

电子

2024 年中期策略报告

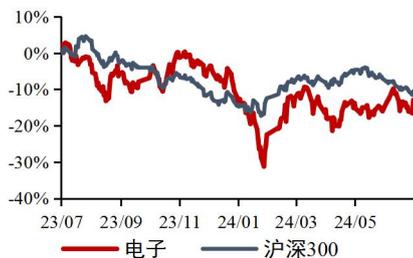
领先大市-A(维持)

AI 供需两旺铸就科技新趋势

2024 年 7 月 10 日

行业研究/行业中期策略

电子板块近一年市场表现



资料来源：最闻

投资要点：

➢ **AI 终端：**2024 年伴随多品牌先后推出 AI PC、AI 手机，AI 端侧加速落地，带动换机潮的同时催生终端全面升级。预计到 2027 年生成式 AI 手机的市场渗透率将达到 43%，生成式 AI 手机的总量将从 2023 年的百万级增长到 2027 年的 12.3 亿部，同时 AI 算力驱动 AI 手机 PCB、散热、电池等硬件配套升级。预计 2025 年 AI PC 渗透率将达到 37%，成为市场主流，2027 年 60% 的 PC 将具备端侧 AI 功能。同时，AI PC 中的关键零部件将与处理器同步进行升级，包括内存、闪存、电池、散热、传感器等环节将有明显提升。

➢ **AI 算力：**英伟达最新旗舰产品 GB200 NVL72 算力较 DGX H100 提升 45 倍，并公布了未来 3 年“一年一代”的迭代节奏，英伟达产业链持续呈现供销两旺情形。与此同时，顶层设计下国内算力网络高速发展，运营商、地方智算中心、互联网企业不断加大算力投资，在外部政策限制和国内政策支持下，国产算力比例将获得快速提升，我们或在大芯片、存储、先进封装、液冷板块挖掘到更多预期差。

➢ **AI 网络：**展望 2024 下半年和 2025 年，我们看到 AI 网络三大主要趋势：一是 网占比将显著提升，带动 网交换机和光模块市场规模快速提升；二是 scaleup 能力持续迭代，铜连接成为 GB200 配套最大增量，同时以 AMD 为引领的 UALink 和以 昇腾为引领的国产算力具备想象空间；三是 scaleout 能力从千卡、万卡到百万卡规模演进，短期看光模块配比有望再增加，未来 800G 向 1.6T、3.2T 演进价值将显著提升。

➢ **投资建议：关注 AI 终端+AI 算力+AI 网络三条主线：**

(1) AI 终端：AI 引领硬件创新升级，建议关注：**鹏鼎控股、立讯精密、东山精密、中石科技、思泉新材、珠海冠宇、洁美科技、顺络电子、三环集团。**

(2) AI 算力：业绩主线，把握海外配套和 机遇，建议关注：国产大芯片**寒武纪、海光信息、景嘉微、龙芯中科**；液冷**英维克**；AI 存储**澜起科技、聚辰股份**；先进封装**长电科技、通富微电、颀中科技、汇成股份、北方华创、中微公司、拓荆科技、芯源微、鼎龙股份、华海诚科。**

(3) AI 网络：光铜共进迎接 scaleup 和 scaleout 能力升级，建议关注：光模块**新易盛、中际旭创、天孚通信、源杰科技、华工科技、铭普光磁**；铜互联**沃尔核材、神宇股份、鼎通科技。**

风险提示：AI 发展不及预期风险，外部制裁升级风险，新技术发展不及预期风险，宏观经济增长乏力风险。

目录

1. AI 终端：AI 端侧加速落地，催生终端全面升级.....	5
1.1 AI 手机：算力驱动手机 PCB、散热、电池等硬件配套升级.....	5
1.2 AI PC：关键零部件与处理器同步进行升级.....	7
2. AI 算力：业绩主线，把握海外配套和 机遇.....	10
2.1 “一超多强”引领，国产大芯片产业链深度受益.....	11
2.2 AI 算力呼唤液冷全链条崛起.....	15
2.3 算力需求爆发，AI 存储大有可为.....	18
2.3.1 AI 存储达成高性能和低功耗的双解.....	18
2.3.2 HBM、DDR5 渗透率持续上升.....	20
2.4 算力时代，先进封装有望迎来加速发展.....	22
2.4.1 封装环节：Foundry 与 OSAT 各有侧重，内资厂商积极布局先进封装.....	23
2.4.2 设备环节：国产设备持续突破，进口替代进程加速.....	24
2.4.3 材料环节：关键材料性能要求升级，高端品类国产化空间巨大.....	25
3. AI 网络：光铜共进迎接 scaleup 和 scaleout 能力升级.....	26
3.1 光模块：海外需求持续性无忧，国内市场有望快速启动.....	26
3.2 铜互联：英伟达引领短距高密度互联新趋势，市场空间广阔.....	29
4. 风险提示.....	32

图表目录

图 1：各品牌 AI 手机发布加速 AI 端侧落地.....	5
图 2：生成式 AI 手机总规模预测.....	6
图 3：全球 AI PC 渗透率曲线.....	8
图 4：具有 AI 功能的处理器线路图.....	9



图 5: 关键零部件与处理器同步进行升级.....	10
图 6: 英伟达在 computex2024 公布三代数据中心芯片路线图.....	11
图 7: 我国智能算力规模及预测.....	11
图 8: 中美拥有最多的 AI 芯片公司, 国内以 为 代表并产生众多互联网自研和创业芯片公司.....	12
图 9: AI Infra 软硬件栈由 5 层架构组成.....	13
图 10: CUDA-X 数据处理平台成为英伟达在垂直行业广泛应用的重要护城河.....	13
图 11: Atlas900 服务器机柜和 Atlas800 训练服务器设计.....	14
图 12: 寒武纪思元 290 核心优势.....	15
图 13: 主力 GPU 厂商芯片功耗快速上升.....	16
图 14: 当机柜密度大于 20kw 时适合采用液冷散热.....	16
图 15: 液冷不同交付模式示意图.....	18
图 16: 美光 AI 存储产品组合.....	19
图 17: 三星存算一体化芯片模型.....	19
图 18: HBM 内部结构图.....	20
图 19: 采用 CoWoS 封装的英伟达 A100.....	22
图 20: 英伟达 A100 CoWoS 封装切面图.....	22
图 21: 全球先进封装市场规模, 亿美元.....	23
图 22: 通信领域 2022-2028 CAGR 17%, 增速最高.....	23
图 23: 2022 年先进封装市场格局集中.....	23
图 24: OSAT 与 IDM、晶圆代工厂技术布局有差异.....	23
图 25: 先进封装设备体系.....	24
图 26: 环氧塑封料是先进封装主要保护材料.....	25
图 27: 垂直互联与异质集成提升 CMP 耗材需求.....	25



图 28: coherent 认为未来几年数通光模块市场将快速增长.....	27
图 29: YOLE 最新预测全球数通光模块将在 2029 年达到 224 亿美金.....	27
图 30: 英伟达将在 2025 全面量产支持 1.6T 端口的网卡和交换机.....	28
图 31: 谷歌 OCS 兼容不同速率可快速适配 TPUv6 对更高速率端口要求.....	28
图 32: 网光模块单位带宽 (Gbps) 价格 (美元) 变化.....	28
图 33: LPO 出货量预计在 2025 年明显增长.....	28
图 34: NVL72 Switch tray 使用的芯片直出跳线.....	30
图 35: NVL72 使用的高速背板 cartridge.....	30
图 36: 服务器使用到的各种高速通信线类型.....	30
图 37: UALINK 拓展通用 scaleup 协议.....	31
图 38: “天成”机柜级算力平台产品.....	31
表 1: 多款 AI 手机处理器的性能对比.....	6
表 2: 微软“Copilot+PC”的 AI 功能.....	7
表 3: 全球数据中心服务器冷板液冷市场规模测算.....	17
表 4: HBM 已经成为 AI 服务器的搭载标配.....	21
表 5: 内存迭代带来内存接口芯片、配套芯片量的增长.....	21

1. AI 终端：AI 端侧加速落地，催生终端全面升级

1.1 AI 手机：算力驱动手机 PCB、散热、电池等硬件配套升级

各品牌 AI 手机陆续发布，驱动新一轮手机换机潮，并加速 AI 端侧落地。2023 年 11 月 13 日，VIVO 率先发布 AI 手机，进入 2024 年，各手机品牌加速发布搭载 AI 功能的智能手机，截至 6 月底，三星、小米、OPPO、荣耀、魅族等各品牌 AI 手机均已发布。与此同时，今年 9 月苹果 AI 功能将正式开放。考虑到 2022、2023 年全球智能手机的平均换机周期被拉长，和 AI 手机陆续发布带来的催化，我们认为 2024 年有望迎来一轮 AI 手机换机潮，并加速 AI 端侧落地。

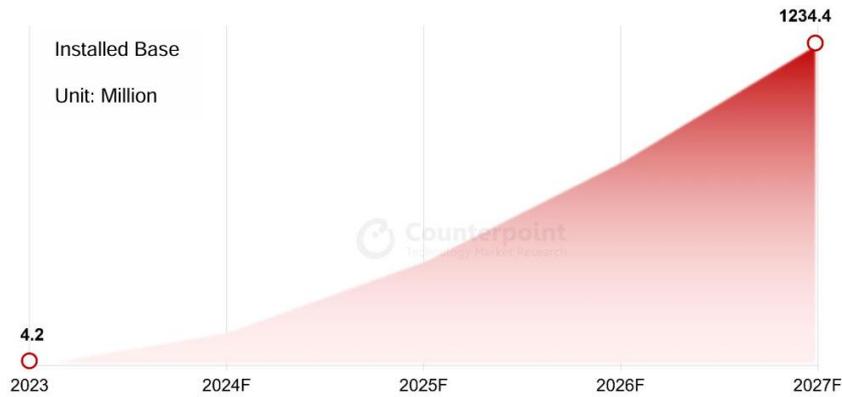
图 1：各品牌 AI 手机发布加速 AI 端侧落地



资料来源：VIVO、OPPO、荣耀、三星、小米、魅族、一加、官网，山西证券研究所

2024 年将成为生成式 AI 手机的元年，预计 2027 年生成式 AI 手机的市场渗透率将达到 43%。根据 Counterpoint 数据，2023 年生成式 AI 手机规模仅有 420 万台，在全年 11.7 亿部手机出货中占比不足 1%。但进入 2024 年，随着各大手机厂商将生成式 AI 功能作为智能手机升级重点，这将加速生成式 AI 手机的普及，预计 2027 年生成式 AI 手机的市场渗透率将达到 43%，生成式 AI 手机的存量规模也将从 2023 年的百万级增至 2027 年的 12.3 亿部。

图 2：生成式 AI 手机总规模预测



资料：Counterpoint 《生成式 AI 手机产业白皮书》，山西证券研究所

AI 算力升级驱动手机 SoC 性能显著提升。对比 AI 手机和传统手机的处理器能看到几大明显变化：（1）AI 手机搭载 NPU，并不断提升 NPU 算力以满足大模型端侧运行。（2）GPU 性能，浮点运算能力实现从传统手机的 GFLOPS 到 AI 手机的 TFLOPS 飞跃。（3）内存容量明显提升，AI 算力带来海量数据，手机内存实现从 4/6/8GB 到 16/20/24GB 的显著升级，预计存储价值量翻 2-3 倍。

表 1：多款 AI 手机处理器的性能对比

		Mediatek Dimensity 9300	Qualcomm Snapdragon 8 Gen 3	Apple A17 Pro	Samsung Exynos 2400	Qualcomm Snapdragon 8S Gen 3	Google Tensor G3
CPU	CORES	8	8	6	10	8	9
GPU	型号	Mali-G720 Immortalis MP12	Adreno 750	Apple A17 GPU	Samsung Xclipse 940	Adreno 735	Mali-G715 MP7
	FLOPS	5990.4 Gigaflops	4435.2 Gigaflops	2147.2 Gigaflops	3407 Gigaflops	3379.2 Gigaflops	2415.8 Gigaflops
存储	最大内存	24 GB	24GB	8GB	24GB	24GB	16 GB
	最大带宽	76.8 Gbit/s	77 Gbit/s	51.2 Gbit/s	68.2 Gbit/s	64 Gbit/s	64 Gbit/s
封装制程	nm 级	4	4	3	4	4	4
AI 性能	Neural processor (NPU)	MediaTek APU 790	Hexagon	Apple Neural Engine	yes	Hexagon	yes
TDP	热功耗	7W	6.3W	8W	6W	6W	6W

资料：NanoReview，山西证券研究所

关注 PCB、散热、电池等硬件配套升级。(1) PCB：随着 AI 加入和 NPU 算力的不断升级，驱动手机 FPC 层数增加、线宽线距更小，进而提升单机 FPC 价值量，同时手机 SLP 板的也需要面积变大、线宽线距更小以容纳更多元器件，加上材料升级，预计单机 SLP 价值量也将提升。(2) 散热：AI 手机算力和功耗的急速增长，使得散热成为确保 AI 手机稳定运行的关键，预计不锈钢 VC 均热板代替铜 VC 均热板、石墨烯材料散热片代替人工石墨片散热的散热方案将在 AI 手机中加速渗透。(3) 电池：为支持大模型运行，不仅手机电池容量要加大，电芯也会更密，电池结构设计用不锈钢外壳代替软包外壳更好散热。

1.2 AI PC：关键零部件与处理器同步进行升级

微软推出专为 AI 设计的新型 Windows PC，称为“Copilot+PC”，开启 AI PC 新时代。与传统 PC 不同，Copilot+PC 具有每秒 45 万亿次操作（TOPS）的神经处理单元（NPU），专门用于加速人工智能和机器学习任务。微软首批 Copilot+PC 搭载高通骁龙 X Elite 和骁龙 X Plus 处理器，较 Surface Pro 9 的性能提高了 90%，并采用定制的高通 Oryon CPU，同时还具有 Copilot 体验、Recall、实时字幕、Surface Studio 相机和 AI 增强等 AI 功能。作为行业领导者，Copilot+PC 的发布将开启 AI PC 新时代。

表 2：微软“Copilot+PC”的 AI 功能

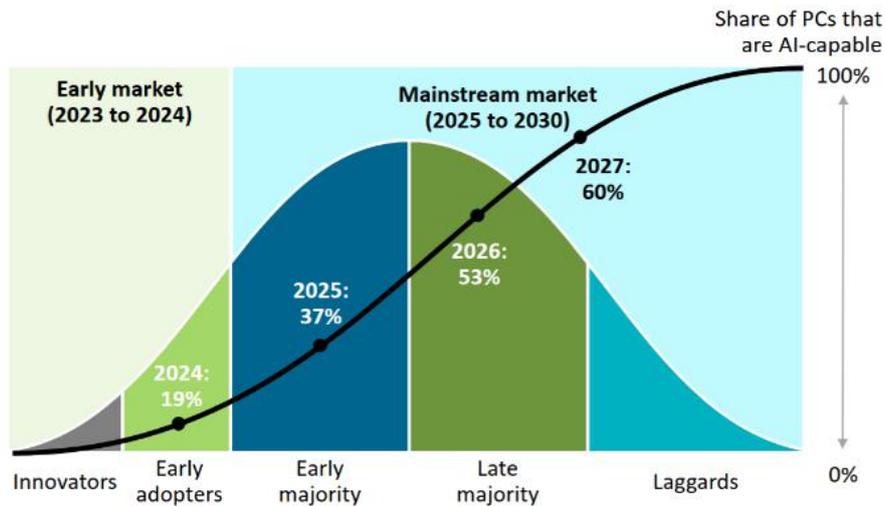
AI 功能	功能介绍
Recall	这项功能利用 NPU 芯片，记录用户在 PC 上的所有操作，包括打开的文件、访问的网站、使用的应用程序等。用户可以通过语音或文本查询，轻松找到特定文件、电子邮件或网页，就如同拥有“过目不忘”的超强记忆力。
实时字幕	支持使用实时字幕提高在视频通话和观看电影时的理解能力，可以在屏幕上实时将口语转换为英文字幕，并支持 40 多种语言翻译。
Copilot 体验	内置键盘包括一个 Copilot 键，可在笔记本电脑上启动 AI 驱动的助手。
Surface Studio 相机	这些相机将使用 Surface Pro 第 11 版上的 Windows Studios 效果和 Surface 笔记本电脑第 7 版上的 Windows Studio 效果，通过 AI 驱动的效果来增强你的相机源。
AI 增强	如 Live Cocreator 让用户能够在设备上实时生成和编辑 AI 图像，用户只需用自然语言描述他们想要创建的图像，Cocreator 就会根据描述生成相应的图像，并允许用户进行实时修改和优化，释放无限创意。

资料：微软官网，山西证券研究所

2025 年 AI PC 将成为主流。Canalys 预计 2024 年出货的 PC 中，19%将具备 AI 功能。这些电脑将包括所有 M 系列 Mac，以及搭载 AMD、英特尔和高通等公司 AI 加速芯片组的首款

Windows PC。Windows 10 服务将于 2025 年 10 月结束，随着客户在此日期之前更新设备，预计具备 AI 功能的 PC 在 PC 总出货量中所占的份额将从 2025 年开始飙升。Canalys 预计 2025 年 AI PC 渗透率将达到 37%，成为市场主流，2027 年 60% 的 PC 将具备端侧 AI 功能。

图 3：全球 AI PC 渗透率曲线



资料：Canalys 《AI PC 的现在与未来》，山西证券研究所

高算力 NPU 是决战 AI PC 性能的关键因素。目前，为了顺应 PC 领域的 AI 应用和趋势，x86 和 ARM 阵营的处理器厂商都已对其产品规格进行了战略性升级。与以往机型相比，一个关键的不同是集成 NPU。传统 PC 处理器由 CPU+GPU 组成，其中 CPU 处理综合计算需求，GPU 处理图像计算需求，而新一代大模型计算需要专业神经网络计算能力，因此引入 NPU 已经成为行业趋势，因此新一代 PC 处理器已经形成 CPU+GPU+NPU 的三核计算架构。虽然没有 NPU 并不妨碍 PC 运行 AI 功能，但集成 NPU 可显著提高计算效率、缩短响应时间并降低 CPU 功耗。在当前的 NPU 技术领域，高通公司的 X Elite 以 45TOPS 算力在所有处理器制造商产品中速度最快，一跃成为行业标杆。

图 4：具有 AI 功能的处理器线路图



资料：Canalys 《AI PC 的现在与未来》，山西证券研究所

AI PC 中的关键零部件将与处理器同步进行升级，其中内存、闪存、电池、散热、传感器等环节将有明显提升。除了处理器将有明显升级，而与处理器配套的芯片及零组件环节也将有直接提升，增量明显的 PC 供应链环节包括：

(1) **内存：容量至少翻倍。**云端模型一般在千亿参数级别，而边缘端模型进行参数蒸馏后参数量级也在百亿级，PC 运行 AI 模型应用，将对密集数据传输和处理有较高的需求，因此内存容量会有显著提升，我们认为未来 AI PC 入门级标配一定是 32GB 内存，当前 16GB 内存会被淘汰，明年则是 64GB PC 开始出货。

(2) **闪存：对固态硬盘性能和容量提出高要求。**AI 大模型预训练算法存储容量较大，同时边缘端模型对用户个性化需求需要进行预训练，训练数据及个性化模型参数同样占据一定容量，因此 AI PC 将对 SSD 性能和容量提出高要求。

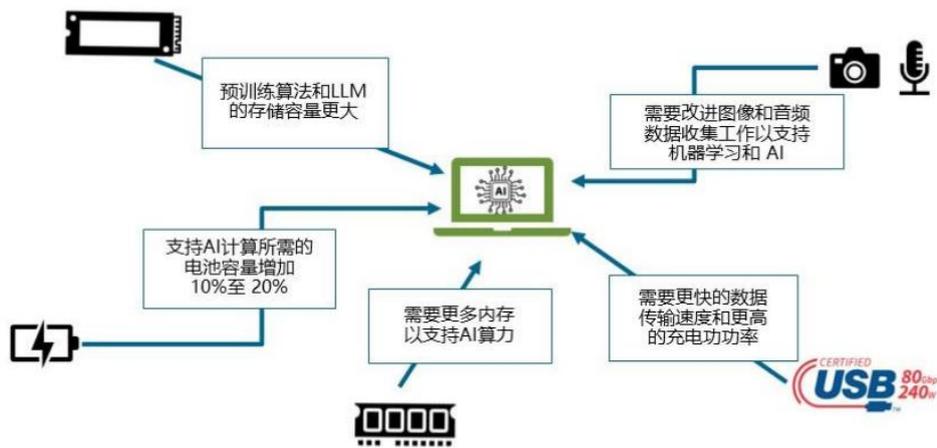
(3) **电池及充电环节：容量、功率均有所增长。**处理器算力升级提高 PC 整体功耗，电能消耗显著提升，因此为适配芯片能耗，电池容量进一步增大，同时充电功率也将随之提升。

(4) **散热：需求提升，单机散热件价值量提升。**由于电池及芯片功耗提升带来的热量提升，散热需求明显提升，传统散热解决方案散热效率有限，在 PC 整体功耗提升的背景下，散热材料的用量增长以及散热方式的多元化已成为趋势，考虑到 PC 散热由多个散热部件组成，随着散热材料用量增长和升级，我们认为单机散热件价值量有望提升。

(5) **传感器：精度升级，数量和种类提升。**AI PC 强调更高维度的用户与 PC 直接交互，而传感器则是输入及输出的窗口，随着 AI PC 对实时环境和用户反馈的需求不断增长，PC 与

外界交互的方式将更加多元化，同时对交互的准确性和及时性也有更高要求。声学、光学以及其他传感方式将持续优化，例如摄像头、麦克风、扬声器等传统光学及声学传感器的数量及准确度将持续提升，新型传感器例如触觉、惯性等也有望导入 PC 等终端，提升终端对外界的感知维度以及信息的输出方式。

图 5：关键零部件与处理器同步进行升级



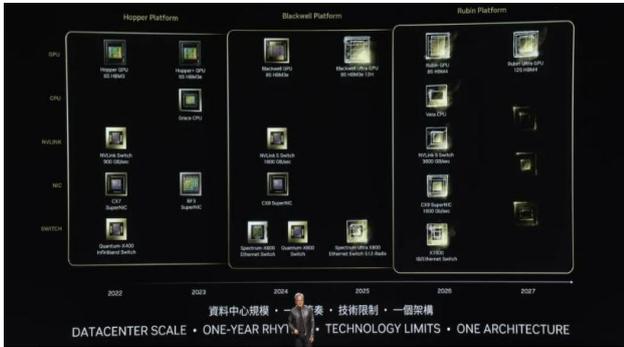
资料来源：Canalys《AI PC 的现在与未来》，山西证券研究所

AI 端侧加速落地，催生终端全面升级，建议关注：鹏鼎控股、立讯精密、东山精密、中石科技、思泉新材、珠海冠宇、洁美科技、顺络电子、三环集团等。

2. AI 算力：业绩主线，把握海外配套和 机遇

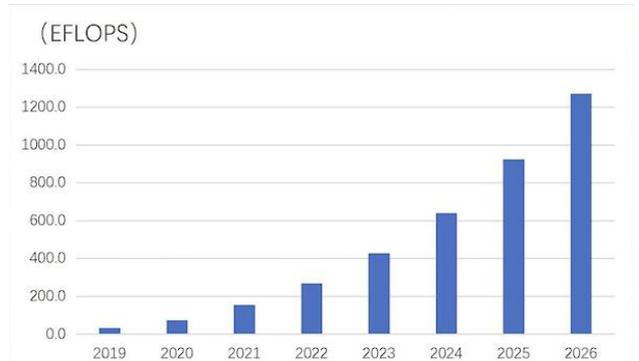
AI 算力需求继续指数级增长。英伟达最新旗舰产品 GB200 NVL72 算力较 DGX H100 提升 45 倍，随着未来 3 年“一年一代”的迭代节奏，英伟达产业链将持续呈现供销两旺情形。与此同时，顶层设计下国内算力网络高速发展，运营商、地方智算中心、互联网企业不断加大算力投资，在外部政策限制和国内政策支持下，国产算力比例将获得快速提升。展望 2024 下半年，我们仍然看好 NV 产业链配套上游的光模块、PCB、服务器代工等环节，但聚焦国产算力，我们或在大芯片、存储、封装、液冷板块挖掘到更多预期差，同时随着国产算力占比提升，我们将看到千亿国产算力大市场带来更多成长股的投资机会。

图 6：英伟达在 computex2024 公布三代数据中心芯片路线图



资料：新智元，山西证券研究所

图 7：我国智能算力规模及预测

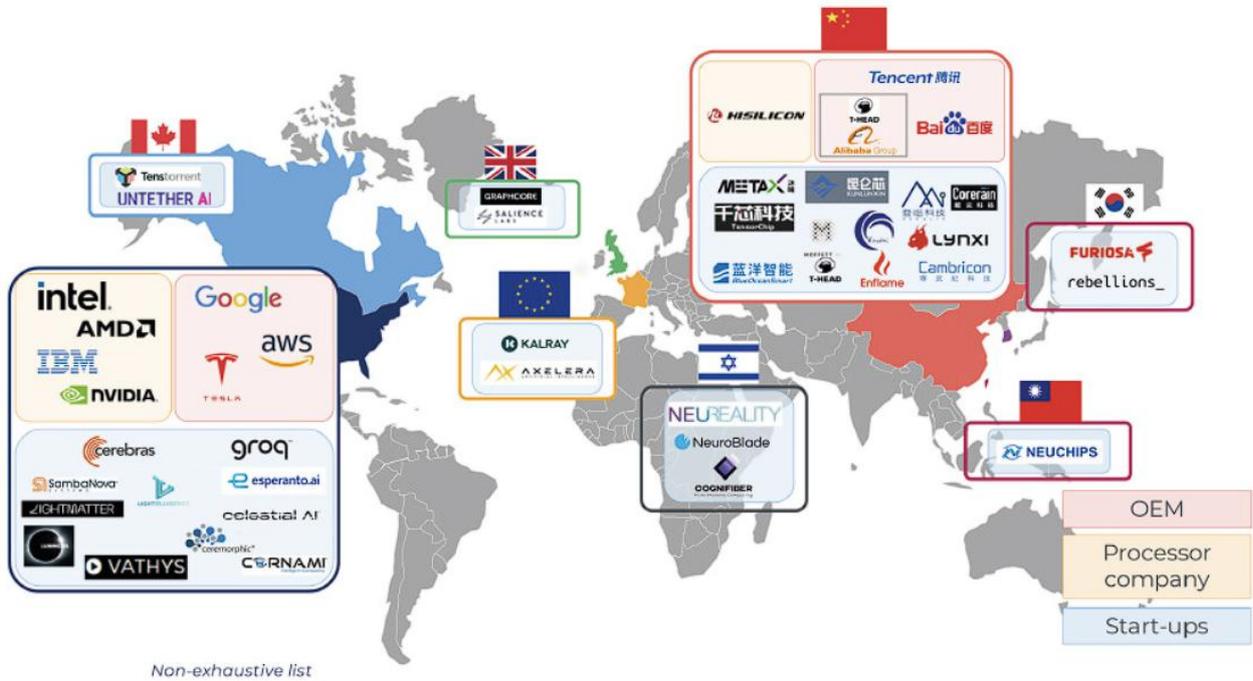


资料：新华三&中国信通院《2023 智能算力发展白皮书》援引 IDC，山西证券研究所

2.1 “一超多强”引领，国产大芯片产业链深度受益

AI 大芯片是最亟需突破的领域，当前已形成以昇腾为引领的“一超多强”局面。AI 大芯片是半导体先进制程、先进封装、计算架构设计、通信接口设计集大成者。英伟达最新芯片 Blackwell B200 FP8 算力达 9PFlops，与国产芯片拉开了较大差距。美国 BIS 在 2023 年 10 月新修订的 AI 芯片出口禁令中将 AI 芯片总处理性能（TPP）大于 4800 或 TPP 大于 1600 的同时性能密度（PD）大于 5.92 作为出口禁令最新标准，几乎涉及到 B100、H100、A100、L40s 以及 AMD、Intel 的所有主力数据中心产品。与此同时，英伟达推出的国内版本 H20 综合性能大幅下降，并不能满足国内用户对于性价比的需求。我们认为，国产大芯片将迎来历史发展机遇，我们将通过 chiplet、加速国产存力运力配套以及集群部署方案来弥补单芯片在性能上的差距。根据 YOLE 最新统计，当前国产 AI 大芯片以昇腾系列为引领，其最新产品 Ascend 910B 目前供不应求，而以百度昆仑芯、阿里含光等互联网自研芯片以及燧原、寒武纪、登临、沐曦、墨芯等为代表的新兴势力未来也将成为市场的有机组成部分。

图 8：中美拥有最多的 AI 芯片公司，国内以 华为 为代表并产生众多互联网自研和创业芯片公司



资料 来源：YOLE，山西证券研究所

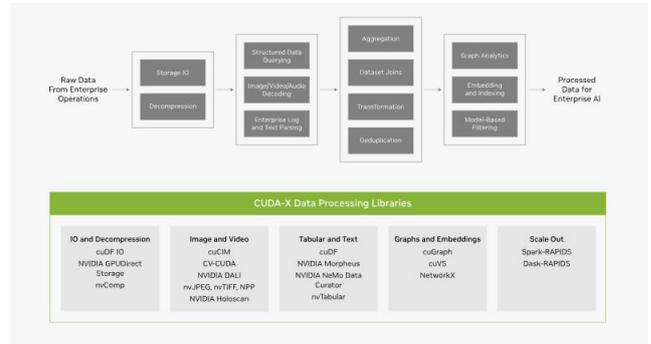
国产大芯片是算力、存力、运力、工具库、软件生态的综合竞争。AI 大芯片的高性能稳定工作需要一系列软件、工具库、平台的支持，我们称之为 AI Infra，根据浪潮信息 OGAI 大模型智算软件栈的介绍，AI Infra 包括基础设施（主要是硬件和运维支撑平台）、系统环境（包括操作系统、Driver、算子库、环境配置、系统监控）、调度平台（包括断点续训、故障定位、算力调度）、模型工具（包括数据治理、预训练、微调的开发工具）以及多模纳管层（实现以应用为目的的可视化、低代码使用），需要巨大的研发投入。此外另一个难点在于“类 CUDA”应用生态的打造，CUDA 内置的丰富应用算子库、函数大大减轻了垂直行业用户开发工作量，形成了巨大使用“粘性”，AI 芯片应用生态非长期研发和迭代更新不能解决。我们认为，国产大芯片如果从“能用”进入“好用”仍需投入大量研发成本和时间，目前丰富 AI infra 的方式一是全方位自主研发并创造开源环境促应用（类似昇腾）；二是搭建“类 CUDA”GPGPU 算子库（类似海光、沐曦、摩尔线程）；三是尽量统一标准接口与第三方协同开发（类似中国移动发起的“全调度 网”在调度平台层做的工作）。

图 9：AI Infra 软硬件栈由 5 层架构组成



资料：浪潮信息官网，山西证券研究所

图 10：CUDA-X 数据处理平台成为英伟达在垂直行业广泛应用的重要护城河

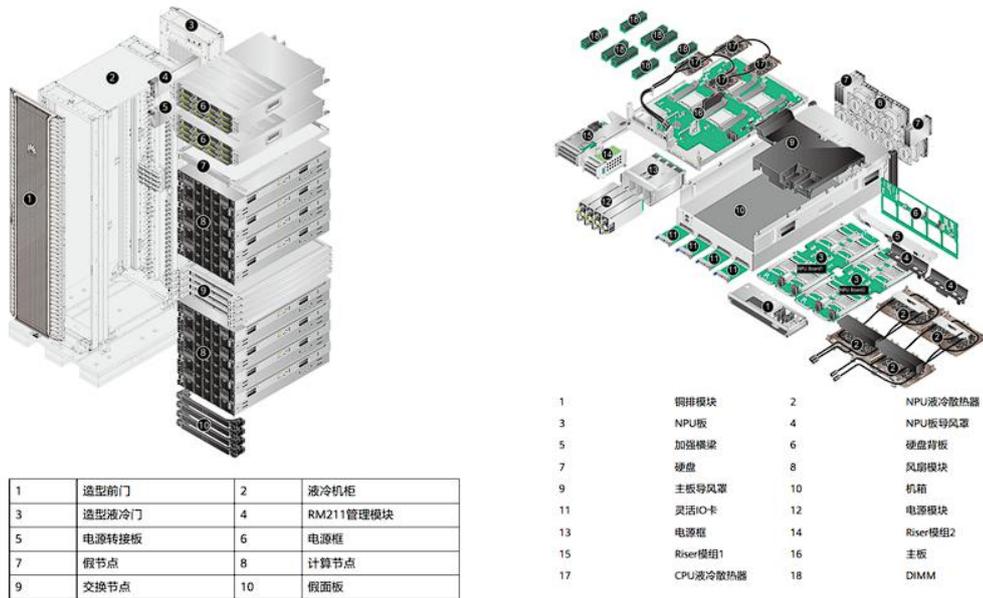


资料：英伟达官网，山西证券研究所

昇腾引领，随着产能提升市场规模有望快速增长，供应链各环节受益。

近年来在昇腾 AI 系列芯片迭代、自主可控产能提升方面取得了长足进展。根据《南华早报》报道，Ascend 910B 芯片在性能上已超越英伟达 A100 GPU，成为运营商、智算中心、互联网企业的主流选择。2024 年 5 月，中国移动发布的 2024 至 2025 新型智算中心 190 亿元集采中，标包 1 中标厂商均为国产化服务器鲲鹏昇腾产业链合作伙伴。根据《2023 智能算力发展白皮书》，假设到 2025 年我国智能算力规模至少要达到 900EFlops，其中昇腾市占率达到 30%，昇腾 910B 算力相当于 A100 (FP16 约 300Tflops)，平均价格为 12 万元，则保守估计昇腾 910B 累计市场空间为 1080 亿元。昇腾产业链配套环节包括服务器组装、电源、散热、高速 IO 连接器、背板连接模组、PCB、铜缆、光模块、DRAM、SSD、RAID 控制器、电源管理等 IC 等，一是配套昇腾出货产值快速增加，二是产品迭代升级加速、价值量或提升，三是对标海外赛道龙头长期成长空间广阔。

图 11: Atlas900 服务器机柜和 Atlas800 训练服务器设计



资料 : 《atlas800 训练服务器用户指南》,《Atlas900 用户指南》, 山西证券研究所

寒武纪新款产品性能提升, 未来有望在智算中心、运营商、互联网多点突破。寒武纪至今在大算力 AI 芯片已发布了四代产品, 2018 年的思元 100、2019 年的思元 270、2020 年的思元 290 以及 2021 年的思元 370 系列。根据官网描述, 寒武纪首款云端训练芯片思元 290 实现了 INT8 算力 512TOPS, 而在研的第五代智能处理器微架构对推荐系统和大语言模型训练推理场景进行了重点优化, 在编程灵活性、能效、功耗、面积等方面大幅提升了产品竞争力。应用落地方面, 寒武纪携手国内头部算法公司推动大模型应用, 在视觉大模型领域与智象未来达成大模型深度适配, 在语言大模型领域与百川智能等头部客户进行了大模型适配。2023 年, 公司智能计算集群系统业务实现营收 6.05 亿元, 在沈阳、台州成功落地了智能计算集群项目, 未来将有更多项目落地。与此同时, 互联网和电信客户基于供应链安全和不同工作负载的需要也有部署多家 AI 芯片的需求, 我们认为公司的客户群体未来会更加多元化。

图 12：寒武纪思元 290 核心优势



资料来源：寒武纪官网，山西证券研究所

未来端侧和云端推理芯片也将更加多元化。随着国内外大模型快速降价、AI agent、数字人、AI 客服等应用快速落地，我们认为 AI 推理工作负载的比重将快速上升。推理芯片与训练芯片的侧重点不同（边缘侧低延时低功耗、针对视觉处理任务的优化、针对不同模型的深度优化），我们看到美国科技大厂（微软、谷歌、亚马逊、苹果）均有量产或自研芯片计划，OPENAI 也表达了自研芯片意向。短期来看英伟达的竞争优势依然明显，但未来云端推理和边缘侧应用或给更多公司带来机会。海光 DCU 在市场上具备三大竞争优势，一是具备全精度各种数据格式算力，二是高速并行数据处理能力，三是兼容“类 CUDA”环境，公司正通过参与开源软件项目加快产品推广速度。景嘉微定增募投项目：“高性能通用 GPU 芯片研发及产业化项目”由图形处理领域拓展至高性能计算领域，“通用 GPU 先进架构研发中心建设项目”主要面向高性能计算和数据处理需求，彰显了公司拓展 AI 大芯片市场的决心。

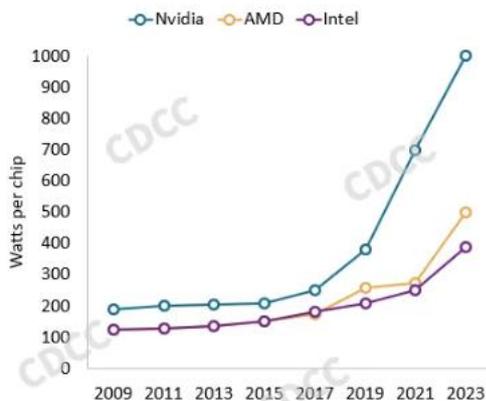
国产大芯片建议关注：寒武纪、海光信息、景嘉微、龙芯中科等。

2.2 AI 算力呼唤液冷全链条崛起

主流大芯片功耗快速提升，液冷从选配逐渐成为标配。全球 AI 算力高增，各大芯片厂商的 TDP 功率也在飞速提升，根据 S&P Global 数据，2017 年主流芯片厂商 TDP 仅为 200w 左右，到 2023 年，一般 GPU 的 TDP 已经达到 350-500W，NVIDIA 新一代 B100 功耗甚至超过 1000W。

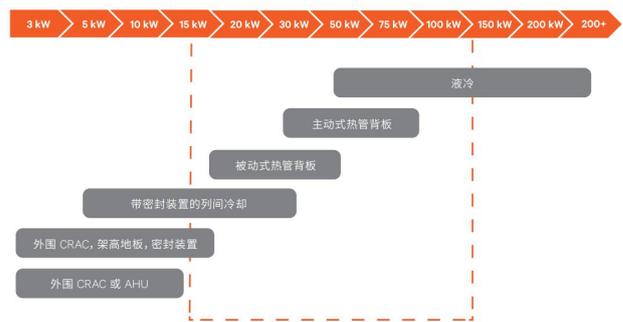
随着芯片 TDP 和机柜容量持续提升，液冷从选配逐渐成为标配。根据 S&P Global 数据，当 TDP 达 200-300W 时，风冷散热系统需要进行特殊升级；当 TDP 达 350-400W 时，风冷逐渐接近极限，液冷将成为必然选择。NVIDIA 新一代 B100 功耗可能超过 1000W，意味 AI 算力所需机柜功率更高，根据《2021-2022 年度中国数据中心基础设施产品市场总报告》和立鼎产业研究网数据，我国约 90% 以上的机柜功率密度在 15kW 以下，但未来主流机柜功率预计会达 25kW-50kW/柜，甚至更高。而液冷技术以其低 PUE（冷板式液冷 PUE<1.2、浸没式液冷技术 PUE<1.1，节能 20%-30% 以上）、满足高密度部署（降低占地和建设成本）、服务器运行更加安全可靠（CPU 温度低至 65℃ 以下）、全年全地域适用等优势，有望解决能耗和散热发展瓶颈。

图 13：主力 GPU 厂商芯片功耗快速上升



资料：CDCC, S&P, 山西证券研究所

图 14：当机柜密度大于 20kw 时适合采用液冷散热



资料：《维谛技术数据中心液冷解决方案白皮书》，山西证券研究所

液冷分为服务器内冷板、机房二次侧（室内）、机房一次侧（室外）等环节，整体市场空间广阔。目前主流液冷应用方式为冷板式液冷和单相浸没式液冷，其中，冷板式液冷需要定制，成本较高，但在可维护性、空间利用性、兼容性等方面较为优秀，由于国内冷板式液冷起步较早，具有先发优势，实际应用场景广泛，因此我们认为冷板式液冷将成为未来主流路线。我们预计，2028 年冷板液冷数据中心市场空间有望超 640 亿元，冷板&管路市场空间有望超 460 亿元。我们预计 2028 年全球服务器 CPU 出货量将超 3600 万片，对应单芯片服务器平均功耗约 290w，冷板液冷渗透率将增长至 25%；全球服务器 GPU 出货量将接近 1500 万片，对应单芯片服务器平均功耗约 1800w，冷板液冷渗透率将增长至 55%；冷板液冷单 kW 投资额下降至 3400 元左右，冷板&管路价值量约 2500 元/Kw；则 2028 年 CPU 服务器冷板液冷市场空间超

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/276212111010010201>