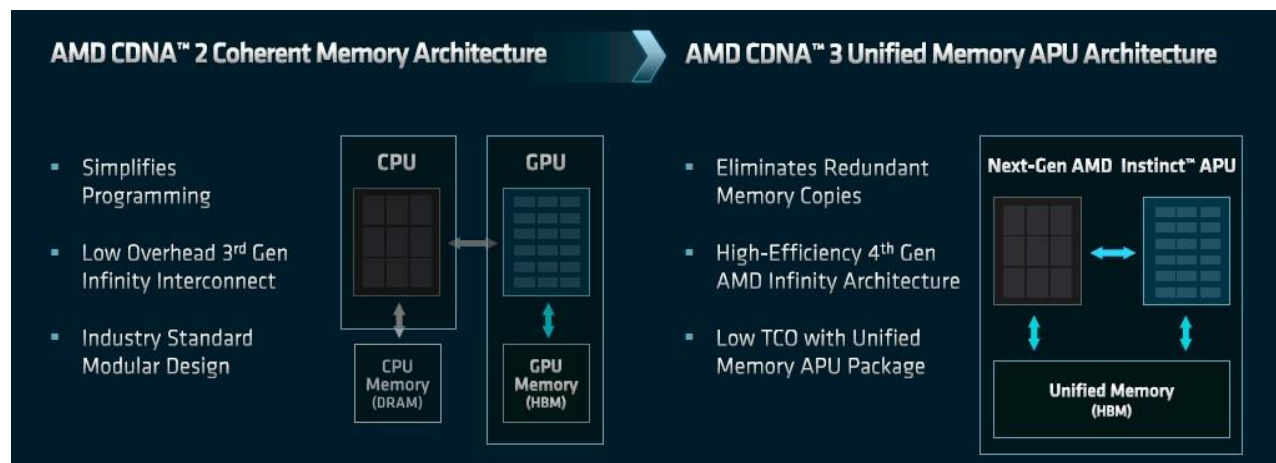
- **Chiplet**：本质是通过提升晶体管密度提升算力。MI300利用3D封装技术将CPU和加速计算单元集成在一起，增强了集成度。
- **存算一体**：缩短了计算和存储的距离，减低传输损耗、提升算力能效。CDNA是AMD的数据中心计算专用GPU架构，最新发展到第三代。CDNA2架构中，CPU和GPU各有“专属”存储，而在MI300采用的CDNA3架构中，CPU和GPU共享一块“统一存储”HBM（Unified Memory）。“共享存储”缩短了计算和存储芯片的“沟通距离”，提升了CPU与GPU间的连接速度。
- **异构计算**：**CPU+GPU+其他**，提升了算力可扩展性。MI300是AMD首款结合了Zen 4 CPU与CDNA 3 GPU的产品，也是市场上首款“CPU+GPU+内存”一体化产品。这种架构能更好满足如今AI大模型的算力需求。

图表：MI300采用AMD旗下最新的3D Chiplets封装



来源：wccfttech.com，中泰证券研究所

图表：MI300采用CDNA3共享内存架构



来源：nextplatform.com，中泰证券研究所

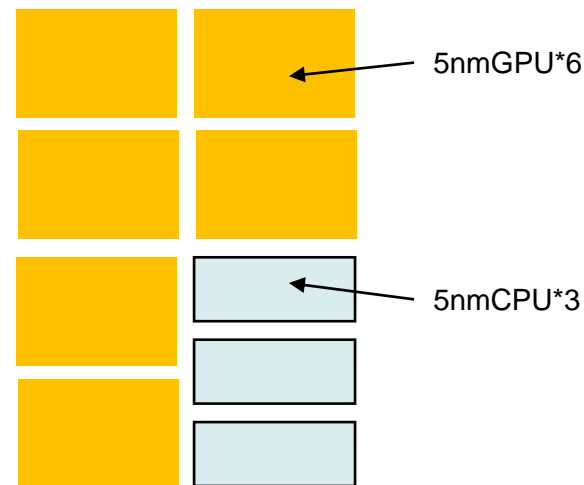
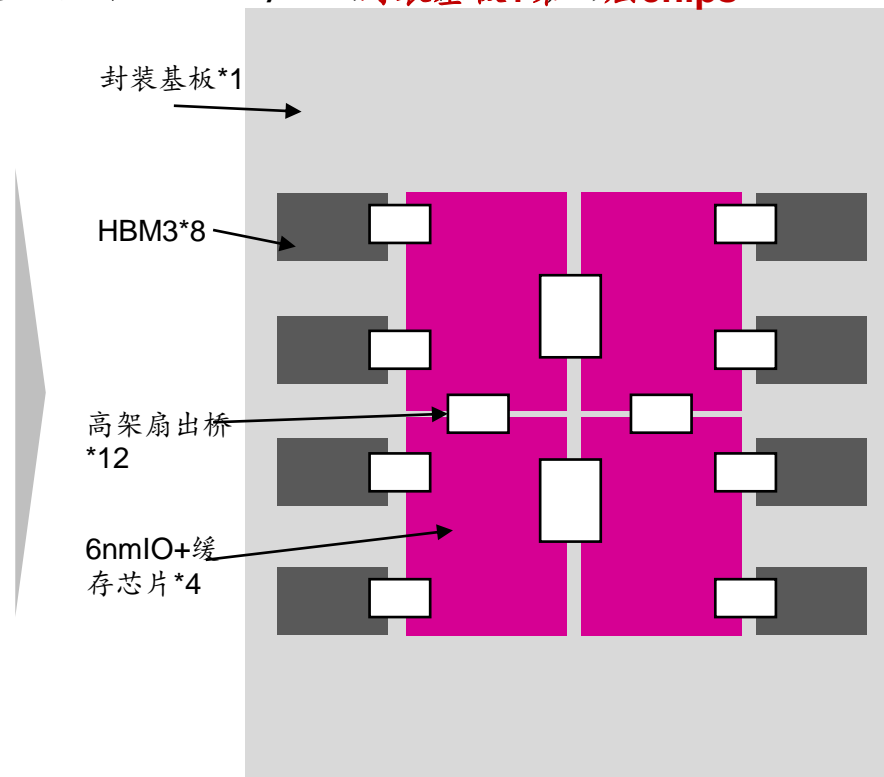
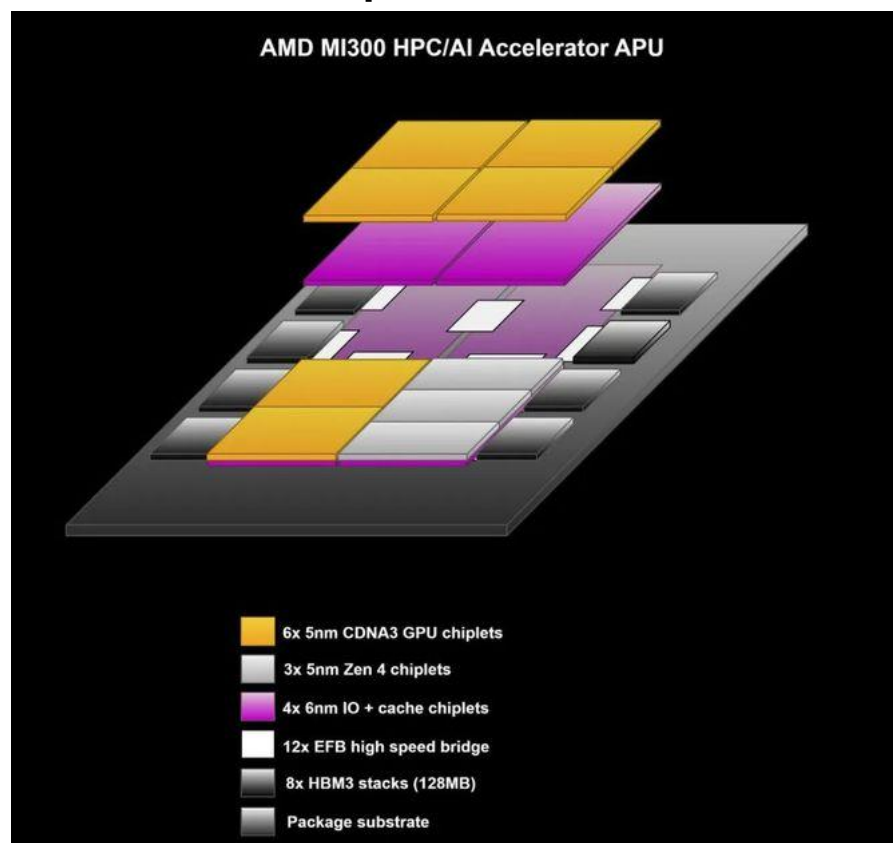


1.2.2 Chiplet: MI300A采用3D Chiplets结构

■ MI300A采用3D Chiplets结构，共有三层：

- 最下面一层-封装基板。封装基板用于承载和连接第一层小芯片（Chips）。
- 第一层Chips: I/O和存储。中间部分为4颗6nm IO/缓存一体的芯片，4颗IO/缓存芯片用高架扇出桥工艺（EFB）两两互联——共需4个连接桥。左右两边各有4颗128MB的HBM3存储芯片，HBM均与相邻的IO/缓存芯片连接，共需8个连接桥。

图表：MI300A的3D chiplets封装结构 第二层Chips覆盖在第一层IO/缓存芯片上方，其中包括5颗GPU、3颗CPU，制程均为5nm。



来源：EETOP半导体社区，中泰证券研究所



1.2.2 Chiplet: 助MI300实现超千亿晶体管集成

- **Chiplet帮助MI300A实现超千亿晶体管集成**：利用3D Chiplets，MI300A集成了6颗GPU、3颗CPU、4颗IO/缓存芯片——共计13个小芯片（Chips），最终晶体管数量达到1460亿个——这一数量远超英伟达数据中心级GPU H100的800亿个。
- **凭借庞大晶体管数量，MI300性能和AI训练能力大幅提升**。据AMD官方，在AI训练能力上，MI300的性能表现是MI250X（MI系列上一代产品）的8倍以上；在单瓦AI性能上，MI300是MI250X的5倍以上。据《Tom's Hardwre》，MI300 能将ChatGPT、DALL·E等大模型的训练时间从几个月缩短至几周，从而替客户节省数百万美元的电费。

图表：MI300的AI性能是MI250X的8倍以上

AMD INSTINCT™ MI300
THE WORLD'S FIRST DATA CENTER APU

- 4th Gen AMD Infinity Architecture: AMD CDNA™ 3 and EPYC™ CPU "Zen 4" Together
CPU and GPU cores share a unified on-package pool of memory
- Groundbreaking 3D Packaging
CPU | GPU | Cache | HBM
- Designed for Leadership Memory Bandwidth and Application Latency
- APU Architecture Designed for Power Savings Compared to Discrete Implementation

Available **2023**

>8X
Expected AI Training Performance
vs. MI250X

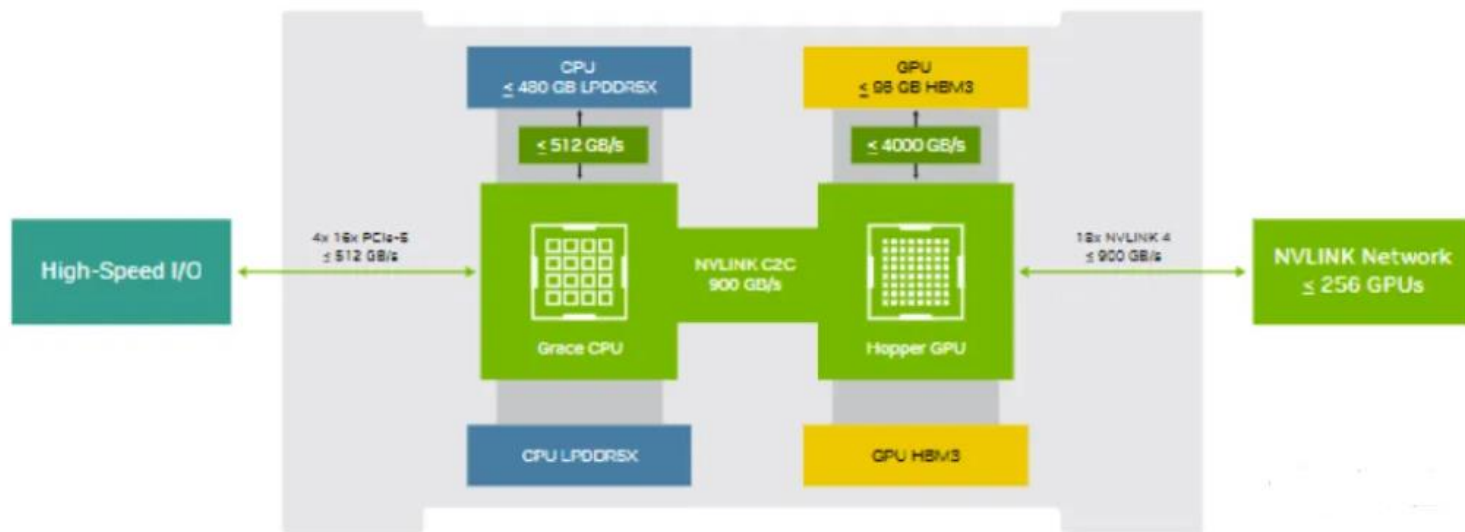
来源：PCEVA评测室，中泰证券研究所



1.2.2 Chiplet: 助英伟达GH200实现更强AI算力

- **CoWoS技术助力GH200实现更强AI算力。** CoWoS是Chiplet技术的一种。英伟达GH200采用CoWoS技术，实现CPU和GPU之间的连接，替换掉传统的PCIe连接方式，并在整体上实现72核Grace CPU+H100 GPU+96GB HBM3+512GB LPDDR5X集成在同一个封装中。与最新的PCIe Gen5技术相比，采用CoWoS技术的GH200其CPU和GPU带宽提高7倍、互连功耗减少5倍以上。而集成了256个GH200的DGX GH200系统，则成为专为高端AI和HPC负载而设计的系统和参考架构，可支持数万亿参数AI大模型训练。

图表：DGX GH200 结构示意图

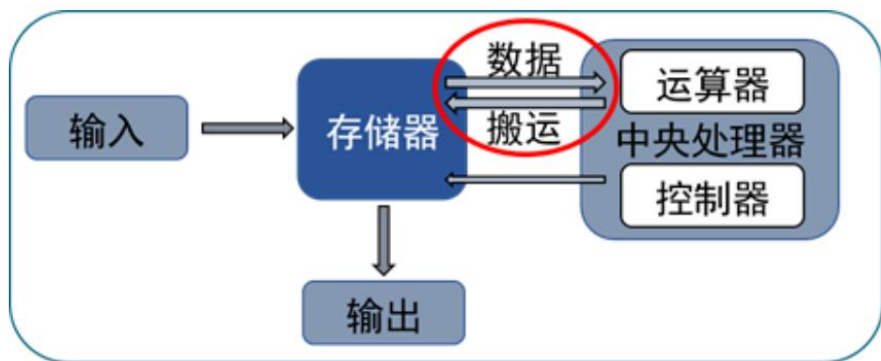


来源：芯榜，中泰证券研究所

1.2.3 存算一体：打破存储墙是AI时代刚需

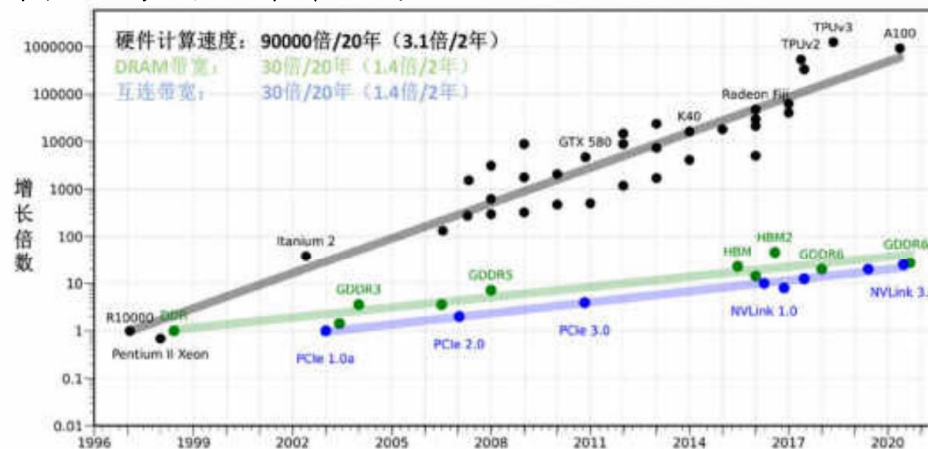
- 传统冯诺依曼架构下，存算剪刀差会造就“存储墙”。传统冯诺依曼架构中，处理器和存储器分别完成计算和存储功能。从发展趋势看，处理器跟随摩尔定律不断提升计算能力，存储器则更注重容量提升和成本优化，处理器和存储器之间性能的“剪刀差”逐渐拉大，导致访存带宽低、时延长、功耗高——即所谓“存储墙”的问题。据九天睿芯董事长，冯诺依曼架构中，处理器从存储器提取数据，搬运时间往往是计算时间的成百上千倍。
- 存算一体可有效克服冯诺依曼架构，迎合AI时代提升算力能效的需求。存算一体是一种新的架构，其核心理念是将计算和存储融合，降低“存储墙”带来的“存储空耗”问题，实现计算能效的数量级提升。从广义而言，存算一体可分为三种：近存计算（PNM）、存内处理（PIM）、存内计算（CIM），狭义的存算一体主要指存内计算。

图表：传统冯诺依曼架构



来源：《中国科学：信息科学》，中泰证券研究所

图表：存储和计算“剪刀差”



来源：中国移动研究院《存算一体白皮书（2022）》，中泰证券研究所

1.2.3 存算一体：三种类型

■ 广义的存算一体有三种类型：

- 近存计算PNM，核心是将存储和计算单元集成，提升计算能效。PNM可分为存储上移，典型产品为HBM（高带宽内存），以及计算下移，典型产品为CS（可计算存储）。
- 存内处理PIM，核心是存储器具备一定计算能力。芯片制造过程中，将存储和计算单元集成在同一颗die上，使得存储器本身具备一定计算能力。典型产品为HBM-PIM，即DRAM中具备一定的计算能力。
- 存内计算CIM，真正意义上实现了同一个晶体管同时具备存储和计算能力。即在芯片设计过程中，存储电路同时具备存储和计算能力，使得计算效能实现数量级提升。典型产品为存内计算（IMC，In-memory Computing）芯片。

图表：存算一体的三种类型

类型	英文全称	英文简称	方法与原理	典型产品	示意图
近存计算	processing near memory	PNM	<p>通过芯片封装和板卡组装，将存储单元和计算单元集成，增加访存带宽、减少数据“搬运”，提升计算效率。</p> <p>又可细分为：</p> <p>→ 存储上移，用先进封装，使得存储器向处理器（CPU、GPU）靠近，增加计算和存储间的链路数量、增加带宽。</p> <p>→ 计算下移，利用板卡集成技术，在存储设备引入计算引擎，承担如数据压缩、搜索、视频转码等本地处理，减少远端处理器负担。</p>	<p>存储上移：HBM（高带宽内存）；</p> <p>计算下移：CSD（可计算存储）；</p>	
存内处理	processing in memory	PIM	<p>在芯片制造时，将存和算集成在同一颗晶粒（die）上，存储器本身即具备一定计算能力。比如在DRAM中内置处理单元，提供大吞吐低延迟片上处理能力。</p>	HBM-PIM、PIM-DIMM	
存内计算	computing in memory	CIM	<p>芯片设计过程中，不再区分存储和计算单元，存储电路重新设计后同时具备存储和计算能力，达到计算能效数量级提升</p>	存内计算（IMC）芯片	



1.2.3 存算一体：MI300和GH200均采用HBM3

- **MI300采用HBM3**，属于存算一体类型中的近存计算PNM。如前文所述，MI300采用8颗HBM3芯片，HBM为近存计算-存储上移类型下的典型产品。就MI300自身而言，通过CPU&GPU-下层IO/缓存芯片-HBM3的路径，实现对HBM的共享。
- **MI300采用近存计算**，充分发挥了**HBM的高带宽性能**。HBM3的带宽高达819GB/s，与英伟达H100 900GB/s的带宽相差不多。由于CPU&GPU共享HBM，二者可以针对HBM中相同的数据同时展开计算，从而减少内存带宽的占用——这种机制有效利用了HBM3高速带宽的特性，增强了CPU与GPU之间的协作效率。
- **存算一体概念在英伟达AI产品上亦有所体现**。英伟达GH200亦采用HBM3，容量高达96GB，其中HBM3与Hopper GPU的带宽最高可达4000GB/s。

图表：四代HBM规格比较

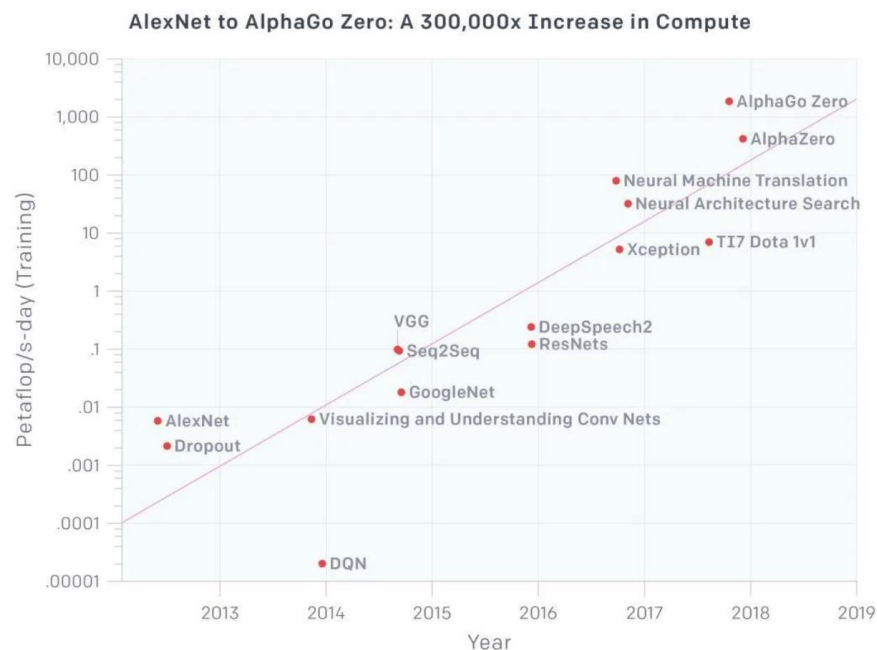
类别	HBM1	HBM2	HBM2E	HBM3
带宽	128GB/s	307GB/s	460GB/s	819GB/s
堆叠高度	4层	4层/8层	4层/8层	8层/12层
容量	1GB	4GB/8GB	8GB/16GB	16GB/24GB
I/O速率	1Gbps	2.4Gbps	3.6Gbps	6.4Gbps

来源：全球半导体观察整理，中泰证券研究所注：以SK海力士产品为例

1.2.4 异构计算：提升AI算力可拓展性

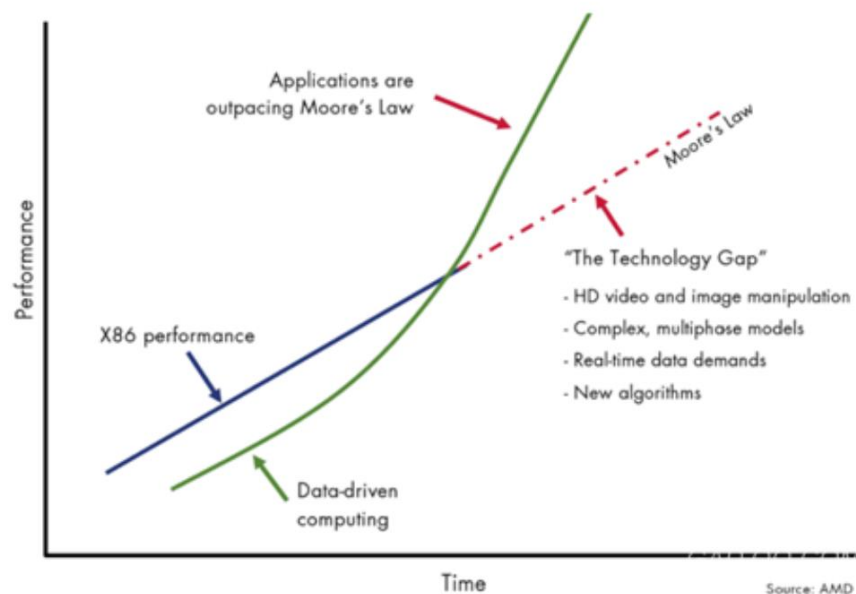
- **AI飞速发展VS摩尔定律趋缓**，当前处理器性能无法充分满足高性能计算需求。据OpenAI统计，2012年以来AI训练中使用的算力每3.5个月翻一倍。AI迅猛发展带来算力需求的急剧提升。与此同时，摩尔定律趋缓使得处理器性能无法再像以前一样高速增长。AI算力要求高增VS摩尔定律下进程进步趋缓，处理器性能与AI计算为代表的高性能运算（HPC）之间的缺口愈发拉大
- **异构计算是满足AI高算力需求的必然选择**。异构计算指用不同类型指令集和体系架构的计算单元，去组成计算系统。异构计算可以灵活采用CPU、GPU、ASIC、FPGA或其他加速器的优势，兼顾“通用计算高算力”与“专用计算高效率”，实现性能、成本和功耗的平衡，是支撑未来更复杂AI应用的必然选择。

图表：2013到2019年AI算力提升了30万倍



来源：OpenAI，中泰证券研究所

图表：处理器性能与算力需求的缺口愈发拉大



来源：AMD，中泰证券研究所



1.2.4 异构计算：AMD收购赛灵思，FPGA成奇兵

- **CPU+GPU**，构成AI时代异构计算的基础类型。CPU芯片侧重指挥统筹其他芯片的计算，其本身对AI计算的能力较弱，需要其他AI加速器（如GPU、ASIC、FPGA等）的协同计算，以实现更好的AI计算性能。GPU芯片由于其在大范围、多任务并行计算领域的高算力，成为AI计算的算力核心。CPU+GPU的组合，兼顾计算统筹+并行高算力突破，较好地满足了AI高算力的需求，故CPU+GPU成为AI时代异构计算的“基础架构”。
- **CPU+GPU+ASIC**，可实现更优的计算能效。与CPU和GPU这类通用芯片不同，ASIC为定制化芯片，可针对某类AI计算场景定制开发最优的算力匹配，故ASIC芯片可在单一领域达到比CPU、GPU更高的计算效率。
- **CPU+GPU+FPGA**，可灵活应对多元化AI计算需求。FPGA芯片与ASIC一样为定制化芯片，但FPGA芯片可以进行重新编程——这种特点使得AI芯片面对不同客户的不同运算场景，可以灵活适配。
- ——2022年底AMD收购赛灵思，补齐FPGA布局，未来CPU+GPU+FPGA异构计算有望进一步增强其AI芯片实力。

图表：不同类型的CPU在异构计算中的原理与区别

芯片类型	全称	基本原理	执行AI计算的表现	异构计算中的作用
CPU	中央处理器，Central Processing Unit	更偏重支持控制流数据，CPU物理核中大部分的硬件资源被做成了控制电路和缓存，用来提高指令兼容性和效率，只有小部分是真正用来做计算的逻辑运算单元（ALU）	能兼容大量指令，但实际的计算效率并不高	指挥统筹，控制核心
GPU	图形处理器，Graphics Processing Unit	GPU把晶体管更多用于计算单元、且存储带宽更高，多核架构，擅长做类似图像处理的高强度并行计算	AI计算往往是大数据流应用，基于GPU的GPGPU针对AI计算比CPU效率更高	AI深度学习、大范围&多任务简单计算
ASIC	特殊应用专用集成电路，Application Specific Integrated Circuit	定制化，编程语言越接近底层硬件，不需要复杂指令，运行速度越快、效能高	可针对某一AI算法开发算力效率最高的ASIC芯片，但通用性不足	针对定制化的计算场景，搭配CPU使用
FPGA	现场可编程门阵列，Field Programmable Gate Array	利用预建的逻辑块和可重新编程来布线资源，可针对定制化需求进行算法开发	针对不同的AI模型和算力需求，进行可重复编程的算法开发，以达到算力效率最优的匹配	针对多元化计算需求，协同CPU进行AI加速

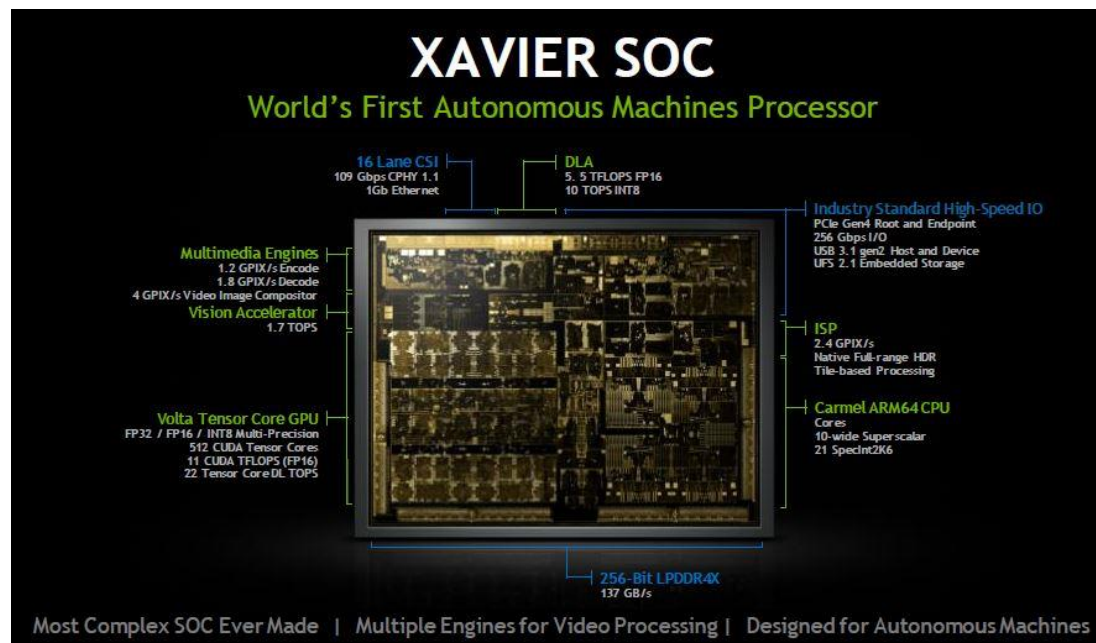
来源：IBM中国公众号，人工智能学家公众号，中泰证券研究所



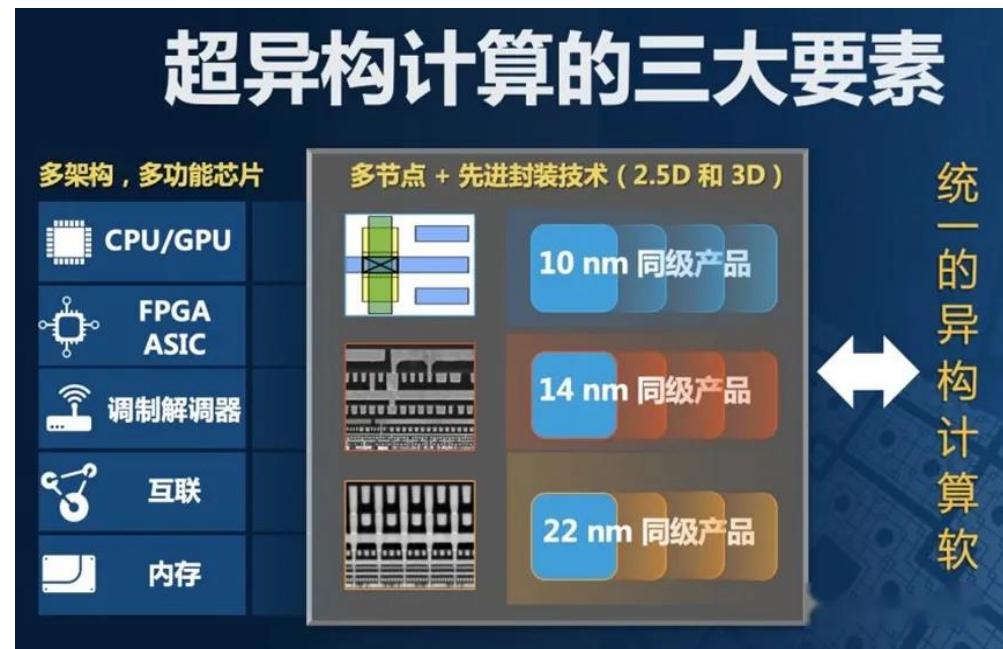
1.2.4 异构计算：英伟达和英特尔均有各自布局

- 英伟达已在自动驾驶上推出**CPU+GPU+ASIC**架构的芯片**Xavier**。英伟达首次在Xavier上采用了CPU+GPU+ASIC芯片混合技术路线，而Xavier是一款12nm工艺、专为自动驾驶设计的芯片，其中GPU包含Volta架构的512颗CUDA Core、数量占比最大；CPU为英伟达自研8核ARM64架构、占比次之，此外搭载两颗专用ASIC——分别为用于推理计算的深度学习加速模块DLA、用于传统视觉算法加速的向量处理单元PVA。整体而言，Xavier可处理视觉距离测量、传感器融合、定位和映射、障碍检测、路径规划等。
- 针对异构计算，英特尔提出**XPU、oneAPI**理念。XPU指多芯片的架构组合，而英特尔目前已布局CPU、GPU、FPGA、VPU等计算芯片，oneAPI指英特尔构建的一套开源的跨平台编程框架，底层可运行CPU、GPU、FPGA或其他DSA加速器，oneAPI可以为应用提供一致性编程接口，应用可实现跨平台复用。

图表：不同类型的CPU在异构计算中的原理与区别



图表：超异构计算的三大要素





1.2.5 Chiplet+存算一体+异构计算：突破AI算力瓶颈

■ **Chiplet+存算一体+异构计算**，有望成为突破**AI**算力瓶颈的主流路径：

- 提升密度-Chiplet异构互联提升单位面积晶体管数量；
- 提升能效-存算一体技术，缩小存储和计算“距离”，减少“空耗”，提升AI计算的能效；
- 提高扩展性-异构计算，CPU+GPU+ASIC/FPGA，兼顾高强度计算+AI算力多元化双重需求。

——此次**MI300**芯片发布，其体现出的“**Chiplet+存算一体+异构计算**”三位一体的理念，为业界提供了突破**AI**算力瓶颈三大方向的思路，其背后反映的技术路径，有望成为未来**AI**芯片发展的主要方向。

目 录

一、逐鹿AI：AMD发布最强AI芯片MI300系列

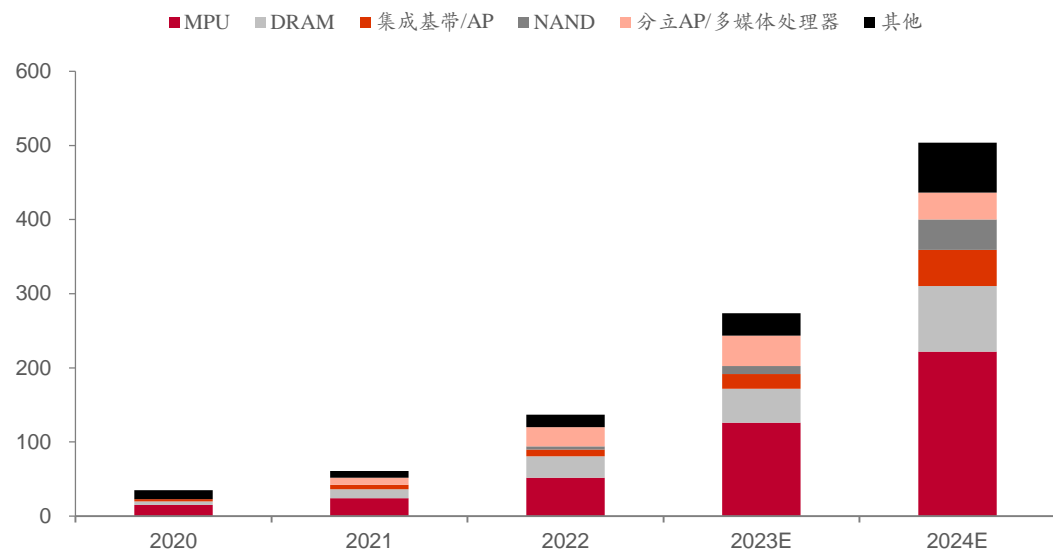
二、**Chiplet**产业链及投资建议

三、AMD产业链及投资建议

2.1 Chiplet空间：500亿美金大市场

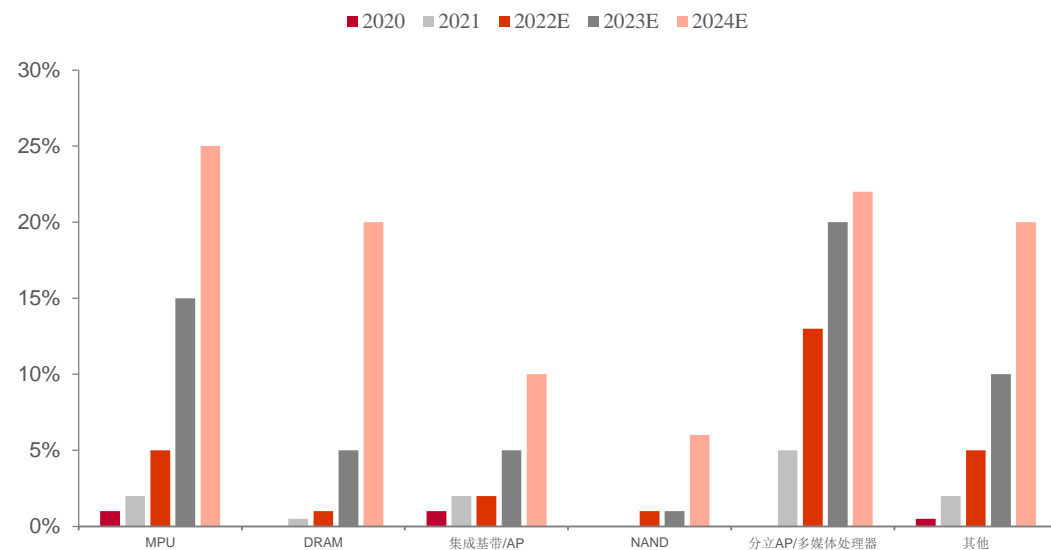
- **2024年全球Chiplet芯片市场空间有望达500亿美金。**据Gartner预测，Chiplet芯片市场在2020年空间为全球33亿美金，2024年全球超500亿美金，2020-24年全球市场CAGR为98%。其背后是Chiplet在MPU、DRAM/NAND、基带芯片上加速渗透。
- **2024年全球Chiplet封测市场空间有望达50亿美金。**2022年全球封测市场空间为469亿美金，对应全球芯片销售额在5735亿美金，封测占芯片产值的比重为8%。由于Chiplet涉及的封测难度更高、相对应地在芯片产值中价值量占比会更高，假设2024年全球Chiplet芯片市场中，Chiplet封测成本占比为10%，则对应50亿美金的市场空间。

图表：2020-24年全球各类Chiplet空间（亿美元）



来源：Gartner，中泰证券研究所

图表：2020-24年各类芯片Chiplet渗透率



来源：Gartner，华经产业研究院，IC Insights，共研网，Yole，QY Research，中泰证券研究所

2.2 封测：国产龙头已量产Chiplet

- 国产封测龙头，在**Chiplet**领域已实现技术布局：
 - 通富微电已为AMD大规模量产Chiplet产品；
 - 长电科技早在2018年即布局Chiplet相关技术，2022年成为中国大陆首家加入UCle产业联盟的封测厂商；公司已通过国内外多家生产型子公司为海内外多家不同客户规模量产Chiplet产品；
 - 华天科技Chiplet技术已实现量产，其他中小封测厂商已有在TSV等Chiplet前期技术上的积累。

图表：国产封测厂商在**Chiplet**方面的布局

公司	Chiplet技术积累	具体商业进展
通富微电	在多芯片组件、集成扇出封装、2.5D/3D等先进封装技术方面的提前布局，可为客户提供多样化的Chiplet封装解决方案	已为AMD大规模量产Chiplet产品
长电科技	1) 2018年启动研发以RDL First为基础的高密度扇出型多芯片Fan-out FCBGA，具备大尺寸、2μm线宽、高密度倒装键合、2-3粒7nm芯片集成等特点 2) 2021年突破FO-Interposer MCM的chiplet封装技术，并进入量产 3) 2022年突破2.5D Si Interposer MEM的大尺寸FCBGA技术，并进入小批量量产，同年成为中国大陆首家加入UCle产业联盟的封测厂商	已量产
华天科技	已具备chiplet封装技术平台，并已量产	已量产
晶方科技	晶圆级TSV技术是chiplet技术重要组成部分，晶方科技在研究该技术方向	技术积累阶段
大港股份	控股孙公司苏州科阳主要从事TSV晶圆级封装业务，截至22年11月未涉及Chiplet相关业务	尚未直接涉及Chiplet，但已布局相关的TSV技术

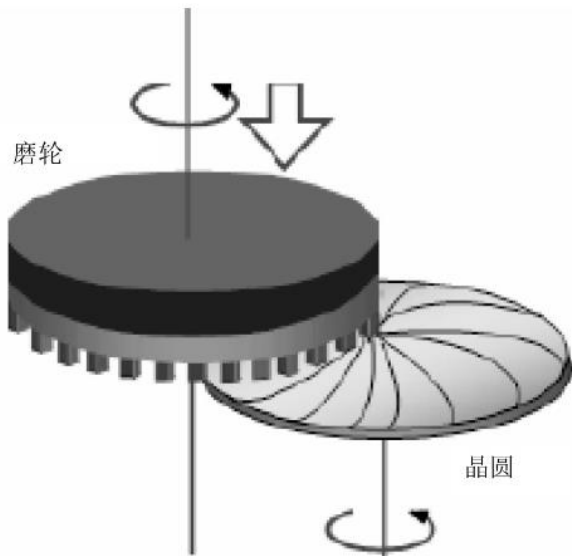
来源：各家官网，互动易平台，中泰证券研究所

2.3 减薄：Chiplet新增晶圆减薄需求

■ Chiplet兴起创造全新的减薄机应用场景：

- 传统减薄机，用于后道封装，主要涉及对芯片背面硅片的减薄、注塑后对塑胶表面的磨平。
- 以Chiplet为代表的3D IC兴起后，前道环节新增对减薄抛光机的需求——因为3D IC涉及到多颗晶圆的堆叠，为了降低堆叠后芯片组厚度，需要对晶圆进行减薄。
- 华海清科减薄机**2023年有望批量出货**。公司针对3D IC领域的减薄抛光一体机已发到客户端进行验证，且验证情况良好，预计该类产品在2023年实现批量出货。除此之外公司还拓展了针对封装领域的减薄机型，目前也处于验证阶段。在技术层面，公司减薄类设备可以对标国际友商的高端机型。

图表：晶圆减薄示意图

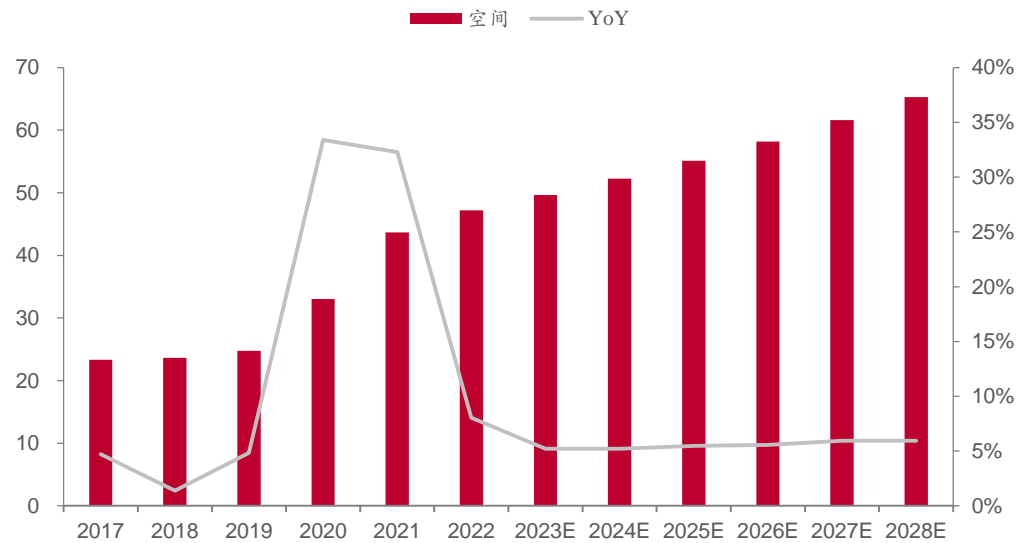


来源：鼎达信公司官网，中泰证券研究所

2.4 ABF载板：Chiplet材料国产化空间广

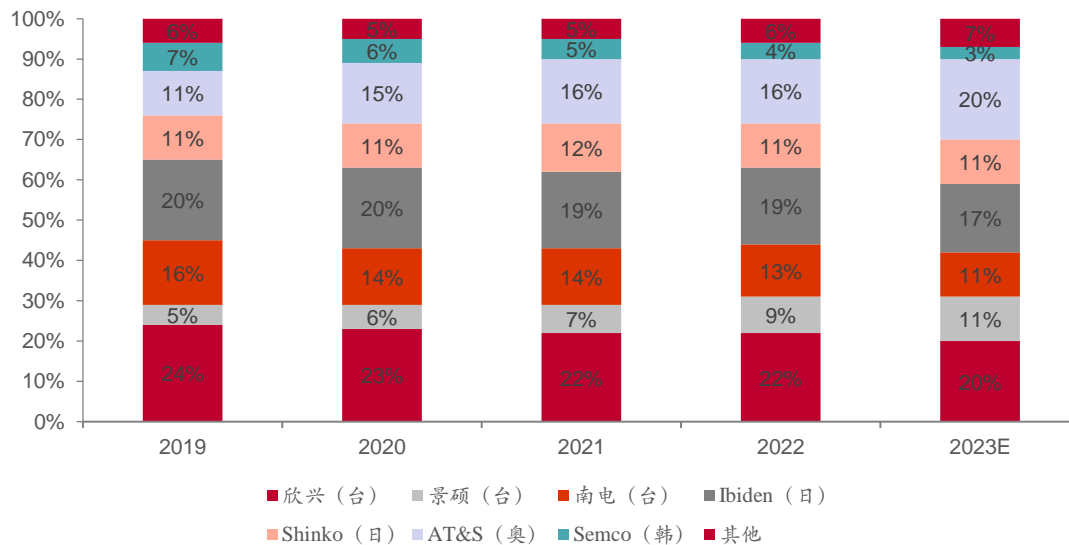
- **Chiplet高增带动ABF载板需求提升：**
- **Chiplet芯片高算力特性，更适合于ABF载板应用。**依据材料，可将IC载板分为BT载板和ABF载板。相较于BT载板，ABF材料可做线路较细、适合针脚数更多的高讯息传输IC，主要用于CPU、GPU、FPGA、ASIC等高算力芯片。
- **2028年全球ABF载板市场有望达65亿美元。**随着HPC发展和Chiplet渗透，ABF载板市场迎来快速增长。据QYResearch，2023年全球ABF载板市场有望达50亿美元，至2028年有望增至65亿美元。
- **中国台湾、日本企业占据全球7成份额，国产化空间广阔。**据半导体行业观察，中国台湾的欣兴、景硕、南电和日本的Ibiden、Shinko共占据全球ABF载板市场近70%份额。中国大陆厂商面临广阔替代空间。

图表：2017-2028年全球ABF载板空间（单位：亿美元）



来源：QYResearch，中泰证券研究所

图表：2019-2023年全球ABF载板竞争格局



来源：半导体行业观察公众号，中泰证券研究所

2.4 ABF载板：Chiplet材料国产化空间广

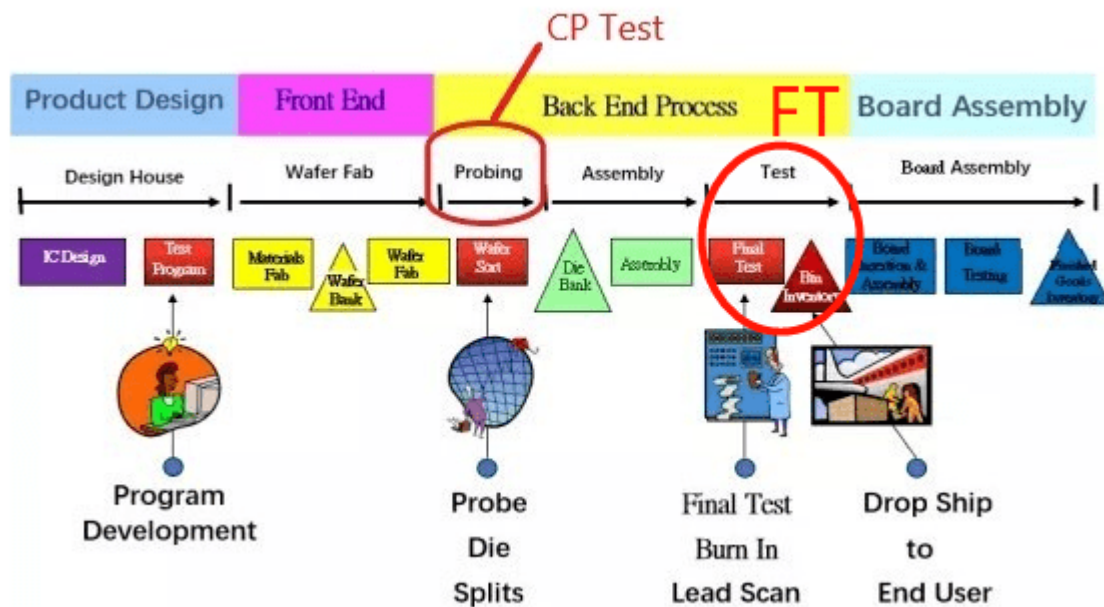
- 国产厂商陆续布局，有望充分受益Chiplet浪潮下的ABF国产化：
 - 兴森科技：ABF载板是战略方向。公司拟在广州投资约60亿元建设FCBGA封装基板生产和研发基地项目，分两期建设月产能为2000万颗的FCBGA封装基板智能化工厂。广州项目计划2023年底进入试产。
 - 深南电路：2021年成立子公司广州广芯，计划投入约人民币60亿元，计划达成FCBGA载板年产能2亿颗；据公司2022年9月30日公告，使用ABF材料的FC-BGA封装基板处于样品研发阶段。



2.5 设备/第三方测试：Chiplet倍增CP测试需求

- **Chiplet倍增CP测试需求。**Chiplet需要对每一颗Chiplet die进行封装前的CP测试（circuit probing），否则封装后任一颗die失效，将会使整个芯片组失效，成本代价高。相比于单芯片SoC测试，Chiplet对芯片的CP测试需求按照小芯片数量成倍增加，进而拉动了对测试服务、测试机的需求。
- **国产第三方测试厂商有望受益：**伟测科技（产能高增的国产第三方测试龙头）、利扬芯片（老牌国产测试龙头）、华岭股份（立足长三角的国产测试龙头）。
- **国产测试机厂商有望受益：**长川科技（国产数字测试机龙头）、华峰测控（国产模拟测试机龙头）、联动科技（功率测试能力突出）。

图表：Chiplet对测试需求的增加主要体现在CP测试环节



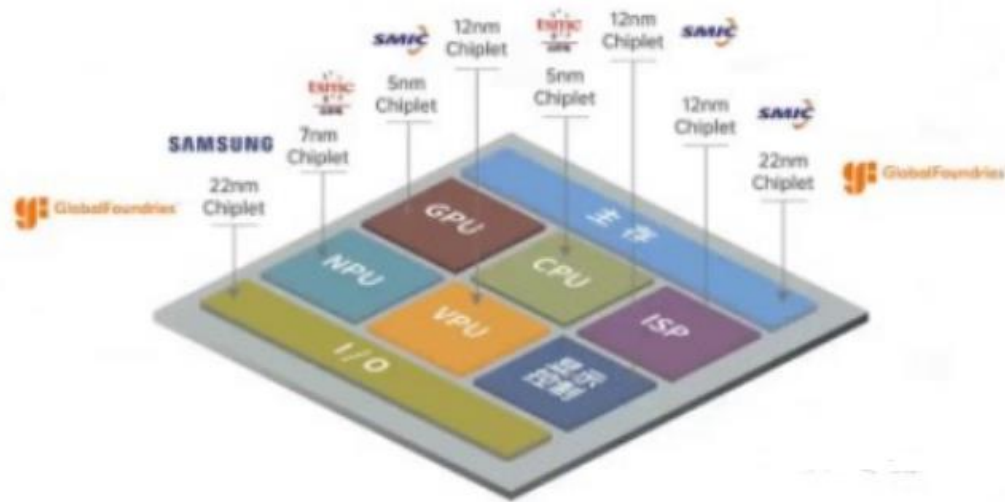
来源：微波射频网搜狐号，中泰证券研究所

2.6 IP: Chiplet释放芯片IP化需求

■ Chiplet释放芯片IP化需求：

- **Chiplet释放IP复用增量需求。**Chiplet将大芯片“化整为零”，单颗Chip本质上是IP硬件化，Chiplet封装可以看作是多颗硬件化的IP的集成。后续Chiplet芯片的升级，可以选择仅升级部分IP单元对应的Chip，部分IP保留——从而实现IP复用，大幅缩短产品上市周期。
- **芯原股份作为国内IP供应龙头，有望受益于Chiplet发展。**据Ipnest，芯原股份是中国大陆排名第一、全球排名第七的半导体IP供应商，拥有图形/神经网络/视频/数字信号/图像/显示六大类处理器IP核，并具备领先的芯片设计能力，近年来一直致力于Chiplet技术和产业的推进。2022年芯原推出了基于Chiplet架构所设计的高端应用处理器平台，该平台12nm SoC版本已完成流片和验证。

图表：基于Chiplet异构架构处理器芯片示意图



来源：芯原股份，中泰证券研究所

2.7 EDA：芯片重构带来国产导入良机

- **Chiplet带来EDA设计的新挑战：**1) 多颗小芯片集成，电、热、力、机械等多种物理量下如何保证芯片设计指标和工作状态正常；2) 需要对多芯片进行整合式EDA验证；3) 设计和封装融为一体，EDA端要对应适配；
- **芯片重构带来国产EDA导入良机：**1) 建立统一的EDA设计工具的国产统一标准；2) 不同国产芯片的协同整合测试；3) 国产封测厂在PCB绕线、数字绕线、模拟绕线上可以给国产EDA提供经验；4) 国产fab厂制定统一的多芯片互联标准。——以上四点，国产EDA厂商均有地理和自主可控方便的优势，也是借助Chiplet打开国产EDA局面的良机。

图表：国产EDA公司一览

公司	公司logo	地区	IPO情况	股东	创始人等	技术背景
华大九天		北京	已上市	CEC、大基金	刘伟平、陈丽洁等	熊猫EDA
概伦电子		上海	已上市	兴橙资本、英特尔	潘建岳（新思系）、倪捷（Synopsys）等	引领存储EDA
国微集团		深圳	港股上市	深圳第一家半导体公司908/909工程	刘志宏	紫光核心企业
国微思尔芯		上海	IPO终止	国微集团	黄学良（国微集团）	技术转让
广立微		杭州	已上市	中芯聚源、武岳峰	国内四大EDA	领先的集成电路EDA软件与晶圆级电性测试设备供应商
芯愿景		北京	IPO终止	宁波梅山保税港区宜安投资、宁波梅山保税港区丰年君和投资	蒋卫军、张军、丁仲等	中科院背景

来源：亿渡数据，中泰证券研究所

目录

一、逐鹿AI：AMD发布最强AI芯片MI300系列

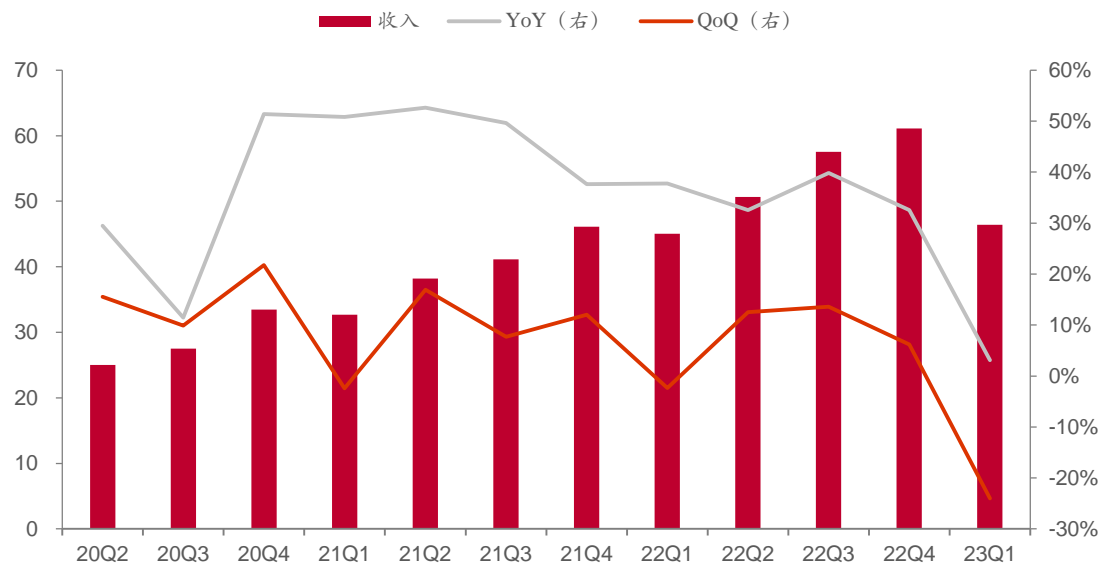
二、Chiplet产业链及投资建议

三、AMD产业链及投资建议

3.1 封测——通富微电

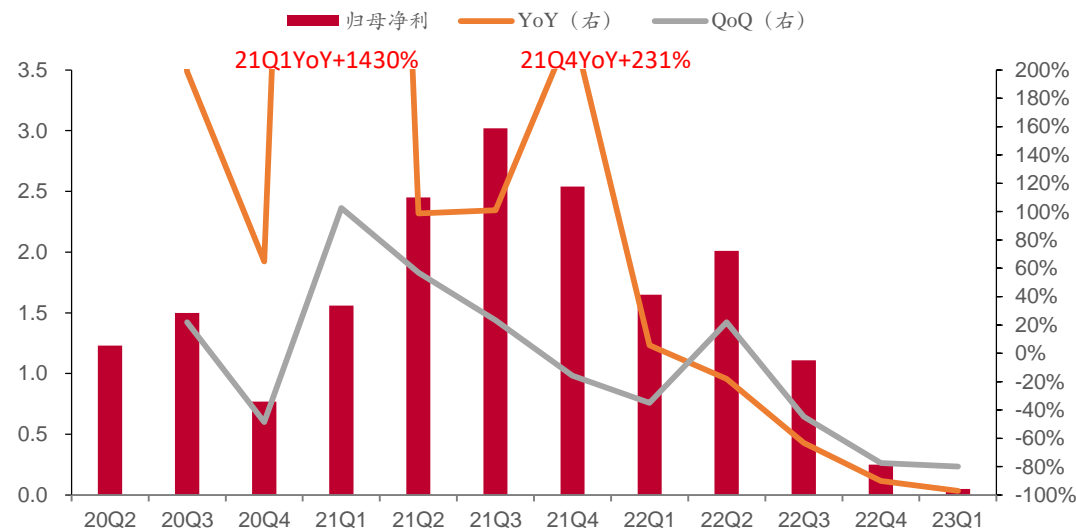
- 通富参与MI300封测，紧跟AMD充分受益AI浪潮。AMD在1月初的CES 2023展会上推出MI300，计算内核为台积电5nm工艺。通富微电在1月11日即表示，公司涉及MI300的封测——可见通富与AMD在前沿封装上合作紧密。3月2日，通富微电表示，公司开启“立足7nm、进阶5nm”的战略，5nm产品完成研发逐步量产，助力大客户高端进阶，并有信心满足大客户制程进阶后的封测需求。——通富微电为AMD最大的封测供应商，占其订单总数80%以上。公司坚持大客户战略，未来有望充分受益于AI芯片Chiplet封装浪潮。
- 稼动率触底，静候景气回升。受行业景气下行影响，公司23Q1稼动率维持22Q4以来的底部位置。行业龙头日月光表示，预期H2稼动率和景气有望进一步回升。封测作为重资产行业，稼动率回升通常带来更大毛利弹性，乐观看待通富微电H2净利环比改善。

图表：20Q2-23Q1通富微电收入及同环比增速（单位：亿元）



来源：Wind，中泰证券研究所

图表：20Q2-23Q1通富微电归母净利润及同环比增速（单位：亿元）




来源：Wind，中泰证券研究所

3.2 PCB ——沪电股份

- 高端PCB龙头公司，服务器+交换机营收占比近60%：公司为数通板龙头供应商，技术能力处于全球第一梯队，我们预计沪电股份服务器+交换机业务营收占比接近60%，深度受益ChatGPT带来的服务器/交换机平台升级，Pcie5.0为代表的通用平台服务器及英伟达/AMD为代表的AI服务器用板更符合其产品定位，预计未来服务器营收占比进一步提升。
- 不断布局高端产品，未来增长动力强：公司应用于EGS级服务器领域的产品已实现规模化量产；HPC领域，公司布局通用计算，应用于AI加速、Graphics的产品，应用于GPU、OAM、FPGA等加速模块类的产品以及应用于UBB、BaseBoard的产品已批量出货；在高阶数据中心交换机领域，应用于Pre800G的产品已批量生产，应用于800G的产品已实现小批量的交付；基于数据中心加速模块的多阶HDI Interposer产品，已实现4阶HDI的产品化，目前在预研6阶HDI产品，同时基于交换、路由的NPO/CPO架构的Interposer产品也同步开始预研；在半导体芯片测试线路板部分重点开发0.35mm以上Pitch的高阶产品，不断布局高端产品，保持公司竞争优势。

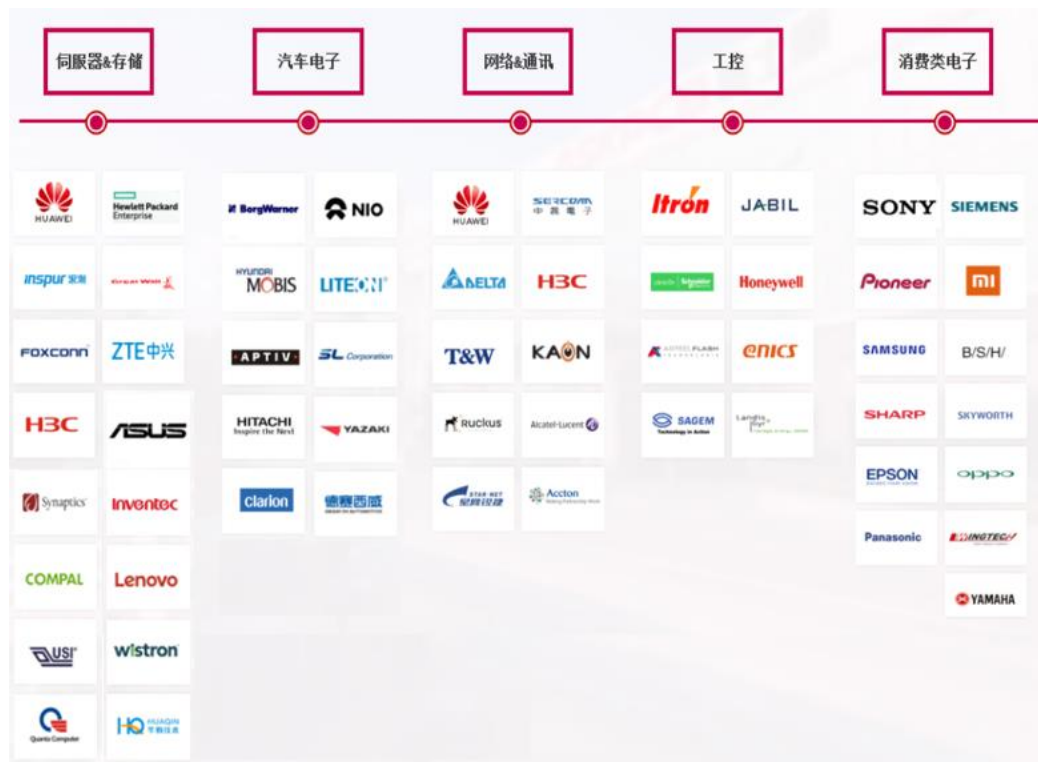
图表：服务器平台升级对PCB要求提升

Intel	Platform	Purley		Whitley		Eagle Stream	
	CPU	Skylake	Cascade Lake	Copper Lake	Ice lake	Sapphire Rapids	Emerald Rapids
	Nano Process	14 nm	14 nm+	14 nm++	10 nm	Intel 7	Intel 7
	PCIe Generation	PCIe 3.0	PCIe 3.0	PCIe 3.0	PCIe 4.0	PCIe 5.0	PCIe 5.0
	MP Time	2017 Q3	2019 Q3	Cancel	2021 Q1	2022 H2	2023
	CCL Material	Mid Loss	Mid Loss	Cancel	Low Loss	Very Low Loss	Very Low Loss
	Layer count	8 to 12	8 to 12	Cancel	12 to 16	16 to 20	16 to 20
AMD	Architecture	Zen	Zen2	Zen3	Zen4		
	CPU	Naples	Rome	Milan	Genoa		
	Nano Process	14 nm (Global Foundries)	7 nm (TSMC)	7 nm (TSMC)	5 nm (TSMC)		
	PCIe Generation	PCIe 3.0	PCIe 4.0	PCIe 4.0	PCIe 5.0		
	MP Time	2017 Q3	2019 Q3	2020 Q4	2022 H2		
	CCL Material	Mid Loss	Low Loss	Low Loss	Very Low Loss		
	Layer count	8 to 12	12 to 16	12 to 16	16 to 20		

3.2 PCB —— 奥士康

- **AMD/英伟达供应商，深度受益算力升级：**公司与AMD已达成战略合作，AMD MI300对标英伟达H100，AMD进一步加强AI布局，公司作为战略合作供应商深度收益；此外奥士康已成功进入英伟达供应链，公司正积极参与英伟达新产品的打样和测试工作，预计后续产品将持续放量，为公司贡献高成长；此外通用服务器公司进展顺利，积极参与新一代服务器平台的打样与认证，预计后续将持续为公司贡献利润。
- **算力提升拉动交换机需求，公司已成为海外龙头AWS供应商：**算力提升带动交换机需求提升，奥士康目前已出货交换机产品，其合作客户包括AWS等，AWS公有云份额全球第一，其在印度云基础设施投资超预期，16-22年在印度共投资了37亿美元，计划到30年之前增加到127亿美元。算力升级交换机进一步放量带动交换机PCB需求激增，公司深度受益。

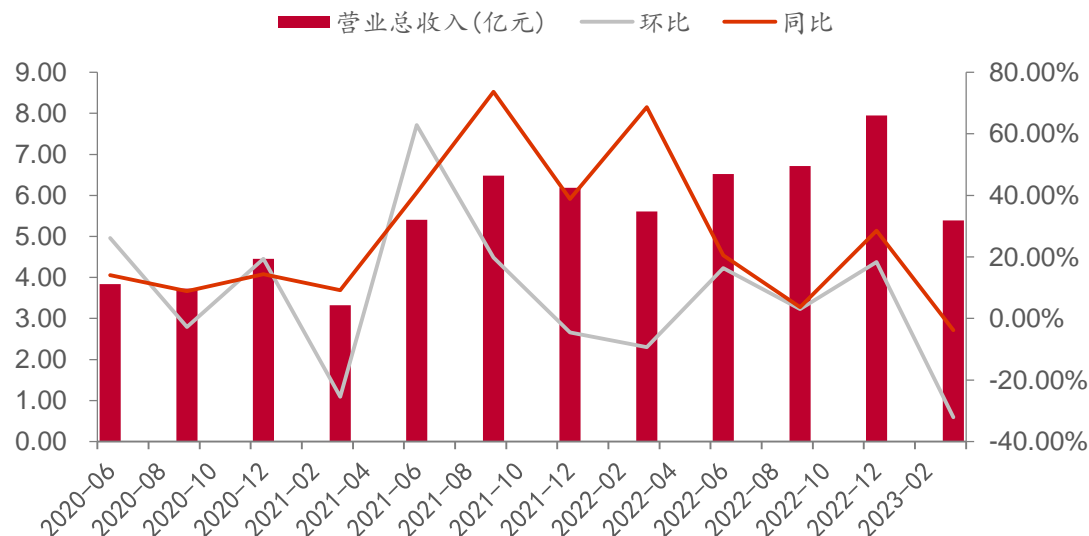
图表：公司各下游客户情况



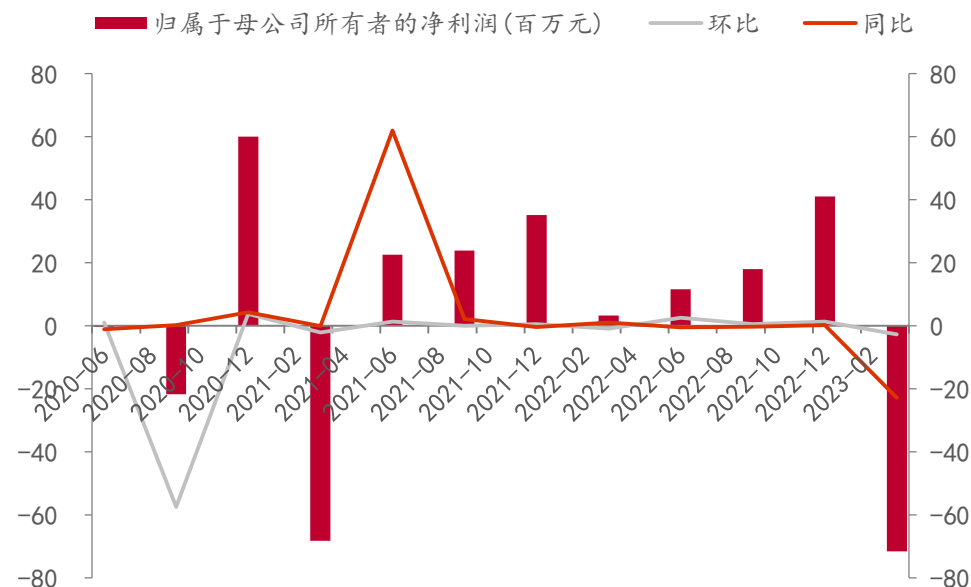
3.3 IP和设计服务——芯原股份

- 芯原股份成立于2001年，总部位于上海，是一家依托自主半导体IP，为客户提供平台化、全方位、一站式芯片定制服务和半导体IP授权服务的企业。通过基于公司自主半导体IP搭建的技术平台，芯原可在短时间内打造出从定义到测试封装完成的半导体产品。芯原拥有6类自主可控的处理器IP，分别为图形处理器IP、神经网络处理器IP、视频处理器IP、数字信号处理器IP、图像信号处理器IP和显示处理器IP，及1,500多个数模混合IP和射频IP。
- 芯原股份通过IP和设计服务支持AMD 成功量产基于5nm 制程工艺的Alveo MA35D ASIC 媒体加速芯片，该芯片可支持AI优化视频质量，显示公司在AI 芯片深厚积累及高市场认可度。2022年3月，AMD及Intel等公司组建了UCle 产业联盟，将推动Chiplet接口规范的标准化的，现在芯原股份也加入了该联盟，成为第一个加入的国产芯片厂商。

图表：芯原股份单季度营业收入及同比、环比增速



图表：芯原股份单季度归母净利润及同比、环比增速

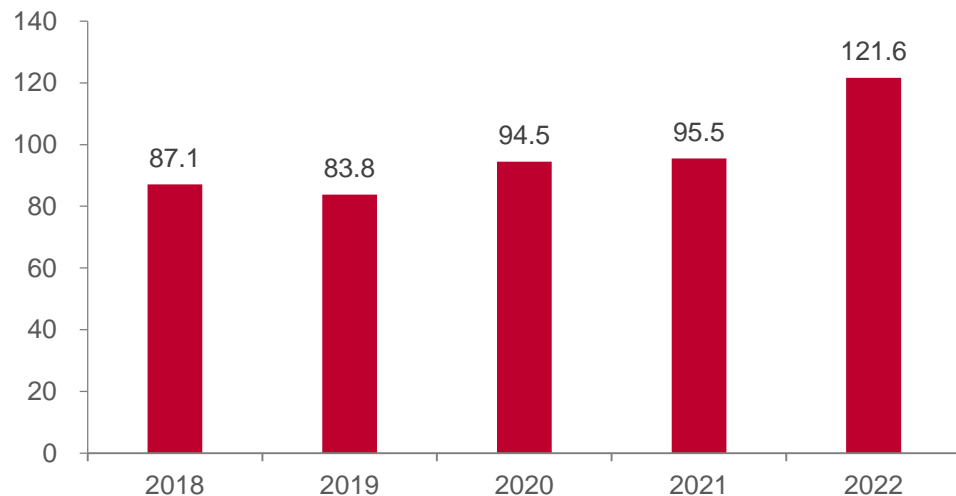


3.4 服务器——工业富联

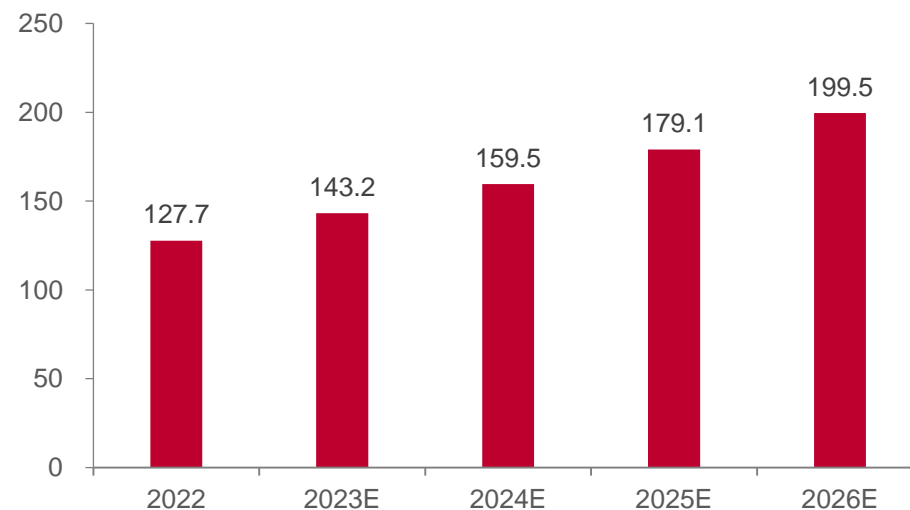
■ 工业富联与海外核心服务器厂商深度合作，将充分受益AI服务器需求的快速增长

- 服务器收入占比高，绑定全球核心大客户。2022年云服务设备收入占比超过40%，其中大部分来自服务器，云计算服务器出货量持续全球第一，客户涵盖微软、谷歌、Amazon，并与英特尔、英伟达紧密合作。公司携手客户不断实现技术突破，2022年10月推出全球第一台 DC_MHS 服务器，并将新一代液冷技术应用在客户最新 HPC 加速器中。
- AI服务器需求快速增长，工业富联充分受益：2022年，全球服务器市场规模达到约1216亿美金，整体出货量达到1516万台，其中AI服务器约为12.8万台。据TrendForce预测，2023年全球AI服务器出货量约为14.3万台，2026年将达到约20万台，2023-26年的CAGR约为11.7%。由于ChatGPT的推出并展现出巨大的商业化应用潜力，全球企业正在加速将生成式AI应用到所有产品、服务和业务流程中，生成式AI应用带来的数据中心需求激增，AI服务器的需求有望超出此前预期，呈加速上量趋势。英伟达、AMD相继推出高性能AI算力芯片，将驱动AIGC行业加速发展，公司与全球主要云服务厂商合作，有望深度受益。

图表：全球服务器市场规模（单位：10亿美元）



图表：全球AI服务器出货量及预测（单位：千台）



- Chiplet相关技术路径尚未定型，存在技术路径被颠覆的风险。
- AMD AI相关产品市场销售不及预期。
- 竞争加剧对行业的盈利性造成影响。
- 研报使用信息数据更新不及时的风险。

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。
- 本报告版权归“中泰证券股份有限公司”所有。事先未经本公司书面授权，任何机构和个人，不得对本报告进行任何形式的翻版、发布、复制、转载、刊登、篡改，且不得对本报告进行有悖原意的删节或修改。