

2025 电子年度策略

AI 策略：中流击水，浪遏飞舟

增持（维持）

总论：全球 AI 产业进入高速发展阶段，技术创新者、云服务厂商与行业应用领导者共同推动 AI 经济格局重塑。巨头加大资本投入，超高性能 GPU 集群、专用 AI 芯片及数据中心等基础设施的建设又进一步推动 AI 应用场景创新。2025 正式进入 AI 端侧“百舸争流”。智能手机、可穿戴、物联网设备、汽车等终端 AI 算力提升，为边缘计算、智能家居、自动驾驶等领域带来全新可能性。

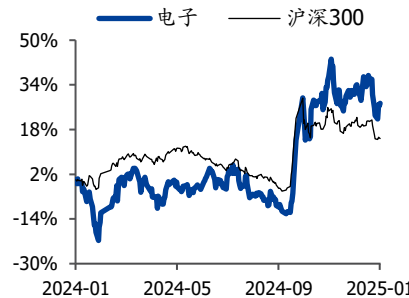
面对地缘政治技术封锁与供应链不确定性，中国 AI 产业亟需推进全面国产化。国产算力、存力、互联等高端芯片的突破，离不开 IC 设计、晶圆制造、先进封装测试、半导体设备、材料及零部件等全产业链配合。此外，还需关注整个配套设施，包括电源管理、PCB、高性能散热方案等关键组件的自主供应链建设。**全球 AI 竞争白热化，通过在核心技术自主创新和完整产业链建设上的持续发力，我们充分看好技术升级迭代、产业积极布局为国产 AI 供应链带来的机遇。**

算力：ASIC 助力定制化，国产芯片加速入局。云厂商持续加码 AI 相关投资，资本开支自 23Q3 开始逐季提升，24Q3 四大 CSP 资本开支 575 亿美元，同比增长 61%，环比提升 11%，25 年资本开支有望持续向上。目前 GPU 仍占据主流 AI 算力市场，英伟达软硬件生态完整，产品持续升级迭代，预计 25-26 年将升级至 Rubin 平台，超大规模训练端核心竞争力稳固。另一方面，云厂商积极布局定制化 ASIC，博通指引 AI SAM 从 2024 年的 150-200 亿美元提升至 27 年的 600-900 亿美元，市场规模高速增长。美国对华科技封锁持续加剧，国产芯片做出性能、性价比，自主可控空间广阔。随着 2025 年端侧 AI 放量，国产算力芯片份额有望加速提升。

存储：云端与边缘侧需求全面扩容，巨头加速扩产 HBM。随着 AI 加速芯片的不断更新换代，其搭载的 HBM 总容量、带宽等规格也持续升级，带动单机 HBM 容量及价值量快速攀升。根据 Intel Market Research，至 2030 年，预计全球 HBM 市场规模将达 489.3 亿美元，2023 至 2030 年 CAGR 达 68.1%。需求端高景气，三大原厂加速扩产 HBM。根据 TrendForce，预计至 2025 年 HBM 芯片每月总产能为 54 万颗，相较于 2024 年增加 27.6 万颗，同比增长 105%。三大巨头扩产将带动 HBM 产业链需求持续高景气。此外，以 AI 服务器、AIPC、AIphone 等为代表的云端和边缘侧 AI 将驱动 3D NAND 市场快速增长。根据 CounterPoint，预计到 2030 年，整体 NAND 闪存市场将超过 930 亿美元，2023 年为 400 亿美元。其中根据 TrendForce，AI 服务器预计将持续推动 SSD 需求的年增长率超过 60%。此外，AI SSD 需求在整个 NAND Flash 市场中的占比预计将从 2024 年的 5% 上升至 2025 年的 9%。未来以企业级 SSD 为代表的大容量高性能存储将持续高速增长，带动产业链相关企业持续受益。

互联：光与铜齐头并进，互连方案向高性能、低成本持续演进。光模块以其高带宽、快传输速率以及长传输距离，在数据中心交换机互连场景广泛应用并持续升级。而在服务器网卡到交换机等之间的短距连接，随着交换芯片、有源铜缆技术等的升级、以及云厂商定制化的 IDC 部署，AEC/ACC 等应用份额及规模有望快速提升。硅光基于硅和硅基衬底材料，通过 CMOS 工艺进行光器件开发和集成，在高速光模块领域可带来低成本、低功耗等优势，同时也能缓解 EML 短缺现状，硅光模块份额或加速提高。LPO 由于

行业走势



作者

分析师 郑震湘
执业证书编号：S0680524120005
邮箱：zhengzhenxiang@gszq.com

分析师 余凌星
执业证书编号：S0680525010004
邮箱：shelingxing1@gszq.com

分析师 钟琳
执业证书编号：S0680525010003
邮箱：zhonglin1@gszq.com

分析师 刘嘉元
执业证书编号：S0680525010002
邮箱：liujiayuan1@gszq.com

分析师 查显彪
执业证书编号：S0680525010001
邮箱：zhaxianbiao1@gszq.com

相关研究

- 《电子：从 CES 前瞻看端侧创新》 2025-01-06
- 《电子：复苏趋势明确，AI 驱动新一轮半导体周期》 2024-05-07
- 《电子：H1 业绩预告陆续发布，需求增温，低点或已过》 2023-07-17

其兼具可插拔性与低功耗优势，预计成为 25-26 年高速光模块的快速落地方案。CPO&OIO 作为更高集成度的下一代高速通信解决方案，台积电、高通等厂商亦积极布局，台积电预计其交换机 CPO 平台 26 年成熟。

电源：AI 技术的精进导致电力消耗剧增，高效能的服务器电源变得至关重要。 AI 服务器电源作为高性能计算和数据中心的基础设备，担负着为服务器集群提供稳定、高效电能供应的任务。高算力需求推升运行功耗亦将显著增加电容用量。

PCB：AI 催生 HDI 浪潮已至。 伴随 AI 服务器的升级，高速高密度互联性能全面提升，功耗和散热要求同步大幅增加。HDI 对比高多层能有效降低 PCB 层数增加布线密度，有效缩减信号在线路板上传输的距离，提高信号传输速率，减少传输时延；从功耗的角度看，HDI 基于其高集成度优势降低系统冗余，减少额外功耗消耗，同时 HDI 高密度短距离布线可以大幅降低信号传输过程中的功耗，同时散热对比高多层也具备显著优势，因此 HDI 有望成为未来 5 年增速最快的 PCB 产品，特别是 4 阶以上的高阶 HDI 产品需求快速增长。根据 Prismark 预计，HDI 将成为 AI 服务器相关 PCB 市场增速最快的品类，2023-2028 年 HDI 的年均复合增速达到 16.3%。从实际生产来看，AI HDI 对应层数阶数及面积更高，因此实际生产过程中无论是产能还是良率均显著降低，产能十分稀缺，全球范围内具备大批量产能能力的厂商有望核心受益于本轮 AI HDI 浪潮。

消费电子：AI 终端创新加速，消费电子产业链景气度持续回升。 2024 年末 AI 创新加速，有望带动终端需求复苏。OpenAI 12 天直播发布，推出完整版 o1、o3 模型、文生视频 Sora；苹果发布 iOS18.2，重点更新 Apple Intelligence；火山引擎冬季 Force 原动力大会上，豆包视觉理解模型首次亮相并升级了豆包主力通用模型。手机端：Canalys 预计 2024 年全球手机出货量达 12.2 亿台，同比增长 6%。苹果发布 Apple Intelligence，全面开启 AI 时代，安卓各大厂商也积极布局 AI。PC 端：AI PC 渗透率持续提升，预计大中华地区在 2028 年达到 73%。可穿戴设备：AI 眼镜浪潮已至，Meta Ray-Ban 2024 年销量有望超 150 万副，随着显示功能的完善，向 AR 发展。政策端，“两新”政策有效激发内需潜力，并且多地开启 3C 数码产品补贴，将进一步加快换机周期。我们认为，随着端侧 AI 多终端落地开启新一轮换机周期，以及在政府补贴政策刺激消费的影响下，消费电子终端市场需求将进一步提升。

风险提示：技术路线演进风险、研发进展不及预期、地缘政治风险。

重点标的

股票代码	股票名称	投资评级	EPS (元)				PE			
			2023A	2024E	2025E	2026E	2023A	2024E	2025E	2026E
688981.SH	中芯国际	买入	0.60	0.51	0.67	0.86	161.30	180.43	136.32	106.57

资料来源：Wind，国盛证券研究所

内容目录

1 算力: ASIC 助力定制化, 国产芯片加速入局	9
1.1 GPU 仍占据主流 AI 算力市场	9
1.2 云厂商积极布局定制化, ASIC 芯片火热	15
1.3 中国算力芯片逐步筑基, 国产替代势在必行	19
1.4 AI 需求涌现, 晶圆代工国产化空间广阔	21
1.5 AI 时代先进封装不可或缺, 本土力量崭露头角	27
2 存储: AI 带动高景气增长, 产业链加速扩产	32
2.1 HBM 高景气, 巨头加速扩产	32
2.2 AI 推动 3D Nand 性能容量持续提升	34
2.3 长鑫存储: 中国 DRAM 龙头产能提升, DDR5 引领国产替代加速	37
2.4 兆易创新: 内生外延并重, DRAM 业务快速放量	38
2.5 普冉股份: AI 端侧放量在即, 普冉有望充分受益	38
3 互联: 光与铜并进, 互连方案向高性能、低成本持续演进	40
3.1 光&铜连接实现数据中心信息互联	40
3.2 你追我赶, 互联技术持续升级	46
3.2.1 AEC——延长铜缆方案传输距离	46
3.2.2 硅光技术——利用 CMOS 工艺的高集成度、高性价比解决方案	47
3.2.3 LPO——兼具可插拔性与低功耗优势的快速落地方案	48
3.2.4 CPO&OIO——高速通信的下一代解决方案	49
3.3 海外龙头引领市场, 国内厂商加速破局	52
3.3.1 Coherent: 800G 光模块需求强劲, 25 年 1.6T 爬坡	52
3.3.2 Marvell: ASIC 需求强劲, 电光产品收入超预期	53
3.3.3 博通: 27 年 AI SAM 有望达 600 亿至 900 亿美元	54
3.3.4 Credo: AEC 收入再创新高, 三大客户需求强劲	55
3.3.5 安费诺: AI 等需求驱动收入超预期	56
3.3.6 沃尔核材: 高速线批量交付, 产能满载积极扩产	56
3.3.7 兆龙互连: 单通道 112G 有源铜缆批量交付, 25 年高速产品需求乐观	57
3.3.8 立讯精密: 24 年服务数据中心量翻四倍, 224G 铜连接器落地在即	57
3.3.9 澜起科技: 运力芯片开启第二成长曲线, PCIe Retimer 需求旺盛	58
4 电源: AI 驱动电源&电容需求攀升	60
5 PCB: AI 引领全面增长, HDI 浪潮已至	64
5.1 PCB 行业长期增长动力充足, 高端产品需求旺盛	64
5.2 AI 需求全面增长, 服务器架构升级带动 PCB 量价跃升	65
5.3 AI 高集成度及高速互联带动 HDI 产业趋势	67
5.4 胜宏科技: AI HDI 升级核心受益, 加速成长区间已至	73
5.5 沪电股份: 业绩持续新高, 前瞻布局强者恒强	74
5.6 深南电路: AI 推动产品结构持续升级	74
5.7 景旺电子: 各项业务有序推进, 高端产能释放引领新增长	75
6 消费电子: 端侧 AI 多终端落地, 新一轮创新周期开启	76
6.1 手机: 景气度持续回升, AI 赋能开启换机潮	76
6.1.1 OpenAI 12 天密集发布, 豆包树立国产 AI 标杆	77
6.1.2 手机为 AI 端侧最佳落地点, 换机周期加速	79
6.2 PC: AI PC 技术突破, 渗透率持续增长	81
6.3 补贴扩大至 3C 产品, 大幅提振换机需求	83
6.4 AI 眼镜: “百镜大战” 愈演愈烈, 逐步迈向 AR 终局	84
6.5 AI 耳机: 字节豆包发布 Ola Friend, 补全 AI 布局拼图	90
6.6 CES 2025 以 AI 为主旋律展开	95
相关标的	103
风险提示	104

图表目录

图表 1: GPU 与 CPU 结构对比..... 9

图表 2: ASIC 芯片架构 (以 AWS 产品为例)10

图表 3: AI 服务器出货量.....10

图表 4: CSP 与品牌高阶 AI 服务器出货量与占比预估10

图表 5: 四大 CSP 资本开支 (单季度, 单位: 亿美元)11

图表 6: 四大 CSP 资本开支 (年度, 单位: 亿美元)11

图表 7: 五大科技公司算力预测.....11

图表 8: 未来五年推动半导体走向 1 万亿美元的主要动力来自于 AI.....12

图表 9: 台积电 24Q3 营收拆分.....12

图表 10: 台积电 24Q3 各应用环比增速.....12

图表 11: AI 服务器收入 22-27 年 CAGR 达 39%.....12

图表 12: 先进节点推动 AI 成长.....12

图表 13: AI 芯片供应商竞争格局.....13

图表 14: 高阶 AI 服务器加速器竞争格局.....13

图表 15: 24 年英伟达 Hopper 出货量预测.....14

图表 16: H100 与 H200 在 1k 输入 token、19k 输出 token 上运行对 token 经济性的改进.....15

图表 17: 英伟达 roadmap.....15

图表 18: A100、H200、B200 参数对比15

图表 19: 推理集群的部署数量极为可观.....16

图表 20: 定制加速计算市场规模扩大逻辑.....16

图表 21: 数据中心定制加速计算市场规模.....16

图表 22: 博通与前两位客户的合作历史.....16

图表 23: 博通获得第三位新客户.....16

图表 24: Marvell 与客户共同开发 XPU.....17

图表 25: Marvell AI 加速器客户进展.....17

图表 26: 24Q3 世芯收入 97% 来自 HPC.....17

图表 27: 24Q3 世芯收入 87% 来自北美.....17

图表 28: 谷歌 Trillium 与 v5e 对比.....17

图表 29: Trainium2 与其他主流 AI 芯片性能对比18

图表 30: BIS 关于中国芯片出口禁令.....19

图表 31: Atlas 800T A2 训练服务器原始视图.....19

图表 32: Atlas 800T A2 训练服务器爆炸图.....19

图表 33: Atlas 900 A2 PoD 集群基础单元爆炸图 (1)20

图表 34: Atlas 900 A2 PoD 集群基础单元爆炸图 (2)20

图表 35: 思元 370 系列板卡与业内主流 GPU 性能对比20

图表 36: 海光信息产品族群.....21

图表 37: 海光产品在大模型的适配与实践21

图表 38: 燧原 S60.....21

图表 39: 全球晶圆代工市场规模.....22

图表 40: AI 芯片产能占比22

图表 41: 24Q3 前十大代工厂排名 (收入单位: 百万美元)23

图表 42: 中国芯片自给率.....23

图表 43: 晶圆代工厂资本支出 (亿美元)24

图表 44: 中芯国际 23Q3-24Q3 分晶圆尺寸营收占比.....25

图表 45: 中芯国际 23Q3-24Q3 分领域营收占比.....25

图表 46: 中芯国际部分技术节点情况.....25

图表 47: 分地区全球 12 英寸晶圆产能占比26

图表 48: 全球晶圆产能制程节点分布.....26

图表 49: 全球六大逻辑晶圆厂先进制程进展情况.....26

图表 50: 先进封装收入预测 2023-2028.....27

图表 51: 各类型先进封装技术在终端的应用情况.....27

图表 52: 台积电系统集成技术平台28

图表 53:	先进封装技术使得连接密度指数级增长.....	28
图表 54:	台积电 2.5D 集成技术赋能下一代 AI 计算.....	28
图表 55:	台积电 3D 堆叠技术发展历程.....	29
图表 56:	移动端的 AI 芯片未来有望从 InFO-PoP 向 SoIC 发展.....	29
图表 57:	车载 AI 芯片有望从 FcBGA 向 CoWoS 发展.....	29
图表 58:	主要厂商的先进封装技术.....	30
图表 59:	封测厂先进封装进展.....	30
图表 60:	台股封测板块月度营收(亿新台币).....	31
图表 61:	“内存墙”——处理器性能提升幅度远超内存.....	32
图表 62:	逻辑-存储带宽升级.....	32
图表 63:	英伟达 GPU HBM 不断升级.....	33
图表 64:	AMD MI300X HBM 容量及带宽与友商对比.....	33
图表 65:	SK 海力士 HBM 产品升级.....	33
图表 66:	英伟达及 AMD AI 芯片发展进程及 HBM 规格比较.....	34
图表 67:	HBM 全球市场规模.....	34
图表 68:	Nand Flash 市场规模及 3D Nand 层数演进趋势.....	35
图表 69:	逻辑-存储带宽升级.....	36
图表 70:	2023 至 2026 年 AI SSD 需求位元占比.....	36
图表 71:	AIphone 出货量预估.....	37
图表 72:	PC 出货量及 AIPC 出货占比预估.....	37
图表 73:	300 层以上 Nand 技术开发的四个方向.....	37
图表 74:	普冉股份产品下游应用领域.....	39
图表 75:	商业模式复用推动 MCU 业务拓展, 核心优势助力公司把握新兴市场机会.....	39
图表 76:	数据中心互连种类.....	40
图表 77:	数据中心互连图例.....	40
图表 78:	光纤通信产业链.....	41
图表 79:	光模块构成.....	41
图表 80:	AI 趋势下互联速率每两年实现翻倍.....	42
图表 81:	交换机及光模块发展趋势预测.....	42
图表 82:	2010-2023 年光模块竞争格局.....	42
图表 83:	2024 年用于 AI 基础设施的光收发器供应链.....	43
图表 84:	2023-2029 年光模块市场规模.....	44
图表 85:	光互连增速将高于 XPU.....	44
图表 86:	以太网光模块按应用分类.....	44
图表 87:	英伟达采购光模块占比.....	44
图表 88:	典型数据中心 CLOS 网络架构示意图.....	45
图表 89:	数据中心服务器接入层链路技术发展趋势.....	46
图表 90:	400G DAC, AOC, ACC, AEC 结构对比.....	46
图表 91:	2024 年和 2029 年 DAC, ACC, AEC 和 AOC 的销售额情况.....	46
图表 92:	2024 年和 2029 年 DAC, ACC, AEC 和 AOC 的销售量情况.....	46
图表 93:	速率提升推动 AEC 需求提升.....	47
图表 94:	Marvell 互连 TAM 不断增长.....	47
图表 95:	1.6T 分立与硅光方案对比.....	48
图表 96:	硅光模块组成.....	48
图表 97:	主流公司硅光技术进展.....	48
图表 98:	LPO 光模块示意图.....	49
图表 99:	LPO 实现能耗节省.....	49
图表 100:	1.6T 光模块能耗对比.....	49
图表 101:	CPO 实现能耗节省.....	50
图表 102:	CPO 技术路线.....	50
图表 103:	台积电 CPO 路线.....	51
图表 104:	从交换机 CPO 到 GPU CPO.....	51
图表 105:	Bailly CPO 能耗节省.....	51

图表 106:	OIO 示意图.....	52
图表 107:	OIO 与 CPO 的差异对比	52
图表 108:	各互连路径量产时间预测	52
图表 109:	Coherent roadmap	53
图表 110:	800G 将占据 Coherent 一半收入	53
图表 111:	各速率光模块占比	53
图表 112:	marvellFY25Q3 数据中心业务亮点.....	54
图表 113:	marvell 数据中心 TAM 及份额展望.....	54
图表 114:	marvell platform.....	54
图表 115:	博通 AI SAM	55
图表 116:	博通 AI 收入占比不断提升.....	55
图表 117:	数据中心中 AEC 的应用	55
图表 118:	安费诺产品.....	56
图表 119:	沃尔核材电线产品.....	56
图表 120:	沃尔核材收入及增速 (单季度)	57
图表 121:	沃尔核材归母净利润及增速 (单季度)	57
图表 122:	兆龙互连收入及增速 (单季度)	57
图表 123:	兆龙互连归母净利润及增速 (单季度)	57
图表 124:	立讯精密数据中心解决方案	58
图表 125:	澜起科技 PCIe Retimer 量产及应用进展	58
图表 126:	澜起科技 PCIe Retimer 示意图	59
图表 127:	PCIe 4.0/5.0 Retimer 芯片的典型应用场景图示	59
图表 128:	AI 电源功率密度持续提升.....	60
图表 129:	CPU/GPU 算力加速提升带动电源需求	60
图表 130:	数据中心电源的架构主要分为保障电路 UPS、机架电源 AC/DC、芯片电源模组 DC/DC 三个层级	60
图表 131:	典型服务器电源架构.....	61
图表 132:	英伟达 GPU 芯片功率持续提升.....	61
图表 133:	GPU/CPU 功率提升对 VRM 和 PoL 提出严峻挑战	62
图表 134:	由多相控制器和 DrMos 组成的拓扑架构是目前公认的 GPU/CPU 供电场景的最佳解决方案	62
图表 135:	2023-2030 年预计服务器市场对 MLCC 需求量将翻 2.2 倍	63
图表 136:	水平供电、垂直供电与集成式稳压器示意图	63
图表 137:	安装在 xPU 下方的垂直电源模块架构	63
图表 138:	全球 PCB 市场规模 (亿美元)	64
图表 139:	国内 PCB 市场规模 (亿元)	64
图表 140:	2024 年 PCB 不同产品同比增速	64
图表 141:	PCB 各下游应用领域.....	65
图表 142:	AI 服务器出货量预测.....	65
图表 143:	八卡架构 AI 服务器的拆解.....	66
图表 144:	八卡加速模块实物.....	66
图表 145:	AI 服务器 PCB 价值量大幅提升	66
图表 146:	英伟达 GB200 NVL72 内部架构	67
图表 147:	12 层 3 阶 HDI 结构示意图	67
图表 148:	HDI 参数特性.....	68
图表 149:	各阶 HDI 示意 (红色为电路层; 黑色为通孔、盲孔、埋孔)	68
图表 150:	HDI 生产工艺流程	69
图表 151:	多层板生产工艺流程.....	69
图表 152:	通孔板和 HDI 板的层数密度对照图.....	69
图表 153:	实现相同功能前提下 HDI 工艺 8 层板减少 40%面积同时降低 33%层数	70
图表 154:	从芯片到 PCB 互联示意图	71
图表 155:	盲孔在内层增加除 BGA 以外 30%的布线.....	71
图表 156:	芯片间距及焊盘尺寸的示意图	71
图表 157:	GB200 集成一颗 CPU 和两颗 B200 GPU	72
图表 158:	GB200 CPU 和 GPU 之间基于 C2C 互联	72

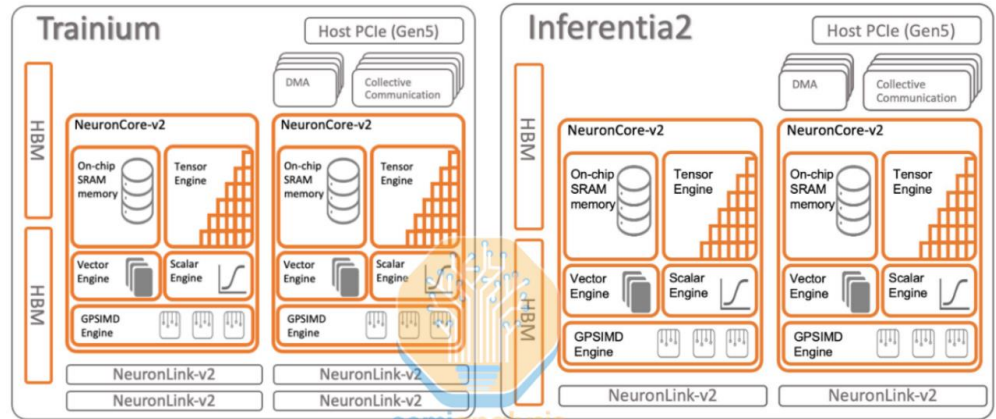
图表 159:	英伟达不同 AI 芯片功耗.....	72
图表 160:	传输速率越快距离越长, PCB 的路径损耗越大.....	73
图表 161:	2021Q1-2024Q3 全球智能手机出货量.....	76
图表 162:	2022Q1-2024Q3 中国智能手机出货量.....	76
图表 163:	2023Q1-2024Q3 全球智能手机厂商市场份额.....	76
图表 164:	2023Q2-2024Q3 中国智能手机厂商市场份额.....	76
图表 165:	写作能力对比.....	77
图表 166:	数学能力对比.....	77
图表 167:	OpenAI 12 天直播内容总结.....	78
图表 168:	豆包通用模型 pro 在主观评测中排名第一.....	78
图表 169:	豆包视觉理解模型在视觉语言模型中排名第二.....	78
图表 170:	大模型角斗场榜单-大语言模型榜单.....	79
图表 171:	豆包通用模型能力全面升级.....	79
图表 172:	iOS18.2 重点更新 Apple Intelligence.....	80
图表 173:	安卓 AI 布局.....	80
图表 174:	2024Q3 全球 AI 手机型号榜单.....	81
图表 175:	2024Q3 中国大陆 AI 手机型号榜单.....	81
图表 176:	生成式 AI 手机渗透率预测.....	81
图表 177:	2018Q3-2024Q3 全球 PC 出货量.....	82
图表 178:	全球 PC 市场各品牌市场份额.....	82
图表 179:	24Q3 全球平板出货量情况.....	82
图表 180:	2023Q4-2024Q3 全球 AI PC 出货量(按操作系统划分).....	83
图表 181:	2024 年-2028 年大中华地区 AIPC 出货量以及预测.....	83
图表 182:	中国部分地区 3C 数码产品补贴活动.....	84
图表 183:	Meta Ray-Ban 外观图.....	85
图表 184:	Ray Ban Meta 智能眼镜主板 BOM.....	86
图表 185:	Ray Ban Meta 智能眼镜 BOM(除主板外).....	87
图表 186:	Meta Ray-Ban 带显示屏效果图.....	87
图表 187:	Meta Orion 效果图.....	87
图表 188:	AR 光学显示方案对比.....	88
图表 189:	2024 年 11-12 月 AI/AR 智能眼镜新品汇总.....	89
图表 190:	闪极 AI 拍拍镜硬件配置.....	90
图表 191:	讯飞录音降噪会议耳机 Pro2.....	91
图表 192:	Ola Friend 智能体耳机.....	91
图表 193:	全球 TWS 耳机出货量(亿部).....	92
图表 194:	24Q3 全球智能耳机市场格局.....	92
图表 195:	全球开放式耳机出货量(百万部).....	93
图表 196:	截止 2024 年 8 月中国 AI 耳机 2024 年月度销售额.....	93
图表 197:	Cleer 私人 AI 运动教练.....	94
图表 198:	截止 2024 年 8 月中国商务+娱乐 AI 耳机 2024 年月度销售额(百万元).....	95
图表 199:	截止 2024 年 8 月中国运动 AI 耳机 2024 年月度销售额(百万元).....	95
图表 200:	英伟达发布 GeForce RTX™ 50 GPU.....	96
图表 201:	英伟达 GB10 Grace Blackwell Superchip.....	96
图表 202:	英特尔发布全新酷睿 Ultra 200HX 和 200H 系列移动处理器.....	96
图表 203:	AMD 锐龙 AI Max 和锐龙 AI Max PRO 系列处理器.....	97
图表 204:	AMD 锐龙 AI 300 和锐龙 AI 300 PRO 系列处理器.....	97
图表 205:	联想发布卷轴屏 ThinkBook Plus Gen 6 AI PC.....	97
图表 206:	CES 展出 AR+AI 眼镜.....	98
图表 207:	Halliday 在镜框部分塞入 Digi Window 微型显示器.....	99
图表 208:	XREAL One 产品亮点.....	99
图表 209:	ThinkAR 首次亮相 AI+AR 眼镜 AiLens.....	100
图表 210:	莫界轻量型全彩 AR+AI 眼镜.....	100
图表 211:	NVIDIA DriveOS.....	101

图表 212: 宝马推出新世代超感智能座舱 BMW 全景 iDrive 102

图表 213: 吉利展示最新智能驾驶系统 102

FPGA 半定制化,但冗余多且能效极致化不如 ASIC,不适用于 AI 计算大规模出货。FPGA,即 Field Programmable Gate Array (现场可编程门阵列),属于半定制芯片,它能够依据用户需求,在制造完成后进行不限次数的重复编程,进而达成所需的数字逻辑功能。ASIC 则是全定制芯片,其功能固定无法更改。相比之下,FPGA 作为半定制芯片,功能具有灵活性,无需进行流片,不过因其可编辑的特性,冗余功能较多,若仅用于单一目的便会造成资源浪费。在大规模生产时,FPGA 的成本高于 ASIC,并且在能效极致化方面也不及 ASIC。因此,用于 AI 计算大规模出货时,一般考虑使用 GPU 与 ASIC 较多。

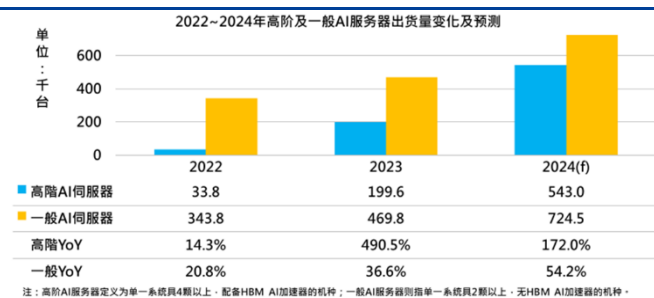
图表2: ASIC 芯片架构 (以 AWS 产品为例)



资料来源: semianalysis, AWS, 国盛证券研究所

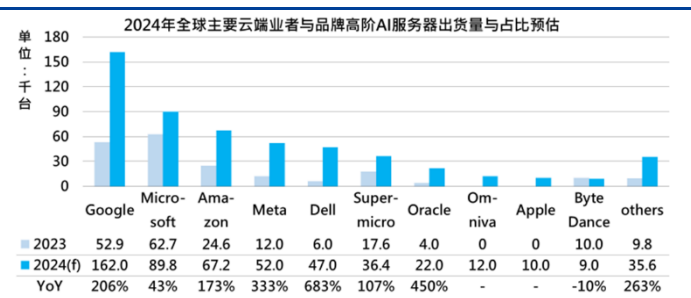
AI 服务器需求旺盛,北美四大 CSP 资本开支上行。从 AI 服务器出货量来看,2023 年高阶 AI 服务器出货 19.96 万台,一般 AI 服务器出货 46.98 万台。2024 年高阶 AI 服务器预期出货 54.3 万台,较 2023 年增长 172%;2024 年一般 AI 服务器预期出货 72.5 万台,较 2023 年增长 54.2%。根据 Trendforce,2024 年 AI 服务器占整体服务器出货的比重预估将达 12.2%,较 2023 年提升约 3.4 个百分点;从产值看,预估 2024 年 AI 服务器产值将达 1,870 亿美元,同比+69%,产值占整体服务器高达 65%。从需求方来看,2023 年前五大业者出货占总体比重达 85.1%,2024 年前五大业者占比则降至 77%,前四大仍为以代工厂直接销售 (ODM-direct) 为服务器主要采购模式的大型 CSP 厂商。

图表3: AI 服务器出货量



资料来源: Digitimes, 国盛证券研究所

图表4: CSP 与品牌高阶 AI 服务器出货量与占比预估

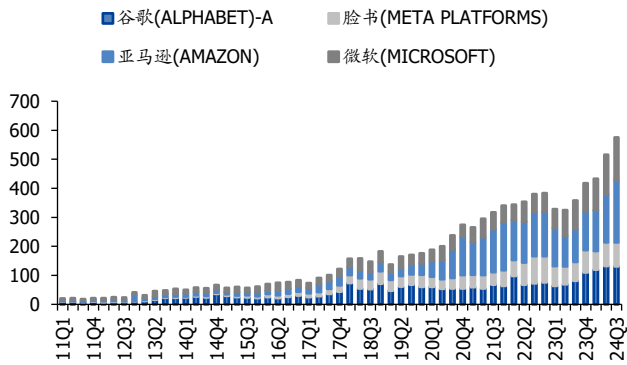


资料来源: Digitimes, 国盛证券研究所

CSP 持续加码 AI 服务器及相关投资。24Q3 四大 CSP (谷歌、meta、亚马逊、微软) 资本开支 (现金流量表口径) 总和为 575 亿美元,同比增长 61%,环比增长 11%,自 23Q3 开始,云厂商资本开支逐季提升。谷歌 24Q3 资本支出约 130 亿美元,其中最大的组成部分是对服务器的投资,其次是数据中心和网络设备,预计 Q4 资本支出将与 Q3 持平;meta24Q3 资本支出 (包括融资租赁的本金支付) 为 92 亿美元,主要系对服务器、数据中心和网络基础设施的投资,预计 2024 年全年的资本支出将在 380-400 亿美元之间,下限较之前的 370-400 亿美元有所增长,预计 2025 年的资本支出将大幅增长;亚马逊 24Q3 资本支出为 213 亿元,预计 2024 年的资本支出约为 750 亿美元,大部分支出于支持对技术基础设施日益增长的需求;微软 FY25Q1 包括融资租赁在内的资本支出

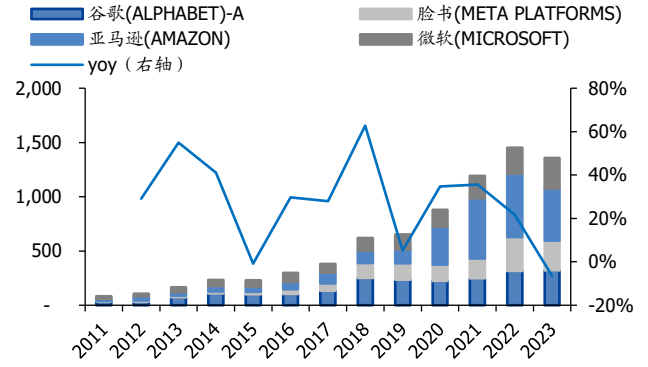
为 200 亿美元，PP&E 支付的现金为 149 亿美元，大约一半的云和 AI 相关支出继续用于长期资产，其余的云和 AI 支出主要用于服务器（CPU 和 GPU）。

图表5: 四大 CSP 资本开支 (单季度, 单位: 亿美元)



资料来源: wind, 国盛证券研究所

图表6: 四大 CSP 资本开支 (年度, 单位: 亿美元)



资料来源: wind, 国盛证券研究所

科技公司紧锣密鼓布局、升级迭代大模型训练，AI 算力芯片需求强劲。在加速计算和生成式 AI 的带动下，算力需求陡增，全球高科技公司纷纷涌入 AI 领域开展军备竞赛。预计 2024 年微软有 75 万-90 万块等效 H100，25 年预计达到 250 万-310 万；谷歌 2024 年有 100 万-150 万块等效 H100，25 年预计达到 350 万-420 万；Meta2024 年有 55 万-65 万块等效 H100，25 年预计达到 190 万-250 万；亚马逊 2024 年有 25 万-40 万块等效 H100，25 年预计达到 130 万-160 万；xAI2024 年有 10 万块等效 H100，25 年预计达到 55 万-100 万。

图表7: 五大科技公司算力预测

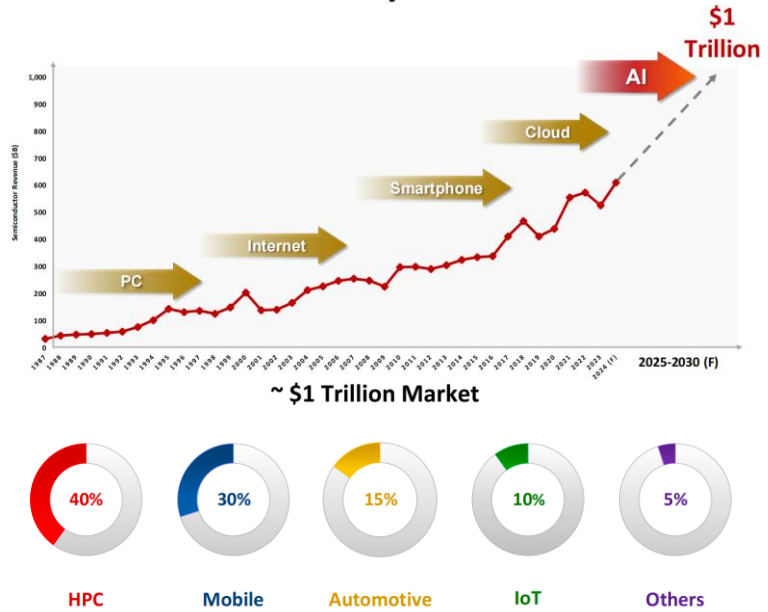
	2024YE(等效H100)	2025 (GB200)	2025YE (等效H100)
MSFT	750k-900k	800k-1m	2.5m-3.1m
GOOG	1m-1.5m	400k	3.5m-4.2m
META	550k-650k	650k-800k	1.9m-2.5m
AMZN	250k-400k	360k	1.3m-1.6m
XAI	~100k	200-400k	550k-1m

资料来源: 新智元, 国盛证券研究所

台积电指引 AI 需求强劲，2024 年 AI 处理器收入同比增长 3 倍以上。24Q3 营收 235 亿美元，超出指引上限（224-232 亿美元），qoq+12.9%，yoy+36.0%，其中，3nm 占比 20%，推动公司业绩增长。台积电在 24Q3 法说会中表示客户对 AI 相关需求极度强劲，带来更高的稼动率。公司预计 2024 年 AI 处理器带来的收入将同比增长 3 倍以上，占总收入比重中值在 15% 左右（mid teens）。台积电认为在未来五年，推动半导体市场规模走向 1 万亿美元的主要动力来自于 AI，从应用端划分，其中 40% 应用于 HPC，移动市场贡献 30%，汽车市场贡献 15%，IoT 贡献则将贡献 10%。AI 服务器市场规模 22-27 年 CAGR 达 39%，技术进步推动更多的计算，更高的存储带宽，更大规模的异构集成需求。

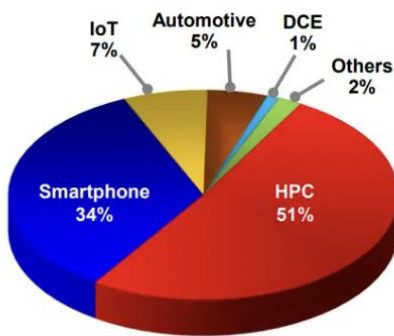
图表8: 未来五年推动半导体走向 1 万亿美元的主要动力来自于 AI

Semiconductor Industry Outlook



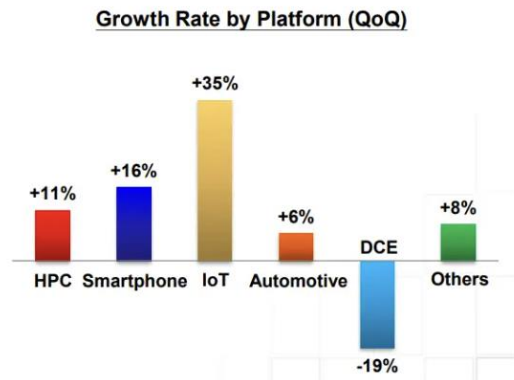
资料来源: 台积电, 国盛证券研究所

图表9: 台积电 24Q3 营收拆分



资料来源: 台积电, 国盛证券研究所

图表10: 台积电 24Q3 各应用环比增速

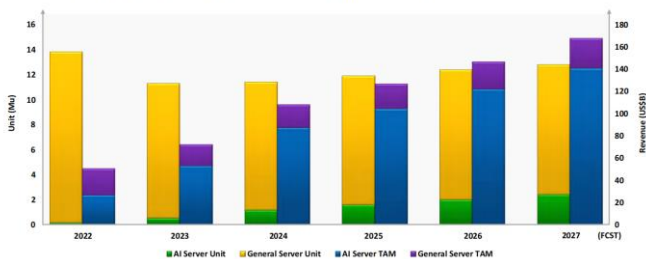


资料来源: 台积电, 国盛证券研究所

图表11: AI 服务器收入 22-27 年 CAGR 达 39%

AI Server Revenue Forecast

- AI Server ASP is much higher than General Server
- AI Server Unit CAGR is 73%, revenue CAGR 39% from 2022 to 2027

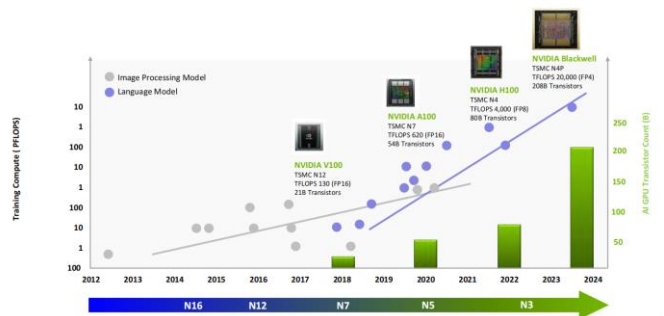


资料来源: 台积电, 国盛证券研究所

图表12: 先进节点推动 AI 成长

Technology Advancement Ignites Explosive AI Growth

More Compute, Higher Memory Bandwidth, Larger Scale Heterogeneous Integration



资料来源: 台积电, 国盛证券研究所

英伟达 GPU 采用率占比第一, 24 年 CSP 自研 ASIC 采用率提升。根据 Trendforce 和 Digitimes, 单看搭载 GPU 的 AI 服务器, 英伟达市占率最高, 逼近 9 成, AMD 市占率则仅约 8%; 从整体 AI 服务器来看, 23 年英伟达占比 65.5%, 仍为第一, 预计 24 年市占

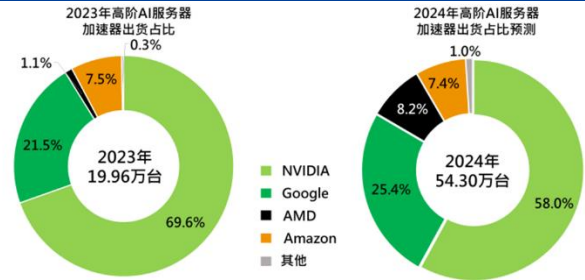
率降至 63.6%，主要系搭载自研 ASIC 芯片的服务器占比提升至 25% 以上。从高阶 AI 服务器来看，2023 年高阶 AI 服务器所采用的加速器中，70% 采用 NVIDIA GPU，出货量达 13.9 万台，产品形态以采用 NVLink 技术、GPU 间互连速度达 600GB/s 以上的 HGX 版本为主，2024 年 AMD 及 CSP 自行研发的 AI 服务器机种成长较快，在高阶 AI 服务器中，预计采用 NVIDIA GPU 机种占比将降至 58%，出货量则将增至 31.5 万台。我们看到，英伟达在 AI 芯片中的龙头地位不变，但 CSP 自研 ASIC 采用率提升，智算方案选择更加丰富。

图表13: AI 芯片供应商竞争格局

Company	2022	2023	2024F
NVIDIA	67.6%	65.5%	63.6%
AMD (incl. Xilinx)	5.7%	7.3%	8.1%
Intel (incl. Altera)	3.1%	3.0%	2.9%
Others	23.6%	24.1%	25.3%
Total	100.0%	100.0%	100.0%

资料来源: trendforce, 国盛证券研究所

图表14: 高阶 AI 服务器加速器竞争格局

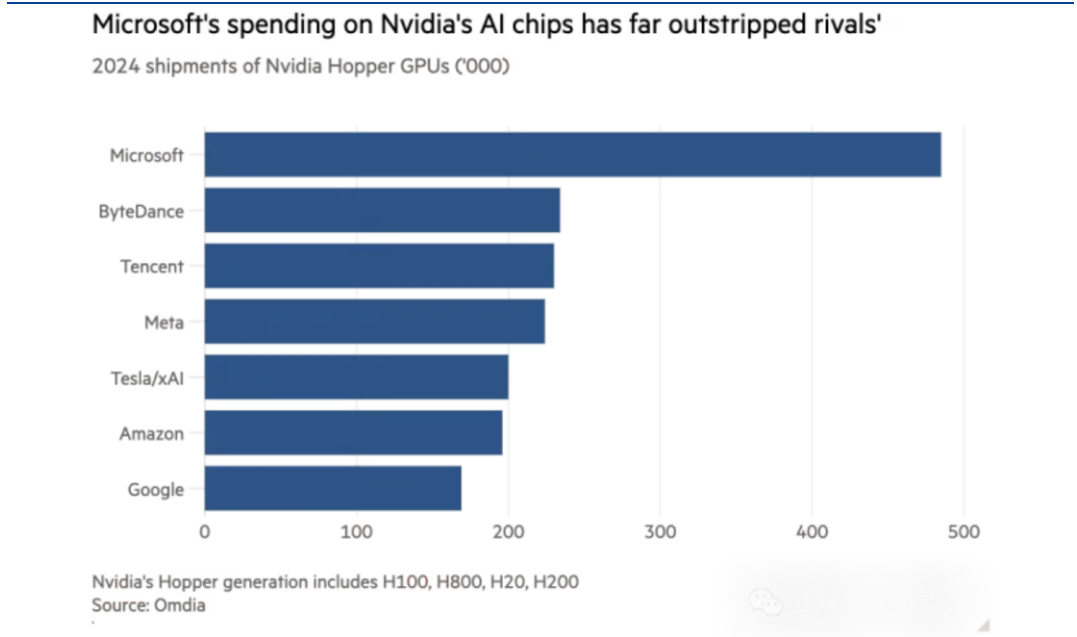


资料来源: Digitimes, 国盛证券研究所

FY25Q3 Blackwell 投产，Hopper 与 Blackwell 系列将在 FY25Q4 并行销售。 英伟达 FY25Q3 数据中心收入为 308 亿美元创记录，yoy+112%，qoq+17%，占比为 88%。Hopper 需求旺盛，H200 销售额环比大幅增长至数十亿美元，是公司历史上增长最快的产品，Hopper 的需求将持续到 25 年的前几个季度。公司表示 Blackwell 全面扩张后，预计 Blackwell 的利润率中值将达到 75% 左右 (mid-70%)，FY25Q3 向客户交付了 1.3 万个 GPU 样品，本季度交付的 Blackwell 数量超预期，预计将为数十亿美元。

24 年英伟达出货主要为 Hopper 平台产品，25 年 GB200 NVL72 占比有望接近 80%。 根据 TrendForce，2024 年 NVIDIA 高端 GPU 出货将主要围绕 Hopper 平台产品。其中，面向北美 CSPs、OEMs 会出货搭载 H100、H200 等机型，而针对中国客户则会以搭载 H20 的 AI 服务器作为主力产品。预计 24Q3 开始 H200 大规模出货成为 NVIDIA 主流产品并持续至 2025 年。Blackwell 系列在 2024 年尚处于前期出货阶段，预计 Blackwell GPU 芯片 2024Q4 仅少量出货，2025Q1 后逐季放量，GB200 整柜式方案或将于 25Q2 放量。在 NVIDIA 大力推动下，预期 GB200 NVL72 将于 2025 年成为主要的采用方案，占比有望接近 80%。

图表15: 24 年英伟达 Hopper 出货量预测

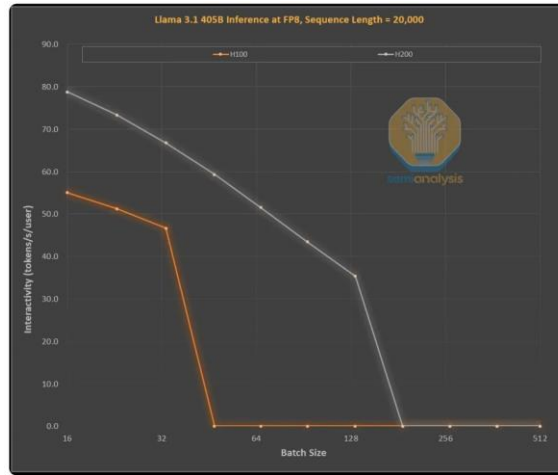


资料来源: 新智元, 国盛证券研究所

英伟达或将于 **B200** 上市后半年推出 **B300**, 内存由 **8-Hi** 升级到 **12-Hi**。据 SemiAnalysis, B300 将在 B200 上市大约半年后上市。1) 算力: B300 基于 TSMC 4NP 工艺节点全新流片, GPU 能够提供比 B200 高 50% 的 FLOPS, 部分性能提升将来自 200W 的额外功率, GB300 和 B300 HGX 的 TDP 分别达到 1.4KW 和 1.2KW (GB200 和 B200 的 TDP 分别为 1.2KW 和 1KW); 2) HBM: 内存从 8-Hi 升级到 12-Hi HBM3E, 每个 GPU 的 HBM 容量增加到 288GB; 3) NIC: 可能采用 800G ConnectX-8 NIC, 带宽是当前 400G ConnectX-7 的两倍, 有 48 个 PCIe 通道, 而前代产品只有 32 个, 这将为新服务器提供显著横向扩展带宽改进, 更有利于大型集群。

B300 内存的改进对于 **OpenAI o3** 风格的 LLM 推理训练和推理至关重要, **NVL72** 助力推理模型。由于长序列长度会增加 KVCache, 从而限制关键批次大小和延迟, 因此内存的改进是 OpenAI o3 风格 LLM 推理训练和推理的关键。以 H200 与 H100 为例, 由于内存带宽更大, 所有可比批次大小的交互性普遍提高了 43% (H200 4.8TB/s vs H100 3.35TB/s)。由于 H200 运行的批处理大小比 H100 大, 每秒可生成 3 倍的 token, 因此成本降低了约 3 倍, 更大的内存容量所带来的动态效益较大。NVL72 系统的性能和成本将大幅提升, 在推理中能够让 72 个 GPU 以极低的延迟处理同一个问题, 共享内存, 更高的交互性使得每个思路链的延迟更低, 72 个 GPU 分散 KVCache, 以实现更长的思维链 (提高智能)。与典型的 8 GPU 服务器相比, 批量大小扩展效果更好, 从而降低成本。使用 NVL72 的 token 经济性要好 10 倍以上, 尤其是在长推理链上。

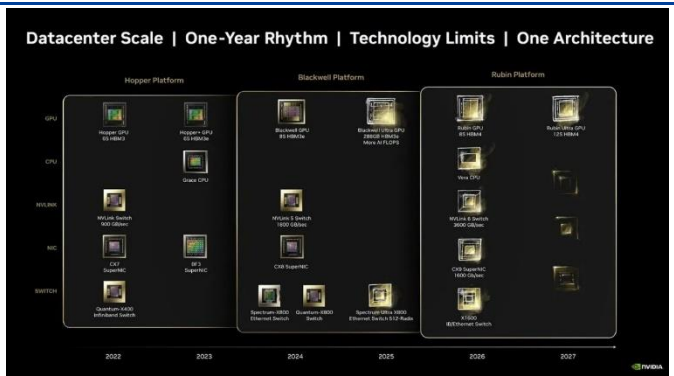
图表16: H100 与 H200 在 1k 输入 token, 19k 输出 token 上运行对 token 经济性的改进



资料来源: semianalysis, 国盛证券研究所

英伟达软硬件生态完整, 预计 25-26 年将从 Blackwell 平台升级至 Rubin 平台。英伟达同时拥有 GPU、CPU 和 DPU 的计算芯片和系统, 通过 NVLink、NVSwitch 和 NVLink C2C 技术将 CPU、GPU 进行灵活连接组合形成统一的硬件架构, 并与 CUDA 一起形成完整的软硬件生态。目前英伟达已从 Hopper 平台迭代到 Blackwell 平台, 预计 25-26 年将演进为 Rubin, 演进周期缩短至一年。

图表17: 英伟达 roadmap



资料来源: 英伟达, 国盛证券研究所

图表18: A100, H200, B200 参数对比

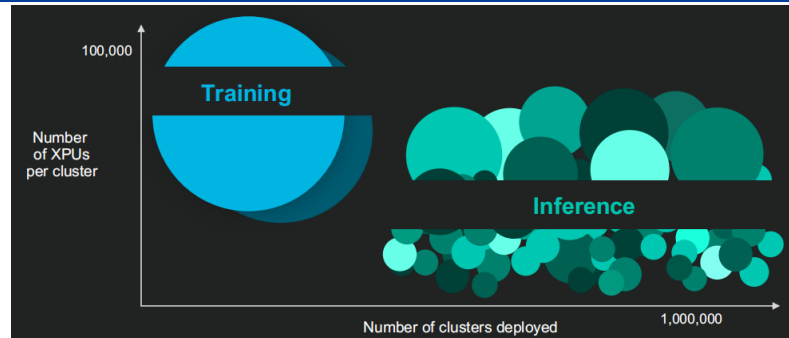
NVIDIA AI Accelerator	A100	H200	B200
Process Node	N7	N5	N4P
Die Size (mm ²)	826	814	N/A
Gate Count (B)	54	80 (1.48X)	104 (1.3X, x2)
Clock (MHz)	1410	1830 (1.3X)	TBA
BF16 TFLOPS	624	1979 (3.2X)	4500 (2.27X)
Memory Type	HBM2 (x5)	HBM3e (x6)	HBM3e (x8)
Advanced Packaging	CoWoS-S	CoWoS-S	CoWoS-L
Memory BW (GBps)	1555	4800 (3.1X)	8000 (1.7X)
TDP (Watt)	400	700	1000
TFLOPS / Watt	1.56	2.83 (1.8X)	4.5 (1.6X)

资料来源: 台积电, 国盛证券研究所

1.2 云厂商积极布局定制化, ASIC 芯片火热

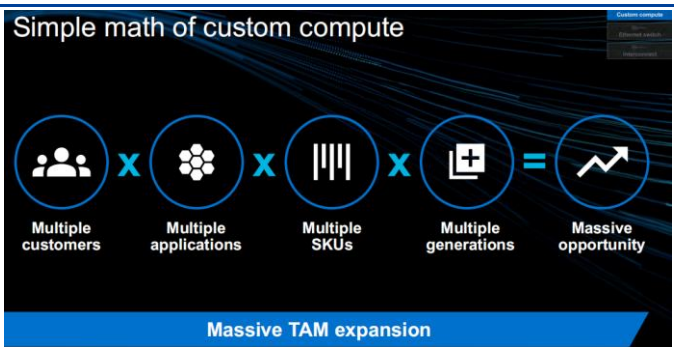
定制加速计算芯片需求涌现, 2028 年市场规模有望达 429 亿美元。目前, 训练阶段的训练集群对加速计算芯片的需求已经提升到万卡数量级。随着 AI 模型训练要求的日益增长, 向 10 万卡级别迈进已然在望。在推理阶段, 因计算量与业务应用息息相关, 单个推理集群所需加速计算芯片数量通常低于训练集群, 但推理集群的部署数量却极为可观, 预计将达到百万级, 远多于训练集群的数量。AI 算力集群特别是推理集群对加速计算芯片的庞大需求, 驱动 ASIC 快速成长。根据 Marvell, 2023 年定制加速计算芯片市场规模为 66 亿美元, 占加速芯片的 16%, 预计到 2028 年定制加速计算芯片市场规模将达 429 亿美元, 占加速芯片的 25%, 2023-2028 年 CAGR 为 45%。

图表19: 推理集群的部署数量极为可观



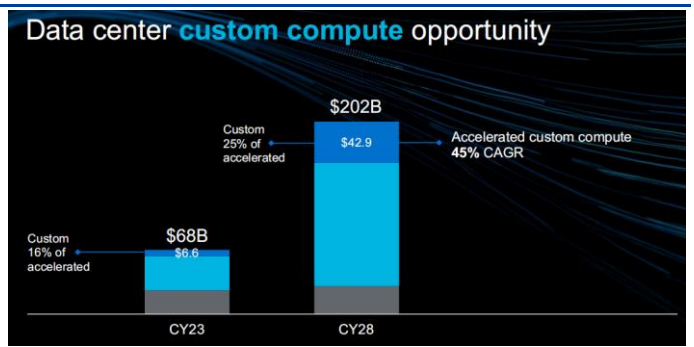
资料来源: Marvell, 国盛证券研究所

图表20: 定制加速计算市场规模扩大逻辑



资料来源: Marvell, 国盛证券研究所

图表21: 数据中心定制加速计算市场规模



资料来源: Marvell, 国盛证券研究所

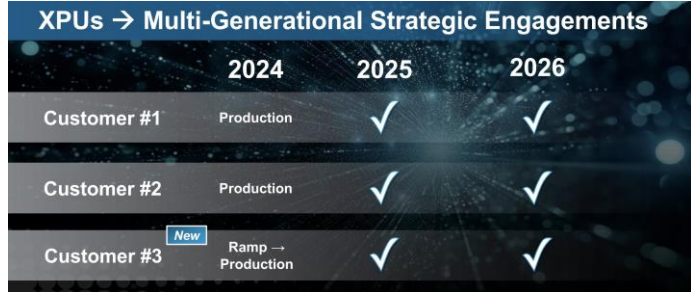
云厂商积极布局, 博通/Marvell/世芯等公司与 CSP 共同开发 XPU。XPU 的开发通常以谷歌、亚马逊等科技巨头为主导, 博通/Marvell 等芯片企业提供辅助配套服务。根据博通, 其与客户 1 有长达 10 年的共同开发历史, 合作了 10 款 XPU, 与客户 2 有 4 年的共同开发历史, 合作了 4 款 XPU, 24 年除了客户 1 与客户 2 之外, 又新增了一家客户, 目前博通 2024 年 AI SAM (XPU+AI 连接) 约 150 亿至 200 亿美元, 预计 27 年 SAM 为 600 亿至 900 亿美元, 目前公司正与另外两家 CSP 开发下一代 AI XPU, 预计 27 年之前发展成创收。根据 Marvell, 公司已经与 4 大 CSP 中的 3 家达成合作, 其中, 公司已与 AWS 达成五年多代全面协议; 世芯与 AWS 和英特尔合作紧密, 24Q3HPC 收入占比达 97%, 87% 收入来源于北美, 公司表示 24 年 AWS AI 芯片成长非常显著, 年增将大于 30%, 24Q3 7nm AI ASIC 的出货量超预期、收入创纪录, 5nm AI 加速器爬坡驱动增长。

图表22: 博通与前两位客户的合作历史



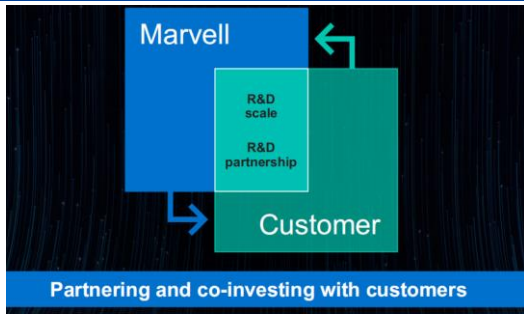
资料来源: 博通, 国盛证券研究所

图表23: 博通获得第三位新客户



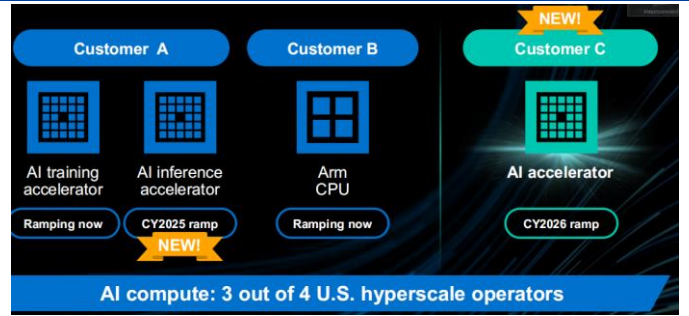
资料来源: 博通, 国盛证券研究所

图表24: Marvell 与客户共同开发 XPU



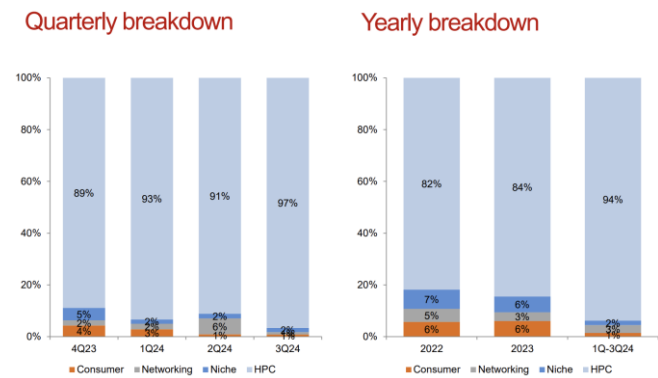
资料来源: Marvell, 国盛证券研究所

图表25: Marvell AI 加速器客户进展



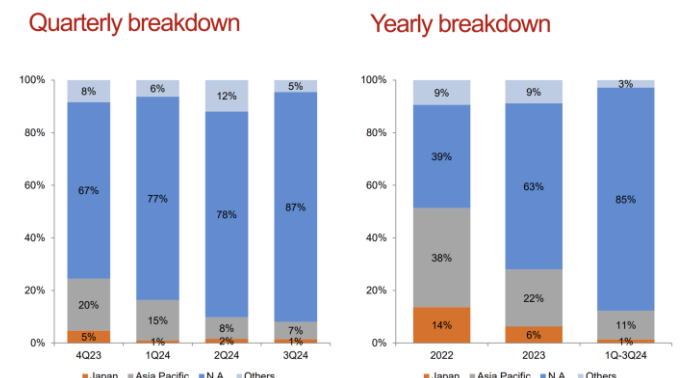
资料来源: Marvell, 国盛证券研究所

图表26: 24Q3 世芯收入 97%来自 HPC



资料来源: 世芯, 国盛证券研究所

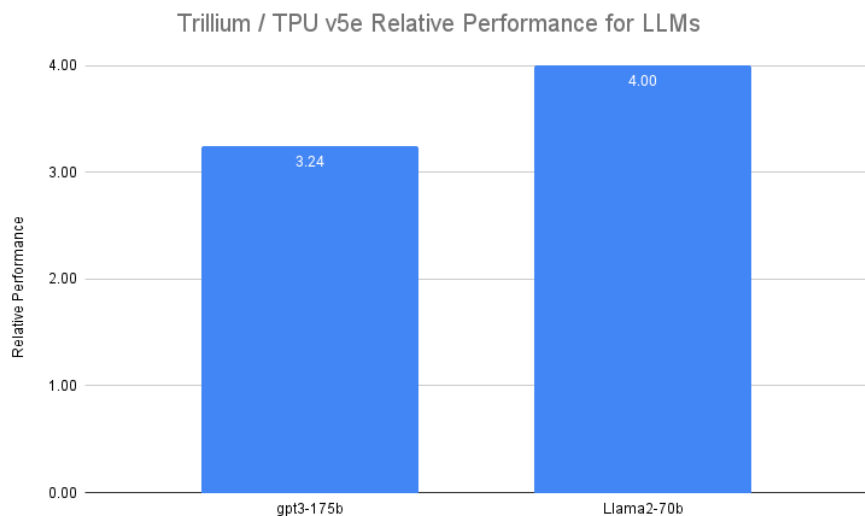
图表27: 24Q3 世芯收入 87%来自北美



资料来源: 世芯, 国盛证券研究所

谷歌推出 Trillium, 算力、内存带宽和互联性能全面提升。Trillium 的规格在多个维度上有重大进步, 与上一代产品相比, 其芯片峰值计算性能提高了 4.7 倍, 高带宽内存容量和芯片间互连带宽增加一倍, 能源效率提高 67%, 这是应对数据中心 AI 训练巨大电力需求的关键指标。在训练 Llama-2-70B 模型时, 测试表明 Trillium 从 4 片 Trillium-256 芯片组到 36 片 Trillium-256 芯片组实现近乎线性扩展, 扩展效率达 99%。每美元的训练性能较上一代芯片提高了 2.5 倍, 可能重塑人工智能开发的经济性, 这对开发大型语言模型的企业和初创公司很重要。

图表28: 谷歌 Trillium 与 v5e 对比



资料来源: 半导体行业观察, 国盛证券研究所

亚马逊推出 **Trainium2**, 投入 **40 亿美元** 部署 **Project Rainier**。亚马逊推出 Trainium2, Trainium2 功耗约 500w, 每芯片具有 650 万亿次浮点运算 (TFLOP/s) 的密集 BF16 性能和 96GB 的 HBM3e 内存容量。亚马逊目前正在进行全球最大的人工智能集群建设之一, 部署了相当数量的 Hopper 和 Blackwell GPU。除了向基于英伟达的集群投入大量资本支出外, AWS 目前正在为 Anthropic 部署一个拥有 20 万+ Trainium2 芯片的集群 Project Rainier, 资本支出投入 40 亿美元。

图表29: Trainium2 与其他主流 AI 芯片性能对比

Trainium2 Overview and Comparison					
	Trn2	Trn2-Ultra	TPUv6e	GB200	H100
Theoretical BF16 Dense TFLOP/s/chip	650	650	918	2500	989
Theoretical 8 bit Dense TFLOP/s/chip	1300	1300	1836	5000	1978
HBM Capacity (GByte/chip)	96	96	32	192	80
HBM Bandwidth (GByte/s/chip)	3200	3200	1640	8000	3350
Arithmetic Intensity (BF16 FLOP per Byte)	203.1	203.1	559.8	312.5	295.2
FLOP per HBM Capacity (BF16 FLOP per Byte)	6770.8	6770.8	28687.5	13020.8	12362.5
Structured Sparsity Support	2.4, 4.8, 4:16	2.4, 4.8, 4:16	None	2.4	2.4
Chip Power (Watts)	500	500	600	1200	700
Cooling Technology	Air Cooled	Air Cooled	Air Cooled	Direct To Chip Liquid (DLC)	Mostly Air Cooled/Some DLC
Scale Up Networking					
Scale Up Technology	NeuronLink3 (PCIe Gen 5.0 Based)	NeuronLink3 (PCIe Gen 5.0 Based)	TPU ICI	NVLink5	NVLink4
Scale Up Bandwidth (GByte/s/chip Unidirectional)	640	640	448	900	450
Scale Up World Size (# of chips)	16	64	256	72	8
Scale Up Topology	4x4 2D Torus Symmetric BW along all axis	4x4x4 3D Torus (With Z axis being half as much BW)	16x16 2D Torus Symmetric BW along all axis	18-Rail Optimized Fat Tree With NVLink SHARP In Network Redux	4-Rail Optimized Fat Tree With NVLink SHARP In Network Redux
Number of Scale Up Neighbors Per Chip	4	6	4	Any to Any	Any to Any
Total HBM Capacity Per Scale Up World	1536	6144	8192	13824	640
# of Physical Servers Per Scale Up World	1	4	32	18	1
Scale Out Networking					
Scale Out Technology	EFAv3 (AWS Flavor of Ethernet)	EFAv3 (AWS Flavor of Ethernet)	Google Flavor of Ethernet	InfiniBand/Ethernet	InfiniBand/Ethernet
Scale Out Bandwidth (Gbit/s/chip Unidirectional)	Up to 800	200	100	800	400
Scale Out Topology	ToR Fat Tree	ToR Fat Tree	ToR Fat Tree	4-Rail Optimized Fat Tree	8-Rail Optimized Fat Tree
Scale Up to Scale Out Bandwidth Ratio	6.4	25.6	35.84	9	9

资料来源: SemiAnalysis, 国盛证券研究所

1.3 中国算力芯片逐步筑基，国产替代势在必行

美国对华的科技封锁持续加剧，AI芯片国产替代势在必行。2022年9月，美国政府通过设置功率门槛和芯片之间通信速度门槛，限制英伟达A100和H100以及AMD MI250对华出口。英伟达根据限制条件为中国研发出了A800和H800芯片。2023年10月，为了进一步限制中国AI产业，美国升级限制力度，要求英伟达禁止向中国出售运行速度达到300 teraflops(即每秒可计算3亿次运算)及以上的数据中心芯片，A800和H800也被禁止出售中国。英伟达面向国内的产品性价比偏低，国产芯片做出性能、做出性价比，自主可控市场空间广阔。

图表30: BIS关于中国芯片出口禁令



资料来源: BIS, 国盛证券研究所

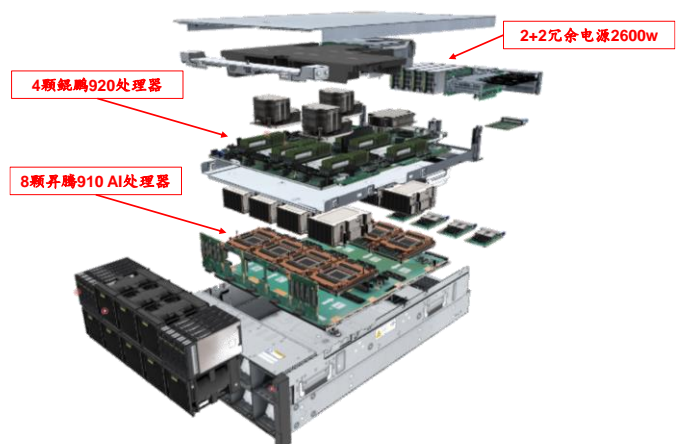
Atlas 800T A2 适用于 AI 训练场景，满足大规模数据中心集群部署。 Atlas 800T A2 训练服务器是基于华为自研鲲鹏 920 处理器和昇腾 910 AI 处理器的 AI 训练服务器，其中，服务器中包括 4 颗鲲鹏 920 处理和 8 颗昇腾 910 AI 处理器，支持 4 个热插拔 2600w 电源模块，支持 2+2 冗余备份。Atlas 800T A2 广泛应用于深度学习模型开发和 AI 训练服务场景。该设备面向公有云、互联网、运营商、政府、交通、金融、高校、电力等领域，具有高计算密度、高能效比、高网络带宽、易扩展、易管理等优点，满足企业机房部署和大规模数据中心集群部署。

图表31: Atlas 800T A2 训练服务器原始视图



资料来源: 华为, 国盛证券研究所

图表32: Atlas 800T A2 训练服务器爆炸图



资料来源: 华为, 国盛证券研究所

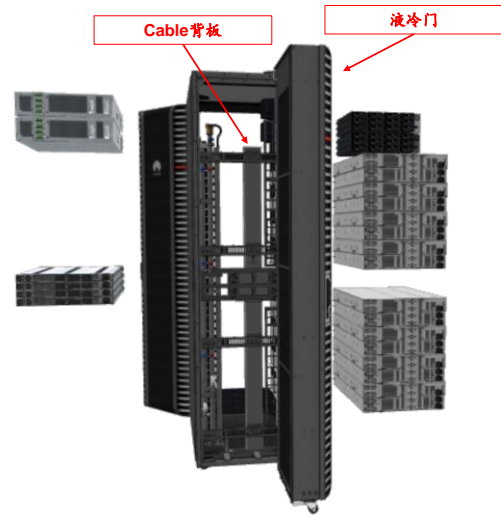
Atlas 900 A2 PoD 集群基础单元支持 8 个 4U 计算节点，配置 4 台 DX511 交换机，支持 24 个 PSU，单 PSU 3kw，最大支持 66kw 供电，具有高密度，高效能，高可靠，易拓展，一体化交付，极简运维，低 TCO 等特点，支持大规模，高性能 AI 训练业务，可应用于人工智能计算中心、运营商、云计算等重算力场景。

图表33: Atlas 900 A2 PoD 集群基础单元爆炸图 (1)



资料来源: 华为, 国盛证券研究所

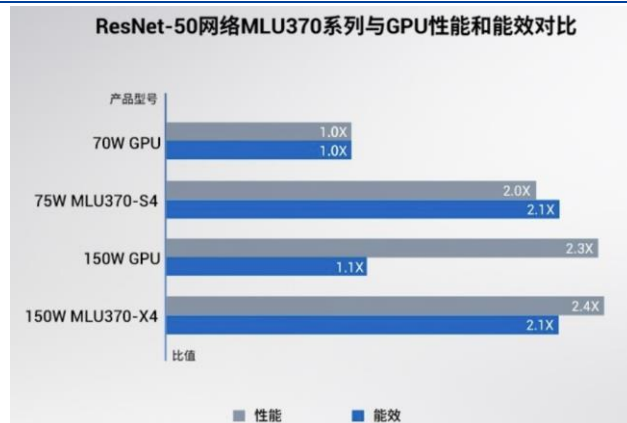
图表34: Atlas 900 A2 PoD 集群基础单元爆炸图 (2)



资料来源: 华为, 国盛证券研究所

寒武纪思元 370 实测性能表现优秀，国产算力再下一城。寒武纪目前智能加速卡有思元 370 系列、290、270 系列，其中，思元 370 系列基于 7nm 制程工艺，是公司首款采用 chiplet(芯粒)技术的 AI 芯片，集成 390 亿个晶体管，最大算力高达 256TOPS (INT8)，是思元 270 算力的 2 倍。凭借最新智能芯片架构 MLUarch03，思元 370 实测性能表现更为优秀。思元 370 是国内第一款公开发布支持 LPDDR5 内存的云端 AI 芯片，内存带宽是上一代产品的 3 倍，访存能效达 GDDR6 的 1.5 倍，搭载 MLU-Link™多芯互联技术，在分布式训练或推理任务中为多颗思元 370 芯片提供高效协同能力。

图表35: 思元 370 系列板卡与业内主流 GPU 性能对比



资料来源: 寒武纪官网, 国盛证券研究所

发布景宏填补 AI 空白，产品矩阵不断完善。景嘉微是国内首家成功研制国产 GPU 芯片并实现大规模工程应用的企业，先后研制成功 JM5 系列、JM7 系列、JM9 系列等具有自主知识产权的高性能 GPU 芯片。公司 2024 年成功研发了“景宏”系列高性能智算模块与整机产品，填补了公司在 AI 训练、AI 推理和科学计算领域的产品空白。景宏支持 INT8、FP16、FP32、FP64 等混合精度运算，支持全新的多卡互联技术进行算力扩展，同时适配国内外主流 CPU、操作系统及服务器厂商，能够支持当前主流的计算生态、深度学习框架和算法模型库，大幅缩短用户适配验证周期。景嘉微作为国产 GPU 领军企业，其首款

面向 AI 领域的算力产品丰富了自身的产品线，AI 系列产品在增加公司竞争力的同时有望成为公司业绩增长的全新动力。

深算二号性能翻倍提升，多角度持续产品优化。海光信息 DCU 产品深算二号于 2023H2 实现了在大数据处理、人工智能、商业计算等领域的商业化应用，能够完整支持大模型训练，同时性能相较于深算一号提升 100% 以上。2023 年公司持续的研发投入使其产品在处理器体系结构设计、处理器核心微结构验证、处理器安全架构等多角度持续优化，其中：在处理器体系结构设计方面，采用高带宽低延时 chiplet 互联技术，不断提升计算性能；在处理器核心微结构验证方面建立了处理器核心功能部件级、处理器核心级、处理器核心簇级、全片多核心簇级、多芯粒级和多芯片级完整的多层次处理器验证环境。

图表36: 海光信息产品族群



资料来源: 海光信息公告, 国盛证券研究所

图表37: 海光产品在大模型的适配与实践



资料来源: 海光信息公告, 国盛证券研究所

燧原科技发布 S60，性能持续提升。2018 年 3 月燧原科技成立，燧原科技在仅仅 18 个月月后，发布“邃思 1.0”，这款 AI 训练芯片一次性流片即收获成功，创造纪录，燧原科技在业界崭露头角。之后，燧原科技对外发布基于“邃思”的云端训练加速卡“云燧 T10”和第一代推理产品“云燧 i10”。2021 年，燧原科技发布第二代训练产品“云燧 T20/T21”和推理产品“云燧 i20”，以及配套的“驭算”软件平台，燧原科技成为当时国内第一家云端训练和推理产品迭代到第二代的企业。2024 年 7 月，新一代推理产品燧原 S60 发布。架构升级，计算效率提升，大内存满足大模型参数增长需求，PCIe5.0 提升传输速率，虚拟化支持灵活的算力分配，覆盖 200+ 主流模型，降低客户应用迁移成本，缩短业务上线周期，大幅提升模型泛化支持能力，软件易用性持续提升。

图表38: 燧原 S60



资料来源: 燧原科技官网, 国盛证券研究所

1.4 AI 需求涌现，晶圆代工国产化空间广阔

全球 AI 布局叠加供应链补库，预计 2025 年晶圆代工产值将年增 20%。据 TrendForce 集邦咨询最新调查: 2024 年由于消费类产品终端市场不景气，零部件厂商备货较为保守。因此晶圆代工厂平均产能利用率低于 80%，只有高性能计算（HPC）产品和旗舰智能手机主流采用的 5/4/3nm 等先进制程能保持满负荷生产。尽管 2025 年消费类终端市场前

景依旧不明朗，但汽车、工控等供应链的库存从2024年下半年开始逐渐见底，再加上边缘人工智能（Edge AI）使单个整机的晶圆消耗量增加，以及云人工智能（Cloud AI）持续布局，预计2025年晶圆代工产值将同比增长20%，好于2024年的16%。

图表39: 全球晶圆代工市场规模

2018-2025晶圆代工产业营收与年增率

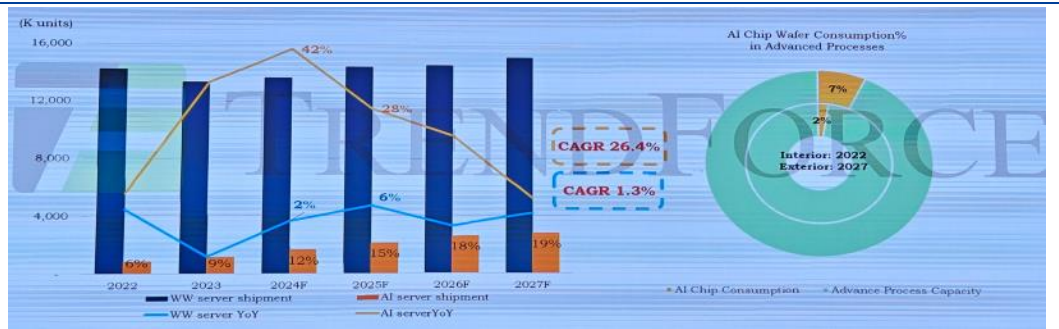


注: 部分资讯取自台积电CSR报告
Source: TrendForce, Sep. 2024

资料来源: Trendforce, 国盛证券研究所

AI需求为代工行业带来新动能。根据TrendForce的预计，AI服务器市场在2024年将实现42%的成长，2025年预计将保持28%的同比增长。以AI服务器所带动的AI芯片需求为例，**从AI芯片在整个先进工艺中的产能占比来看，2022年的占比仅有2%，2024年预计将会达到4%，预计到2027年占比将会达到7%，**对整个晶圆代工产业的产值贡献正在快速增长。2025年智能手机、笔记本电脑市场在端侧生成式AI需求的推动下，出货量有望分别实现同比2%和5.7%的增长。同时端侧生成式AI还将进一步推动AI手机和AI PC对于DRAM容量需求的显著提升。地缘政治影响，台积电等海外代工厂为中国大陆客户提供7nm及以下更先进制程的芯片代工能力可能受限，中国大陆先进制程亟需进步，解决卡脖子问题，实现国产替代。

图表40: AI芯片产能占比



资料来源: TrendForce, 国盛证券研究所

24Q3全球晶圆代工市场产能利用率持续改善，先进制程及先进封装供不应求。根据TrendForce，2024Q3受益于下半年智能手机、PC/笔电新品带动供应链备货，加上AI服务器相关HPC需求持续强劲，整体晶圆代工产能利用率较第二季改善，Q3全球前十大晶圆代工业者产值季增9.1%，达349亿美元，这一成绩部分归功于高价的台积电3nm制程大量贡献产出。2024Q4，TrendForce预计先进制程将继续推动前十大厂商产值增长，不过环比增幅将略有缩小，且运营表现将两极分化。预计到年底，人工智能、旗舰

智能手机、PC 主芯片将带动 5/4nm、3nm 制程的需求，CoWoS 先进封装仍将供不应求。至于 28nm（含）以上的成熟制程，由于终端销售情况不明朗，且进入 2025Q1 淡季，消费类产品在 2024 年第三季度备货后，TV SoC、LDDI、Panel PMIC 等外围集成电路的备货需求大幅降低，这些因素将被中国智能手机品牌年底冲量以及中国以旧换新补贴刺激供应链紧急订单的效应抵消，预计 Q4 成熟制程产能利用率将与环比持平或略有增长。

图表 41: 24Q3 前十大代工厂排名 (收入单位: 百万美元)

Ranking	Company	Revenue			Market Share	
		3Q24	2Q24	QoQ	3Q24	2Q24
1	台积电 (TSMC)	23,527	20,819	13.0%	64.9%	62.3%
2	三星 (Samsung)	3,357	3,833	-12.4%	9.3%	11.5%
3	中芯国际 (SMIC)	2,171	1,901	14.2%	6.0%	5.7%
4	联电 (UMC)	1,873	1,756	6.7%	5.2%	5.3%
5	格芯 (GlobalFoundries)	1,739	1,632	6.6%	4.8%	4.9%
6	华虹集团 (Huahong Group)	799	708	12.8%	2.2%	2.1%
7	高塔半导体 (Tower)	371	351	5.6%	1.0%	1.1%
8	世界先进 (VIS)	366	342	6.9%	1.0%	1.0%
9	力积电 (PSMC)	336	320	4.9%	0.9%	1.0%
10	合肥晶合 (Nexchip)	332	300	10.7%	0.9%	0.9%
Total of Top 10		34,869	31,962	9.1%	96%	96%

资料来源: Trendforce, 国盛证券研究所

中国芯片自给率急需提高。根据 TechInsights 的统计数据，2020 年中国大陆芯片市场规模约为 1460 亿美元，而中国本土生产的芯片规模约为 242 亿美元，自给率约 16.6%，2021 年芯片自给率约 17.6%，2022 年约 18.3%，芯片自给率提升空间大。

图表 42: 中国芯片自给率



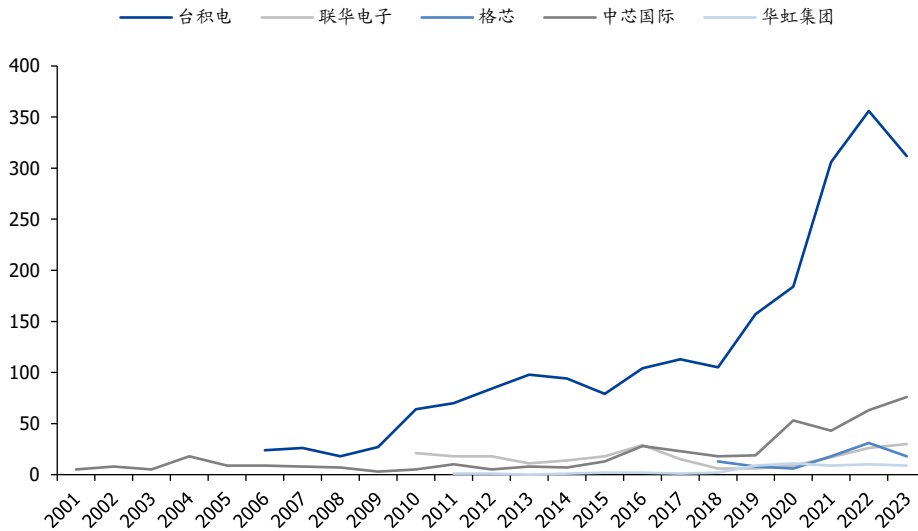
资料来源: 半导体产业纵横, TechInsights, 国盛证券研究所

中国大陆晶圆厂发展速度全球领先，资本开支力度强劲:

- 台积电: 适度收紧资本开支规划，2023 年相较于 2022 年资本开支收窄，2023 年资本开支 304.5 亿美元，计划 2024 年资本支出介于 280-320 亿美元之间，其中约 70%-80% 用于先进制程技术，10%-20% 用于成熟和特色工艺，10% 用于先进封装测试及掩模生产等设备采购。

- 联华电子：2023年资本支出约30亿美元，2024年预计约33亿美元，其中95%将投入12英寸厂，5%用于8英寸厂。
- 中芯国际：资本开支并未削减，2023年资本开支为75亿美元，同比增长17.3%，主要用于扩充产能，计划在2024年继续推进12英寸工厂及产能建设计划，预计2024年资本支出与2023年大致持平。
- 华虹：目前公司单一最大的资本开支即为新的12英寸产线，总体约67亿美元的投资计划。该投资从2023年年中开始，贯穿2024-2026年，平均来看，约20亿美元/年，2025年的资本开支会在20-25亿美元左右。

图表43: 晶圆代工厂资本支出(亿美元)

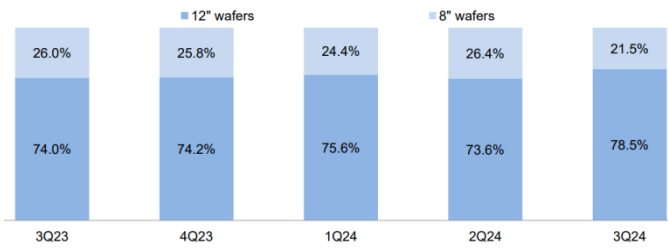


资料来源: Wind, 国盛证券研究所

中芯国际位居全球晶圆代工企业销售额第三。中芯国际是世界领先的集成电路晶圆代工企业之一，也是中国大陆集成电路制造业领导者，拥有领先的工艺制造能力、产能优势、服务配套。根据全球各纯晶圆代工企业最新公布的2023年销售额情况排名，中芯国际位居全球第四位，在中国大陆企业中排名第一。单季度来看，根据Trendforce数据，自24Q1起，受益于消费性库存回补及国产化趋势加速，24Q1中芯国际销售额超越格芯与联电，居全球晶圆代工企业第三，且受益于产品组合优化及12英寸新增产能开出，公司24Q2，24Q3稳居第三，全球市场份额24Q3达6.0%。

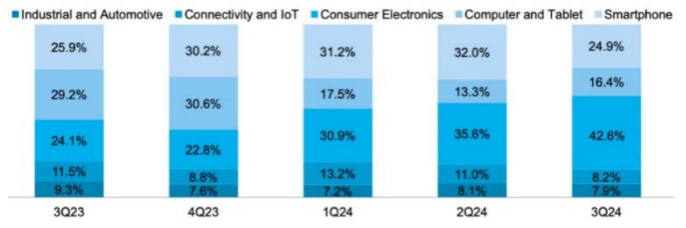
产能持续扩建，产品结构进一步优化。从晶圆尺寸来看，中芯国际12英寸晶圆产品营收占比逐步提升，24Q3占比达78.5%，且根据公司24Q3法说会，公司24Q3新增2.1万片12英寸晶圆产能，预计Q4再释放3万片每月12英寸产能，当前12英寸产能持续接近满载，且对应节点附加值较高，公司提出自**2023年起，未来5-7年将新增12英寸产能34万片/月，主要集中于中芯东方、中芯深圳、中芯西青、中芯京城**，伴随公司产能扩充，产品结构进一步优化，ASP提升，且12英寸产能进一步释放有助于增强覆盖客户需求能力，提升竞争力。分领域来看，自23Q4起，消费电子领域营收占比逐步提升，24Q3占比达42.6%，受益于下游市场温和复苏及消费电子厂商提前备货。

图表44: 中芯国际 23Q3-24Q3 分晶圆尺寸营收占比



资料来源: 中芯国际公告, 国盛证券研究所

图表45: 中芯国际 23Q3-24Q3 分领域营收占比



资料来源: 中芯国际公告, 国盛证券研究所

成熟制程覆盖全面, 大陆首具备量产 14nm FinFET 能力本土晶圆厂商。中芯国际拥有全面一体的集成电路晶圆代工核心技术体系, 成功开发了 0.35 微米至 14 纳米的多种技术节点, 其中公司 28nm 具备高介电常数金属栅极、锗硅应力提升技术和超低电介质材料铜互联工艺, 可用于高性能应用处理器、移动基带及无线互联芯片领域, 14nm 应用 FinFET 新型器件, 高性能/低功耗, 支持超低工作电压, 应用多重曝光图形技术, 集成度超过 3×10^9 个晶体管/平方厘米, 可应用于高性能低功耗计算及消费电子产品领域, 例如智能手机、平板电脑、机顶盒、AI、射频、车载和物联网等领域。

图表46: 中芯国际部分技术节点情况

技术节点	表征及特点	应用领域	先进程度
14 纳米	①应用 FinFET 新型器件, 高性能/低功耗, 支持超低工作电压; ②应用多重曝光图形技术, 集成度超过 3×10^9 个晶体管/平方厘米; ③应用高介电常数金属栅极技术, 提供三种不同阈值电压的核心器件; ④低介电常数介质的铜互连技术, 支持最多 13 层金属互联。	高性能低功耗计算及消费电子产品领域, 例如智能手机、平板电脑、机顶盒、AI、射频、车载和物联网等领域。	国际领先
28 纳米	①具备高介电常数金属栅极、锗硅应力提升技术和超低电介质材料铜互联工艺; ②运用了 193 纳米浸润式两次微影技术和形成超浅结的毫秒级退火工艺; ③核心组件电压 0.9V, 具有三种不同阈值电压。	高性能应用处理器、移动基带及无线互联芯片领域, 例如智能手机、平板电脑、电视、机顶盒和互联网等领域。	国际领先
45/40 纳米	①核心组件电压 1.1V, 涵盖三种不同阈值电压; ②运用了先进的浸润式光刻技术, 应力技术, 超浅结技术以及低介电常数介质等技术。	手机基带及应用处理器、平板电脑多媒体应用处理器、数字电视、机顶盒、游戏及其他无线互联应用等领域。	国际领先
65/55 纳米	①基于完备的设计规则、规格及 SPICE 模型; ②核心元件电压: 1.2V, 输入/输出电压: 1.8V, 2.5V 和 3.3V。	高性能、低功耗的应用领域, 如移动应用领域和无线应用等领域。	国际领先
90 纳米	①低介电常数介质的铜互连技术; ②支持客户定制, 达到各种设计要求, 包括高速, 低耗, 混合信号, 射频以及嵌入式和系统集成等方案。	低能耗, 卓越性能及高集成度领域, 如无线电话、数字电视、机顶盒、移动电视、个人多媒体产品、无线网络接入及个人计算机应用芯片等。	国内领先
0.13/0.11 微米	①采用全铜制程技术; ②使用 8 层金属层宽度仅为 80 纳米的门电路, 核心元件电压: 1.2V, 输入/输出电压: 2.5V 和 3.3V。	低成本领域, 如闪存控制器、媒体播放器和其他各种应用产品等领域。	国际领先
0.18/0.15 微米	①采用铝制程技术, 特点是每平方毫米的多晶硅门电路集成度高达 100,000 门; ②有 1.8V、3.3V 和 5V 三种不同电压。	低成本领域, 如智能卡、移动/消费应用和汽车和工业应用产品等领域。	国际领先
0.35/0.25 微米	①采用铝制程技术; ②有 2.5V、3.3V 和 5V 三种不同电压。	智能卡、消费性产品以及其它多个领域。	国内领先

资料来源: 中芯国际公告, 国盛证券研究所

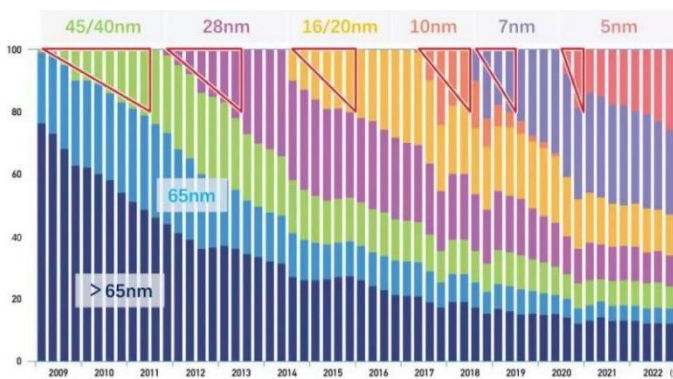
本土优势+品牌效应，打造晶圆代工国产替代龙头。相对于国内竞争对手，公司已与境内外领先芯片设计厂商建立了长期稳定的合作关系，以便掌握行业、产品最新技术动态，及时了解和把握客户最新需求，准确地进行芯片产品更新升级，确保公司产品在市场竞争中保持先发优势，同时积累产品行业应用经验，完善产品性能，提高产品质量水平。相对于国外竞争对手，公司一方面更加贴近、了解本土市场，能够快速响应客户需求，提供充分的服务支持；另一方面，公司与国内芯片设计厂商在企业文化、市场理念和售后服务等方面更能相互认同，业务合作通畅、高效，形成了密切且相互依存产业生态链。根据 SEMI 数据，中国大陆 12 英寸晶圆产能全球份额将从 2022 年 22% 增加到 2026 年的 25%，产能达到每月 240 万片，当前全球晶圆产能制程节点向 28nm 及以下集中，先进制程晶圆进口存在极大不稳定性，公司以本土优势+品牌效应为基础，在国内需求+国产替代需求双重上升拉动下，晶圆代工国产替代龙头地位进一步巩固。

图表47：分地区全球12英寸晶圆产能占比



资料来源：SEMI，国盛证券研究所

图表48：全球晶圆产能制程节点分布



资料来源：头豹研究院，国盛证券研究所

图表49：全球六大逻辑晶圆厂先进制程进展情况

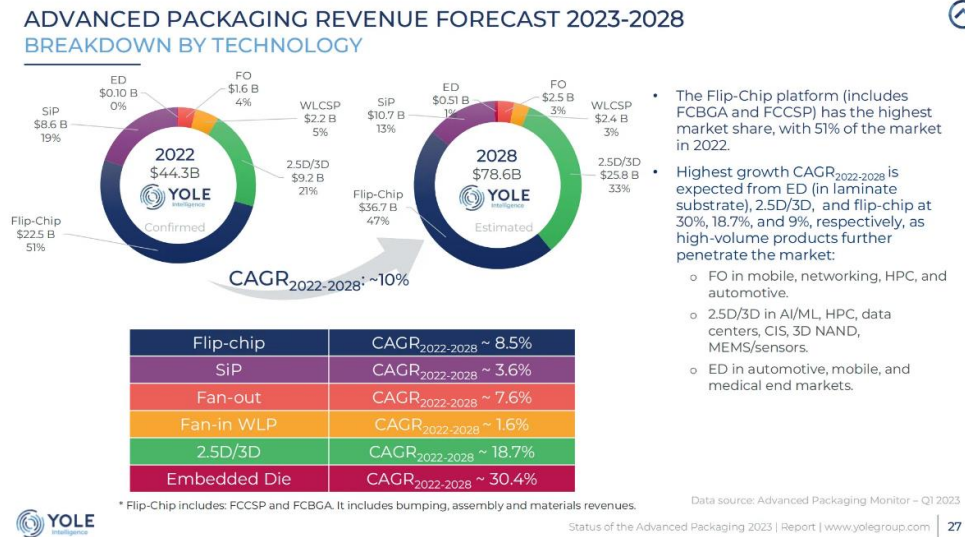
	2016	2017	2018	2019	2020	2021	2022
Intel	14nm+	10nm (limited) 14nm++	10nm	10nm+	10nm++	10nm++	7nm EUV
Samsung	10nm	10nm	8nm	7nm EUV 6nm EUV	18nm FDSOI 5nm	4nm	3nm GAA
TSMC	10nm	12nm	7nm EUV	7nm+ EUV	5nm	6nm	5nm+ 4nm 3nm
GlobalFoundries			22nm FDSOI 12nm finFET	12nm FDSOI	22nm+ FDSOI 12nm+ finFET		
SMIC				14nm finFET	12nm finFET	8-10nm finFET	
UMC		14nm finFET			22nm platform		

资料来源：IC insight，国盛证券研究所

1.5 AI时代先进封装不可或缺，本土力量崭露头角

AI时代先进封装不可或缺，2.5D/3D封装增速领先。2022年，先进封装市场的价值达443亿美元，预计到2028年将超过786亿美元，2022至2028年间年复合增长率(CAGR)为10%。先进封装技术主要包括Flip Chip倒装、WLCSP晶圆级封装、Fan Out扇出、2.5D/3D封装等，22年占比最高的还是Flip Chip倒装，占比超过50%。但随着AI和HBM的应用近年来迎来爆发，市场对高性能封装解决方案的持续需求，以及新兴技术如芯片互连和异构集成的推动下，2.5D/3D封装技术增速较快，2022-2028年间的复合增速达18.7%。

图表50: 先进封装收入预测 2023-2028



资料来源: yole, 国盛证券研究所

图表51: 各类型先进封装技术在终端的应用情况

应用领域	CPU/GPU	APU	DPU	MCU	ASIC	FPGA	存储	传感器	模拟	光电子
人工智能					FC, FO					
智能驾驶	FC, 2.5D/3D, FO, SIP	FC, FO, ED		FC, WB, QFN, WLCSP		FC, 2.5D/3D, FO		FC, FO, WB, QFN, WLCSP, SIP	FC, FO, WB, QFN, ED, SIP	
AR/VR										
HPC			FC, FO, ED			FC, 2.5D/3D, FO				FC, 2.5D/3D, WB, SiP
IoT				FC, WB, QFN, WLCSP			FC, 3D, WB, QFN, WLCSP, SIP	FC, FO, WB, QFN, WLCSP, SIP		
5G	FC, 2.5D/3D, FO, SIP									
手机通信		FC, FO, ED						FC, FO, WB, QFN, WLCSP, SIP	FC, FO, WB, QFN, ED, SIP	FC, 2.5D/3D, WB, SiP
区块链	FC, 2.5D/3D, FO				FC, 2.5D/3D, FO					

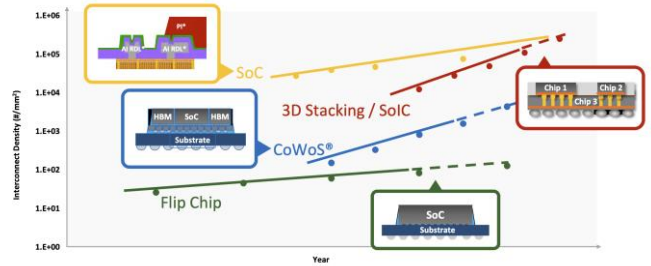
资料来源: 集微网, 国盛证券研究所

从 CoWoS 走向 SoIC, 互联密度成倍提升。台积电的系统集成技术可分为两大平台, 一是 3D 堆叠技术平台 SoIC; 二是先进封装技术平台, 涵盖了 InFO、CoWoS 和 SoW 技术。CoWoS 技术使得单位面积上的互联密度比倒装技术提高了 10 倍以上, 而 3D 堆叠技术 SoIC 则可以把互联密度提升 100 倍以上。

图表52: 台积电系统集成技术平台

图表53: 先进封装技术使得连接密度指数级增长

System Integration Technologies

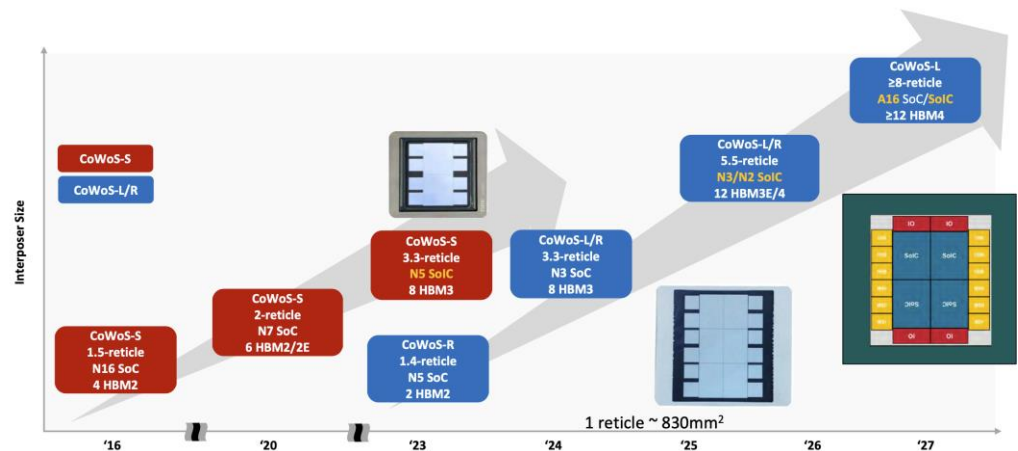


资料来源: 台积电, 国盛证券研究所

资料来源: 台积电, 国盛证券研究所

CoWoS-L 进一步扩大掩模尺寸, 容纳更多的 HBM。 CoWoS 作为台积电面向高端的先进封装解决方案, 目前在 AI 芯片的封装中得到了广泛的应用。随着算力需求的提升, 封装技术需要能集成更多的 HBM 和更大的 GPU, 因此 CoWoS 技术一直致力于提升掩模尺寸。2016 年的 CoWoS-S 技术实现了 1.5x 掩模板面积的硅中介层, 而到 2027 年 CoWoS-L 将通过局部硅互连 (local silicon interconnect, LSI) 和全局重布线 (global redistribution layers), 形成一个重组的中介层以替代 CoWoS-S 中的单片硅中介层, 从而能做到 8 倍掩模板面积。

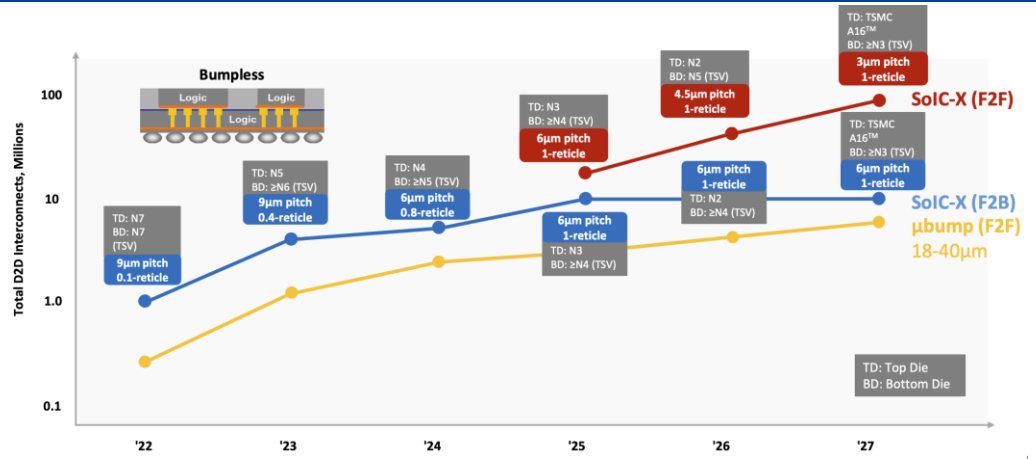
图表54: 台积电 2.5D 集成技术赋能下一代 AI 计算



资料来源: 台积电, 国盛证券研究所

3D 堆叠技术进展迅速, 混合键合助力互联密度提升。 采用 3D 堆叠技术的首个应用于数据中心的 CPU 在 2022 年实现出货, 首个 GPU 则在 2023 年实现出货。3D 堆叠技术实现了互联密度大规模的提升, 主要归功于混合键合工艺, 从而使 bump pitch 微缩到 10um 以下。

图表55: 台积电 3D 堆叠技术发展历程



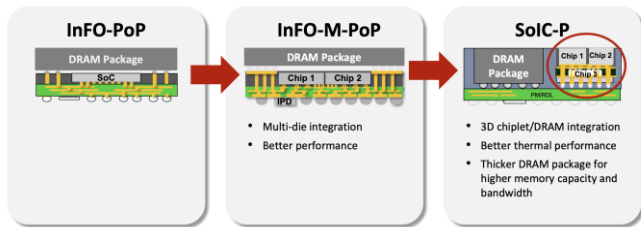
资料来源: 台积电, 国盛证券研究所

移动端的 AI 应用: 目前主要采用 InFO-PoP 技术将 DRAM 和 SoC 封装在一起, 未来的 InFO-M-PoP 技术将通过异构集成实现计算芯片和存储芯片的集成。台积电给出的终极解决方案则是 SoIC-P 技术, 该方案将提供更好的散热性能, 同时兼顾更厚的 DRAM 封装, 以带来更大的存储容量和带宽。

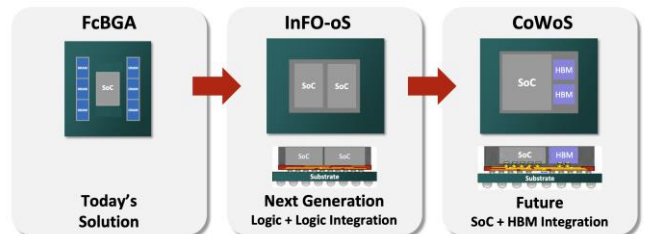
车载 AI 芯片: 目前仍以 FcBGA 为主, 未来将使用 InFO-oS 技术以实现 2 个逻辑芯片的集成, 使用 CoWoS 技术以实现逻辑芯片和 HBM 的集成。Chiplet 技术的运用将大幅提升 ADAS 的性能。

图表56: 移动端的 AI 芯片未来有望从 InFO-PoP 向 SoIC 发展

图表57: 车载 AI 芯片有望从 FcBGA 向 CoWoS 发展



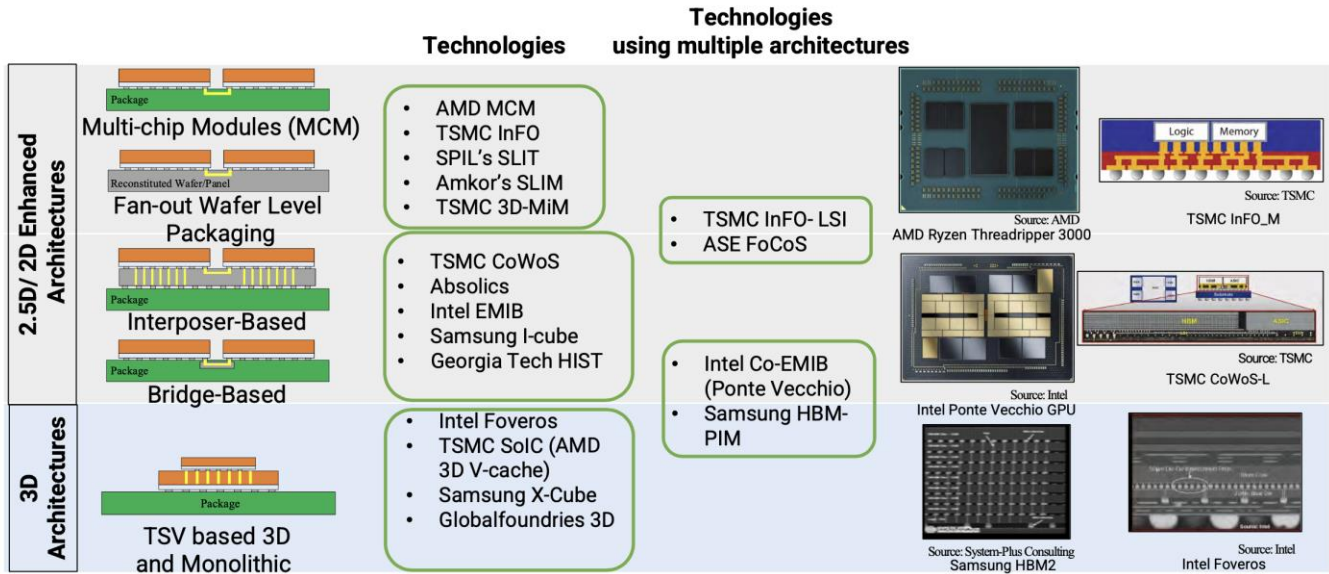
资料来源: 台积电, 国盛证券研究所



资料来源: 台积电, 国盛证券研究所

先进封装技术百花齐放: 封测厂扎根于载板, 主要聚焦于 CoWoS 中的后道 oS 工艺, 提供的解决方案也以 2D/2.5D 为主。由于 CoW 工艺与前道工艺相近, 代工厂更擅长对晶圆的加工能力, 因此代工厂在 2.5D/3D 技术上的布局更为前瞻。

图表58: 主要厂商的先进封装技术



资料来源: IEDM 2024, 国盛证券研究所

中国封测厂在先进封装领域逐渐崭露头角:

- 长电科技: XDFOI® Chiplet 高密度多维异构集成系列工艺已进入稳定量产阶段。2024年资本开支60亿,持续聚焦高性能先进封装和测试产的扩张,以满足客户在云计算、数据中心、高性能存储、高密度电源管理芯片方面的需求。应用于高端服务器的高密度电源模块需求旺盛,产品占比显著提升。
- 通富微电: 基于高端处理器和AI芯片需求旺盛,公司上半年高性能封装业务保持稳步增长。公司大力开发扇外型、圆片级、倒装焊、Chiplet、2D+、基于玻璃芯基板和玻璃转接板的FCBGA芯片封装技术等技术。
- 甬矽电子: 先进封装产线核心设备全部move in,正在通线和验证阶段。公司多维异构项目主要包括: 1)用于激光雷达、消费类AI产品的Fan-out产品,已进入前期打样阶段; 2)不含interposer的RDL产品,应用领域广泛,性价比高; 3)针对运算和服务器主芯片的大芯片产品。

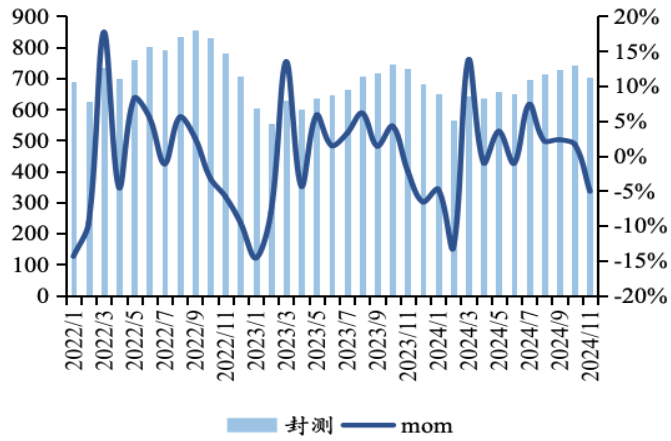
图表59: 封测厂先进封装进展

公司名称	先进封装进展
长电科技	XDFOI® Chiplet 高密度多维异构集成系列工艺已进入稳定量产阶段。2024年资本开支60亿,持续聚焦高性能先进封装和测试产的扩张,以满足客户在云计算、数据中心、高性能存储、高密度电源管理芯片方面的需求。应用于高端服务器的高密度电源模块需求旺盛,产品占比显著提升。
通富微电	基于高端处理器和AI芯片需求旺盛,公司上半年高性能封装业务保持稳步增长。公司大力开发扇外型、圆片级、倒装焊、Chiplet、2D+、基于玻璃芯基板和玻璃转接板的FCBGA芯片封装技术等技术。
甬矽电子	先进封装产线核心设备全部move in,正在通线和验证阶段。公司多维异构项目主要包括: 1)用于激光雷达、消费类AI产品的Fan-out产品,已进入前期打样阶段; 2)不含interposer的RDL产品,应用领域广泛,性价比高; 3)针对运算和服务器主芯片的大芯片产品。

资料来源: Wind, 国盛证券研究所

台股封测板块月度营收趋势良好,进入淡季11月营收环比小幅回落。台股封测板块11月营收合计为703亿新台币,环比下滑5.1%,主要受进入淡季扰动。此前2024年2-10月板块营收趋势良好,稼动率逐步攀升。作为重资产行业,稼动率提升有望拉动行业利润率提升,规模效应下利润显著增厚。

图表60: 台股封测板块月度营收(亿新台币)



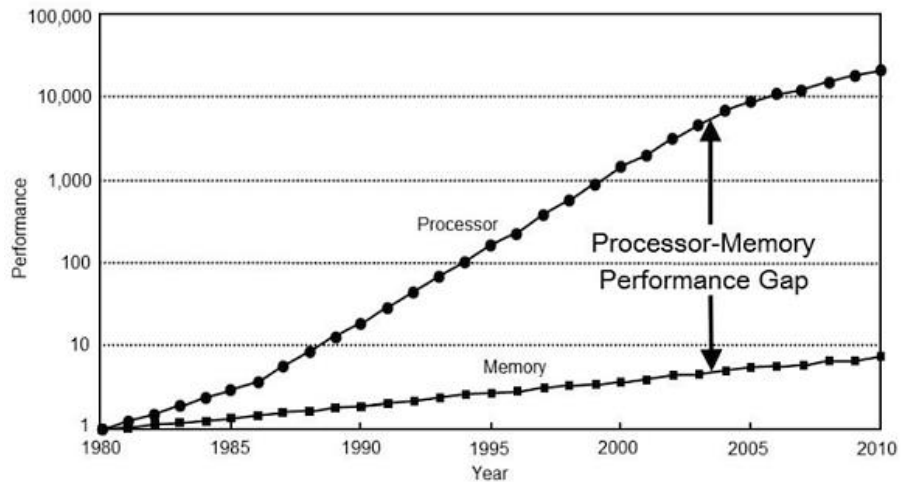
资料来源: Wind, 国盛证券研究所

2 存储：AI 带动高景气增长，产业链加速扩产

2.1 HBM 高景气，巨头加速扩产

现今大多数计算机均基于冯·诺依曼体系结构制造。该结构将程序和数据存储在内存中，CPU 从内存中获取指令和数据并依次执行。但是，将数据从主内存传送至 CPU 可能导致较长的延迟和较高的能耗。随着处理器性能的快速提升，而内存性能的提升速度相对较慢，导致内存的存取速度严重滞后于处理器的计算速度。这种严重阻碍处理器性能发挥的内存瓶颈被命名为“内存墙”。

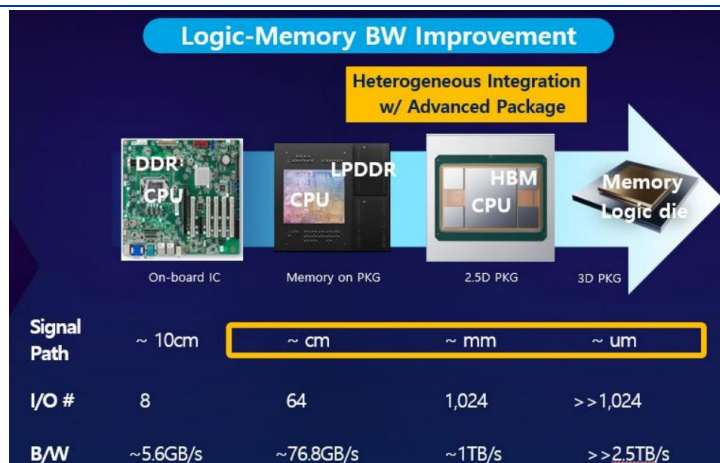
图表61：“内存墙”——处理器性能提升幅度远超内存



资料来源：ResearchGate, 国盛证券研究所

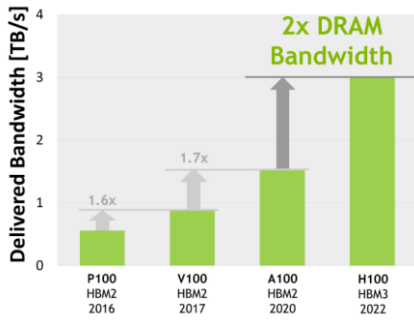
高带宽内存（HBM）克服了内存带宽和功耗的制约因素。HBM的核心优势，在于采用了3D堆叠技术，将多个DRAM芯片垂直堆叠在一起。通过硅通孔（TSV）技术实现芯片间的高速信号传输，大幅缩短了数据传输的距离和延迟，从而能以极高带宽为处理器提供数据支持。

图表62：逻辑-存储带宽升级



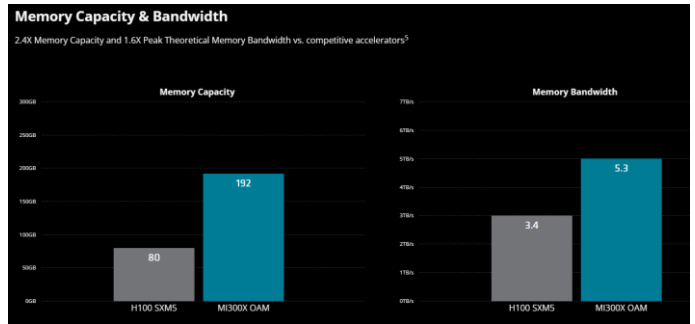
资料来源：三星, 国盛证券研究所

图表63: 英伟达 GPU HBM 不断升级



资料来源: 英伟达, 国盛证券研究所

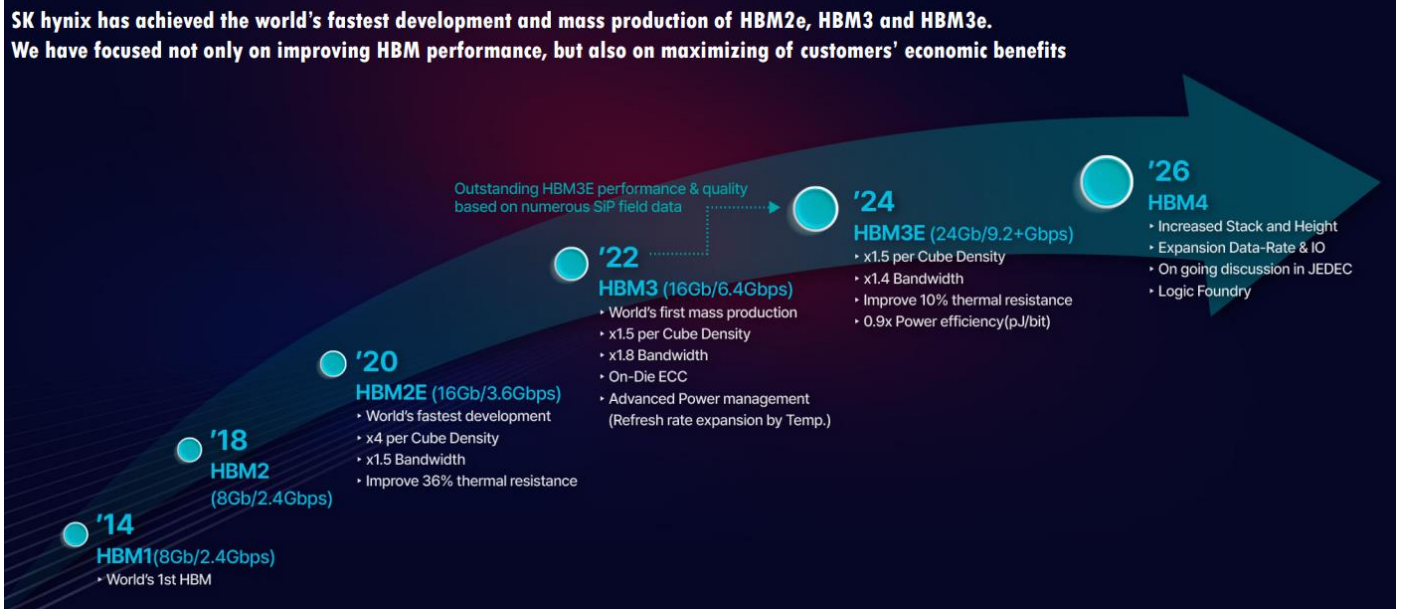
图表64: AMD MI300X HBM 容量及带宽与友商对比



资料来源: AMD, 国盛证券研究所

高性能计算驱动 HBM 加速升级迭代, HBM4 即将到来。根据全球半导体观察, HBM4 标准预计支持 2048 位接口和 6.4GT/s 的数据传输速率。相比 HBM3E, HBM4 的单个堆栈带宽已达到 1.6TB/s, 极大地提升了内存系统的数据吞吐能力, 能够更高效地满足人工智能、深度学习、大数据处理和高性能计算等领域对内存性能日益苛刻的需求。

图表65: SK 海力士 HBM 产品升级



资料来源: SK 海力士, 国盛证券研究所

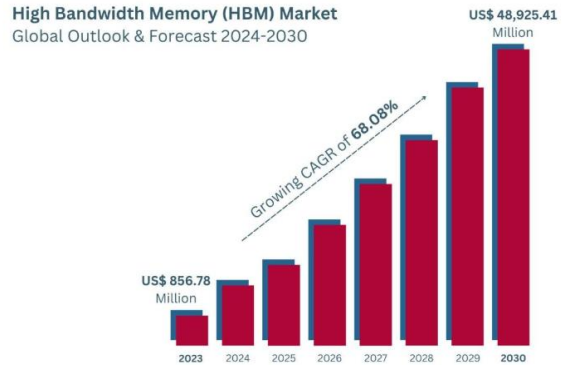
HBM 市场规模高速增长。随着 AI 加速芯片的不断更新换代, 其搭载的 HBM 总容量、带宽等规格也持续升级, 带动单机 HBM 容量及价值量快速攀升。根据 Intel Market Research, 2023 年全球 HBM 市场规模约为 8.6 亿美元, 而至 2030 年, 预计全球 HBM 市场规模将达 489.3 亿美元, 2023 至 2030 年全球年均复合增速达 68.08%。

图表66: 英伟达及 AMD AI 芯片发展进程及 HBM 规格比较

Company	AI Chips	2022	2023				2024(F)				2025(F)						
			1Q23	2Q23	3Q23	4Q23	1Q24	2Q24	3Q24	4Q24	1Q25	2Q25	3Q25	4Q25			
NVIDIA	H100	HBM3 80GB															
	GH200 (CPU+GPU)					HBM3e 80GB											
	H20					HBM3 96GB											
	H200					HBM3e 80GB											
	B100									HBM3e 80GB							
	GB200 (CPU+GPU)									HBM3e 80GB							
AMD	M200	HBM2e 80GB															
	M300A (CPU+GPU)					HBM3 80GB											
	M300K					HBM3 80GB											
	M350									HBM3 120GB							
	M375 (CPU+GPU)																

资料来源: TrendForce, 国盛证券研究所

图表67: HBM 全球市场规模



资料来源: Intel Market Research, 国盛证券研究所

需求端高景气，三大原厂加速扩产 HBM。

预计到 2024 年底三大原厂 HBM(含 TSV)产能合计达到约 270K/月。根据 TrendForce, HBM 的芯片尺寸通常比采用相同工艺节点和容量的 DDR5 大约 35%至 45%。HBM 的整体良率(包括 TSV 封装过程)比 DDR5 低 20%至 30%,而生产周期(涵盖 TSV 工艺)则比 DDR5 长约 1.5 至 2 个月。在产能方面, TrendForce 预测到 2024 年底,三星的 HBM 总产能(含 TSV)将达到约 130K, SK 海力士的产能预计约为 120K。Trendforce 预计 2025 年 HBM 芯片每月总产能为 54 万颗,相较于 2024 年增加 27.6 万颗,同比增长 105%。

三星:2024 年 HBM 产量将同比增长 2.5 倍, 2025 年再翻倍。三星电子计划在 2024 年年底将其 HBM 产能提高至每月 14 万至 15 万片晶圆,并在 2025 年年底前进一步提高至每月 17 万至 20 万片晶圆。为了扩大 HBM 产能,三星电子此前已投资 105 亿韩元收购了位于韩国天安市的 Samsung Display 工厂及相关设备。此外,公司还计划投资 7000 亿至 1 万亿韩元(约合 5040 万至 7200 万美元)用于建设新的封装生产线。

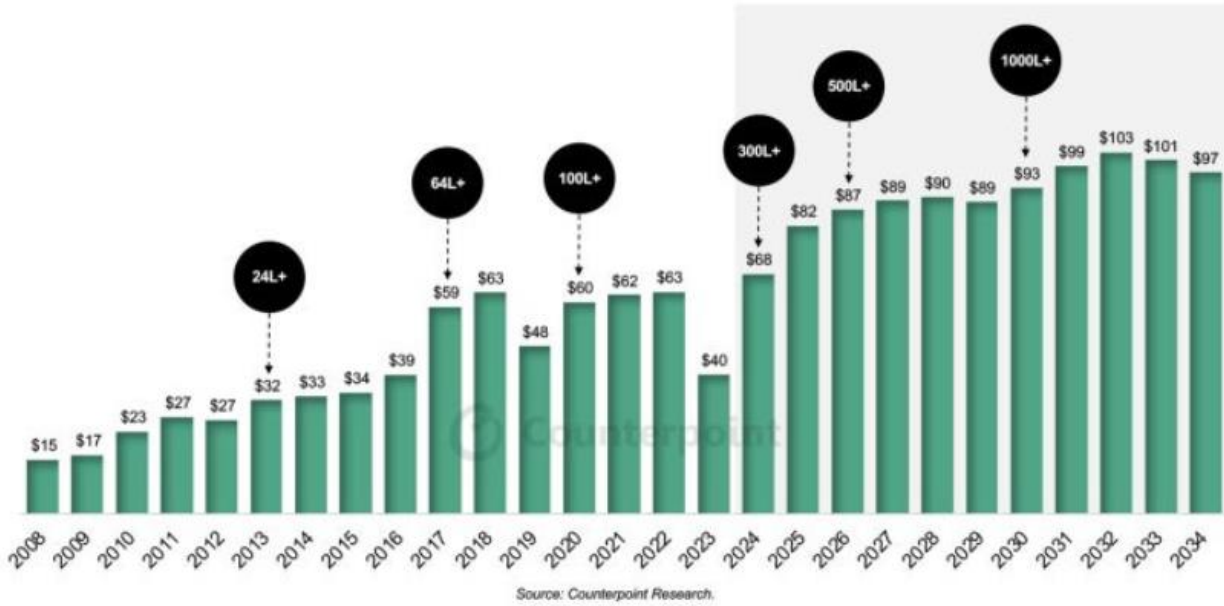
SK 海力士:2024 年 HBM 产能翻倍。公司在财报交流会中曾表示计划在 2024 年增加资本支出,重点是高端存储产品如 HBM。预计到 2024 年 HBM 的生产能力同比增长将超过一倍。SK 海力士的目标是到 2025 年实现每月 14 万片晶圆的 HBM 产能。除此之外,SK 海力士还计划在位于美国印第安纳州兴建一座顶尖的制造工厂。《金融时报》报道称,SK 海力士将在该工厂生产 HBM 堆栈,这些产品将应用于台积电代工生产的 NVIDIA GPU 上。

美光:押注最新一代 HBM3e 扩产, 2025 年 HBM 目标市场份额 20%。美光此前计划在 2024 年早些时候开始大规模出货 HBM3e。同时美光已经发布了其下一代 HBM 存储器,暂时命名为 HBM Next,预计 HBM Next 将提供 36GB 和 64GB 的容量,可用于各种配置。美光首席执行官 Mehrotra 表示:“我们的目标是 2025 年将 HBM 市场份额提高到 20% 以上。”他强调了该公司积极的扩张计划。目前,美光的 HBM 产能预计到 2024 年底将达到 20000 片晶圆左右,目标是到 2025 年底将月产量提高到 60000 片晶圆。

2.2 AI 推动 3D Nand 性能容量持续提升

大容量存储需求陡增,推动 3D Nand 层数不断攀升。云端和边缘计算在训练和推理生成式 AI 模型方面的持续进步,加上通过智能手机摄像头生成的高质量用户内容不断增加,将在未来至少二十年内推动对速度更快、容量更大的存储的需求。AI 智能手机、AIPC、企业中的 AI 服务器、自动驾驶汽车和机器人将成为 3D NAND 增长的主要终端市场驱动力。根据 CounterPoint,预计 2024 年 Nand Flash 量产层数将达到 300 层,而至 2030 年,1000+层数的 Nand 产品有望成为市场主流。Counterpoint 预计到 2030 年,整体 NAND 闪存市场将超过 930 亿美元,2023 年为 400 亿美元。

图表68: Nand Flash 市场规模及 3D Nand 层数演进趋势



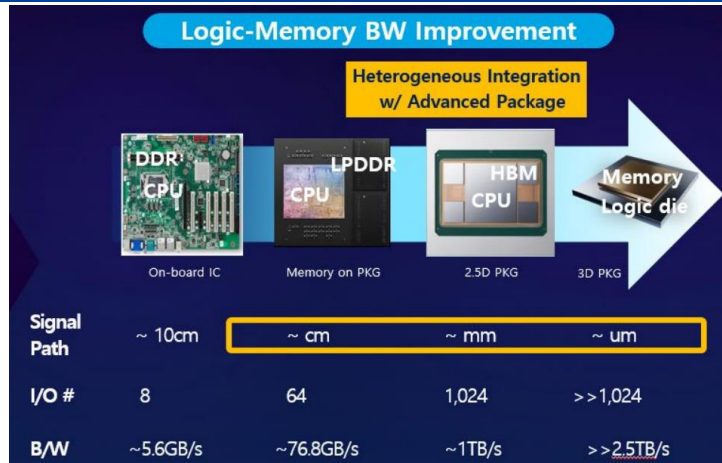
资料来源: CounterPoint, 国盛证券研究所

SSD 在 AI 发展中扮演重要角色, SSD 不仅在 AI 模型训练期间存储模型参数, 包括不断更新的权重和偏差, 而且还负责创建检查点, 以便定期保存 AI 模型的训练进度, 确保在训练意外中断时能够从特定点恢复。由于这些功能高度依赖于高速传输和写入耐用性, 客户通常会选择 4TB/8TB TLC (三层单元闪存) SSD 产品, 以满足 AI 训练过程中严苛的性能要求。

根据 TrendForce, SSD 在 AI 推理服务器中的应用发挥着重要作用, 能够在推理过程中协助调整和优化 AI 模型。特别是 SSD 可以实时更新数据, 从而对推理模型的结果进行微调。AI 推理主要提供检索增强生成 (RAG) 和大型语言模型 (LLM) 服务, SSD 可以存储 RAG 和 LLM 所需参考的相关文档和知识库, 以生成包含更丰富信息的响应。此外, 随着越来越多的生成信息以视频或照片的形式呈现, 数据存储需求也随之增加。因此, TLC/QLC 16TB 及以上的大容量 SSD 成为了 AI 推理领域的主要选择。

2024 年 AI 服务器 SSD 市场呈现出强劲的增长态势, 尤其是 16TB 以上的产品需求从第二季度开始显著上升。这一增长势头得益于 NVIDIA H100、H20 和 H200 等新系列产品的上市, 促使相关客户增加了对 TLC 4.8TB Enterprise SSD 的订单需求。根据 TrendForce, 2024 年 AI 相关 SSD 的采购容量有望超过 45EB。在未来几年里, AI 服务器预计将持续推动 SSD 需求的年增长率超过 60%。此外, AI SSD 需求在整个 NAND Flash 市场中的占比预计将从 2024 年的 5% 上升至 2025 年的 9%。

图表69: 逻辑-存储带宽升级



资料来源: 三星, 国盛证券研究所

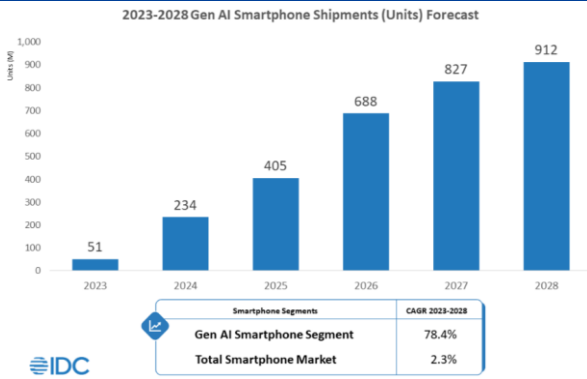
图表70: 2023 至 2026 年 AI SSD 需求位元占比



资料来源: Trendforce, 国盛证券研究所

此外在端侧, 消费者对 AI 助手和端侧处理等增强功能的需求不断增加, 推动了 AI 手机和 AI PC 市场的快速增长。根据 IDC, 2024 年开始全球 AI 手机将加速放量, 到 2027 年底, 内置 AIGC 功能的智能手机出货量预计将超过 8 亿部。根据 Canalys 预测, 2024 年全球 AI PC 出货量将达到 4800 万台, 占 PC 总出货量的 18%, 并在未来几年内加速渗透。增加了 AI 功能的移动终端将对 NAND 的存储容量和性能提出更高的要求, 将成为带动 Nand 市场需求增长的另一重要驱动力。

图表71: iPhone 出货量预估



资料来源: IDC, 国盛证券研究所

图表72: PC 出货量及 AIPC 出货占比预估

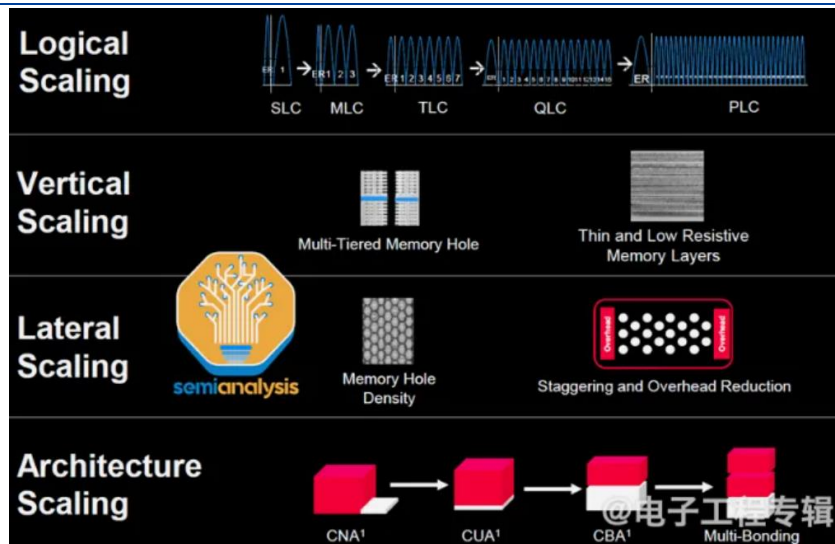
150 million+ AI-capable PCs to ship by end of 2025



资料来源: canalys, 国盛证券研究所

AI 引领的强劲需求驱动 Nand Flash 向大容量方向发展。从技术创新的角度, 3D NAND 的未来发展仍然聚焦在增加层数和提升密度两大方向, 且呈齐头并进的发展态势。在 CFMS 2024 上, 根据铠侠 CTO 柳茂知, 300 层以上的 NAND 技术研发工作可以分为以下四个方向: 一是横向缩放, 即平面方向的微细化; 二是纵向缩放, 即层数; 三是架构缩放, 即 CUA→CBA 等; 四是逻辑缩放, 即 TLC→QLC→PLC。

图表73: 300 层以上 Nand 技术开发的四个方向



资料来源: Semianalysis, 国盛证券研究所

2.3 长鑫存储: 中国 DRAM 龙头产能提升, DDR5 引领国产替代加速

国产 DRAM 先锋。长鑫存储成立于 2016 年, 已成为中国最大的 DRAM 厂商, 公司此前主要产品是 19nm 工艺的 DDR4/LPDDR4/LPDDR4X 芯片。2023 年 11 月 28 日, 长鑫存储宣布推出 LPDDR5 系列 DRAM 产品, 正式进军移动终端市场, 并在小米、传音等国产手机品牌完成上机验证。长鑫存储的 DDR5 芯片量产和 LPDDR5 系列产品的推出, 将助力中国在全球 DRAM 市场地位的提升, 突破海外高端 DRAM 垄断, 促进国内存储产业发展。

DDR5 已量产，良率有望于 2025 年底达 90%。12 月 24 日根据韩媒 ZDNetKorea 报道，长鑫存储已成功实现 DDR5 内存芯片的量产。根据报道长鑫 DDR5 良率约为 80%，与韩国企业良率接近，预计到 2025 年底，良率会进一步拉升到 90%。中国第三方存储器模组制造商已开始向市场销售采用长鑫存储生产的 DDR5 存储器芯片的 DDR5 DRAM 存储器模组。

技术能力持续进步，产能持续扩张。公司在生产能力上取得显著进步。目前长鑫存储在合肥营运两座晶圆厂，其中 Fab 1 专门生产 DDR4 芯片，采用 19 纳米制程技术，月产能为 10 万片晶圆。另外，Fab 2 使用了 17 纳米制程，用于生产 DDR5 芯片，月产能为 5 万片晶圆。产能快速扩张推动长鑫存储在全球 DRAM 市场份额从三年前的不到 2% 迅速攀升至 2024 年的约 10%，并跃居为全球第四大 DRAM 供应商。预计到 2025 年底，长鑫存储的月产能将进一步扩大至 30 万片，占全球总产能的 15%。

2.4 兆易创新：内生外延并重，DRAM 业务快速放量

2024 年前三季度公司实现营收 56.5 亿元，同比+28.56%；实现归母净利润 8.32 亿元，同比+91.87%；实现扣非归母净利润 7.77 亿元，同比+128.31%。2024 年前三季度公司实现毛利率 39.46%，同比+5.08pcts；实现净利率 14.73%，同比+4.86pcts。2024Q3 末公司存货规模达 20.38 亿元，较 2024Q2 末提升了 1.02 亿元。

分季度来看，2024Q3 公司实现营收 20.4 亿元，同比+42.83%，环比 2.97%；实现归母净利润 3.15 亿元，同比+222.55%，环比+0.93%；实现扣非归母净利润 3.04 亿元，同比+367.35%，环比+4.94%。2024Q3 公司实现毛利率 41.77%，同比+5.41pcts，环比+3.62pcts；实现净利率 15.44%，同比+8.6pcts，环比-0.31pcts。

DRAM 快速成长。2024 年公司自研 DRAM 市场拓展效果明显，总成交客户数量进一步增加。公司发布公告预计 2024 年与长鑫存储关联交易金额约为 9.95 亿元，较 2023 年同比大幅增长约 175%，同时公司公告预计 2025 年 1 月至 6 月关联交易金额为 5.89 亿元，反应公司 DRAM 业务开始加速兑现，且业务发展优于预期。同时公司公告募投项目调整增加开发 LPDDR5 产品，顺应存储芯片行业发展规律，满足当下及未来市场需求，符合公司长期发展战略。

收购苏州赛芯，拓展模拟芯片领域布局。兆易创新发布公告，公司拟与合肥石溪资本、合肥国投、合肥产投共同以现金方式收购苏州赛芯电子全体股东合计持有的苏州赛芯 70% 股份。本次交易完成后，石溪资本、合肥国投、合肥产投将与兆易创新构成一致行动人，兆易创新将成为苏州赛芯的控股股东。模拟芯片是公司的重要战略，本次交易是推动公司模拟战略的重要举措。苏州赛芯的模拟芯片与公司现有产品与业务在技术、产品、市场、销售与供应链等层面存在较强的战略协同性。同时苏州赛芯的晶圆代工与封装测试供应商与公司存在协同性，收购完成后可强化协同合作，发挥规模优势，增强客户粘性，提升盈利能力。

风险提示：下游需求不及预期，研发进度不及预期，行业竞争加剧。

2.5 普冉股份：AI 端侧放量在即，普冉有望充分受益

普冉股份深耕利基存储。普冉股份成立于 2016 年，是国内知名存储及 MCU 芯片供应商。公司产品包含存储及“存储+”两大条线，分别为 NOR Flash 和 EEPROM；微控制器芯

片以及模拟产品四大类别，公司 NOR Flash 及 MCU 产品均基于 SONOS 底层工艺，具备成本及性能优势。

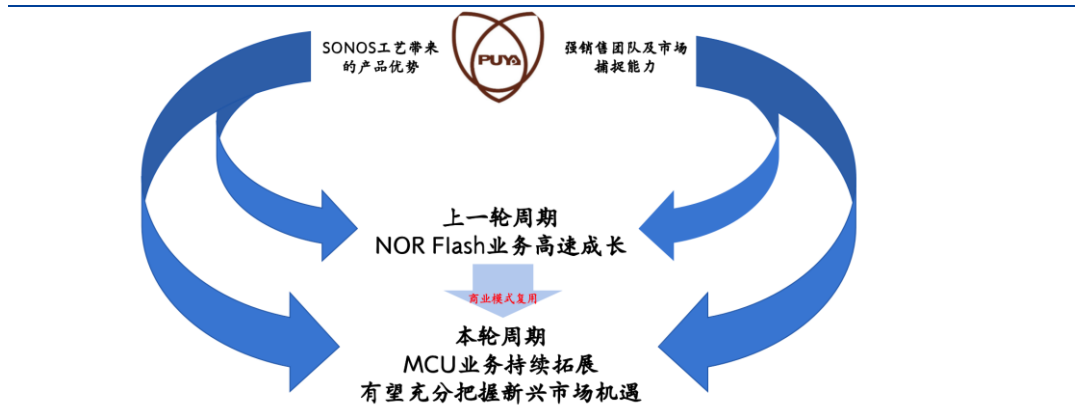
图表74: 普冉股份产品下游应用领域



资料来源: 普冉股份公告, 国盛证券研究所

商业模式复用推动 MCU 业务拓展，核心优势助力公司把握新兴市场机会。复盘公司发展历史我们认为在上一轮半导体周期中，普冉股份依托 SONOS 工艺带来的产品优势以及自己强大的销售能力及市场嗅觉实现了 NOR Flash 业务的快速成长。在本轮半导体周期中，普冉股份将 SONOS 工艺的产品优势拓展到 MCU 领域，并在 2023 年年末开始把握住了电子烟市场机遇，相关产品实现较快上量。我们认为未来公司 MCU 业务有望复用 NOR Flash 商业模式顺利拓展，公司也有望凭借对消费领域的深厚理解及市场嗅觉充分受益新兴消费市场机会。

图表75: 商业模式复用推动 MCU 业务拓展，核心优势助力公司把握新兴市场机会



资料来源: 国盛证券研究所绘制

AI 端侧放量在即，普冉股份有望充分受益。AI 端侧落地推动消费电子革新趋势明朗，未来随 AI 耳机、AI 眼镜、AI 服务机器人、AI 宠物等各类 AI 终端百花齐放，Nor Flash 市场有望持续增长。以字节跳动豆包 Ola Friend AI 耳机为例，单只耳机内部配备两颗 128Mb NOR Flash，总存储容量为 256Mb。相比之下，普通 TWS 耳机的存储容量一般较低，通常为 64Mb 到 128Mb。普冉股份下游覆盖各大知名蓝牙、音频、WIFI 等主控 SoC 芯片厂商，未来有望充分受益 AI 端侧发展。

风险提示: 下游需求不及预期，研发进度不及预期，行业竞争加剧。

3 互联：光与铜并进，互连方案向高性能、低成本持续演进

3.1 光&铜连接实现数据中心信息互联

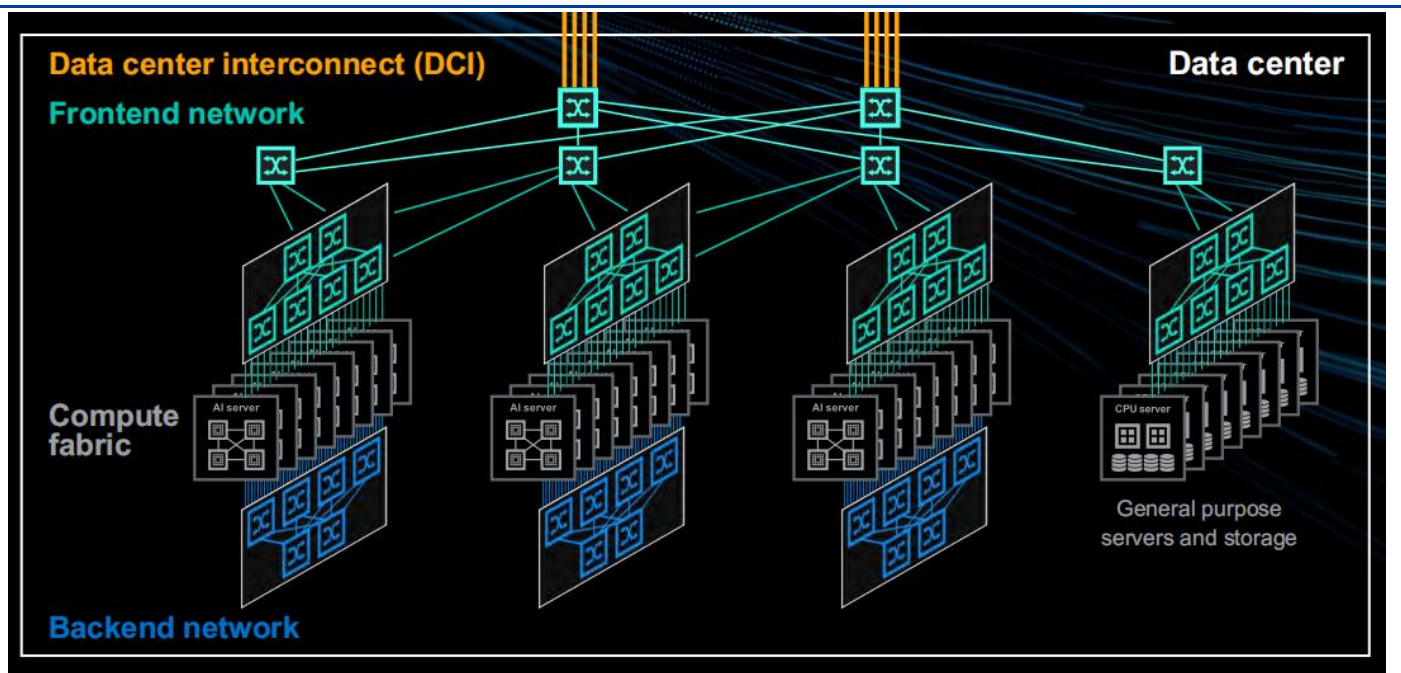
光与铜助力数据中心信息互联。随着云计算、大数据、超高清视频、人工智能、5G 行业应用等快速发展，网络访问频率和接入手段不断增加，网络数据流量迅猛增长，对数据中心互连提出更高挑战。以 Spine-leaf CLOS 架构数据中心为例，前三类为数据中心内部互连，距离从 2m-2km，相对较短，第四类为数据中心之间互连 (DCI)，距离可达 120km。根据 Marvell，随着集群的扩大，新的 DCI 机会甚至长达 1000km。数据中心内部互连以基于直调直检方案的光模块 (典型场景为 TOR 到 Leaf 以及 Leaf 到 Spine) 和 AOC/铜缆 (典型场景为服务器到 TOR) 为主，在数据中心之间的互联场景中，主要采用相干光模块进行连接。

图表76: 数据中心互连种类

互连场景		典型范围		典型需求			
				当前	下一代	再下一代	
场景一	数据中心内部	服务器到TOR	2m (机架内) /30-50m (跨机架)	机房内	25G AOC/DAC	100G AOC/DAC	200G AOC/DAC?
场景二		TOR到Leaf	>70/100m	楼栋内	100G SR4	400G SR8/SR4.2	800G PSM8/PSM4?
场景三		Leaf到Spine	500m/2km	楼栋间	100G CWDM4	400GFR4/DR4	800G FR4/PSM4?
场景四	数据中心之间		80-120km	园区间	100G DWDM	400G ZR/ZR-	800G ZR

资料来源:《5G 承载与数据中心光模块白皮书》，国盛证券研究所

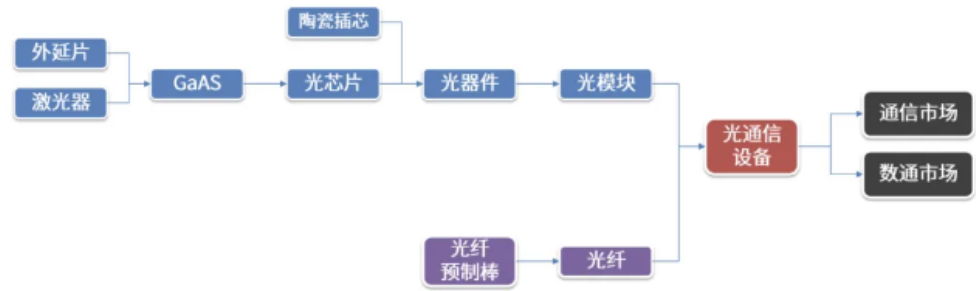
图表77: 数据中心互连图例



资料来源: Marvell, 国盛证券研究所

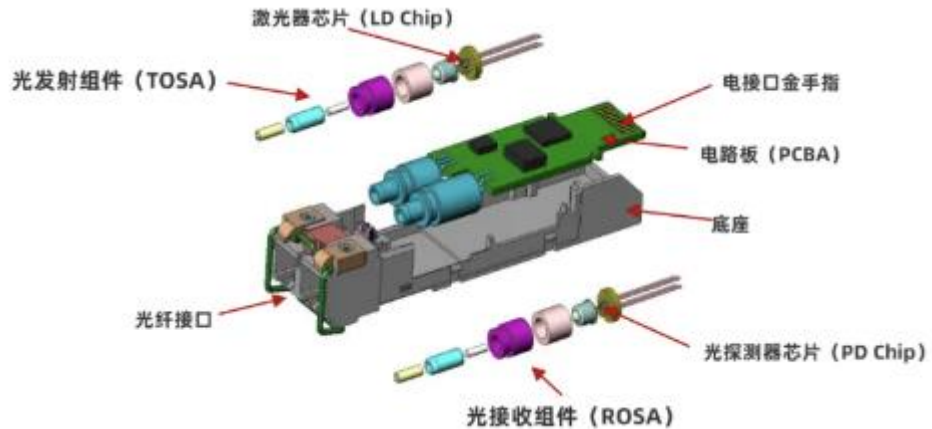
光模块是光通信中实现电光和光电转换的核心器件，光器件成本占据大头。光模块 (Optical Modules) 是光通信系统的核心器件之一，由各种无源器件以及光电芯片组合封装。光模块构成了数据中心互连、5G 承载网络和全光接入网络的基础单元，主要作用就是实现光纤通信中的光电转换和电光转换功能。光模块行业上游主要包括光芯片、电芯片、光组件企业。光芯片是光模块中完成光电信号转换的直接芯片，占据成本大头 (占光模块成本的 30%-60%)，且高端光芯被海外所垄断，国产产量小，暂不能满足市场需求；而电芯是光模块性能表现的关键器件，它的成本 (占光模块成本的 18%) 占比仅次于光芯，且技术壁垒高，国内自给率仅达 5%。

图表78: 光纤通信产业链



资料来源: 深圳市光学光电子行业协会, 国盛证券研究所

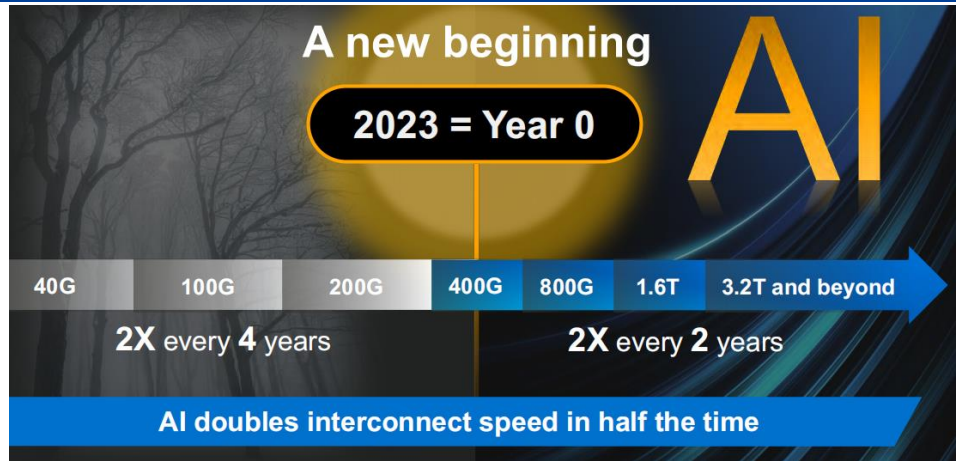
图表79: 光模块构成



资料来源: 深圳市光学光电子行业协会, 国盛证券研究所

AI 加速互联速率提升, 每两年实现翻倍。数据中心内光互联模块与交换机交换芯片串行-解串行器 (Serdes) 发展紧密相连。2023 年交换芯片 Serdes 速率达 112Gbit/s, 吞吐量为 51.2Tbit/s, 依其演进趋势、市场需求与技术成熟度, 预计 2025 年吞吐量将达 102.4Tbit/s, 2027 年达 204.8Tbit/s, 光互联模块需相应演进至 1.6Tbit/s 和 3.2Tbit/s 予以支撑。AI 算力网络与常规数据中心相比, 对计算的需求量每 18 个月增长 10 倍, 对高带宽、低时延的光互联需求更加迫切。根据 Marvell, 此前互联速率每四年实现翻倍, 在 AI 趋势下, 互联速率每两年实现翻倍。目前, 业界正将 800G 链路的单通道速率从 100Gb/s 提高到 200Gb/s 以降低功耗和成本, EML 和 CW-DFB 器件已为 200G/lane 应用做好准备, 而 200G/lambda VCSEL 预计将于 2026 年投入量产。

图表80: AI趋势下互联速率每两年实现翻倍



资料来源: Marvell, 国盛证券研究所

图表81: 交换机及光模块发展趋势预测

类型		2010	2012	2014	2016	2018	2020	2023	2025 (预)		2027 (预)	
交换机	带宽/ (Tbit/s)	0.64	1.28	3.2	6.4	12.8	25.6	51.2	102.4		204.8	
	Serdes通道数	64	128	128	256	256	512	512	1024/512		1024	
	Serdes速率/ (Gbit/s)	10	10	25	25	50	50	112	112/224		224	
光互联模块	速率/ (Gbit/s)	200	40	100	100	400	400	400	800	800	1600	3200
	通道数	2	4	4	4	8	8	4	8	4	8	16
	单通道速率/ (Gbit/s)	100	10	25	25	50	50	100	100	200	200	200

资料来源: 《数据中心光互联模块发展趋势及新技术研究》宋梦洋、朱虎、江毅, 国盛证券研究所

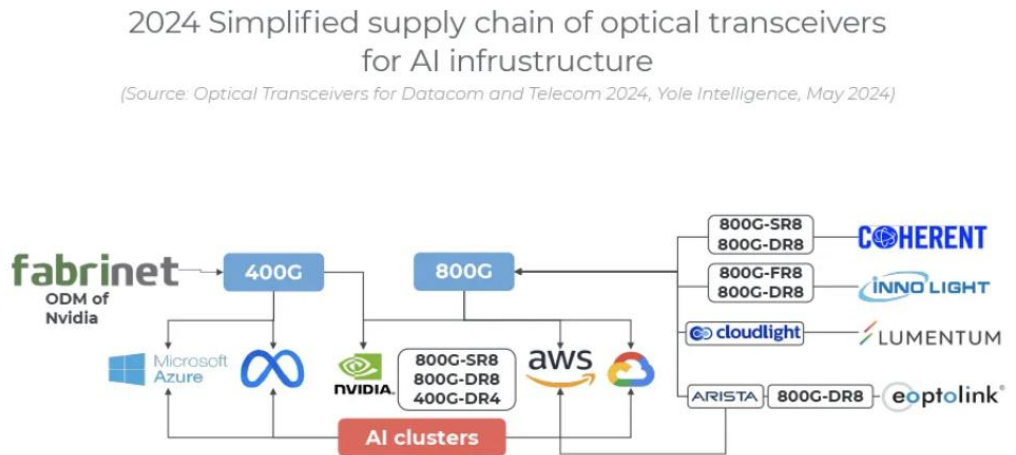
2023年中际旭创光模块排名升至全球第一。Coherent和旭创科技分别在多模和单模应用领域处于领先地位。2018年,大部分日本和美国供应商退出了这一市场,而以中际旭创为首的中国供应商的排名有所提升,2023年中际旭创首次排名第一,Coherent(Finisar)排名第二,主要系衰退的电信市场拖累。

图表82: 2010-2023年光模块竞争格局

2010	2016	排名	2018	2023
Finisar	Finisar	1	Finisar	中际旭创
Opnext	海信	2	中际旭创	Coherent
Sumitomo	光迅科技	3	海信	华为 (HiSilicon)
Avago	Acacia	4	光迅科技	Cisco (Acacia)
Source Photonics	FOIT (Avago)	5	FOIT (Avago)	光迅科技
Fujitsu	Oclaro TC	6	Lumentum/Oclaro	海信
JDSU	中际旭创	7	Acacia	新易盛
Emcore	Sumitomo	8	Intel	华工科技
WTD	Lumentum	9	AOI	Source Photonics
NeoPhotonics	Source Photonics	10	Sumitomo	Marvell

资料来源: lightcounting, 国盛证券研究所

图表83: 2024 年用于 AI 基础设施的光收发器供应链



资料来源: yole, 国盛证券研究所

中际旭创: 25 年硅光预计将导入主要客户, 1.6T 逐步增长。重点客户进一步增加资本开支加大算力基础设施领域投资, 400G 和 800G 光模块需求显著增长, 并加速了光模块向 800G 及以上速率的技术迭代。24 年公司 400G 和 800G 等高端产品出货比重的快速增加, 预计 25 年 800G 需求强劲, 新增 1.6T 预期也会带来逐步增长。在高端产品进展方面: **1) 1.6T 光模块:**公司已全面完成前期的送测和认证工作, 预计 24 年 12 月开始出货; **2) 硅光:**公司通过投入大量研发资源, 在硅光芯片设计研发和技术储备方面已取得显著进展, 并已成功应用到 400G、800G 甚至 1.6T 等更高级别的光模块产品上, 目前主要客户对公司的硅光产品均反馈不错, 预计主要的客户会在 25 年导入公司的硅光模块; **3) CPO:**公司重视 CPO 研发, 公司已具备一定的技术基础去研发 PO 相关产品。

新易盛: 高速率产品出货占比提升, LPO 产品进展顺利, 前瞻布局 AEC。在高端产品以及新品进展方面, 公司目前已成功推出基于 VCSEL/EML、硅光及薄膜铌酸锂方案的 400G、800G、1.6T 系列高速光模块产品, 和 400G 和 800G ZR/ZR+ 相干光模块产品、以及基于 100G/lane 和 200G/lane 的 400G/800G LPO 光模块产品。800G 产品 24 年已经实现批量出货, 随着公司高速率产品出货占比提升, 公司毛利率由 2023 年的 31% 提升至 24 年前三季度的 42%, 利润端实现大幅优化。LPO 方面, 公司系 LPO 系列光模块的引领者, 是业界首家宣布推出创新的 800G LPO 收发器的厂商, 正在积极推进相关项目的进展; AEC 方面, 根据公司 23 年年报, 公司正研发 800G 高速带电缆模块 AEC, 完成项目前期样品测试, 功耗低于 11W, 满足 IEEE 802.3ck 协议要求。

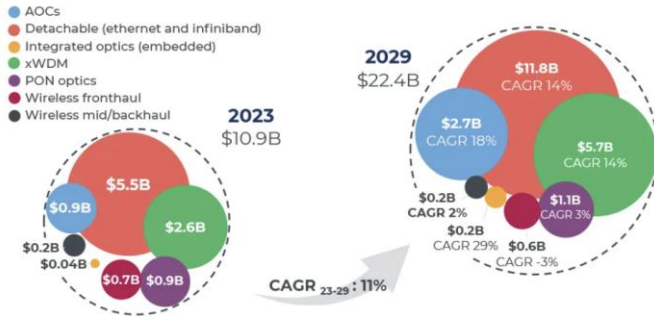
天孚通信: 积极扩产高速率&硅光, 持续投入 CPO 配套研发。公司立足光通信领域, 长期致力于各类中高速光器件产品的研发、生产、销售和服务, 为下游客户提供垂直整合一站式解决方案, 包括高速率同轴器件封装解决方案, 高速率 BOX 器件封装解决方案, AWG 系列光器件无源解决方案、微光学解决方案等。24 年高速率产品需求较为旺盛, 尤其是为 400G、800G 速率配套的光无源和有源器件, 公司 24 年上半年主要为高速率、硅光配套的相关无源器件扩产较多。公司持续保持研发投入, 主要的研发项目有: 适用于各种特殊应用场景的 Fiber Array 器件产品开发、适用于硅光模块特殊光纤器件及单通道高功率激光器产品开发、适用于 CPO-ELS 模块应用的多通道高功率激光器的开发、单波 200G 光发射器件的开发、适用于高速光模块的多通道接收波长解复用器件 (POSA) 的开发等, 公司针对硅光技术平台和 CPO 相关配套产品有多个研发项目在正常推进。

AI 驱动光模块市场规模增长, 光连接增速或将快于 XPU。光模块市场的整体收入从 2022 年的 110 亿美元略降至 2023 年的 109 亿美元, 但 Yole 预计到 2029 年将达到 224 亿美元。2024 年在数通细分领域, AI 驱动的光模块市场将出现同比 45% 的增长, 对 400G 和 800G 数通光模块的需求, 尤其是来自英伟达、谷歌和亚马逊的需求, 推动了市场规模增

长。根据 Marvell, 由于端口数和端口速率的增长, 网络将在乘数效应下实现超高速增长, TAM 大规模扩张, 光互连可实现大型 AI 集群, 发展速度超过 XPU。

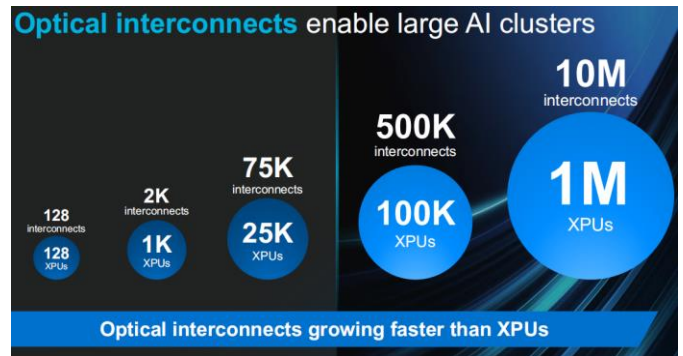
图表84: 2023-2029 年光模块市场规模

2023-2029 optical transceiver revenue growth forecast by segment
(Source: Optical Transceivers for Datacom and Telecom 2024, Yole Intelligence, May 2024)



资料来源: yole, 国盛证券研究所

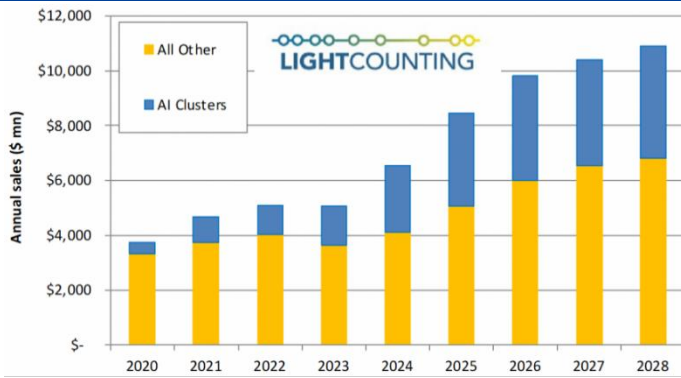
图表85: 光互连增速将高于 XPU



资料来源: Marvell, 国盛证券研究所

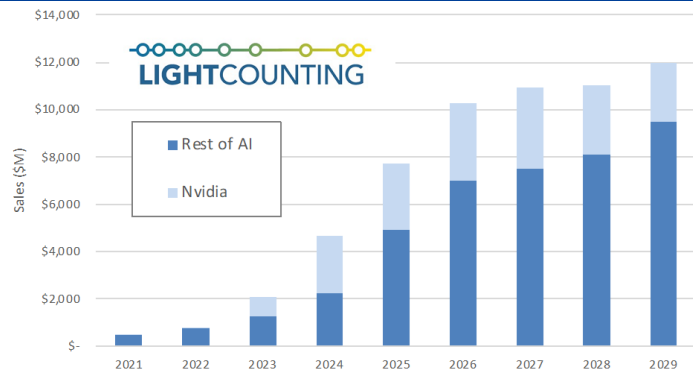
亚马逊、谷歌、Meta、英伟达光模块需求量庞大, 400G/800G 光模块供不应求。按客户来看, 亚马逊、谷歌、Meta、英伟达等对光模块需求很大, 因此每家公司几乎都是一个独立细分市场, 其中亚马逊 24 年资本开支预计约为 750 亿美元, Meta 预计约为 380-400 亿美元, 相关供应商将受益于高涨的资本开支。按速率来看, 4x100G 和 8x100G 光模块的需求量超过供应量的 100%, 许多客户需要等到 2025 年才能交货, LightCounting 对 2025 年 4x100G 光模块的预测提高了 5 亿美元, 2026 年提高了 10 亿美元, 预计在 2026 年达到 40 亿美元以上的峰值; 对 2025 年 8x100G 光模块的预测增加 20 亿美元, 预计 2026 年销售额将超过 70 亿美元。此外, 首批 4x200G 和 8x200G 光模块将于 2024 年底发货。

图表86: 以太网光模块按应用分类



资料来源: lightcounting, 国盛证券研究所

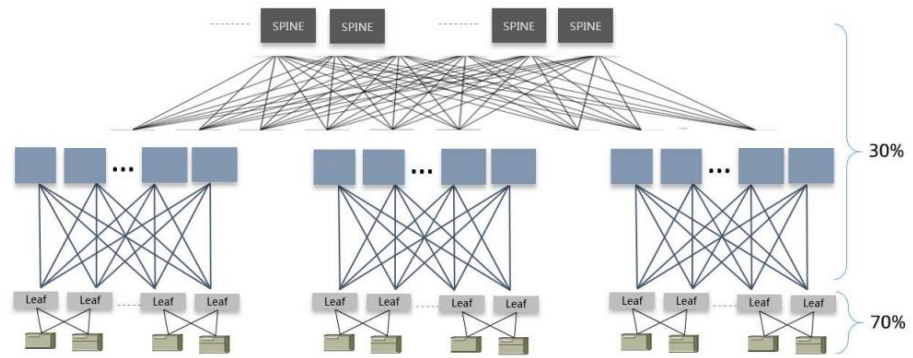
图表87: 英伟达采购光模块占比



资料来源: lightcounting, 国盛证券研究所

短距离场景下可通过高速铜缆 (DAC/ACC/AEC) 和有源光缆 (AOC) 进行互连。在一个典型的系统中芯片对芯片、芯片对光模块、光模块对光模块之间的连接, 通过连接器、PCB 传输线、铜电缆或者光纤互连。这些互连可以是光的或电的, 可以支持一定范围的数据速率, 设计中的考虑因素包括: 链路性能、成本、工作温度、可加工可装配性、长期可靠性等。互联网 70% 的流量发生在数据中心内部, 在当前比较流行的 CLOS 的数据中心网络架构下, Leaf 至 Spine 层之间采用短距至长距光模块的链路数量占比约为接入层交换机至服务器之间链路总数的三分之一左右, 而对于占据主要互联用量的 Server 和 TOR 交换机, 则可以通过高速铜缆 (DAC/ACC/AEC) 和有源光缆 (AOC) 进行互连, 覆盖 20 米以下的距离。

图表88: 典型数据中心 CLOS 网络架构示意图

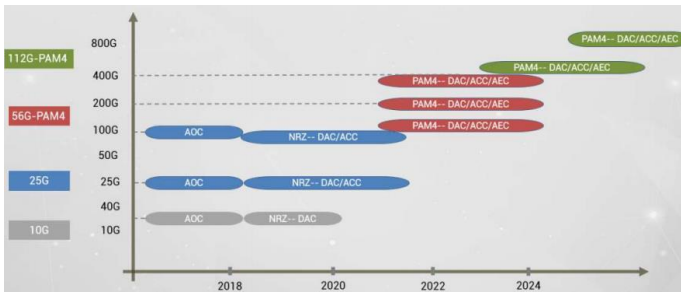


资料来源: ODCC, 国盛证券研究所

以 400G 连接解决方案为例, 总结来看相比于有源光缆 AOC, 铜缆具备失效率低、功耗低以及成本低的优势, 是短距离场景性价比最高的选择:

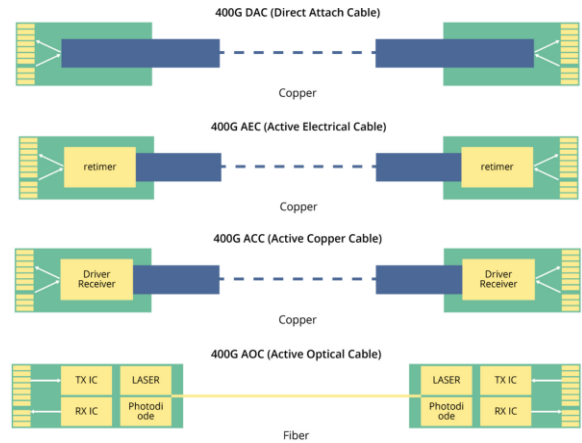
- **400G 直连电缆 (DAC)** 主要是指无源直连电缆。它使用导电铜线在两端之间直接连接。DAC 的内部结构通常采用屏蔽双芯设计, 不包括有源元件, 功耗小于 0.1W, 成本是四种方案中最低的。DAC 铜缆主要用于系统内机架连接, 将计算服务器连接到存储子系统, 最大长度为 3 米;
- **400G 有源铜缆 (ACC)** 是指有源铜缆, 其连接介质与 DAC 相同。但是, 电缆内部添加了线性放大器, 它不会重新调节信号, 只是放大信号, 同时也放大了噪音, 与 400G DAC 电缆相比, 400G ACC 电缆的功耗和成本更高, 但线性放大器相较于重定时器 (AEC 使用的有源器件) 成本更低。有源芯片可以补偿铜缆传输造成的部分损耗, ACC 的传输距离是 DAC 的 2 到 3 倍;
- **400G 有源电缆 (AEC)** 是一种特殊类型的有源 DAC。400G AEC 作为 400G AOC 的可靠即插即用替代方案, 然而, 它的功耗只有 400G AOC 的一半, 而且成本更低, 介于 400G DAC 和 400G AOC 之间。它在电缆组件内包含多个重定时器, 解决数据丢失的问题——一个放在电缆的开头, 一个放在末端, AEC 的重定时器可产生更清晰的信号, 消除噪音并放大信号, 以实现更清晰、更清晰的数据传输, 从而确保更远距离的高速传输, 传输距离为 3-7 米;
- **400G 有源光缆 (AOC)** 利用光纤作为传输介质, 将电信号转换为光信号。由于 AOC 的光纤芯由绝缘材料制成, 因此在数据传输过程中不易受到电磁、闪电或无线电信号的干扰, 功耗约为 10W, 由于与铜缆相比, 光纤的成本普遍较高, 因此 400G AOC 的价格通常高于 400G DAC (假设水平或封装相似)。400G AOC 的最大传输距离可达 100 米。

图表89: 数据中心服务器接入层链路技术发展趋势



资料来源: ODCC, 国盛证券研究所

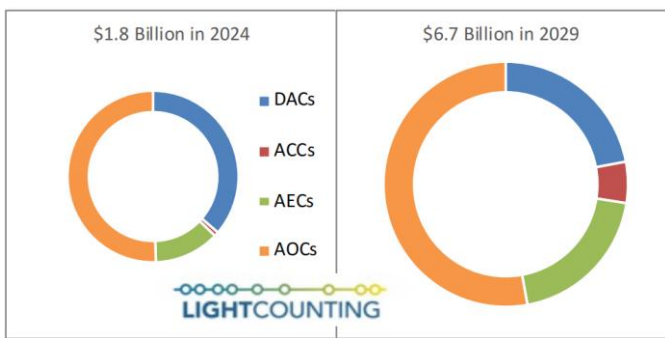
图表90: 400G DAC, AOC, ACC, AEC 结构对比



资料来源: FS, 国盛证券研究所

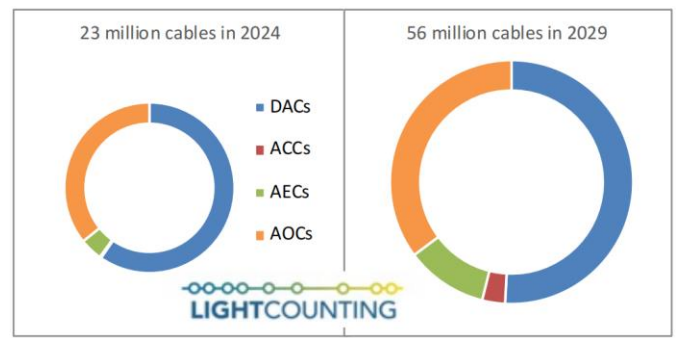
未来五年高速线缆销售额有望增长超两倍，AEC/ACC 份额提升。lightcounting 预计未来五年高速线缆销售额将增长超两倍，至 2029 年达 67 亿美元。AOC/DAC/ACC/AEC 的 2024-2029 年销售额 CAGR 分别为 32%/18%/90%/43%。AEC 和 ACC 传输距离长、厚度薄，ACC 延迟低对人工智能集群重要。由于 DAC 不耗电，是数据中心提高能效默认连接方案，2029 年 DAC 仍占高速电缆总出货量的 50%。以 Nvidia NVL72 服务器为例，DAC 也在与 AI 服务器中的背板连接竞争，单个 NVL72 的背板内包含 5000 根铜缆，支持 72 个 GPU 之间的 NVLink 连接，背板还使用双轴铜缆。

图表91: 2024 年和 2029 年 DAC, ACC, AEC 和 AOC 的销售额情况



资料来源: lightcounting, 国盛证券研究所

图表92: 2024 年和 2029 年 DAC, ACC, AEC 和 AOC 的销售量情况



资料来源: lightcounting, 国盛证券研究所

3.2 你追我赶，互联技术持续升级

3.2.1 AEC——延长铜缆方案传输距离

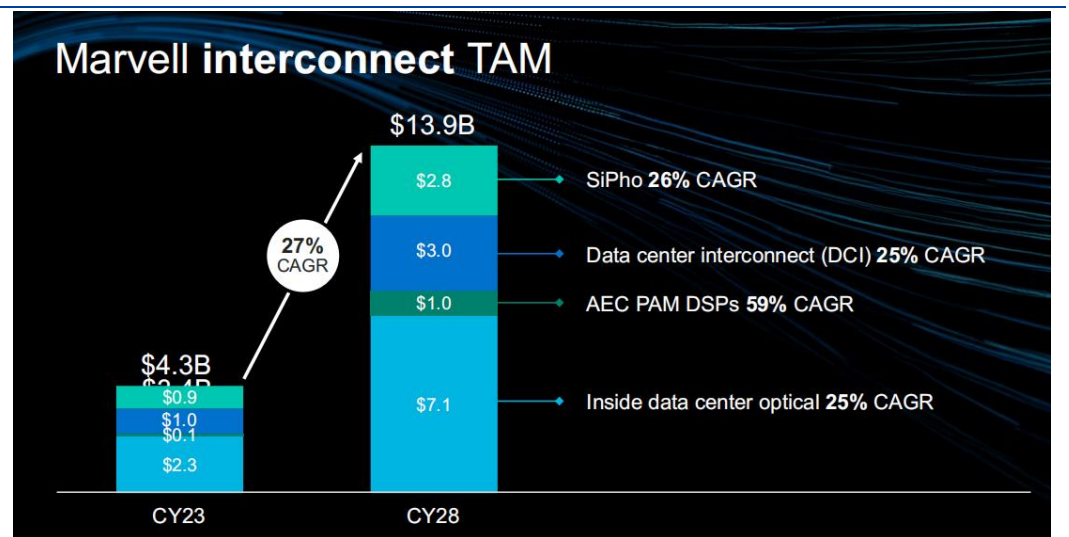
AEC 是一种高性能铜缆解决方案，两端均集成了 Retimer，随着速率的提升，DAC 有效距离被大幅缩短，AEC 需求大幅提升。AEC 优势明显：1) 功耗方面：400G AEC 单端功耗约 5 瓦，相比 400G 光模块，功耗降幅约 50%；2) 物理特性与部署便利性：较笨重的无源 DAC 更为轻巧、纤细，可弯曲和卷盘，在机架部署时，其轻便属性利于布线，且不影响机架内空气流通与散热，提升了机架管理和使用效率；3) 寿命与可靠性：不含光学组件，寿命长且可靠性极高，可达无源 DAC 的 100 倍。有源设计使其无易碎部件，坚固耐用，在安装及日常使用中的弯曲、盘卷、踩踏等情况下仍能维持性能稳定。**根据 Marvell 预测，AEC PAM DSPs 市场空间在 2028 年或将达到 10 亿美元，2023-2028CAGR 达 59%，增速领先于其他互连市场。**

图表93: 速率提升推动 AEC 需求提升



资料来源: Marvell, 国盛证券研究所

图表94: Marvell 互连 TAM 不断增长



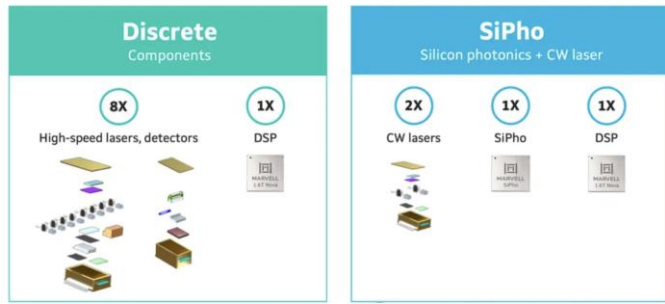
资料来源: Marvell, 国盛证券研究所

3.2.2 硅光技术——利用 CMOS 工艺的高集成度、高性价比解决方案

硅光模块在高速率领域具有高集成度、低成本、低功耗的显著优势。硅是用量最大的半导体晶圆材料，具有低成本和加工工艺成熟的优势。硅光基于硅和硅基衬底材料，通过 CMOS 工艺进行光器件开发和集成。硅光模块产生的核心理念是以光代电，即利用激光束代替电子信号进行数据传输。硅光模块将激光器、调制器、探测器等光/电芯片都集成在硅光芯片上，传统器件中的透镜和大型组件都被取代，陶瓷、铜等材料用量大幅降低，晶圆、硅光芯片等电子材料占比上升，光模块价值向硅光芯片、硅光引擎转移。此外，由于硅材料本身不具备发光条件，无法直接传输和探测光信号，需要有其他发光材料提供光源，与传统的光学模块不同，硅光子学使用共波长 (CW) 激光器，这种激光器成本更低，更容易制造。

图表95: 1.6T 分立与硅光方案对比

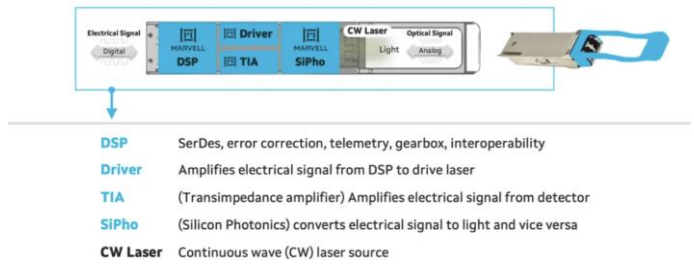
Discrete vs SiPho (1.6T pluggable)



资料来源: Marvell, 国盛证券研究所

图表96: 硅光模块组成

What's inside silicon photonics module?



资料来源: Marvell, 国盛证券研究所

硅光有望缓解光芯片紧张带来的供应问题，硅光模块占比有望快速提升。光芯片短缺下硅光模块有望缓解目前的供应问题，提高供应链安全性。目前台积电携手博通、英伟达等共同开发硅光子技术、共同封装光学元件等新产品，制程技术从 45 nm 延伸到 7 nm，台积电已组建一支由约 200 名专家组成的研发团队，专注于利用硅光子技术开发未来芯片。英特尔在硅光的布局领先，截至 24 年 8 月，英特尔已出货超过 800 万个硅光子集成电路，包含多达 3200 万个片上集成激光器；国内中际旭创、华工科技、光迅科技等也在硅光上有较好布局。我们认为硅光模块占比有望快速提升。

图表97: 主流公司硅光技术进展

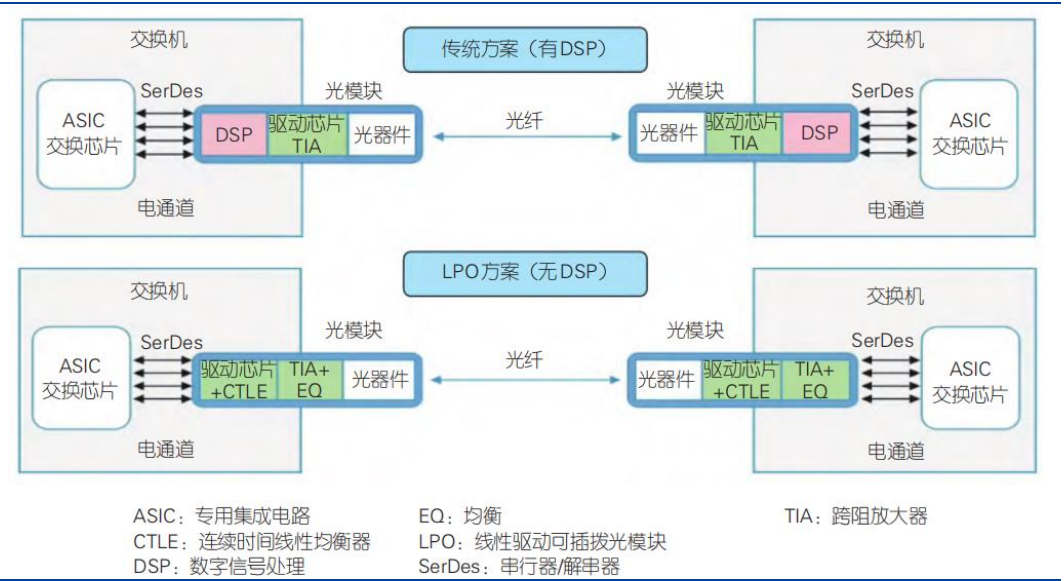
公司	硅光技术进展
Marvell	2022年推出业界首款800 Gbit/s 多模平台解决方案,同年量产用于数据中心的400G DR4 硅光子平台
中际旭创	400G 硅光模块已进入市场导入阶段,正在接受海外客户认证;800G 硅光模块已开发成功,并向海外客户送样;400G/800G 的硅光模块已采用自研的硅光芯片
熹联光芯	100G 硅光模块已实现规模化量产,400G 光学引擎及硅光模块正处于多个客户认证测试中;在张家港经开区建设中国第一条硅光芯片及封测生产线,项目总投资达20亿元
华工科技	2021年400G 硅光芯片已实现量产,2022年发布800G 硅光芯片并实现小批量生产
光迅科技	2018年发布100G 硅光收发芯片并正式投产使用,200G/400G 硅光数通模块已经开始出货,800G 产品已开始向客户送样
华为	2012和2013年分别收购英国光子集成公司CIP和比利时硅光子公司Caliopa,2019年后累计投资10余家光芯片产业链相关企业,比如熹联光芯、微源光子、长光华芯、芯视界等

资料来源:《数据中心光模块技术及演进》张平化、王会涛、付志明, 国盛证券研究所

3.2.3 LPO——兼具可插拔性与低功耗优势的快速落地方案

LPO 通过线性直驱技术替换传统的 DSP，将功能集成到交换芯片中，只留下驱动芯片（Driver）和跨阻放大器（TIA）芯片。LPO 光模块中用到的 Driver 和 TIA 芯片性能也有所提升，从而实现更好的线性度。在 400G 光模块中用到的 7 nm DSP，其功耗约为 4W，占到了整个模块功耗的 50%左右，从成本的角度来看，400G 光模块中，DSP 的 BOM 成本约占 20%~40%。LPO 具有低功耗、低成本、低延时、易维护等突出的优势。

图表98: LPO 光模块示意图



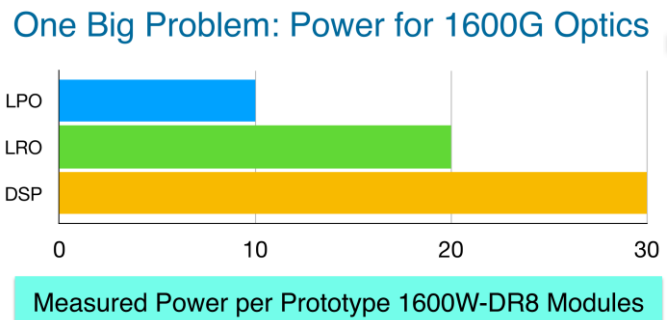
资料来源:《数据中心光模块技术及演进》,张平化、王会涛、付志明,国盛证券研究所

图表99: LPO 实现能耗节省

Configuration	DSP Optics	LPO Optics	Delta	LPO/DSP %
1M XPU	3 GW	3 GW	0	
32M Optics	960 MW	320 MW	640 MW	-66.7%
TOTAL Power	3960 MW	3320 MW	640 MW	-16%
Optics Power/XPU %	32%	11%		-66%
Power/Y @ 0.25/kWh	\$8.67B	\$7.27B	\$1.4B	-16%
Power over 5 Years	\$43.35B	\$36.35B	\$7B	-16%

资料来源: Arista, 国盛证券研究所

图表100: 1.6T 光模块能耗对比



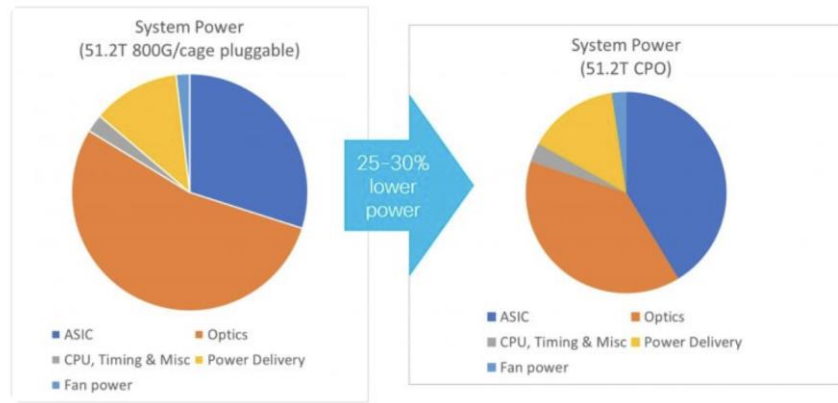
资料来源: Arista, 国盛证券研究所

LPO 有望于 25-26 年实现放量。 LightCounting 预计, 英伟达可能在 2025 年进行少量每通道 200G 的 LRO 和 LPO 部署; 预计字节有可能在 2025 年成为第一大光模块消费者, 是认真考虑 8x100G LPO 的潜在客户之一, 字节跳动的 Yinxing Zhang 对 2025 年-2026 年部署 8x100G LPO 持乐观态度, 但目前并未打算转向每通道 200G。从厂商布局来看, Arista 在 OFC 2023 上首次展示了有关降低 LPO 功耗的数据; 新易盛、剑桥科技等已发布相关产品; 中际旭创已进行相关产品开发; 海信宽带已推出 800G 线性互联光缆。

3.2.4 CPO&OIO——高速通信的下一代解决方案

CPO 是在成本、功耗、集成度各个维度上优化数据中心的光电封装方案。 CPO 将光模块不断向交换芯片 (ASIC 芯片) 靠近, 缩短芯片和模块之间的走线距离, 最终将光引擎和电交换芯片封装成一个芯片。在理想情况下, CPO 可以逐步取代传统的可插拔光模块, 将硅光子模块和超大规模 CMOS 芯片以更紧密的形式封装在一起, 从而使系统成本、功耗和尺寸都得到进一步优化。

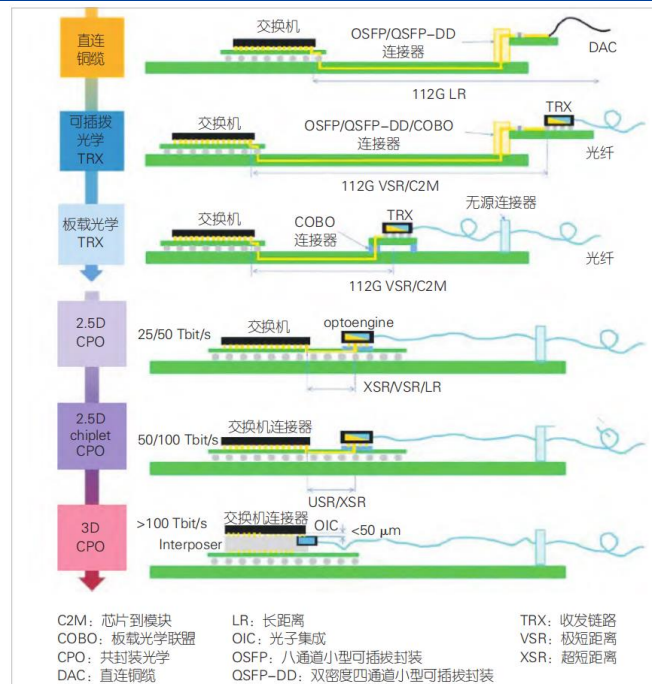
图表101: CPO 实现能耗节省



资料来源: cisco, 国盛证券研究所

按照物理结构分类，CPO 可分为 3 种：2D 平面 CPO、2.5D CPO 和 3D CPO。1) 2D 平面 CPO: 基于 2D 封装的 CPO 技术是光子集成电路 PIC 和集成电路并排放置在基板或 PCB 上，通过引线或基板布线实现互连；2) 2.5D CPO: 2.5D 封装将 EIC 和 PIC 均倒装在中介层 (Interposer) 上，通过中介层上的金属互连 PIC 和 EIC，中介层与下方的封装基板或 PCB 板相连；3) 3D CPO: 3D 封装技术将光电芯片进行垂直互连，可以不仅能实现更短的互连距离、更高的互连密度和更好的高频性能，还能实现更低的功耗、更高的集成度和更紧凑的封装。

图表102: CPO 技术路线

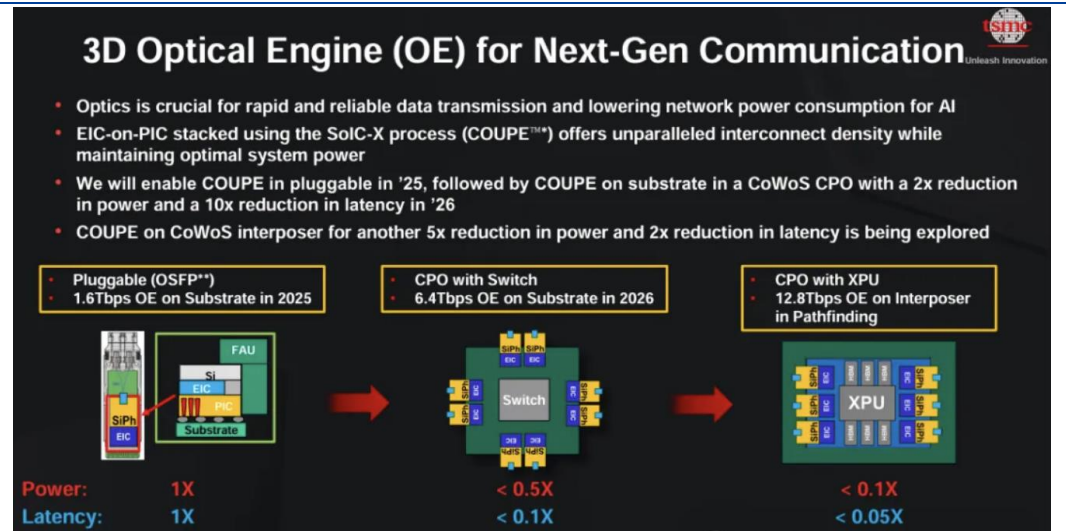


资料来源: 《数据中心光模块技术及演进》，张平化、王会涛、付志明，国盛证券研究所

台积电交换机 CPO 平台 26 年有望成熟，未来有望进入计算芯片侧。根据台积电，1) 第一代 COUPE: 2025 年完成基于 COUPE 技术的可插拔 1.6T 光模块的技术验证；2) 第二代 COUPE: 2026 年推出基于 CoWoS 封装的 CPO 模块，将产品和 Switch 封装在一起，功耗降低 50% 以上，延迟降低 90% 以上，这是主板级别的光互连，数据不需要通过外部交换机传输；3) 第三代 COUPE: 将光连接放在 XPU (各种处理器) 边，数据直接在芯片内部通过光学进行传输。根据博通，使用 Bailly CPO 的光互连 (交换机侧 CPO) 功耗

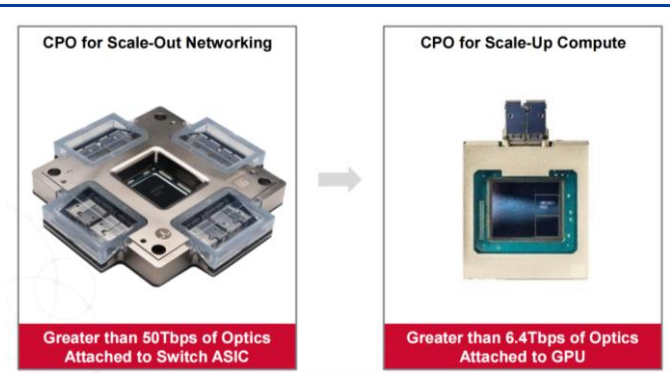
降低 70%，总开关箱使用功耗降低约 30%，对于 32k GPU 集群，可以实现大于 1MW 的功耗节省，未来有望实现计算芯片侧的 CPO。

图表103: 台积电 CPO 路线



资料来源: 台积电, 国盛证券研究所

图表104: 从交换机 CPO 到 GPU CPO



资料来源: 博通, 国盛证券研究所

图表105: Bailly CPO 能耗节省

Power Measurement Results

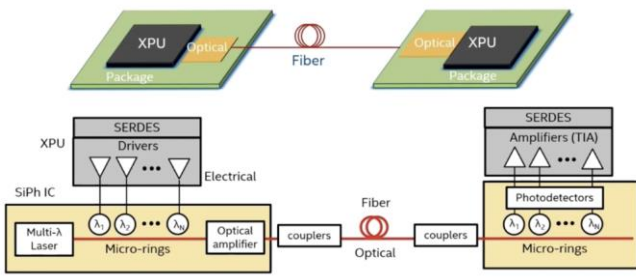
	51T Switch Box Total Power		
	Bailly CPO	Pluggable LPO	Pluggable w/DSP
8x OE Chiplets	241	630	1024
16x PLS	118		
ASIC, CPU, Other	975	975	975
Total (Watts)	1,334	1,605	1,999

- Optical Interconnect using Bailly CPO is 70% lower power
- Total Switch Box is ~30% lower power using Bailly CPO

资料来源: 博通, 国盛证券研究所

OIO 性能卓越，未来有望向全光互连演进。 OIO 是一种基于芯片的光互连解决方案，与计算芯片（中央处理器、图形处理器、张量处理器）集成在同一封装中，旨在解决计算芯片（如 CPU、GPU、XPU 等）之间的互联问题（chip to chip interconnect），凭借光互连低功耗、高带宽、低延迟的特性，替代传统 electrical IO 方案，使芯片输入输出的 IO 转换为光信号，以此构建分布式计算网络。OIO 性能优越，单个 CPO 模块的带宽为 1.6-3.2Tbps，带宽密度为 50-200Gbps/mm，能效为 15pJ/bit，而 Optical IO 的总带宽为 40Tbps，带宽密度为 5Tbps/mm，能效为 3pJ/bit。

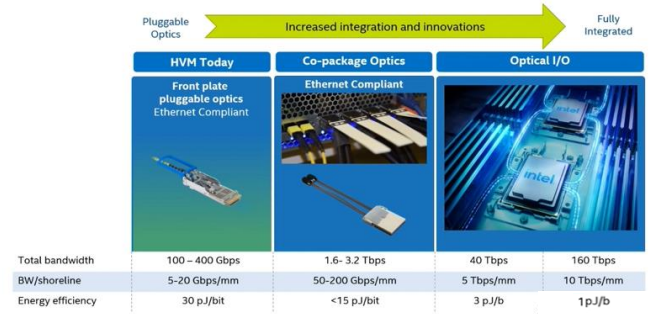
图表106: OIO 示意图



资料来源: cadence, 国盛证券研究所

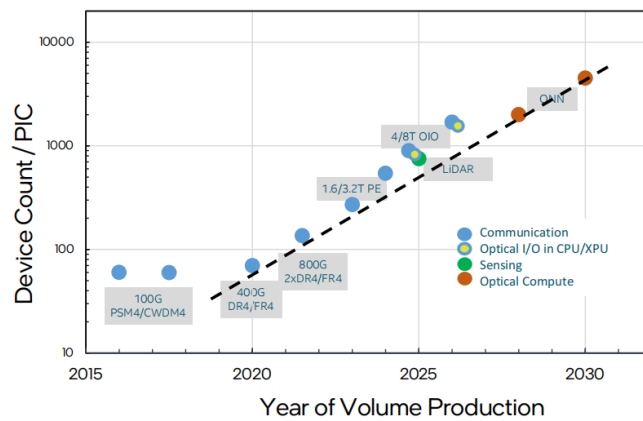
图表107: OIO 与 CPO 的差异对比

Disrupting the trajectory with Optical I/O (OIO)



资料来源: cadence, 国盛证券研究所

图表108: 各互连器量产时间预测



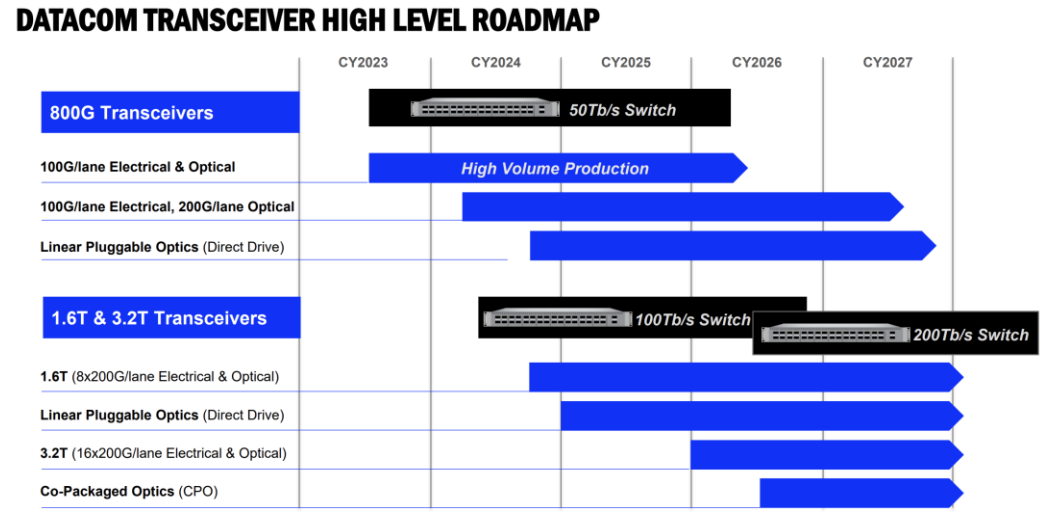
资料来源: intel 2024OCP, 国盛证券研究所

3.3 海外龙头引领市场，国内厂商加速破局

3.3.1 Coherent: 800G 光模块需求强劲，25 年 1.6T 爬坡

800G 光模块收入持续增长，25 年 1.6T 光模块收入有望逐步提升。 CoherentFY25Q1 数据通信收入环比增长约 16%，同比增长 89%，此增长主要得益于 AI 数据中心需求，800G 光收发器的呈持续增长态势，在更广泛客户群中应用。公司预计未来几个季度仍将保持环比增长，1.6T 收发器方面亦持续取得显著进展，自上一季度交付首批样品以来，**Coherent 预计于 FY25 开始逐步提升 1.6T 数据通信收发器的销售额。**公司投资于涵盖 VCSEL、EML 和硅光子学等广泛的收发器组件技术组合。凭借广泛的技术组合，针对每位客户及应用部署最优技术解决方案。公司预计随着 800G 爬坡，该速率光模块将占据公司一半的收入，800G/1.6T 将引领未来五年的市场。

图表109: Coherent roadmap



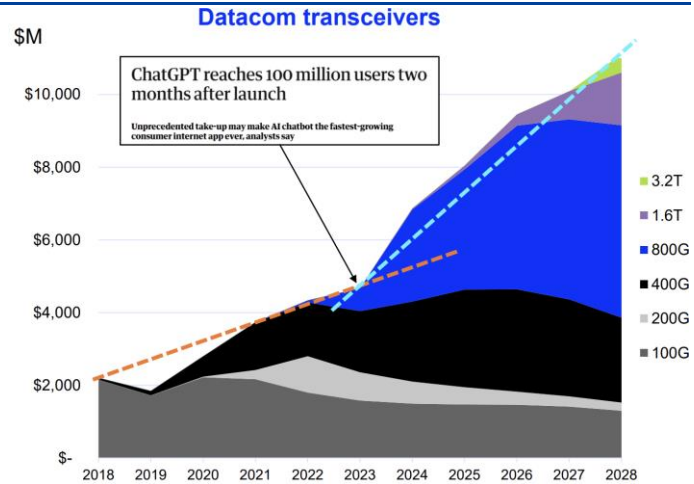
资料来源: Coherent, 国盛证券研究所

图表110: 800G 将占据 Coherent 一半收入



资料来源: Coherent, 国盛证券研究所

图表111: 各速率光模块占比



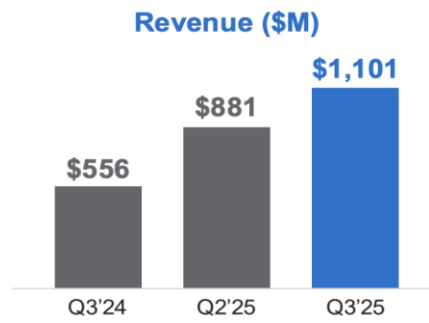
资料来源: Coherent, 国盛证券研究所

3.3.2 Marvell: ASIC 需求强劲, 电光产品收入超预期

ASIC 需求强劲, 电光产品收入超预期。 MarvellFY25Q3 宣布与 AWS 达成多代五年全面协议, 涵盖定制 AI 芯片产品等多种数据中心半导体产品, 还将合作在云中实现 EDA 加速芯片设计。数据中心 FY25Q3 营收创纪录达 11 亿美元, 同比增长 98%、环比增长 25%, 定制 AI 芯片需求强劲且已确保供应链产能。公司成功将复杂芯片从初始样品提升至大批量生产, 为客户提供上市时间优势。电光产品收入超预期, **800GB PAM 产品预订量强劲**, 已出货首款 **1.6T PAM DSP** 和 **5 纳米工艺技术产品**, 设计中标, 增长势头强, 明年产量预计加快。推出业界首款 3 纳米 1.6T DSP 降低功耗提升能效。有源电缆方面, **每通道 100GB 800GB DSP 产量加速**, 开始对首款每通道 **200GB 1.6T AEC DSP 采样**。预计第四季度数据中心环比增长率在 20% 中低范围, 定制 AI 芯片收入将因项目量产大幅增加, 以太网交换机及互连产品组合增长将进一步推动增长。

图表112: marvellFY25Q3 数据中心业务亮点

Data center end market Q3 FY25



Up 98% y/y and 25% q/q

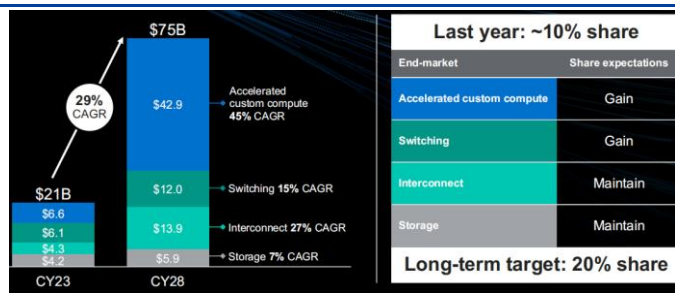
Strong demand for interconnects and custom silicon

Highlights

- Announced 5-year, multi-generational agreement with Amazon Web Services, expanding relationship
- Strong ongoing demand from AI applications
- Significant step up in custom AI silicon ramp
- Began shipments of 5nm 1.6T PAM4 DSPs
- Strong bookings for 800G PAM4 DSPs
- Launched industry-first 3nm 1.6 Tbps PAM4 optical DSP
- Demonstrated industry-first PCIe Gen 7 connectivity
- Expect Q4 revenue to grow greater than 20% sequentially on a percentage basis, driven by high-volume production ramp of AI custom programs

资料来源: Marvell, 国盛证券研究所

图表113: marvell 数据中心 TAM 及份额展望



资料来源: marvell, 国盛证券研究所

图表114: marvell platform

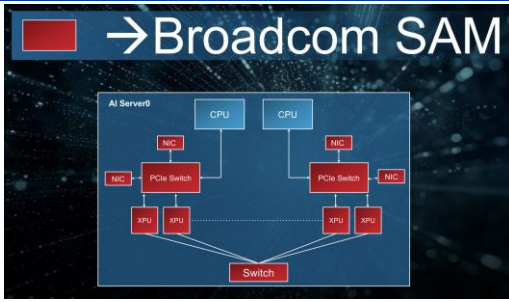


资料来源: marvell, 国盛证券研究所

3.3.3 博通: 27年 AI SAM 有望达 600 亿至 900 亿美元

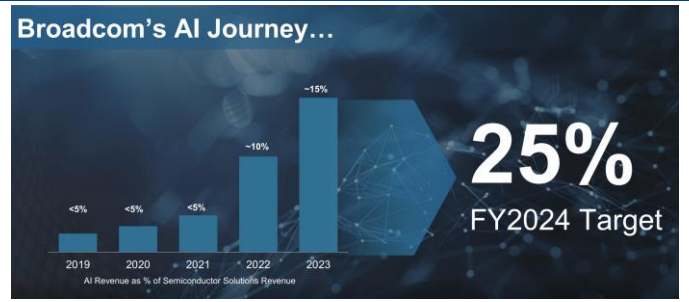
27年 AI SAM 有望达 600 亿至 900 亿美元, 连接占比显著提升。FY24Q4 博通网络业务收入为 45 亿美元, 同比增长 45%, AI 网络收入占网络收入的 76%, 同比增长 158%, 主要系对三个超大规模客户的 AI XPU 出货量翻倍, 以及 Tomahawk 和 Jericho 出货量推动的 AI 连接收入增长 4 倍。公司预计 FY25Q1 随着越来越多的超大规模企业在其结构中部署 Jericho3-AI, AI 连接持续强劲, 公司下一代 XPU 为 3 纳米, 将是进入市场的第一款 XPU, 有望在 FY25 下半年为超大规模客户批量出货。根据目前已经达成合作的三家 CSP, 博通 2024 年 AI SAM (XPU+AI 连接) 约 150 亿至 200 亿美元, 其中连接占比 5%-10%, 27 年 SAM 为 600 亿至 900 亿美元, 其中连接占比 15%-20%, 目前公司正与另外两家 CSP 开发下一代 AI XPU, 预计 27 年之前发展成创收客户。

图表115: 博通 AI SAM



资料来源: 博通, 国盛证券研究所

图表116: 博通 AI 收入占比不断提升



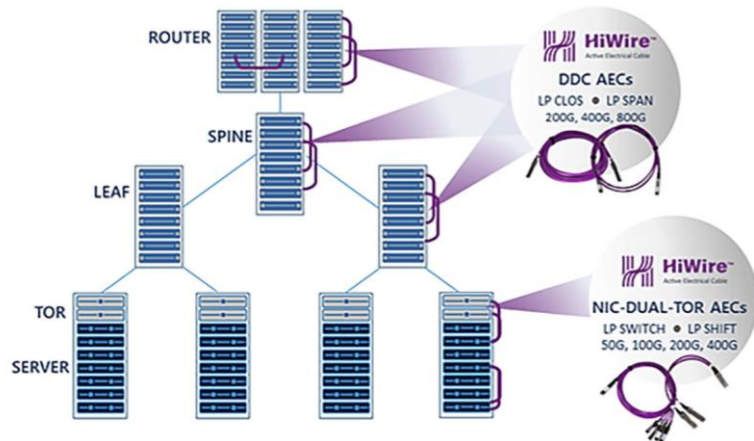
资料来源: 博通, 国盛证券研究所

3.3.4 Credo: AEC 收入再创新高, 三大客户需求强劲

FY25Q2 公司收入 7200 万美元, 环比增长 21%, 同比增长 64%, 三个主要产品线的收入创下历史新高, 并开始看到出货量的增加: 1) AEC: AEC 是目前最大的产品线。AEC 收入再次创下历史新高, 主要系前两大客户和一家新兴超大规模企业的强劲需求, 公司认为 AEC 仍处于采用的早期阶段, 近期推出了适用于 AI 后端网络的 800 GB ZeroFlaps AEC, 支持 7 米; 2) 光学 DSP: 光学 DSP 是公司的第二大市场, 主要针对以太网应用。光学 DSP 收入创纪录, 主要系 50 GB 和 100 GB 每通道解决方案的推动 (光模块与 AOC), 支持 100G-800G, 目前近期完成了 3 纳米每通道 200GB 设计的流片, 重点关注电源效率, 预计该市场将在 FY26 之后的更长时期内发展, 在 FY27 达到高产量; 3) 铜解决方案: 包括线路卡 retimer、chipllets 等, 其中, 线路卡 retimer 400G 和 800G 应用领域收入实现季度新高。

AI 驱动收入增长, 公司预计未来所有美国 CSP 均将采用 AEC。公司预计 FY26 实现 50% 增长, 主要的增长驱动力是 AEC, 目前大多数 AEC 产品的出货量均基于 AI 应用。从趋势来看, AEC 从每个 GPU 连接一个, 逐步过渡到每个 GPU 连接多个, GPU 部署的绝对数量显著增长。下一代 AI 设施将需要液冷技术, 随着液冷技术的应用, AI 节点的密度增加, GPU 之间的距离变短, 将使得电气连接 (如铜线/AEC) 更加普及。从客户来看, 公司对北美四大 CSP 以及 Oracle、XAI 和 Omniva 均有出货, FY25Q2 有三家客户收入占比超过 10%, 其中微软收入占比达 11%, 第二家 AEC 大型客户收入占比达到 33%, 本季度有一家新的大型客户导入, 收入占比达 14%。公司预计所有美国超大规模客户从长远来看都会以某种形式使用 AEC。

图表117: 数据中心的 AEC 的应用

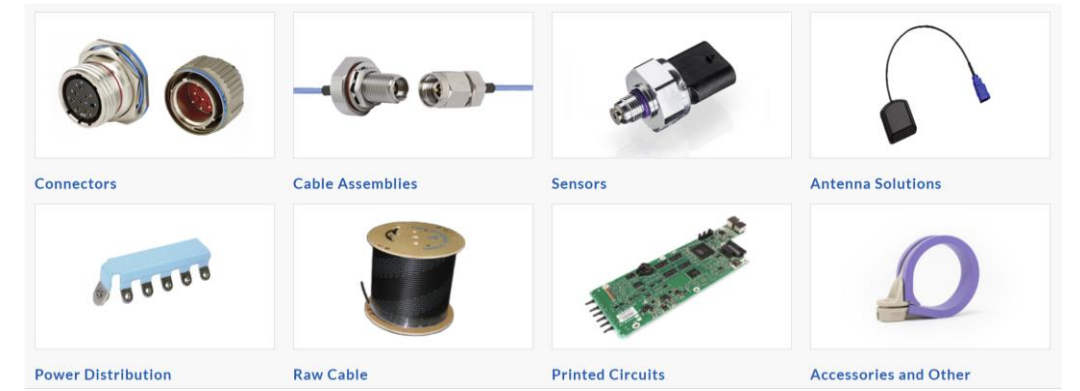


资料来源: credo, 国盛证券研究所

3.3.5 安费诺：AI 等需求驱动收入超预期

AI 等需求驱动 Q3 IT 数据通信业务超预期。24Q3 安费诺 IT 数据通信市场约 11 亿美元，占比 25%，以美元计价同比增长 60%，环比增长 15%，超预期，主要系人工智能应用及基础 IT 数据通信业务的强劲增长。展望 24Q4，公司预将进一步实现中等个位数的环比增长，预计 2024 年全年 IT 数据通信的销售总额同比增长 50% 以上。

图表118: 安费诺产品



资料来源：安费诺，国盛证券研究所

3.3.6 沃尔核材：高速线批量交付，产能满载积极扩产

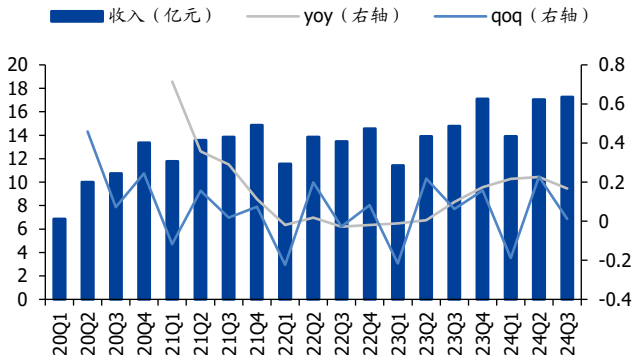
产能满载、高端线批量交付，25 年关键设备陆续到货。沃尔核材电线系列产品由乐庭智联经营，乐庭智联在电线行业深耕多年，拥有强大的研发技术团队，具有丰富的产品开发经验和制程控制经验，通过多年的技术积累，乐庭智联掌握全部重点产品的核心技术，2024 年实现了 800G 高速通信线、单通道 224G 高速通信线的批量交付。公司的高速通信线主要客户包括安费诺、豪利士、莫仕、泰科等国际客户；以及立讯、英飞联等国内多家连接器企业，公司的高速通信线供应给下游客户群体后配套给国际、国内各大服务器厂商终端使用。目前，公司高速通信线订单饱和，生产满载。产能及扩张方面，乐庭智联已拥有绕包机近四百台，芯线机超三十台，有超两百台绕包机和几十台芯线机已下单；随着新采购的高速通信线设备陆续交付，其产能正在逐步提升，预计 2025 年将有多台关键设备陆续到货。

图表119: 沃尔核材电线产品

主要产品名称	电线横截面图示	应用领域
<ul style="list-style-type: none"> 高速通信线 	<p>400G/800G 线缆 112G 背骨线</p>	数据中心交换机、工业路由器设备、服务器等
<ul style="list-style-type: none"> 汽车线 	<p>车载通信线 汽车花线</p>	汽车领域
<ul style="list-style-type: none"> 工业线 	<p>机器人线缆 工业线缆</p>	工业机器人、服务机器人、特种机器人及工业各类设备、机床等设备
<ul style="list-style-type: none"> 消费电子线 	<p>电器装备线缆 单芯屏蔽线 电源线</p>	工业、消费电子设备、影音通讯系统等领域

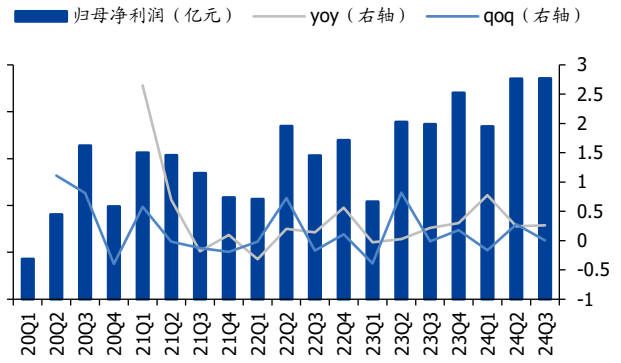
资料来源：公司公告，国盛证券研究所

图表120: 沃尔核材收入及增速(单季度)



资料来源: wind, 国盛证券研究所

图表121: 沃尔核材归母净利润及增速(单季度)

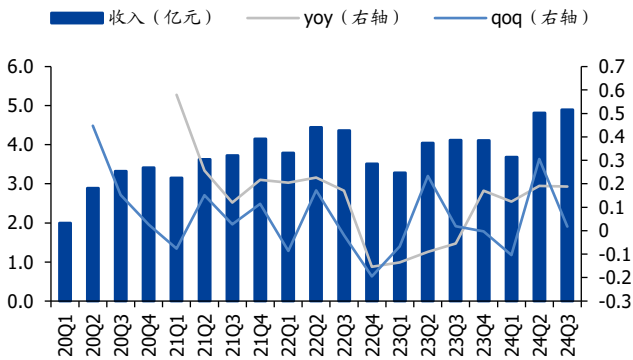


资料来源: wind, 国盛证券研究所

3.3.7 兆龙互连: 单通道 112G 有源铜缆批量交付, 25 年高速产品需求乐观

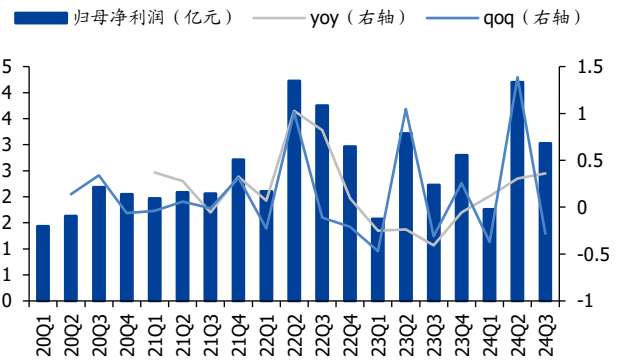
高速产品订单饱和积极扩产, 指引 25 年乐观需求。高速电缆组件是公司重点发展方向之一, 主要应用于大型数据中心、服务器群等对传输速率要求高的领域。公司在服务器内外高速连接产品上拥有全套自主设计能力, 包括 PCBA、线端连接器等。公司无源铜缆和有源铜缆产品针对不同的情况提供了不同的解决方案。目前, 公司有源铜缆单通道 112G 产品已经陆续批量交付中; 高速产品订单处于相对饱和的状态, 公司预计 25 年高速产品需求乐观; 高速产品产能紧张, 公司正在通过购买新设备、改造旧设备等持续稳步扩大生产能力, 新设备正在陆续交付与调试。

图表122: 兆龙互连收入及增速(单季度)



资料来源: wind, 国盛证券研究所

图表123: 兆龙互连归母净利润及增速(单季度)



资料来源: wind, 国盛证券研究所

3.3.8 立讯精密: 24 年服务数据中心量翻四倍, 224G 铜连接器落地在即

224G 铜连接器落地在即, 448G 产品进入预研阶段。在数据中心高速互联领域, 公司协同头部芯片厂商前瞻性为全球主流数据中心及云服务厂商共同制定 800G、1.6T 等下一代高速连接标准, 部分细分领域产品如外部高速铜缆、线缆及连接器组件、背板连接器及背板线缆、HSIO、SSIO 等更是以领先的技术、成本优势和优异的产品表现赢得了客户的认可。High speed 铜连接器以及线束在海外的业务正在快速增长。公司认为从目前准备落地的 224G 产品到未来预研的 448G 产品来看, 铜连接未来在整个架构的应用上会有更大的空间, 特别是无源连接部分会有更好的机会。在光连接方面, 目前在海外主流客户暂时还有些挑战, 但也已快速进入前三大数据中心客户项目, 有的是在测试阶段, 有的已经小批量交货, 公司采取“电连接带动光连接”的策略, 要先让电连接得到客户充分的认可。24 年所服务的国内外中小型数据中心的量大概是 23 年的 4 倍, 且市场还在逐渐扩大, 增量空间广阔。

图表124: 立讯精密数据中心解决方案



资料来源：立讯精密官网，国盛证券研究所

3.3.9 澜起科技：运力芯片开启第二成长曲线，PCIe Retimer 需求旺盛

PCIe Retimer 24Q3 出货量超 60 万颗，高性能运力芯片开启第二成长曲线。24Q3 公司互连类芯片产品线销售收入为 8.49 亿元，同比+46.02%，环比+1.83%，创产品线单季度销售收入历史新高，毛利率为 62.21%。其中：DDR5 内存接口芯片单季度出货量超过 DDR4 内存接口芯片；在新产品方面，PCIe Retimer 芯片第三季度出货量超 60 万颗，推动公司三款高性能运力芯片新产品合计销售收入环比小幅增长。

PCIe4.0/5.0 Retimer 已实现量产，自研核心底层技术 SerDesIP。在 PCIe 4.0 时代，澜起是全球量产 PCIe 4.0 Retimer 芯片的三家企业之一；在 PCIe5.0 时代，公司于 2023 年 1 月量产 PCIe 5.0/CXL 2.0 Retimer 芯片，是全球第二家量产该产品的厂家。公司 PCIe Retimer 产品已于 24Q1 开始出货，24Q2 出货量翻倍增长，24Q3 出货量再次翻倍增长。预计 PCIe Retimer 芯片将在未来几年为公司贡献新的业绩增长点，主要系：

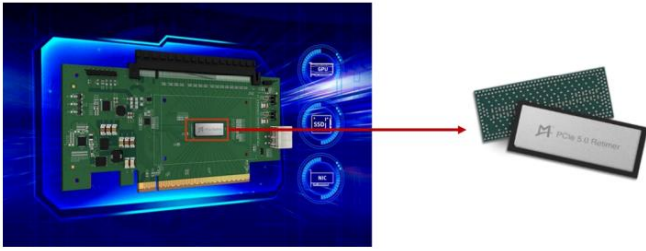
- ✓ AI 服务器需求增加：一台典型的配置 8 块 GPU 的主流 AI 服务器需要 8 至 16 颗 PCIeRetimer 芯片，PCIeRetimer 芯片的市场空间将随着 AI 服务器需求量的增加而扩大；
- ✓ 市场份额提升：澜起自研该产品核心底层技术 SerDesIP，因此在产品时延、信道适应能力等方面具有竞争优势，澜起的 PCIeRetimer 芯片正在获得越来越多客户及下游用户的认可；
- ✓ PCIe 协议持续迭代：PCIeRetimer 芯片是未来数据中心领域重要的互连芯片，可用于 CPU 与 GPU、NVMeSSD、Riser 卡等典型高速外设的互连。PCIe 协议每次迭代将带来数据传输速率翻倍，PCIeRetimer 芯片的作用是提升信号完整性，增加高速信号的有效传输距离，因此未来需要用到 PCIeRetimer 芯片的场景会越来越多。

图表125: 澜起科技 PCIe Retimer 量产及应用进展

PCIe 4.0/5.0 Retimer芯片产品	应用
8通道PCIe 4.0 Retimer	服务器、存储设备和硬件加速器
16通道PCIe 4.0 Retimer	服务器、存储设备和硬件加速器
16通道PCIe 5.0/CXL 2.0 Retimer	服务器、存储设备和硬件加速器

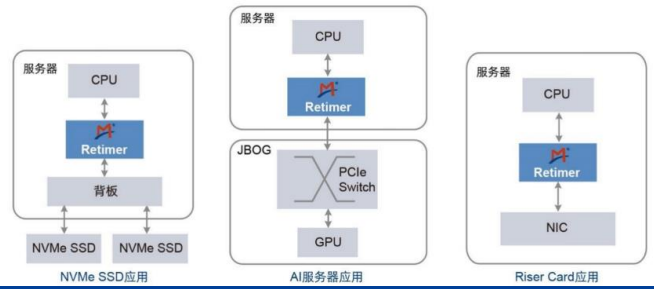
资料来源：澜起科技公告，国盛证券研究所

图表126: 澜起科技 PCIe Retimer 示意图



资料来源: 澜起科技公告, 国盛证券研究所

图表127: PCIe 4.0/5.0 Retimer 芯片的典型应用场景图示

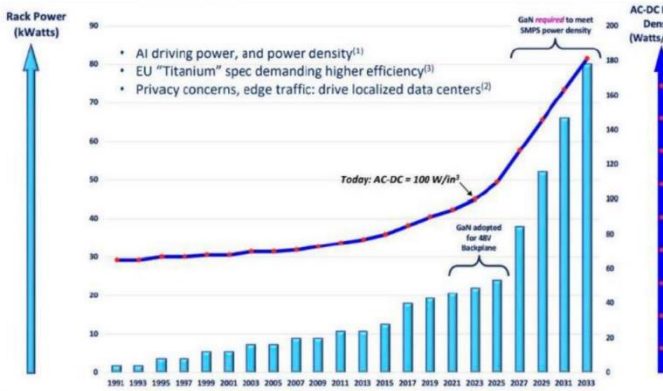


资料来源: 澜起科技公告, 国盛证券研究所

4 电源：AI 驱动电源&电容需求攀升

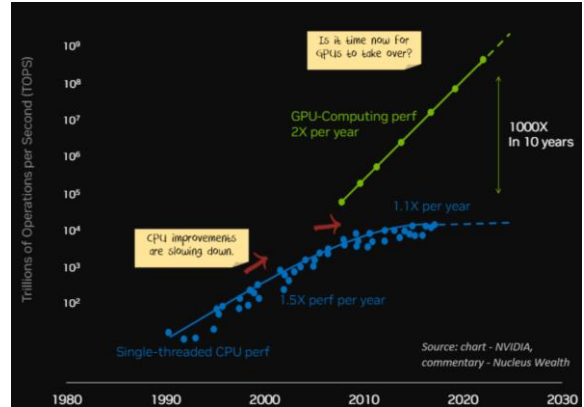
AI 技术的精进导致电力消耗剧增。大模型的训练及推理响应均需要消耗大量电能。训练谷歌于 2022 年发布的大语言模型 PaLM 需要消耗 3436 兆瓦时的电量，约等于 11.8 万美国普通家庭日耗电量（美国普通家庭日均耗电量约为 29 千瓦时），据《纽约客》杂志援引国外研究机构报告，ChatGPT 每天要响应大约 2 亿个请求，在此过程中消耗超过 50 万度电力，而随着生成式 AI 的广泛应用，预计到 2027 年，整个人工智能行业每年将消耗 85 至 134 太瓦时（1 太瓦时=10 亿千瓦时）的电力。

图表128: AI 电源功率密度持续提升



资料来源: *Electronicdesign*, 国盛证券研究所

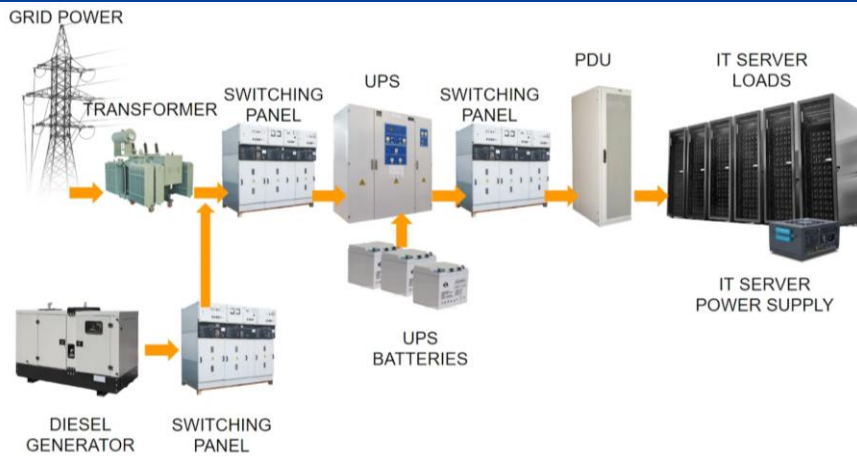
图表129: CPU/GPU 算力加速提升带动电源需求



资料来源: *nucleuswealth*, 国盛证券研究所

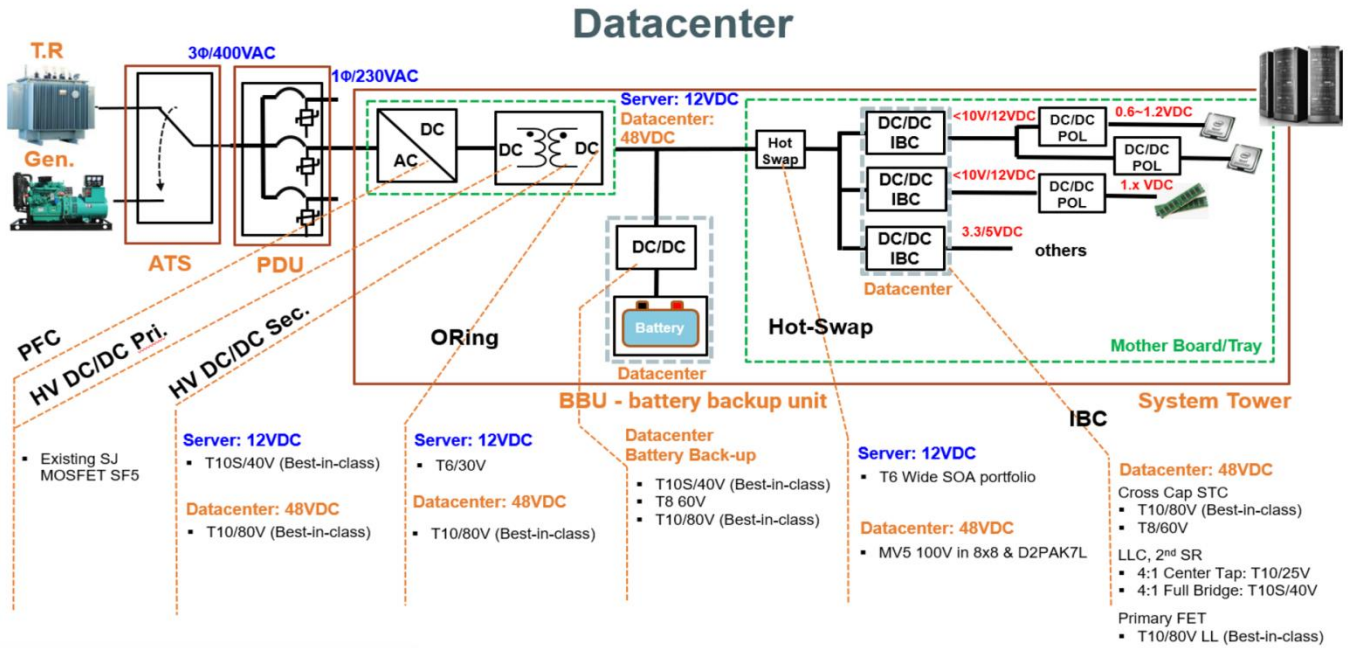
AI 服务器电源是数据服务器集群的电力保障。AI 服务器电源作为高性能计算和数据中心的基础设备，担负着为服务器集群提供稳定、高效电能供应的任务。高压电从电网进入数据中心后，服务器电源供应器会先将交流电转为直流电，并降压到 48V；接着主板上的 DC/DC 转换器，再将电压转换成 12V、5V、3V3 和 0.8V 等。

图表130: 数据中心电源的架构主要分为保障电路 UPS、机架电源 AC/DC、芯片电源模组 DC/DC 三个层级



资料来源: *densenetworks*, 国盛证券研究所

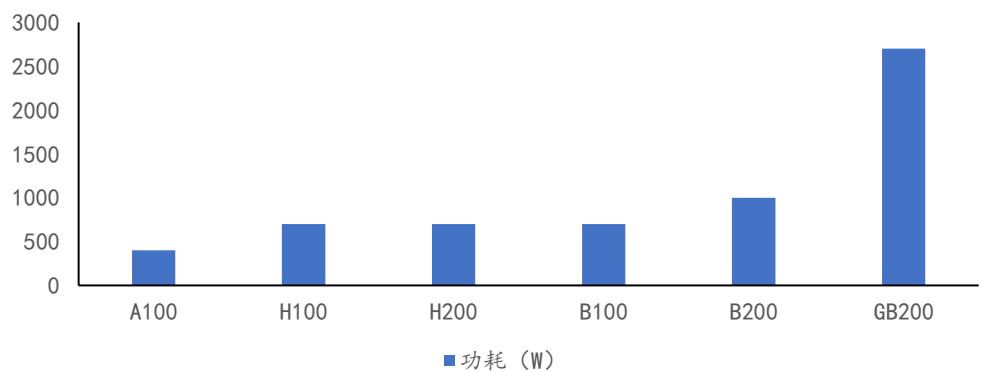
图表131: 典型服务器电源架构



资料来源: 安森美, 国盛证券研究所

英伟达 GPU 功率持续提升, AI 服务器电源市场规模稳步增长。英伟达芯片额定功率持续提升, A100 为 400W, H100\H200\B100 均为 700W, 2024 年的 GTC 上, 英伟达发布的基于 Blackwell 架构的 B200 GPU 功率则首次达到了 1000W, 超级芯片 GB200 功率最高达到 2700w。通用型服务器只需要 2 颗 800W 服务器电源, 而 AI 服务器则需要 4 颗 1800W 高功率电源, AI 服务器对于电源的需求大大提升。未来随 GPU 功率提升, AI 服务器电源市场规模亦将持续提升。

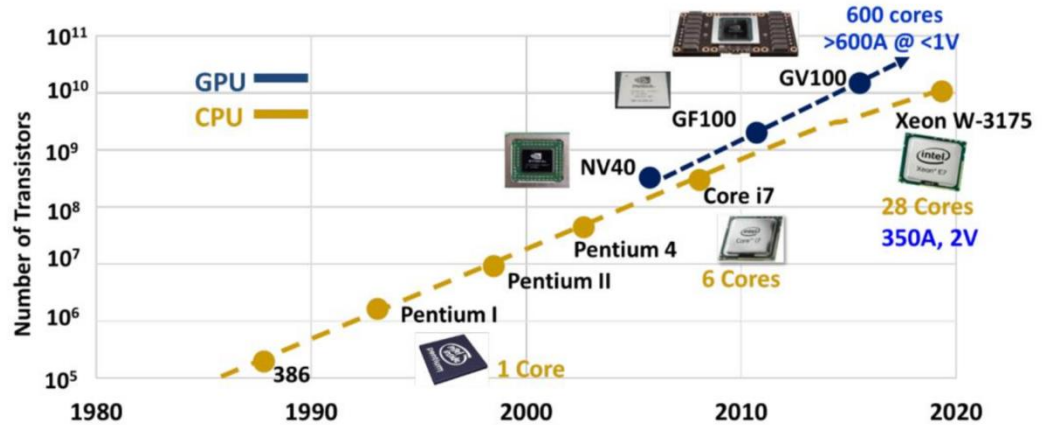
图表132: 英伟达 GPU 芯片功率持续提升



资料来源: 快科技, 英伟达官网, 国盛证券研究所

板卡层面, GPU、CPU 等算力芯片朝着更高性能升级也对供电系统提出更高要求。算力的提升需要更大的耗电, 这对供电电压调节器模块 (VRM/Vcore) 和负载点电源 (PoL) 提出了严峻挑战, 包括更高的效率、更高的功率密度, 以及同时满足处理器 $di/dt > 1000A/us$ 瞬态响应要求。

图表133: GPU/CPU 功率提升对 VRM 和 PoL 提出严峻挑战



资料来源: IDC 导航, 国盛证券研究所

由多相控制器和 DrMos 组成的拓扑架构是目前公认的 GPU/CPU 供电场景的最佳解决方案。每相 Buck 对应的半桥 MOSFET 可由包含驱动和温度/电流检测的 DrMOS 代替, 由一个控制器采集反馈的电压、电流、温度/错误等信号, 并发出各 PWM 波实现功率的闭环控制。控制器可通过特定协议的通信接口和信号指示 IO 口, 与系统上位机或负载处理器进行信号交互。多相控制器连接 DrMOS 可实现多相拓扑结构的大电流 DC/DC 系统, 可大幅提升供电功率, 优化能耗, 提升整体性能, 简化系统电源设计, 是大功率供电芯片的核心, 也是 GPU 的主流供电形式。

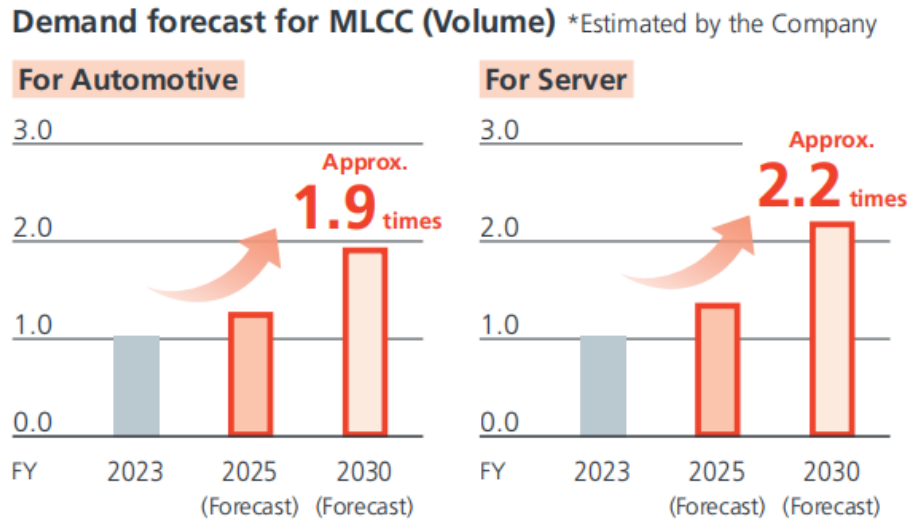
图表134: 由多相控制器和 DrMos 组成的拓扑架构是目前公认的 GPU/CPU 供电场景的最佳解决方案



资料来源: eda365, 国盛证券研究所

高算力需求推升运行功耗, 显著增加电容用量。以 MLCC 为例, 服务器供应电压是 48V 的直流电源, 但服务器中承担算力重任的 GPU、CPU 供应电压主要是 1V, 中间需要 MLCC 护航多路电源转变; 由于 CPU 工作时有轻载和重载切换, 电压会跌落, 且有高频谐波和杂音, 此时也都需要电容发挥稳定电压作用。在服务器中, 由于算力需求的提升, GPU 和 CPU 所需晶管数量增多, 功耗也快速上升, 需更多 MLCC 来保障电路的安全。并且由于高算力 GPU/CPU 需要的电容数量更多, 容量更大, 在面积有限的板子上, 更小的大容量电容才能满足需求。而 AI 服务器中较大的电流和更高的 GPU/CPU 工作频率, 也对 MLCC 提出了更加耐高温、超低的 ESR、更低 ESL、更高的自谐振频率 SRF 等要求。根据太阳诱电预测, 2023-2030 年服务器市场对 MLCC 的需求量增速超过 100%, 相比汽车市场更快。

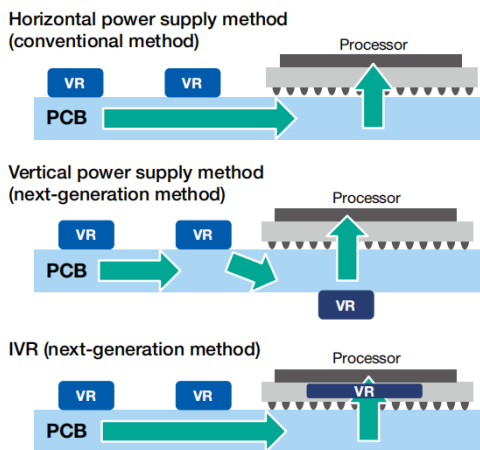
图表135: 2023-2030年预计服务器市场对MLCC需求量将翻2.2倍



资料来源: 太阳诱电, 国盛证券研究所

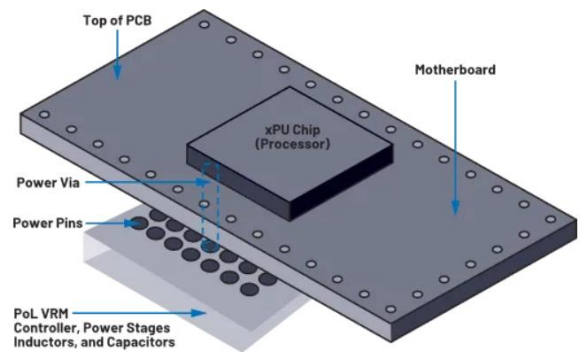
垂直供电或成为未来 AI 服务器供电架构, 进一步提高电感等被动元件的超小型、超薄型要求。目前, 将电子元件布置在处理器周围的水平的供电方式一直是 AI 服务器主流的供电架构, 但是在这种模式下, 传统 DC/DC 模块和芯片同样位在系统板上, 电流采取水平方式传输, 电流会因为传输距离而产生损耗。随着人工智能服务器所用处理器性能的提升, 电源的耗电量以及产生的热量预计也会大幅增加。因此为提高能源效率, 垂直供电网络或成为未来的发展趋势。在这种新架构下, DC/DC 电源和电子元件被直接放置在处理器下方, 传输距离只有一块主板的厚度, 让电流流动损耗降到最低, 提升效率, 以及需要在电路板上安装集成式稳压器 (IVR)。根据台达客制化设计部门研发总监皮柏 (Ralf Pieper) 在 GTC 演讲的内容, 4000 瓦的处理器在 0.8 伏特电压下, 垂直供电可以比水平供电提升 23% 效率。但垂直供电也对电路中的电子元件提出超小型、超薄型的要求, 以 TDK 为例, 这家全球电感龙头也在为这一新应用场景研发采用绕线、多层化和薄膜技术的下一代电感器。

图表136: 水平供电、垂直供电与集成式稳压器示意图



资料来源: TDK, 国盛证券研究所

图表137: 安装在 xPU 下方的垂直电源模块架构



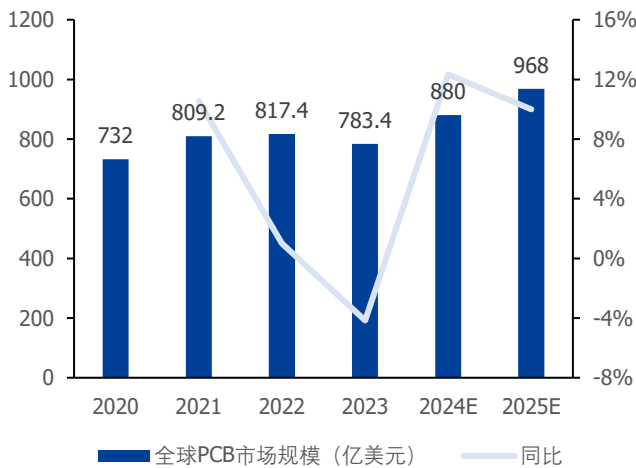
资料来源: ADI, 国盛证券研究所

5 PCB: AI 引领全面增长, HDI 浪潮已至

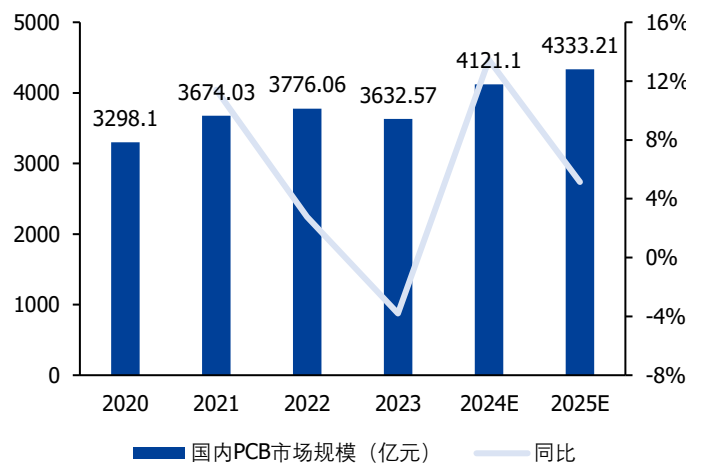
5.1 PCB 行业长期增长动力充足, 高端产品需求旺盛

伴随 AI 算力需求持续增长以及智能汽车持续升级, AI 服务器和汽车电子相关的 PCB 需求显著提升, 成为驱动 PCB 行业成长的主要驱动力。根据中商产业研究院, 2024 年全球 PCB 市场规模将达到 880 亿美元, 2025 年达到 968 亿美元, 2025 年同比增速达 10%。中商产业研究院分析师预测, 2024 年中国 PCB 市场将回暖, 市场规模将达到 4121.1 亿元, 2025 年达到 4333.21 亿元, 同比增速为 5.1%。

图表138: 全球 PCB 市场规模 (亿美元)



图表139: 国内 PCB 市场规模 (亿元)



资料来源: 中商情报网, 国盛证券研究所

资料来源: 中商情报网, 国盛证券研究所

从产品角度来看, AI 引领高端 PCB 需求长青, 高阶 HDI 增速迅猛。根据 Prismark, 预计 2024 年 PCB 行业同比增速为 5.0%。以 AI 算力硬件为代表的需求爆发, 提升了高端 PCB 产品的需求。其中 18 层以上的高多层板预计同比增速 21.1%, HDI 板同比增速为 10.4%, 同比增速均显著高于行业整体平均增速。

往后展望, 随着人工智能技术和应用的快速发展, 未来五年 AI 系统、服务器、存储、网络设备是 PCB 需求增长的主要动能。随着 AI 服务器的升级, GPU 主板也将由高多层板逐步升级为 HDI, 因此 HDI 将是未来 5 年增速最快的 PCB 产品, 特别是 4 阶以上的高阶 HDI 产品需求增速快。HDI 将成为 AI 服务器相关 PCB 市场增速最快的品类, 根据 Prismark 预计, 2023-2028 年 HDI 的年均复合增速达到 16.3%。

图表140: 2024 年 PCB 不同产品同比增速

产值复合增长率	多层板			HDI	封装基板	柔性板	其他	总计
	4-6层	8-16层	18层以上					
美洲	2.0%	2.8%	6.6%	3.9%	19.8%	3.7%	1.5%	3.3%
欧洲	1.1%	1.3%	4.6%	0.6%	12.9%	1.5%	0.8%	1.2%
日本	1.9%	2.5%	3.1%	2.2%	-0.4%	2.1%	1.0%	0.9%
中国	2.0%	3.7%	21.3%	11.5%	10.4%	6.7%	0.7%	5.3%
亚洲	2.6%	7.0%	45.6%	10.4%	5.9%	3.5%	3.1%	6.4%
总计	2.1%	4.1%	21.1%	10.4%	5.4%	4.8%	1.1%	5.0%

资料来源: 沪电股份公告, 国盛证券研究所

从下游应用来看，服务器及汽车增速领先。根据 PrismaMark，2023~2028年 PCB 多层板整体复合增速为 4.4%。下游各应用领域中，服务器行业成长速度较高，2023~2028年 cagr 达 8.7%，主要增长驱动力为 AI 服务器带动的 PCB 需求的量价齐升；汽车行业 2023~2028年 cagr 达 4.4%，主要增长动力来自于汽车智能化的发展，包括自动驾驶以及智能座舱带动的 PCB 需求的持续增长；此外，军事/航空航天 2023~2028 年均复合增速为 5.4%。

图表141: PCB 各下游应用领域

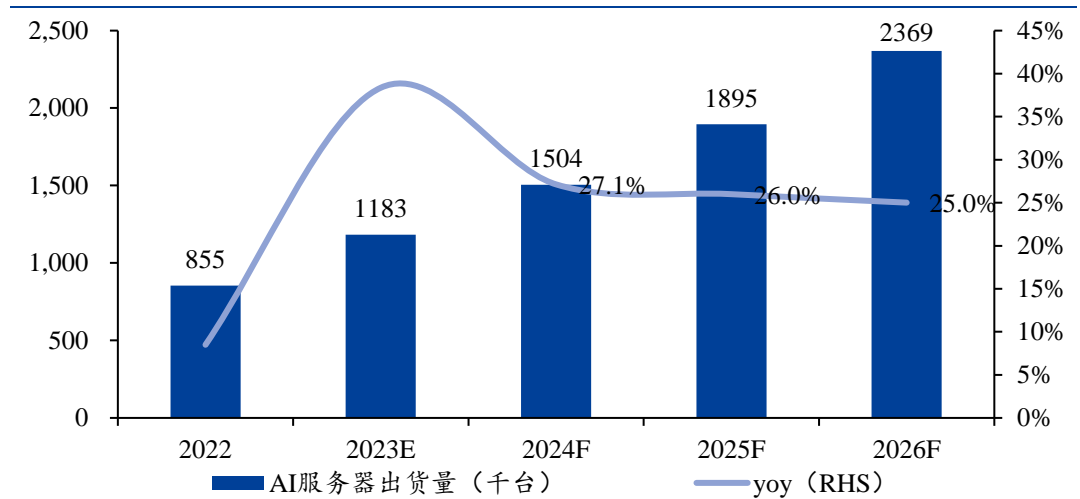
多层板应用领域	2023-2028年均复合增长率
计算机: PC	1.7%
服务器/数据存储	8.7%
其他计算机	1.9%
手机	3.2%
有线基础设施	4.8%
无线基础设施	2.0%
其他消费电子	3.2%
汽车	4.4%
工业	3.8%
医疗	3.5%
军事/航空航天	5.4%
合计	4.4%

资料来源: 沪电股份公告, 国盛证券研究所

5.2 AI 需求全面增长，服务器架构升级带动 PCB 量价跃升

AI 服务器出货量加速扩张。根据 TrendForce，预计 2023 年 AI 服务器出货量为 118.3 万台，占整体服务器比重为 9%左右，2024 年 AI 服务器出货量将再次增长 27%，达 150.4 万台，且往后展望至 2026 年出货同比增速达 25%。

图表142: AI 服务器出货量预测



资料来源: TrendForce, 国盛证券研究所

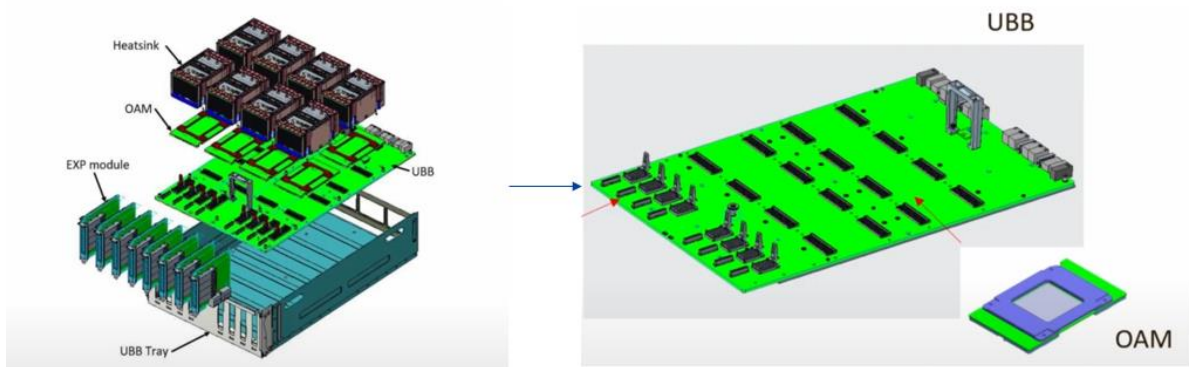
AI 服务器价值量跃升。八卡架构服务器 PCB 的主要构成为 GPU 母板 UBB (Universal Base Board)、GPU 加速卡 OAM (Open Accelerator Module)、CPU 主板以及网卡、内存、电源等产品。以 DGX H100 为例，其内部 PCB 价值量较高的 GPU 加速模块，包含一片 UBB 板，8 片 OAM 加速卡，其中 UBB 通常为高多层板，加速卡为 HDI PCB。

从材料端来看, AI 服务器升级对应 PCB 的材料持续升级, AI 产品的 CCL 价格对比 whitley 服务器的 CCL 单价提升了 2~2.5 倍;

从 PCB 层数来看, 伴随 AI 服务器的 PCB 层数提升, 其价值量也对应有显著提升。传统服务器层数一般不超过 20 层, 单价约为 800~2000 元; 而 AI 服务器 PCB 层数通常在 20 层以上, 其单价超过 5000 元, 较传统服务器 PCB 有成倍提升;

从制造工艺来看, AI 加速卡的 PCB 通常采用 4 阶及以上的 HDI 工艺, 高阶 HDI 制造其工序复杂生产难度大, 伴随阶数提升价值量亦有显著增长。因此 AI 服务器将带动 PCB 的量价齐升, 尤其是伴随高速高密度互联需求崛起, 高等级材料的高层高阶 HDI 需求有望持续增长, 对应价值量有望持续增长。

图表143: 八卡架构 AI 服务器的拆解



资料来源: OCP REGIONAL SUMMIT, 国盛证券研究所

图表144: 八卡加速模块实物



资料来源: servethehome, 国盛证券研究所

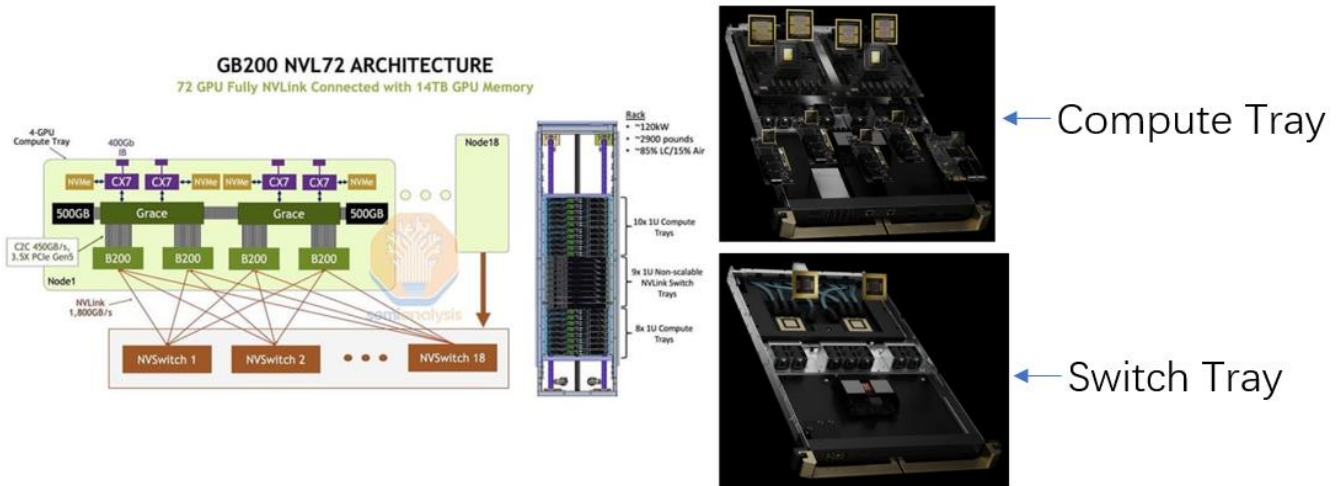
图表145: AI 服务器 PCB 价值量大幅提升

	Whitley	Eagle Stream	AI Level
PCB层数	12~14	16~20	20+
CCL材料	Low loss	Very Low loss	Ultra Low loss
传输速度	10Gbps	25Gbps	56Gbps
损耗 (Df)	0.006~0.009	0.005~0.006	0.0015~0.005
CCL ASP	1x	1.25x~2x	2x~2.5x
单价/单片	800元	2000元	5000元以上

资料来源: 优分析, 国盛证券研究所

GB200 服务器架构创新, 其与 8 卡 AI 服务器架构最大的区别在于其内部集成度再次提升, CPU 和 GPU 同时集成在一片 Compute Board 上, GPU 和 CPU 之间数据传输速率大幅提升。同时, GB200 服务器将负责实现各个 GPU 之间高速互联的 Switch 芯片独立出来形成 Switch Tray, 实现整个机柜间 GPU 互联效率的提升。

图表146: 英伟达 GB200 NVL72 内部架构

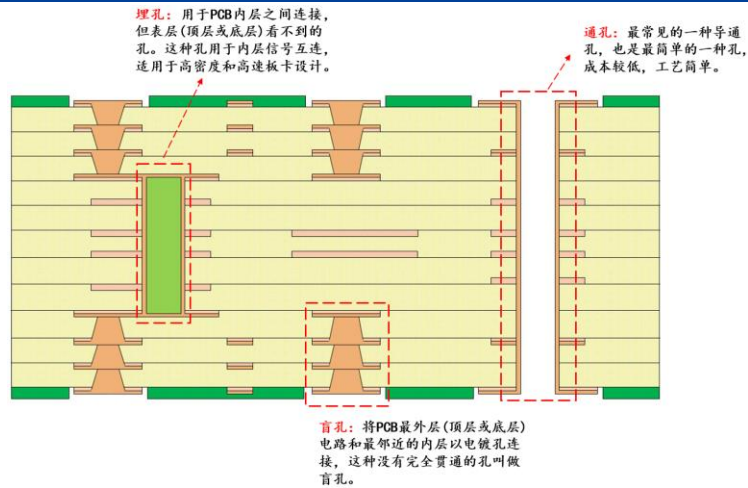


资料来源: 广东连接器协会, 英伟达, 国盛证券研究所

5.3 AI 高集成度及高速互联带动 HDI 产业趋势

HDI PCB 是高密度互连 (High Density Interconnect) 印制电路板的简称, 也称微孔板或积层板。HDI 线路板的主要特征包括埋孔、盲孔的互联结构、多阶压合的制作工序以及更高密度的线宽线距。

图表147: 12层3阶 HDI 结构示意图



资料来源: 景旺电子, 国盛证券研究所

从定义上来看, HDI 板高密度的特征主要体现在孔、线路、焊盘密度、层间厚度等方面。从微导孔的孔径来看, HDI 埋盲孔的孔径通常小于 150 μm , 而伴随电子终端集成度的日益提升, HDI 的微导孔孔径也在日益缩小。从线宽线距的角度来看, 通常 HDI 产品的线宽线距不超过 76.2 μm ; 从焊盘密度的角度看, HDI 线路板对应的焊盘密度通常为每平方厘米大于 50 个; 最后从介质厚度的薄型化的角度来看, HDI 层间介质厚度向 80 μm 及以下的趋势发展, 并且对厚度均匀性要求越来越严格。

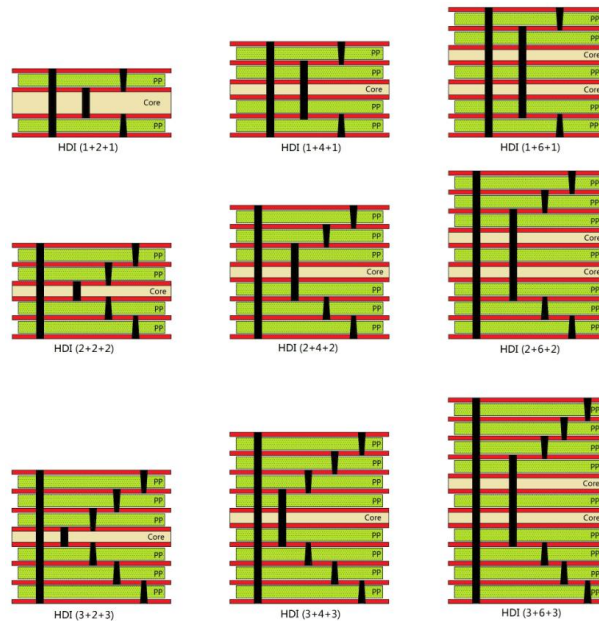
图表148: HDI 参数特性

参数	特性
微导孔	HDI 板内含有盲孔等微导孔设计，其主要表现在孔径小于 150um 的微孔成孔技术以及成本、生产效率和孔位精度控制等方面的高要求化。而传统的多层电路板中只有通孔而不存在微小的埋盲孔。
线宽与线距的精细化	主要表现在导线缺陷和导线表面粗糙度要求越来越严格，一般线宽和线距不超过 76.2um。
焊盘密度高	焊接接点密度每平方厘米大于 50 个。
介质厚度的薄型化	层间介质厚度向 80um 及以下的趋势发展，并且对厚度均匀性要求越来越严格，特别对于具有特性阻抗控制的高密度板和封装基板。

资料来源：华秋 PCB，国盛证券研究所

HDI 根据盲孔 PCB 层堆叠的次数可分为一阶、二阶、高阶、以及任意阶 HDI。一阶 HDI 通常表述其层数结构为“1+N+1”。N 代表内层板层数，内层板由芯板和铜箔压合而成；1 代表在外层电路的第一层通过激光钻孔实现盲孔互联，也即一阶 HDI。同理，二阶 HDI 层数结构为“2+N+2”，是在一阶 HDI 的基础上，再增加一层盲孔层，8 层二阶 HDI 也可表述为“2+4+2”的结构。通常阶数越高，线路的布线密度越高，电路板的功能越复杂。

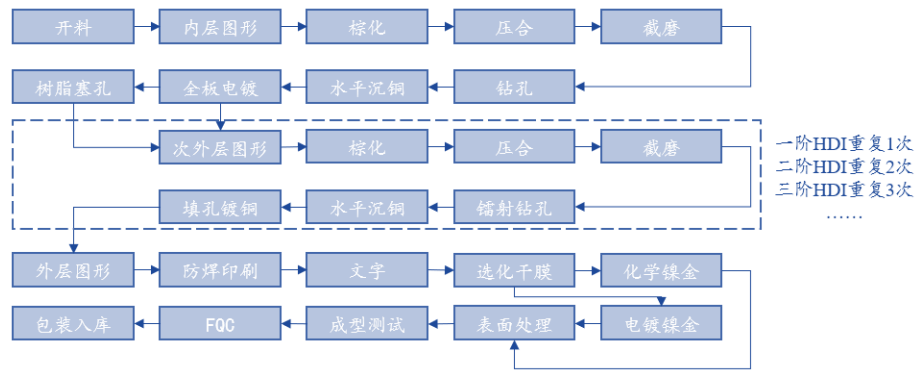
图表149: 各阶 HDI 示意 (红色为电路层; 黑色为通孔、盲孔、埋孔)



资料来源：深泽多层电路，国盛证券研究所

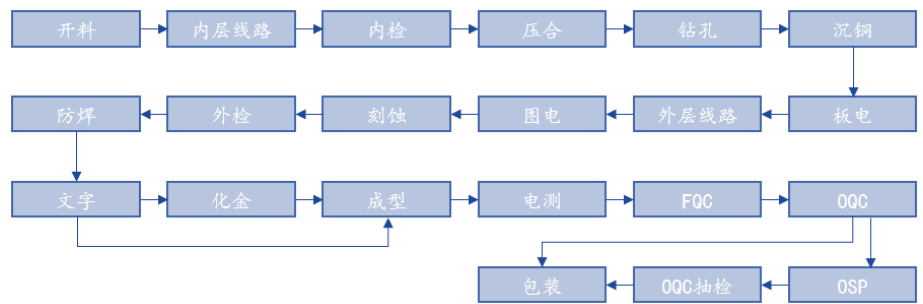
从具体生产流程来看，高阶 HDI 产品对加工产能的消耗显著增加。多层板的生产流程较 HDI 更短，主要是多层 PCB 压合以及机械钻孔等相关的后续工序。HDI 对比多层板，主要区别在内层板制作完成后，制造次外层所增加的压合、减铜、镭射等工序，这些次外层加工环节的重复次数即对应 HDI 的阶数，如压合、减铜、镭射等工序走完一遍即为一阶，因此同等面积的二阶 HDI 板产品相比一阶 HDI 板产品，在次外层加工环节所消耗的产能要增加一倍，三阶或任意阶 HDI 需要的产能为一阶的三倍以上。

图表150: HDI 生产工艺流程



资料来源：华经产业研究院，国盛证券研究所

图表151: 多层板生产工艺流程



资料来源：EETChina，国盛证券研究所

HDI 对比高多层 PCB 有效降低 PCB 层数增加布线密度。HDI 以埋盲孔的形式实现灵活布线，提高布线密度，对应降低 PCB 的必要层数和使用面积。从下图可知，18 层的高多层通孔板可以通过 10 层 1 阶的埋盲孔 HDI 实现，对应 18 层通孔板的布线密度为 100 (p/si，每平方英寸引脚数)，而对应 10 层 1 阶的埋盲孔 HDI 的布线密度为 210 p/si，HDI 对比多层板布线密度提升 110%。

图表152: 通孔板和 HDI 板的层数密度对照图

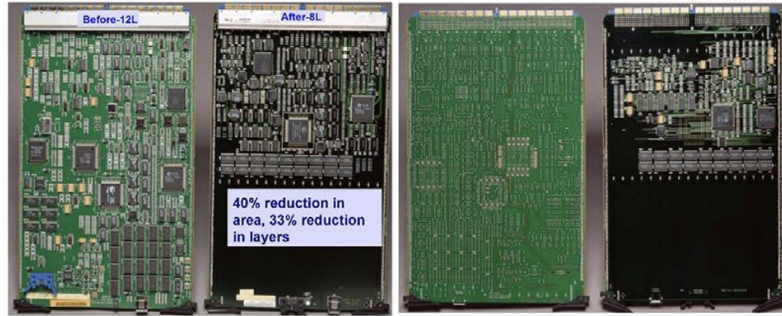
N Layers	A THRU-HOLE		B HDI BLIND		C HDI BL/BU		D 1BU BLIND		E 2BU BLIND		F 2BU BL/BU		G 2BU BL/BU	
	N		1+N+1		1+bN+1		1+N+1		2+N+2		2+bN+2		2+bN+2	
	blind via*	buried via	L1-L2	none	L1-L2	L2-L(N-1)	skip via L1-L3	L2-L(N-1)	staggered L1-L2, L2-L3	none	skip via L1-L3	L2-L(N-1)	staggered L1-L2, L2-L3	L3-L(N-2)
	RCI	DEN	RCI	DEN	RCI	DEN	RCI	DEN	RCI	DEN	RCI	DEN	RCI	DEN
4L	0.62	20	0.83	40	1.05	80	1.26	135	1.38	135	--	--	--	--
6L	0.78	20	0.99	60	1.24	160	1.46	200	1.60	200	1.74	260	1.91	280
8L	1.00	30	1.21	120	1.49	180	1.74	240	1.90	240	2.06	300	2.25	320
10L	1.30	40	1.51	200	1.83	210	2.11	260	2.30	260	2.50	400	2.73	440
12L	1.70	60	1.92	210	2.31	230	2.62	300	2.85	300	3.10	600	3.37	650
14L	2.24	70	2.48	220	2.95	250	3.32	360	3.61	360	3.91	800	4.25	860
16L	2.97	80	3.22	260	3.81	300	4.25	420	4.61	420	5.00	1000	5.43	1100
18L	3.92	100	4.21	300	4.95	400	5.47	480	5.93	480	6.42	1250	6.96	1350
20L	5.14	105	5.48	360	6.41	500	7.04	600	7.62	600	8.23	800	8.90	900
22L	6.67	110	7.08	400	8.23	600	8.99	700	9.70	700	10.45	1000	11.27	1100
24L	8.53	125	9.03	460	10.41	700	11.32	800	12.17	800	13.07	1200	14.04	1300
26L	10.68	130	11.30	500	12.92	800	13.96	900	14.96	900	16.00	1400	17.11	1500
28L	13.09	135	13.83	540	15.65	900	16.82	1000	17.93	1000	19.09	1600	20.32	1700
30L	15.63	140	16.50	580	18.47	1000	19.73	1100	20.94	1100	22.18	1800	23.48	1900
32L	18.17	145	19.17	620	21.21	1100	22.53	1200	23.79	1200	--	--	--	--
34L	20.59	150	21.69	660	23.73	1200	25.09	--	--	--	--	--	--	--
36L	22.79	160	23.96	700	25.94	--	--	--	--	--	--	--	--	--
38L	24.68	180	25.91	740	--	--	--	--	--	--	--	--	--	--
40L	26.26	200	--	--	--	--	--	--	--	--	--	--	--	--

资料来源：《haihandbook》，国盛证券研究所

以下图为例进一步说明，对比 12 层通孔板，8 层 HDI 方案在减少 33% 层数的基础上，同时减少了 40% 的 PCB 面积，可以有效降低 PCB 生产成本，同时提供了更高的互联密度。

伴随芯片性能的持续提升，封装集成度要求越来越高。HDI 线路板能够实现电路、元器件、材料、设计以及生产过程五大关键变量的平衡，因而在高速信号及高集成度领域前景广阔。

图表153: 实现相同功能前提下 HDI 工艺 8 层板减少 40% 面积同时降低 33% 层数



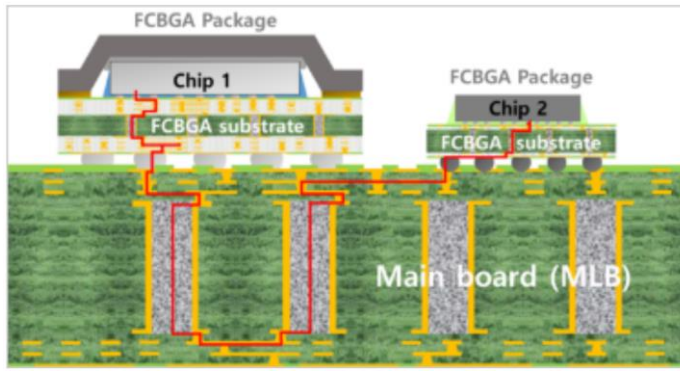
资料来源:《hdlhandbook》, 国盛证券研究所

HDI 线路板一方面基于其高密度布线的优势，可以在有限的面积上更好承接高速互联的多组 I/O 线路；同时由于 HDI 基于盲孔可以实现不同层之间的灵活布线，因此可以有效缩减信号在线路板上传输的距离，进而提高信号传输速率，减少传输时延。

BGA 间距定义为从中心到中心测量的两个相邻焊球之间的距离。随着芯片引脚数量的增加，焊盘之间的间距会减小。这种间距的减小使得 BGA 扇出变得复杂，一般来说，当 BGA pitch 间距小于或等于 0.5mm 的状态下，此时 BGA 便难以扇出打孔，因此需要添加更多的内层布线来满足要求。HDI 的激光钻孔孔径一般在 3-5mil (0.076-0.127mm) 之间，线路宽度一般为 3-4mil (0.076-0.10mm)。其较小的过孔孔径可以大幅增加过孔密度，以及扇出布线的灵活性，从而满足 AI 产品领域极高的 BGA 封装数量要求，同时，通过将盲孔打在焊盘上，直接实现芯片引脚与内部线路的连接，可以有效增加布线效率并减少信号走线距离，实现更高速的信号传输。与之对应的，通孔板过孔孔径较大，通常当孔径达到 0.15mm 时，成本显著上升，且难以进一步提升。因此机械钻孔的通孔板无法在有限的面积内大规模打孔，且无法实现有效的布线扇出，因而无法直接满足 AI 芯片领域高密度互联的要求。

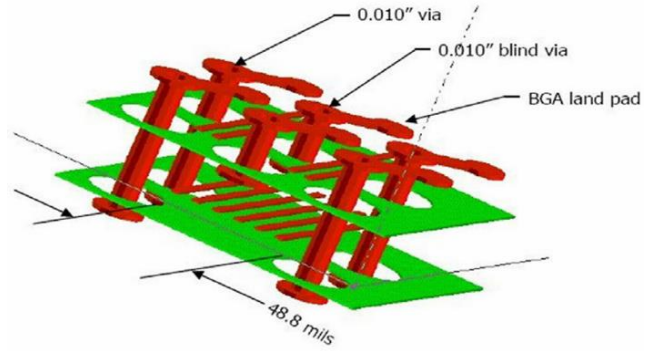
从布线效率的角度来看，通常 HDI 微盲孔的孔径要远小于通孔的孔径，同时 HDI 非通孔的设计也节省了 PCB 内层的空间，因此可以大幅提升布线效率以及布线密度。以下图为例，可以直观看到内部空间盲孔下层可以仍可以实现布线，在内层增加了除 BGA 以外 30% 的布线。而通孔则由于一次全部导通，即使在某一层不需要连接，也强制占用了对应的内部空间，且孔径较大，布线效率相对更低。

图表154: 从芯片到 PCB 互联示意图



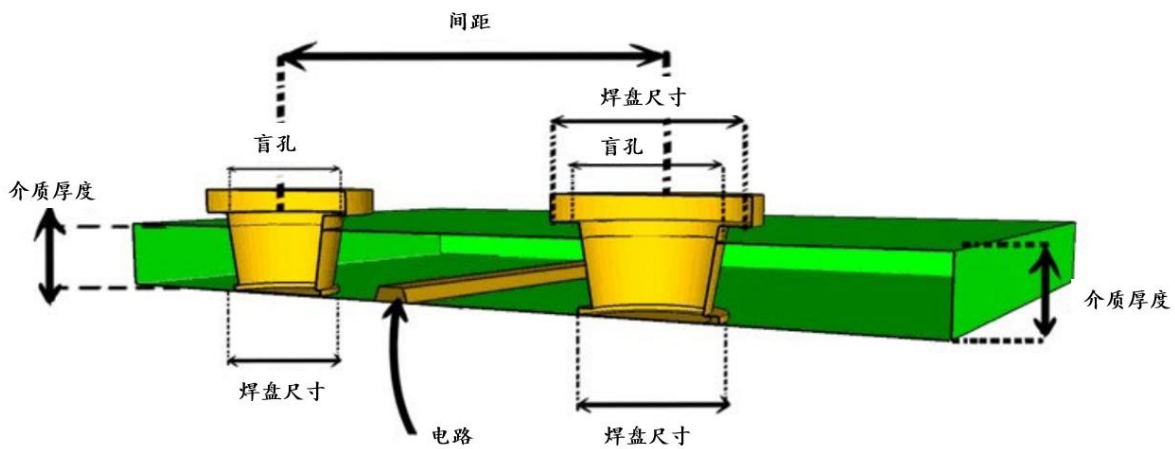
资料来源: daeduck, 国盛证券研究所

图表155: 盲孔在内层增加除 BGA 以外 30%的布线



资料来源: 《hdihandbook》, 国盛证券研究所

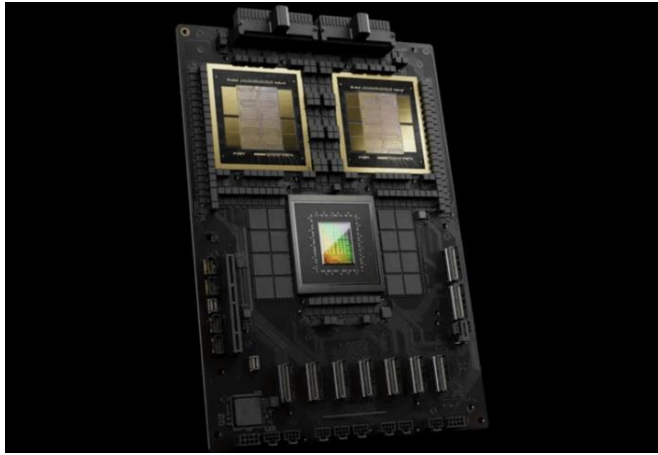
图表156: 芯片间距及焊盘尺寸的示意图



资料来源: 百芯 EMA, 国盛证券研究所

以 GB200 为例, 从集成和高速互联的角度来看, HDI 高密度互联的优势助力实现 CPU 与 GPU 的统一集成。八卡架构的 AI 服务器中 CPU 和 GPU 分属于不同模块, 从集成度、互联、功耗等角度来看会带来一些低效。英伟达推动算力服务器架构的整合与升级, 在 GB200 Superchip 上集成了两颗 Blackwell GPU 和一颗 Grace CPU。由于直接集成多颗高性能芯片, 并基于 NVLink 等进行高速互联, 高频高速以及高集成度需求下 HDI 是最优解, 推动 Compute Board 采用高阶 HDI 方案, CPU 和 GPU 互联传输速率的大幅提升; CPU 和 GPU 功耗及散热得到统一管理, 并且集成度大幅提升, 也推动了整机解决方案的成本下降。

图表157: GB200 集成一颗 CPU 和两颗 B200 GPU



资料来源: Fibermall, 国盛证券研究所

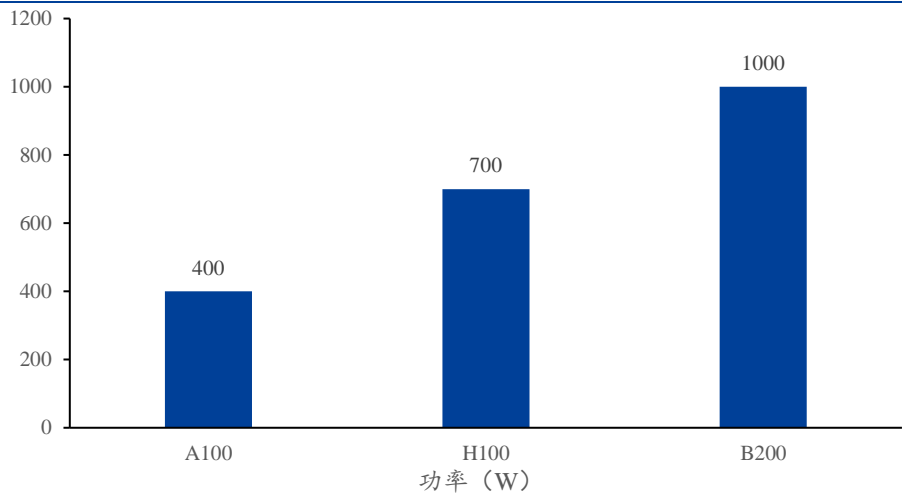
图表158: GB200 CPU 和 GPU 之间基于 C2C 互联



资料来源: cnbeta, 国盛证券研究所

从功耗的角度来看，HDI 基于其高集成度优势降低冗余，减少额外功耗消耗。一方面前文已述及，英伟达推动 CPU 和 GPU 集成在同一片 HDI 电路板上，CPU 和 GPU 的电源功率得到统一管理，电源管理效率得到显著提升。同时 CPU 和 GPU 模块的融合也大幅降低了整机系统其他部件的冗余度，由此降低冗余部件带来的额外功耗消耗。

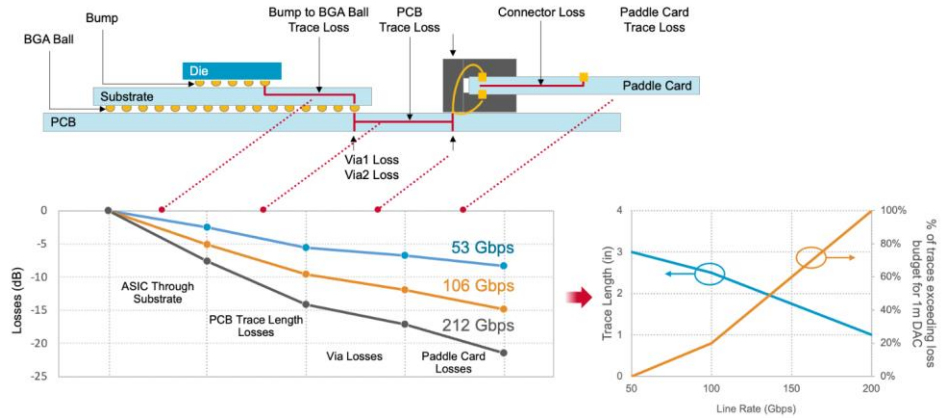
图表159: 英伟达不同 AI 芯片功耗



资料来源: 纵横云计算, 科普中国, 电子工程专辑, 国盛证券研究所

HDI 高密度短距离布线可以大幅降低信号传输过程中的功耗，同时散热具备显著优势。得益于 Compute Board 实现了高集成度的器件布局，各主被动元器件之间的信号传输距离相对更短，由于信号传输速率越高，信号在传输过程中的能量损耗也越大，因此 HDI 的短距离传输优势可以显著降低传输功耗，并且提升信号传输质量。而从散热的角度来看，HDI 高密度的铜互联线路以及相对通孔板显著的层数减少也可以更有效实现散热，确保内部信号不会因过热产生稳定性的问题。

图表160: 传输速率越快距离越长, PCB的路径损耗越大



资料来源: 博通, 国盛证券研究所

总结来看, HDI 的发展趋势出现了新的变化。上一轮 HDI 产业趋势, 核心升级点在于 HDI 的“轻薄短小”特征满足了智能手机为代表的终端集成度提升的核心诉求, 带动了消费类 HDI 需求的高速增长。而本轮 HDI 的产业趋势, 核心升级点在于 HDI “高密度布线”特征满足了算力硬件高速互联的核心诉求, 且在信号完整性、集成度、成本、散热等多方面具备优势, 由此带来 AI HDI 需求的持续增长。并且在云端算力向边缘算力下沉的过程中, 我们看到 HDI 的产业趋势也从云端向边缘侧 AIPC、智能汽车等产品加速渗透, HDI 有望成为 AI 时代增长最快速的 PCB 产品。从实际生产来看, 由于产品规格更高, 消费类 HDI 产线无法共线生产 AI HDI 产品, 而 GB200 的 HDI 板较常规 HDI 板面积更大, 层数阶数更高, 因此实际生产过程中无论是产能还是良率均显著降低, 产能十分稀缺, 全球范围内具备大批量量产能力的厂商有望率先受益于本轮 AI HDI 浪潮。

5.4 胜宏科技: AI HDI 升级核心受益, 加速成长区间已至

业绩持续新高, 加速成长区间已至。2024 年前三季度公司实现营收 76.98 亿元, 同比+34.0%; 实现归母净利润 7.65 亿元, 同比+30.5%; 实现扣非归母净利润 7.78 亿元, 同比+33.2%。2024 年前三季度公司实现毛利率 21.6%, 同比-0.49pcts; 实现净利率 9.93%, 同比-0.27pcts。分季度来看, 公司 2024Q3 实现营收 28.42 亿元, 同比+37.1%, 环比+15.36%; 实现归母净利润 3.1 亿元, 同比+26.70%, 环比+22.52%; 实现扣非归母净利润 3.2 亿元, 同比+36.54%, 环比+23.98%。2024Q3 公司实现毛利率 23.17%, 同比-0.77pcts, 环比+1.48pcts。24Q3 公司实现净利率 10.75%, 同比-0.88pcts, 环比+0.63pcts。

AI 布局多点开花, 工艺能力提升带动良率提升。公司针对 AI 算力、AI 服务器产品, 下一代 PCIe 6.0 传输协议与芯片, Oak stream 平台技术; 800G/1.6T 光传输在光模块与交换机上单通道 112G & 224G 的传输技术, 下一代 6G 通讯技术, L3/L4 级自动驾驶等多个高端技术领域进行研发突破, 并顺利落地到产品应用。在生产工艺环节, 公司持续针对 AI 服务器、AI 算力、光传输交换机等产品的技术能力改造, 完成对高多层精密 HDI 5.0 mm 和高多层 PCB 8.0 mm 厚板的设备优化与改造, 大孔径盲孔填孔能力与超薄芯板能力建制, 为下一代 AI 服务器、算力、通讯产品研发及生产打下坚实基础。

胜宏科技抢占 HDI 制高点, 把握机遇快速导入。AI 芯片高速高密度互联催生高阶 HDI 产业趋势, 胜宏科技是全球领先的高阶 HDI 厂商, 有望受益于 AI HDI 升级浪潮。2024 年以来公司在 AI 领域全面突破, 多款数据中心 AI 算力 HDI 产品实现大批量量产。同时公司加强 AI 相关产品布局研发, 产品线持续拓宽, 在交换机、AI 服务器、高端通信领域

与多家全球顶级客户深度合作。伴随客户合作深化及海外产能释放，公司 AI 业务有望全面开花，助力公司打开新成长曲线。

风险提示：下游需求不及预期，研发进度不及预期，行业竞争加剧。

5.5 沪电股份：业绩持续新高，前瞻布局强者恒强

沪电股份 2024 年前三季度实现营收 90.1 亿元，同比+48.15%，实现归母净利润 18.5 亿元，同比+93.94%，实现扣非归母净利润 18.1 亿元，同比+105.8%。2024 前三季度公司实现毛利率 35.86%，同比+5.29pcts，实现净利率 20.31%，同比+4.89pcts。

沪电股份加码研发，AI 算力&网络 PCB 双线受益。公司基于 PCIe 6.0 的下一代通用服务器产品已开始技术认证；800G 交换机 PCB 已出货，且支持 224Gbps 速率（102.4T 交换容量 1.6T 交换机）产品主要技术已完成预研；OAM/UBB2.0 产品已批量出货；GPU 类产品已通过 6 阶 HDI 认证准备量产。公司持续进行前瞻技术布局，长期竞争力持续强化。

加码 AI 资本开支，彰显长期发展信心。公司 2024 年 10 月 24 日公告，启动新的人工智能芯片配套高端印制电路板扩产项目。该项目预计总投资约为 43 亿元人民币，对应年产约 29 万平方米的高端印制电路板，以满足持续增长的 AI 需求。项目全面完成后，预计将新增年营收约 48 亿元人民币。此次扩产充分展现了管理层对人工智能行业长期需求以及公司自身竞争力的坚定信心。

风险提示：下游需求不及预期，研发进度不及预期，行业竞争加剧。

5.6 深南电路：AI 推动产品结构持续升级

AI 驱动业绩新高。公司 2024Q1-Q3 实现营收 130.49 亿元，同比+37.92%；实现归母净利润 14.88 亿元，同比+63.86%，实现扣非归母净利润 13.76 亿元，同比+86.67%。其中 2024Q3 公司实现营收 47.28 亿元，同比+37.95%，环比+8.45%。实现归母净利润 5.01 亿元，同比+15.33%，环比-17.60%。实现扣非归母净利润 4.72 亿元，同比+51.53%，环比-16.98%。

封装基板产能爬坡进展顺利，构筑未来成长动力。公司广州封装基板一期已于 2023 年 Q4 连线，2024 年产线能力持续提升，产能爬坡处于前期阶段，主要聚焦平台建设与客户各阶产品认证。另一方面，公司在南通基地尚有土地储备，具备新厂房建设条件，南通四期项目已有序推进基建工程，拟建设为具备覆盖 HDI 等能力的 PCB 工艺技术平台。

AI 推动产品结构升级。在通信领域，2024 年上半年，受人工智能相关需求推动，400G 及以上高速交换机和光模块的需求呈现出显著增长态势，这不仅促使有线通信产品在公司整体业务中的占比提升，还进一步优化了产品结构，使公司的盈利能力得到了显著增强。在数据中心领域，2024 年上半年公司收获了同比显著增长的订单量，这一成绩主要得益于 AI 加速卡、Eagle Stream 平台产品等的持续放量，这些产品需求的提升为公司数据中心业务的增长提供了有力支撑。

公司是国内数通 PCB 领先企业，将持续受益于国内及海外 AI 算力发展热潮。AI 带动的交换机、AI 服务器、光模块等产品的旺盛需求，将持续驱动公司产品结构升级。

风险提示：下游需求不及预期，研发进度不及预期，行业竞争加剧。

5.7 景旺电子：各项业务有序推进，高端产能释放引领新增长

业绩稳健增长。公司 2024Q1-Q3 实现营收 90.78 亿元，同比+17.11%；实现归母净利润 9.04 亿元，同比+29.14%，实现扣非归母净利润 8.28 亿元，同比+20.30%；实现毛利率 23.77%，同比-0.85pcts；实现净利率 9.87%，同比+0.99pcts。其中 2024Q3 单季度实现营收 32.11 亿元，同比+15.06%，环比+2.76%；2024Q3 公司实现净利润 2.48 亿元，同比-16.38%，环比-26.80%。

持续加强研发投入，数通产品布局进展显著。通用服务器领域，公司 EGS/Genoa 平台高速 PCB 已稳定量产，同时在 Birth stream 平台高速 PCB 等产品技术上取得重大突破。与 EGS/Genoa 平台相比，Birth stream 平台对信号传输速率、数据传输损耗、布线密度等方面的要求提升，PCB 性能持续升级，Birth stream 平台高速 PCB 技术突破有助于公司进一步拓展服务器产品，抢抓高端数据中心市场机遇。AI 服务器领域，公司成功开拓了软板和软硬结合板在数据中心的应用产品，同时在高阶 HDI、高多层 PTFE 板等产品上实现了重大突破。高速通信领域，公司 800G 光模块、通信模组高阶 HDI 等产品实现批量出货，112G 交换路由 PCB 取得重大技术突破。

产能有序释放，高端产能释放引领未来增长。珠海景旺 HLC、HDI（含 SLP）工厂各项业务持续稳步推进，引进了下游行业的头部客户，面向客户的高端产品开发取得新的突破性进展，产量产值稳步提升，呈现出较好的竞争力；江西景旺三期已顺利投产，产能不断爬坡，为公司业绩带来新增量；信丰高多层电路板生产项目稳步推进，致力于打造自动化智能化程度高、生产效率高、成本优势突出的高多层电路板研发生产基地；泰国生产基地按计划逐步建设，项目已完成土地购买、子公司增资等工作。

风险提示：下游需求不及预期，研发进度不及预期，行业竞争加剧。

6 消费电子：端侧 AI 多终端落地，新一轮创新周期开启

6.1 手机：景气度持续回升，AI 赋能开启换机潮

全球智能手机市场连续四季度同比反弹，景气度逐步回暖。根据 Canalsys，我们看到 24Q3 全球智能手机出货量达到 3.1 亿台，同比增长 5%，是自 2021 年以来出货量最强劲的第三季度，且已连续四个季度保持正增长，预计 2024 全年出货量达 12.2 亿台，同比增长 6%；24Q3 国内智能手机市场出货量同比增长 4%，尽管增速较上一季度的 10% 有所放缓，但整体表现出稳定的回暖趋势。IDC 预计，2025 年中国智能手机市场出货量将达到 2.89 亿，同比增长 1.6%。整体来看，我们认为随着 AI 端侧落地开启新一轮换机周期，以及在政府补贴政策刺激消费的影响下，智能手机市场需求将进一步提升。

图表161: 2021Q1-2024Q3 全球智能手机出货量



资料来源: Canalsys, 国盛证券研究所

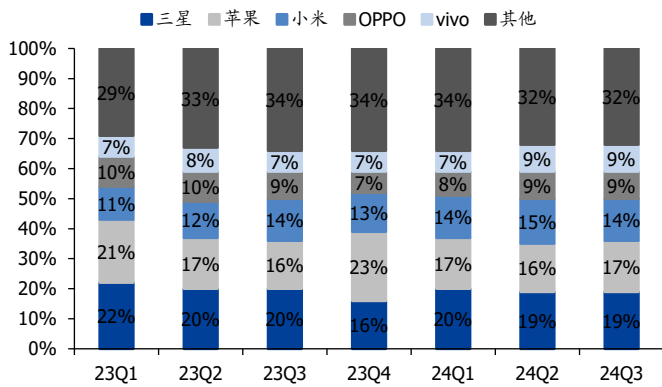
图表162: 2022Q1-2024Q3 中国智能手机出货量



资料来源: Canalsys, 国盛证券研究所

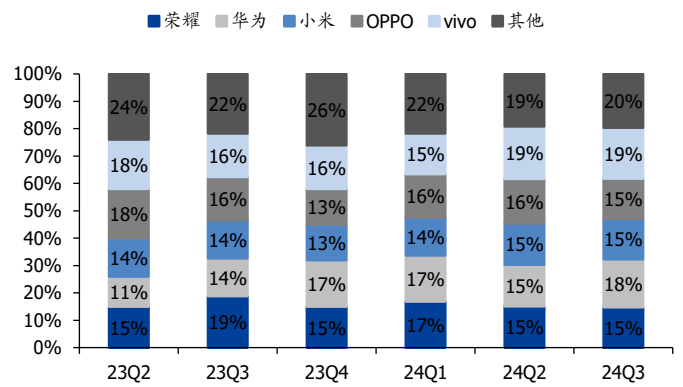
根据 Counterpoint Research, 23Q1-24Q2 全球前五大手机厂商格局较稳定, 分别为三星、苹果、小米、OPPO、vivo, 24Q3 对应份额分别为 19%、17%、14%、9%、9%。从中国智能手机市场份额来看, 华为的市场份额从 23Q2 的 11% 逐步攀升至 24Q3 的 18%, 得益于 Pura 70 和 Mate 60 系列的推动; 24Q3 vivo 以 19% 的份额稳居榜首, 其产品组合丰富, 涵盖所有价格区间。

图表163: 2023Q1-2024Q3 全球智能手机厂商市场份额



资料来源: Counterpoint Research, 国盛证券研究所

图表164: 2023Q2-2024Q3 中国智能手机厂商市场份额



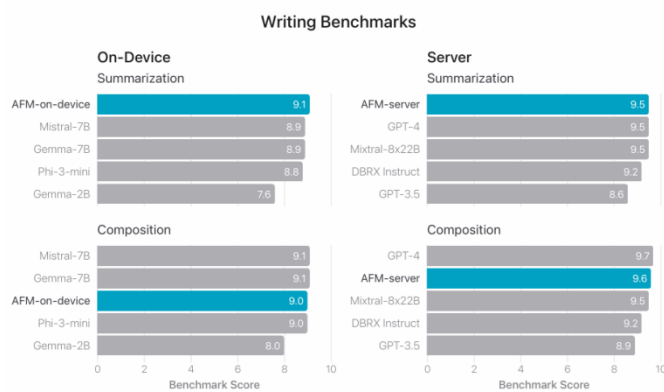
资料来源: Counterpoint Research, 国盛证券研究所

6.1.1 OpenAI 12 天密集发布，豆包树立国产 AI 标杆

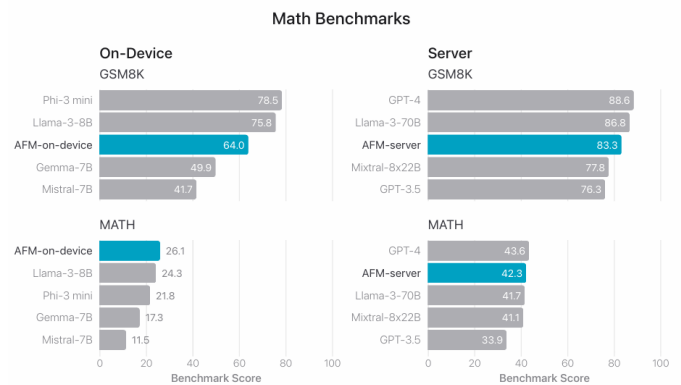
➤ 苹果 AFM 模型测评结果好，创新降低硬件依赖

苹果论文披露 **AFM 模型多维度测评结果**。Apple Intelligence 上的两种基础语言模型包括 AFM-on-device 和 AFM-server，AFM-on-device 为在设备端运行的大约 30 亿参数的语言模型，AFM-server 为在私有云计算上运行的大规模服务器语言模型。根据该论文，在预训练评估中，苹果开发的端侧大模型在语言理解、指令跟随、推理、写作与工具使用等多个任务上表现出色；在后训练评估中，1) 苹果的 AFM 模型在指令遵循层面皆优于其他大模型；2) 在摘要总结方面 AFM 模型无论是端侧还是私有云均好于其他；3) 在安全性评估时，AFM 模型也比其他模型要更为负责；4) 但 AFM 模型的数学能力整体上来看较为一般。

图表165: 写作能力对比



图表166: 数学能力对比



资料来源:《Apple Intelligence Foundation Language Models》, Apple, 国盛证券研究所

资料来源:《Apple Intelligence Foundation Language Models》, Apple, 国盛证券研究所

➤ **OpenAI 12 天: 全方位重塑 AI 体验的技术盛宴**

在 12 天的密集发布中，OpenAI 全面展现了技术创新与应用落地的强大能力。从第一天的完整版 o1 和 ChatGPT Pro 模型发布，到最后一天的 o3 全新模型亮相，OpenAI 完成了一次次技术突破。强化微调技术的推出为定制化 AI 助手提供了高效方案；Sora Turbo 是目前全球生成时长最高的视频模型之一；ChatGPT Canvas 内置在 ChatGPT 中，提供一个功能强大的协作平台。此外，OpenAI 与苹果的 iPhone、iPad、Mac 和 Siri 实现深度集成，可以原生使用 ChatGPT。这 12 天不仅展示了 OpenAI 在技术、功能和体验上的进步，也为未来 AI 技术普及和应用创新奠定了坚实基础。

图表167: OpenAI 12 天直播内容总结

日期	内容
第一天	完整版 o1 和 ChatGPT Pro 发布, 提供更强大的数学、代码和科学问题解决能力。ChatGPT Pro 订阅费用 200 美元/月。
第二天	发布强化微调技术, 少量数据即可定制领域 AI 助手, 提升特定任务准确性。
第三天	发布 Sora Turbo 视频生成模型, 可生成 20 秒 1080P 视频, 成为全球生成时长最高的视频模型。
第四天	推出 ChatGPT Canvas, 作为独立协作平台, 用于代码编写、写作调整和任务执行。
第五天	苹果设备全面接入 ChatGPT, 实现与 iPhone、iPad、Mac 和 Siri 的深度集成。
第六天	发布高级语音和屏幕共享功能, 支持实时互动并新增圣诞老人语音模式。
第七天	推出 ChatGPT Projects, 整合文件上传、网络搜索和实时协作等多项功能。
第八天	发布 ChatGPT Search 功能, 提供实时搜索、语音对话和地图集成功能。
第九天	完整版 o1 模型 API 上线, 降低成本并支持 WebRTC, 简化实时 AI 开发流程。
第十天	新增 ChatGPT 电话功能, 通过 1-800 号码支持电话提问, 与 WhatsApp 集成。
第十一天	ChatGPT 与 MacOS 原生应用集成, 支持代码分析和开发环境高效协作。
第十二天	发布新模型 o3, 在 ARC AGI 测试中超过人类评测分数, 具备顶级数学能力。

资料来源: AIGC 开放社区, 国盛证券研究所

➤ 豆包大模型: 国产 AI 标杆, 全方位领先与普及之路

豆包通用模型 Pro 主观评测排名第一, 展现卓越中文能力。豆包通用模型 Pro (Doubao-pro-32k-preview) 在智源研究院 FlagEval 的大语言模型主观评测榜单中以 77.75 分位列榜首, 成为表现最突出的模型。凭借知识运用 (91.14 分)、简易理解和推理能力的优异表现, 豆包全面领先国内外多款开源与闭源模型, 展现了其在中文理解和应用上的卓越能力, 成为国产大模型中的顶尖代表。

多模态模型评测榜单国内第一, 豆包展现视觉语言综合实力。豆包·视觉理解模型 (Doubao-Pro-Vision-32k-241028) 在多模态模型评测榜单中排名第二, 仅次于 GPT-4o, 是得分最高的国产模型。豆包在图文理解、复杂信息处理等能力上表现突出, 得分达到 4.43, 成为多模态领域的标杆模型。这一成绩充分体现了豆包在视觉与语言融合任务中的技术领先性, 适配了更多实际应用场景。

图表168: 豆包通用模型 pro 在主观评测中排名第一

排名	模型名称	模型版本	评测方式	综合得分	知识运用	简易理解	推理能力	数学能力	任务解决
1	Doubao-pro-32k	Doubao-pro-32k-preview	主观评测	77.75	86.48	77.72	91.14	73.81	73.78
2	ERNIE 4.0 Turbo	ERNIE-4.0-Turbo-06-Preview	主观评测	77.26	84.22	77.23	83.54	85.48	77.76
3	o1-preview	o1-preview-2024-09-12	主观评测	76.53	91.63	76.73	84.71	83.38	62.11
4	o1-mini	o1-mini-2024-09-12	主观评测	76.07	91.06	76.79	81.01	88.35	81.26
5	Claude 3.5 Sonnet	claude-3.5-sonnet-20241022	主观评测	74.44	87.35	77.57	80.00	71.43	62.32
6	GPT-4o	gpt-4o-2024-11-20	主观评测	73.91	87.74	79.70	86.71	70.73	67.10
7	Qwen-Max	qwen-max-0919	主观评测	71.90	83.47	76.73	85.44	72.62	66.63

资料来源: 智源研究院, 国盛证券研究所

图表169: 豆包视觉理解模型在视觉语言模型中排名第二

排名	模型名称	模型版本	评测方式	综合得分	中文能力	英文能力	通用能力	数学
1	GPT-4o-20241120	GPT-4o	主观评测	4.00	6.50	3.00	57.21	30.8
2	Doubao-Pro-Vision-32k-241028	主观评测	4.43	2.00	5.40	55.47	40.5	
3	Claude-3.5-Sonnet-20241022	主观评测	4.43	8.50	2.80	57.71	45.2	
4	Qwen2-VL-72B-Instruct	主观评测	73.48	3.00	6.40	52.67	32.5	
5	GPT-4o-20240806	主观评测	6.00	8.50	5.00	53.71	29.8	
6	Qwen-VL-Max	主观评测	6.00	6.00	6.00	50.72	33.9	
7	Gemini-1.5-Pro	主观评测	6.71	14.50	3.60	54.48	50.8	

资料来源: 智源研究院, 国盛证券研究所

大模型角斗场榜单中跻身第一梯队, 用户偏好国产标杆。豆包通用模型 Pro (Doubao-pro-32k-240828) 在智源研究院的 FlagEval 大模型角斗场榜单中排名第二, 仅次于 OpenAI 的 o1-mini, 位列第一梯队。豆包凭借强大的用户偏好评分和实际对战表现, 成为用户高度认可的国产模型代表。这一榜单反映了豆包在应用落地和用户需求满足上的出色表现。

图表170: 大模型角斗场榜单-大语言模型榜单

模型等级	评分排名	模型	竞技场评分	95%置信区间	有效投票数
1	1	o1-mini-2024-09-12	1149.16	+11.62/-13.04	3207
1	2	Doubao-pro-32k-240828	1135.29	+10.09/-10.70	3092
1	3	Nanbeige2-Turbo-0611	1132.28	+11.26/-11.53	3494
1	4	GLM-4-Plus	1124.81	+14.28/-11.28	2103
2	5	Yi-Lightning	1112.56	+10.84/-13.21	2306
2	6	DeepSeek-V2.5	1094.9	+11.63/-11.72	4344
2	7	Hunyuan-Turbo	1090.5	+12.61/-12.73	2091

资料来源: 智源研究院, 国盛证券研究所

图表171: 豆包通用模型能力全面升级



资料来源: 智源研究院, 国盛证券研究所

豆包能力全面提升, 多维度表现远超以往。豆包通用模型 Pro 在 2024 年完成全方位升级, 综合能力提升 32%, 各项指标均显著增强。推理能力提升 13%, 指令遵循能力提升 9%, 代码能力提升 58%, 数学能力提升 43%, 专业知识能力提升 54%。这些升级显著提高了豆包在高阶任务和复杂问题解决上的能力, 进一步夯实了其在多场景应用中的技术优势。

价格策略领先行业, 豆包以低成本推动 AI 普及。豆包通过极具竞争力的价格策略, 每千 tokens 仅 0.003 元, 比行业平均价格低 85%, 在市场中占据优势。例如, 1 元即可处理 284 张 720P 图片, 极大降低了多模态模型的使用门槛。豆包的高性价比和强性能为企业和开发者提供了高效的解决方案, 加速了 AI 技术的普及与创新场景的落地。

6.1.2 手机为 AI 端侧最佳落地点, 换机周期加速

苹果 AI 时代开启, 软硬件全面布局。2024 年, 苹果全面开启 AI 时代, 通过软硬件的深度融合不断提升用户体验: 9 月, 苹果发布的 iPhone 16 Pro 系列搭载潜望长焦镜头与 48MP 超广角摄像头, 并配备 A18 Pro 芯片, 大幅提升拍摄与处理性能; 10 月, 苹果发布 iOS 18.1 正式版, 升级 Siri 自然语言处理能力, 并新增写作工具与通话录音功能, 这些 AI 功能依托 AFM 端侧与云端模型, 展现卓越性能; 12 月, 苹果推送了 iOS 18.2 正式版, 更新了 Apple Intelligence, 带来包括 AI 文字生成、AI 图片生成、AI 表情生成, 另外 Siri 还集成了 AI ChatGPT 等新功能, 有效增强 iPhone 的生产力。我们预计, iPhone 将提升摄像头像素, 为 AI 驱动的图像识别与 AR 应用提供更强支持。苹果通过软件和硬件的不断创新, 稳步推动 AI 技术普及, 进一步巩固其行业领先地位。

图表172: iOS18.2 重点更新 Apple Intelligence



资料来源: 芝麻科技讯, 国盛证券研究所

安卓 AI 手机: vivo、OPPO、荣耀、华为、小米等各大手机厂商纷纷发布了 AI 大模型, 向生成式 AI 手机进化。2024 年 2 月, 魅族决定 All in AI, 停止传统智能手机新项目; OPPO 宣布正式进入 AI 手机时代, 在除夕当天将超过 100 项 AI 实用功能推送至 10 余款手机系列; 华为宣布 HarmonyOS NEXT 将 AI 与 OS 深度融合, 构筑了全新的鸿蒙原生智能框架; 小米宣布小爱大模型将迎来全量升级, 全部免费。

图表173: 安卓 AI 布局

品牌	AI 大模型	参数量级	智能助手
华为	盘古大模型	十亿级-万亿级	小艺
荣耀	魔法大模型	70亿	YOYO助手
小米	MiLM	13亿	小爱同学
OPPO	AndesGPT	70亿	小布助手
Vivo	蓝心大模型	70亿	蓝心小V

资料来源: 电子工程专辑, 量子位, 半导体产业纵横, 国盛证券研究所

苹果在全球 AI 手机市场中占据绝对领先地位。根据 Canalsy 数据, 2024 年第三季度, 苹果以 3210 万台的出货量和 61% 的市场份额领先; 三星位列第二, 出货量为 1300 万台, 市场份额为 25%; 小米排名第三, 出货量为 360 万台, 占全球市场的 7%。中国大陆 AI 手机市场, 苹果以 800 万台的出货量占据 52% 的国内市场份额; 小米以 320 万台的出货量 (21% 的市场份额) 排名第二; vivo 以 220 万台的出货量占据 14% 的市场份额, 位列第三。

图表174: 2024Q3 全球 AI 手机型号榜单



图表175: 2024Q3 中国大陆 AI 手机型号榜单

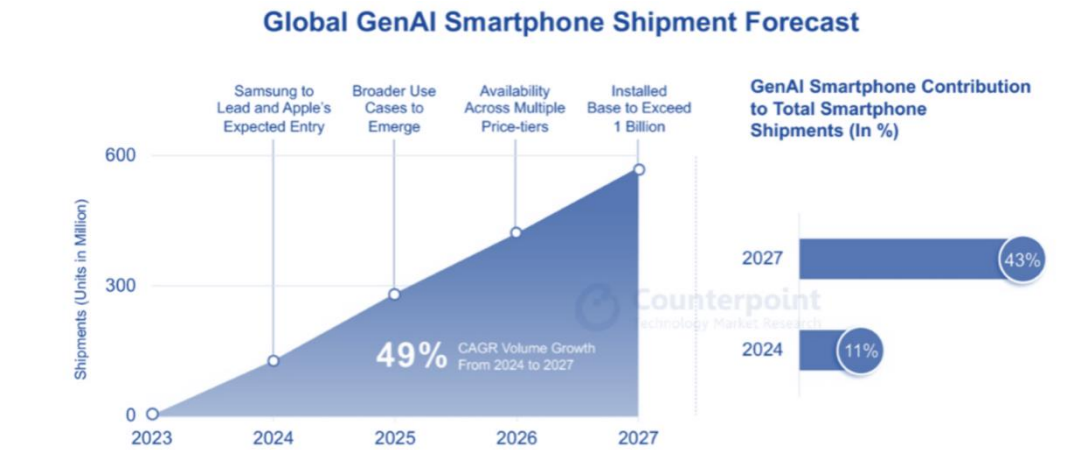


资料来源: Canalis, 国盛证券研究所

资料来源: Canalis, 国盛证券研究所

AI 端侧落地开启新一轮换机潮，预计 2027 年全球 AI 手机渗透率达 43%。今年下半年以来，随着安卓厂商第二代 AI 旗舰手机陆续推出和模型算法的迭代，端侧小模型的运行效果已有长足进步，构建开放的 AI 服务生态体系已成为众多安卓厂商下一阶段 AI 战略重心。随着行业领头玩家的相继入局，将说服并吸引更多开发者为移动端开发更多 AI 应用与服务，进一步完善目前初具雏形的手机 AI Agent 应用场景。根据 Counterpoint Research 的预测，2024 年全球生成式人工智能 (AI) 手机渗透率将达到 11%，并随着更多次旗舰以及中高端机型将配备更强大的端侧 AI 能力，进一步增长到 43%，届时出货量将超过 5.5 亿台，增长近 4 倍。根据 IDC 的预测，2027 年中国 AI 手机出货量将达到 1.5 亿台，渗透率超过 50%。

图表176: 生成式 AI 手机渗透率预测



资料来源: 电子工程专辑, 国盛证券研究所

6.2 PC: AI PC 技术突破，渗透率持续增长

全球 PC 出货量连续四个季度增长，市场逐步走出低谷企稳回升。2021 年，在远程办公和在线教育需求的推动下，PC 出货量达到历史高点，接近 5.2 亿台。23Q1 出货量降至低点，同比大幅下降超过 30%。随后市场开始回暖，到 24Q3，全球 PC 出货量恢复至约 6640 万台，同比小幅增长 1%。

分品牌来看：24Q3 联想以 1649 万台的出货量继续稳居第一，同比增长 2.8%，占比 24.8%；惠普出货量为 1357 万台，同比小幅增长 0.4%，占 20.4% 的市场份额；戴尔排名第三，出货量为 984.7 万台，占 14.8% 的市场份额；华硕出货量达到 551.3 万台，同比增长 15.8%，市场份额提升至 8.3%；苹果同比下降 17.5%，出货量为 511.4 万台，市场份额从 23Q3 的 9.5% 下降至 7.7%。整体来看，市场份额集中于联想、惠普与戴尔。

图表177: 2018Q3-2024Q3 全球 PC 出货量



资料来源: Canalsys, 芯智讯, 国盛证券研究所

图表178: 全球 PC 市场各品牌市场份额

全球台式机 and 笔记本出货量(市场份额和年增长率)					
Canalsys PC 分析统计数据: 2024 年第三季度					
厂商	2024 年	2024 年	2023 年	2023 年	年增长率
	第三季度	第三季度	第三季度	第三季度	
	出货量	市场份额	出货量	市场份额	
联想	16,490	24.8%	16,036	24.5%	2.8%
惠普	13,572	20.4%	13,513	20.6%	0.4%
戴尔	9,847	14.8%	10,255	15.6%	-4.0%
华硕	5,513	8.3%	4,762	7.3%	15.8%
苹果	5,114	7.7%	6,197	9.5%	-17.5%
其他	15,848	23.9%	14,796	22.6%	7.1%
合计	66,384	100.0%	65,557	100.0%	1.3%

注: 出货量单位为“万台”
由于四舍五入, 百分比可能无法达到 100%
来源: Canalsys PC 分析统计数据 (出货量), 2024 年 10 月

资料来源: Canalsys, 芯智讯, 国盛证券研究所

全球平板市场快速增长。2024 年第三季度全球平板电脑市场出货量达 3960 万台，同比增长 20.4%，呈现出强劲复苏的态势。苹果以 1260 万台的出货量稳居市场第一，市场份额达到 31.7%，但相比去年同期的 37.7% 有所下降；三星以 710 万台的出货量位列第二，同比增长 18.3%；亚马逊出货量从 2023 年的 220 万台激增至 460 万台，增长率高达 111.3%，市场份额从 6.6% 提升至 11.6%，得益于其低价策略和节日促销（如 Prime Day）的强大推动力；华为以 330 万台的出货量排名第四，同比增长 44.1%，市场份额从 6.9% 上升至 8.2%，体现出其在国内外市场的竞争力正在恢复；联想出货量为 300 万台，同比增长 14.7%，市场份额为 7.6%。

图表179: 24Q3 全球平板出货量情况

Top Five Tablet Companies, Worldwide Shipments, Market Share, and Year-Over-Year Growth, Q3 2024 (Preliminary results, combined company view for the current quarter only, shipments in millions)

Company	3Q24 Unit Shipments	3Q24 Market Share	3Q23 Unit Shipments	3Q23 Market Share	Year-Over-Year Growth
1. Apple	12.6	31.7%	12.4	37.7%	1.4%
2. Samsung	7.1	17.9%	6.0	18.2%	18.3%
3. Amazon.com	4.6	11.6%	2.2	6.6%	111.3%
4. Huawei	3.3	8.2%	2.3	6.9%	44.1%
5. Lenovo	3.0	7.6%	2.6	8.0%	14.7%
Others	9.1	22.9%	7.4	22.6%	21.8%
Total	39.6	100.0%	32.9	100.0%	20.4%

Source: IDC Worldwide Quarterly Personal Computing Device Tracker, November 1, 2024

资料来源: IDC 咨询, 国盛证券研究所

AI PC 出货量环比大幅增长，Windows 平台市场份额持续扩大。AI PC 市场需求持续扩大，进入快速发展阶段。根据 Canalsys 统计，2024 年第三季度全球 AI PC 出货量达到 1330 万台，占本季度 PC 总出货量的 20%，环比增长 49%，表现十分强劲。从操作系

统市场格局来看，Windows AI PC 的市场份额从 2024 年第二季度的 41% 进一步提升至 53%，环比增长 12 个百分点。

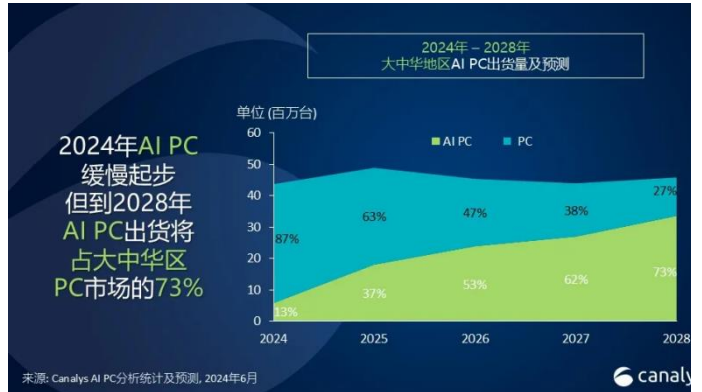
AI PC 加速渗透，在 2028 年占据 PC 市场主导地位。根据 Canalsy 的数据，2024 年 AI PC 在大中华地区渗透率为 13%，AI PC 的发展还处于起步阶段。但随着 AI 技术的普及和用户需求增加，AI PC 的市场份额将逐年攀升，预计 2025 年占比达到 37%，2026 年过半达到 53%，2027 年进一步升至 62%，并在 2028 年达到 73%。AI 技术的落地和功能扩展将在未来几年彻底改变 PC 市场格局，为 PC 厂商带来了新的增长机遇。

图表180: 2023Q4-2024Q3 全球 AI PC 出货量 (按操作系统划分)



资料来源: Canalsy, 国盛证券研究所

图表181: 2024年-2028年大中华地区 AI PC 出货量以及预测



资料来源: Canalsy, 国盛证券研究所

6.3 补贴扩大至 3C 产品，大幅提振换机需求

“两新”政策有效激发内需潜力，成效显著。2024年3月，国务院印发《推动大规模设备更新和消费品以旧换新行动方案》，对实施设备更新、消费品以旧换新、回收循环利用、标准提升四大行动等作出具体安排；7月，国家发展改革委、财政部联合印发《关于大力支持大规模设备更新和消费品以旧换新的若干措施》，统筹安排3000亿元左右超长期特别国债资金，加力支持大规模设备更新和消费品以旧换新。“两新”政策成效显著，根据商务部数据，截至12月13日零时，汽车以旧换新带动乘用车销售量超520万辆，其中报废更新超251万辆，置换更新超272万辆；家电以旧换新带动八大类产品销售量超4900万台。

补贴品类扩大至 3C 产品，进一步加快换机周期。2024年8月，商务部等四部门印发关于进一步做好家电以旧换新工作的通知，明确各地自主确定补贴“8+N”类家电品种，每件最高补贴2000元。自此全国各地家电以旧换新补贴政策陆续落地，并且消费补贴品类从家电逐步扩大至手机、平板电脑等3C数码产品。广东省、浙江省、江苏省、贵州省等地陆续新增3C补贴活动，河北省已出台文件通知商户开始报名2025年3C数码产品促消费活动。综合来看，补贴力度在10%-20%区间，单品上限在600-2000元。

2025年“两新”超长期特别国债资金加大，将实施手机等3C数码产品购新补贴。在2024年12月11日召开的中央经济工作会议后，中央财办解读采访中提到，2025年支持“两新”的超长期特别国债资金将比今年大幅增加，会将更多市场需求广、更新换代潜力大的产品和领域纳入政策支持范围。2025年1月3日，在国务院新闻办公室举行的“中国经济高质量发展成效”系列新闻发布会上，国家发展改革委有关负责人表示2025年，要加力扩围实施“两新”政策，将实施手机、平板、智能手环等3类数码产品购新

补贴。我们预计，明年实施手机等三类数码产品购新补贴后，中国智能手机、平板电脑等市场需求有望大幅增长，上游IC设计和零部件厂商或将表现出更大的业绩弹性。

图表182: 中国部分地区3C数码产品补贴活动

活动时间	省市区	补贴品类	补贴内容
2024年9月28日-2024年12月31日	广东省广州市	手机类、平板类（平板电脑、平板学习机）、智能穿戴设备类（智能手表、手环、耳机、智能眼镜、虚拟现实眼镜）	1) 对个人消费者购买手机，订单金额超500元，按照销售价格的10%享受一次性立减补贴，每位消费者每类产品可补贴1件，每件补贴不超过1000元； 2) 对个人消费者购买平板、智能穿戴设备，订单金额超500元，按照产品销售价格的15%享受一次性立减补贴，每位消费者每类产品可补贴1件，每件补贴不超过2000元。 3) 手机、平板等易变现产品需现场激活。
2024年10月24日-2024年12月31日	浙江省杭州市西湖区	手机类、平板类（平板电脑、平板学习机）、智能穿戴设备类（智能手表、手环、耳机、智能眼镜、虚拟现实眼镜）	实际成交价的10%立减补贴，每件商品最高补贴1000元。
2024年11月27日-2024年12月31日	江苏省	手机、平板电脑、数码相机（含机身、套机，不含镜头等配件）、智能手表、学习机、翻译机、无线蓝牙耳机	产品成交价格的15%，最高可优惠1500元。
2024年12月1日-2024年12月31日	贵州省	国产手机、平板电脑	对消费者购买3000元（含3000元）以下国产手机、平板电脑这两类数码产品，按照产品购置发票金额的20%给予一次性补贴
2024年12月12日起，每天上午10点开始领券	江西省	手机、平板电脑	立减15%，补贴不超过1000元/笔； 省级资金对个人消费者前往活动商户购买手机、平板电脑，按照产品最终成交价格的15%给予补贴，每笔补贴不超过1000元
2024年12月12日-2024年12月22日	四川省成都市	手机、平板电脑、智能手表、学习机、翻译机、耳机	9折消费券：适用于参与活动的手机品类，至高可减1000元； 8.5折消费券：适用于参与活动的平板电脑、智能手表、学习机、翻译机、耳机品类，至高可减1500元

资料来源：本地宝，杭州网，贵州省人民政府，江苏省发展和改革委员会，快科技，国盛证券研究所

6.4 AI眼镜：“百镜大战”愈演愈烈，逐步迈向AR终局

Meta Ray-Ban 2024年销量有望超150万副。回顾Meta Ray-Ban的成功，我们看到第一代智能眼镜Ray-Ban Stories在2021年9月发布后，直至2023年2月也才售出了30万副，月活用户只有27000名，不到设备量的10%。但2023年9月发布的第二代产品Meta Ray-Ban，在造型基本保持不变且价格维持与第一代持平的299美元基础上，据The Verge统计，仅23Q4一个季度的销售量就超过了第一代眼镜全生命周期的出货量；2024年5月，销量突破100万副；预计2024年全年出货量有望超过150万副。而对比行业方面，IDC数据显示2023年全球AR眼镜销量为48万台，智能眼镜销量为101万台，我们预计2024年行业销量增长的大部分仍由Meta贡献。

图表183: Meta Ray-Ban 外观图



META AI

CAMERA

资料来源: Ray-Ban 官网, 国盛证券研究所

回想国内外厂商初期推出的智能眼镜产品, 我们很容易发现一个共通点, 那就是过度关注“智能”, 总是研究着如何把智能酷炫的技术加到眼镜上, 而忽视了如何先做好“眼镜”这个品类去让消费者接受并习惯于长期佩戴, 最起码要先满足“款式、外观、重量和价格”的基本要求, 再去研究如何将智能技术嵌入其中, 而 Meta Ray-Ban 的成功便是有利借助了 Ray-Ban 在“眼镜”品类上的优秀设计理念和时尚形象。

虽然 Ray-Ban Stories 第一代生命周期只卖了 40 万副, 但相比于很多只有几万副销量的其他竞品来说已经算得上成功, 而第二代产品在保持基本形态、外观、重量、价格等基本一致的基础上, 对以下核心功能及体验的全面升级是其放量的关键:

- 1) 相机质量质的飞跃:** 12MP 摄像头+1080P 视频+60 FPS 的拍照和视频录制效果很好地满足了普通用户的日常使用需求; 且一代是方形长宽比进行拍摄和录制, 不太适合在社交媒体发布, 而二代的拍摄格式和效果完美适配了现在的社交媒体; 同时二代支持在 Facebook 和 Instagram 上直播。
- 2) 音质效果的大幅提升:** 3 个麦克风阵列拓展到 5 个, 且支持空间音频录制; 最大音量增加 50%, 低音、漏音、消音进一步优化。
- 3) 款式和舒适度进一步升级:** 进一步减轻重量并缩小体积, 同时拓展了可选择的框形和颜色, 加上镜片的款式可以达到 150 多种不同的组合, 满足了消费者对于眼镜美观度和自定义的诉求。
- 4) Meta AI:** 二代眼镜发售的时候并没有 Meta AI 功能, 直到 2024 年 4 月才正式登陆, 虽然当前的功能还并不是十分丰富, 包括询问天气、时间等日常基础信息, 以及拍照并识别物体等, 但我们认为基于开源 Llama 模型的 AI 能力将是后续产品迭代最值得期待的地方。
- 5) 交互和续航:** 二代产品交互延迟更低, 响应速度更快; 续航时间和充电速度都得到了 30% 以上的提升, 而且蓝牙连接更稳定, 电池寿命也更长。

图表184: Ray Ban Meta 智能眼镜主板 BOM

器件名称	品牌	型号/零件号/丝印	单价 (美元)	数量	总金额 (美元)	占比
处理器SoC	高通	AR1 gen 1	55	1	55	62.1%
存储器	佰维	BWCK1EZC-32G	11	1	11	12.4%
WiFi芯片	高通	WCN7851	1.2	1	1.2	1.4%
电源管理芯片	高通	PMAR2130	1	1	1	1.1%
电源管理芯片	高通	PM3003A	0.7	1	0.7	0.8%
电源管理芯片	Analog Devices	Max77813	0.7	1	0.7	0.8%
音频芯片	Analog Devices	MAX98388	0.6	1	0.6	0.7%
MCU	NXP Semiconductors	MIMXRT685SF	2	1	2	2.3%
射频芯片	其他	2500EMN1/2JS	1	2	2	2.3%
MCU	其他	Y70307	1.5	1	1.5	1.7%
过流过压保护芯片	其他	1BD3XC	1	1	1	1.1%
LED驱动芯片	其他	A26AET	0.8	1	0.8	0.9%
存储器	其他	Z7ELW	0.5	1	0.5	0.6%
电源管理芯片	其他	20F8	1	1	1	1.1%
板载天线	其他		0.1	1	0.1	0.1%
其他元器件	其他		3.4	1	3.4	3.8%
PCB	华通		6	1	6	6.8%
合计					88.5	

资料来源: Wellsenn XR, 国盛证券研究所

根据 Wellsenn XR 对 Ray Ban Meta 智能眼镜 BOM 表的拆解数据, 我们看到在主板上, 处理器 SoC 和存储器分别为 55 美金和 11 美金的 ASP 牢牢占据了整个主板较大部分的价值量, 占比分别达到 62% 和 12%, 两者合计占比高达 75%。其他芯片如电源管理芯片、MCU、射频芯片、WiFi 芯片等价值量差异不大, 不过华通提供的 PCB 也有 6 美金的价值量, 占比达到 7%。在主板之外, 合计 38 美金的价值量中, 雷朋提供的镜片/镜架和索尼提供的摄像头芯片分别拥有 13 美金和 9 美金的 ASP, 占比分别为 34% 和 24%。其他器件如喇叭、麦克风、触摸条、电池的价值量分别有 3、2.5、2.5、1.5 美金, 占比分别为 7.9%、6.6%、6.6%、3.9%。

综合主板以及主板之外所有器件的价值量分布来看, 主板上的处理器 SoC 和存储器, 以及主板之外的镜片/镜架和摄像头芯片, 这 4 类器件的价值量最为靠前, 因此我们认为在智能眼镜产业链上, 处于以上 4 个器件环节的相关厂商在未来下游需求起量时将最为受益, 我们看好相关国产供应商在以上环节的布局和卡位。

图表185: Ray Ban Meta 智能眼镜 BOM (除主板外)

器件名称	品牌	型号/零件号	单价 (美元)	数量	总金额 (美元)	占比
镜片/镜架	雷朋		13	1	13	34.2%
麦克风	歌尔		0.5	5	2.5	6.6%
LED灯	其他		0.1	2	0.2	0.5%
摄像头	索尼	IMX681	9	1	9	23.7%
IMU	STMicroelectronics	LSM6DS3TR	1.2	1	1.2	3.2%
IMUPCB	其他		0.2	1	0.2	0.5%
音频芯片	Analog Devices	MAX98388	0.6	1	0.6	1.6%
电池	德赛		1.5	1	1.5	3.9%
触摸条	其他		2.5	1	2.5	6.6%
喇叭	歌尔		1.5	2	3	7.9%
其他	其他		4.3	1	4.3	11.3%
合计					38	

资料来源: Wellsenn XR, 国盛证券研究所

带显示的 **Ray-Ban** 眼镜最早将于 **25H2** 亮相。我们看到，继 Meta Ray-Ban 的成功之后，Meta 并未停下探索的脚步，这款不带显示的智能眼镜在 Meta 的元宇宙商业蓝图中终究只是一个过渡性产品，根据映维网 Nweon 12 月 24 日消息，Meta 和 EssilorLuxottica (Ray-Ban 母公司) 计划为 Ray-Ban 智能眼镜加入显示屏用来显示 Meta AI 的通知和回复，这款升级版产品最早可能会在 2025 年下半年亮相。

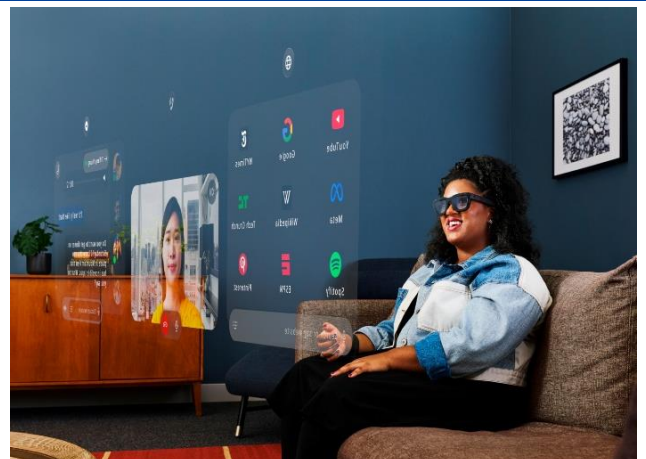
Meta Orion “全彩 Micro-LED+碳化硅衍射光波导”展示 AR 眼镜未来方案雏形。在 9 月 26 日的 Meta Connect 2024 上，Meta CEO 扎克伯格揭晓了公司秘密研发十年的 AR 眼镜——Orion，虽然这款原型机高达 1 万美元的生产成本还远不足以使其成为一款消费品，但其采用全彩 Micro-LED 光机+碳化硅材料的衍射光波导方案，提供了 70 度的超大视场角，让我们看到了 Meta 在产品技术上的不懈追求。

图表186: Meta Ray-Ban 带显示屏效果图



资料来源: ifanr, 国盛证券研究所

图表187: Meta Orion 效果图



资料来源: Meta 官网, 国盛证券研究所

光波导方案逐步成为一致性选择。我们看到，目前市场上发布的 AR 眼镜所采用的主要光学显示技术方案包括: LCoS+棱镜、Micro-OLED+自由曲面反射/BirdBath、DLP/Micro-LED+衍射光波导、LCoS/Micro-OLED/Micro-LED+几何光波导等组合方案。早期的光学技

术方案，存在一个原理性的技术矛盾，即伴随着视场角的扩大，会使镜片变厚、体积增大。另外，大部分光学方案的透光率比较低，无法看清现实画面，难以成为AR方案的理想技术。随着技术的不断发展，光波导方案以其同时兼备大视场角、小体积、高透光率、高清画质等特性，已经逐步成为AR眼镜一致性的终极解决方案。

图表188: AR光学显示方案对比

	棱镜	自由曲面	BirdBath	衍射光波导	几何光波导
技术壁垒	★	★★	★★	★★★	★★★
镜片厚度	>10mm	>10mm	>8mm	<2mm	<2mm
视场角	~15°	~30°	30° -50°	40° -50°	40° -55°
透光度	~50%	~50%	<50%	>80%	>90%
光效	10%-20%	30%-50%	10%-15%	0.3%-1%	>15%
产品尺寸	较大	较大	小	小	小
显示屏幕	LCoS	LCoS/Micro-OLED	Micro-OLED	DLP/Micro-LED	LCoS/Micro-OLED/ Micro-LED
主要缺点	体积大	厚度相对较大	厚度大、亮度低	彩虹色散、光效低	二维扩瞳技术和键合 工艺门槛高
主要代表	Google Glass	EPSON	耐德佳	Microsoft HoloLens WaveOptics	理湃光晶 Lumus

资料来源：映维网，国盛证券研究所

中短期看几何光波导技术更为占优。聚焦光波导技术，其根据不同的光学技术原理和加工工艺，可以分为衍射光波导和几何光波导。几何光波导技术是通过几何阵列反射原理来实现图像的无损输出和画面画幅的扩大，其光效超过15%，是衍射光波导的数十倍以上。此外，借助高清微显示技术可实现高亮度、色彩丰富、景深融合的全彩显示。由于几何光波导的色散控制较好，不存在杂色、彩虹效应等问题。并且几何光波导在显示图像时，正面漏光率低至1%以下，有效地保护了用户的隐私。因此，几何光波导是目前情况下助力消费级AR眼镜落地较为理想的技术方案，其劣势主要在于量产难度较大。

中期仍要关注基于SiC材料的蚀刻衍射方案，以及远期体全息方案在材料和工艺上的突破。衍射波导可分为表面浮雕光波导和体全息光波导，其中基于光刻技术的表面浮雕波导因其光栅显示的原理，使得在显示彩色图像时会出现彩虹色散，导致色彩的丰富性、画面的均匀性等性能很难有所提升，图像颜色失真较为严重。此外，由于光栅的衍射特性，其光效不足1%，在目前AR光学方案中光效最低。但衍射的优势在于其量产的难度明显低于几何方案，因此规模优势较为突出，因此国内外逐渐将目光聚焦于衍射方案。我们认为未来随着AR眼镜出货量级的不断提升，表面浮雕衍射方案在量产上的优势或将逐步显现。此外，体全息光波导技术虽然受限于在材料制备和生产工艺上面临的高壁垒，目前仍处于实验室阶段，但远期来看，如果材料和工艺问题能够逐步突破，其将成为最为理想的终极方案。

智能眼镜发展路径：传统眼镜—>音频眼镜—>拍摄+音频眼镜—>多模态AI眼镜—>AR眼镜。从Meta一系列的动作中我们可以看到智能眼镜行业一条合理的发展路径：首先在传统眼镜的形态基础上叠加少量的科技功能吸引用户无感平替，例如拍照和摄像功能就可以方便消费者在短视频时代记录生活；然后再通过AI等高附加值的功能增加用户的使用时长和粘性，虽然早期的AI功能相对较弱，但现如今大模型的快速发展使得AI对产品的赋能越来越显著；最后再加入显示功能，耐心等待一种高亮度+低成本+小体积+高显示质量的光学方案来完成“最后一公里”的挑战。

国内“百镜大战”愈演愈烈。聚焦国内市场，Meta Ray-Ban 的火爆也成功引发了国产厂商的“百镜大战”，而且 Meta Ray-Ban 由于多种原因还未在国内市场开售，因此也给国产厂商留出了竞争空间。24年的11-12月，各类 AI/AR 眼镜新品频出，闪极科技在12月19日发布的 AI 拍拍镜“闪极 A1”更是热度不断，其硬件配置几乎完全对标 Ray-Ban Meta，而价格仅为 Ray-Ban Meta 的一半——前五万台共创版的价格为 999 元，因此发布会后各平台的预定链接已经全部卖空下架。不过我们也注意到，虽然硬件较 Ray-Ban Meta 更便宜，但闪极也留下了软件使用付费的商业模式的空间——两个 AI 功能“AI 闪记和 AI 云盘”，年订阅价格 299 元，首年免费。目前闪极 A1 已登录京东、天猫、抖音、拼多多、微信商城等各大渠道的闪极官方店铺开启预售，产品将于 2025 年 1 月 15 日左右正式发货。

图表189: 2024年11-12月AI/AR智能眼镜新品汇总

发布时间	品牌	新品	大模型	主要亮点
2024.11.12	小度	AI眼镜	文心一言	融合“全球首款搭载中文大模型的原生AI眼镜”，搭载百度文心大模型与DuerOSAI,具备第一视角拍摄、边走边问、文心一言识物百科、视听翻译、智能备忘等功能,重45g,1600万像素超广角摄像头、 配备自研的AI防抖算法 ,结合4阵列麦克风 and 定制的开放式防漏音扬声器单元,可识别声源方向。
2024.11.16	Looktech	Looktech AI智能眼镜	未知	融合眼镜、耳机、相机和AI功能, 整机37g,配备了1300万像素的摄像头 ,支持2K视频录制和4K照片拍摄,并具备防抖和水平校正功能。 AI助手Memo是最大亮点 ,Memo基于大型语言模型,具备自然对话和记忆功能,可全天候为用户提供新闻、天气、电影导览、翻译、会议记录、饮食健康管理等A功能。
2024.11.18	Rokid	Rokid Glasses AI+AR眼镜	通义千问	采用一体化结构布局, 重量49g,配备1200万像素专业级摄像头 ,支持竖屏模式下高清拍照和视频录制功能,能完成物体识别、文字翻译,具备接收应用程序的通知提醒等功能。搭载了第一代骁龙@AR1平台,配合2GB RAM+32GB ROM存储组合。
2024.11.25	谷东科技	Star1 Star1S	未知	Star1系 业内首款全彩双目光波导分体式AI+AR眼镜 ,配置AI算力和自研大模型应用,主要面向工业设计、教育培训、医疗辅助等。Star1S搭载自研高亮高透 全彩光波导显示系统,入眼亮度超过700尼特,1600万像素高清快速变焦摄像头和语音系统能实时感知AI大模型辅助 。Star1S可自由安装APP,支持语音翻译、外语学习、会议记录等AI功能。
2024.11.29	影目科技	INMO Air 3 智能眼镜	未知	搭载自研的“IMAR光学显示引擎”, 索尼0.44英寸Micro OLED屏幕,是全球首个1080P全高清分辨率的一体式AR眼镜 。搭载了IMOS 3.0空间操作系统,号称“下一代空间计算设备”设计。迄今, 影目科技是全球唯一实现阵列光波导量产的企业 。
2024.11.29	影目科技	INMO GO 2 翻译眼镜	未知	搭载行业领先的流式同声传译技术,初衷是全面代替翻译机,为商务人士提供流畅、高效的翻译体验。支持40种语言在线翻译和90种口音识别,也是 全球唯一支持离线翻译的智能眼镜,支持8个语种离线翻译 。
2024.12.5	XREAL	XREAL One	未知	全球首款搭载自研空间计算芯片X1的AR眼镜,赋予眼镜本体强大的空间计算能力,具备原生3DoF空间定位能力支持“3DoF悬停”和“0DoF云台”两种空间屏模式。 M2P延迟降低至3ms,达到了XR行业最低水平 。
2024.12.19	闪极科技	闪极AI拍拍镜	未知	官方称 首款量产的AI拍摄眼镜,业界首款16MP摄像头,123° 仿人眼超广角,旗舰级ARM平台,2GB RAM+32GB ROM, 重约30g 。
2024.12.21	界环	界环AI音频眼镜	未知	采用眼镜耳机二合一, 重38.1g ,支持多框型、多配色, 支持超长待机15天、连续聆听11小时、连续通话9小时 。支持AI通知播报、AI畅聊、面对面翻译等,有“时尚眼影款”“时尚腮红款”,售价799元。

资料来源: 亿欧, 国盛证券研究所

闪极 A1 首年销量有望超 50 万台。对于闪极 A1 的预期，其 CEO 张波表示：预期第一年能够卖 50 万台，就能做到盈亏平衡，目前手里已经掌握大量订单，已经完成了明年的大部分销量。从该表述中我们可以发现闪极 AI 眼镜的产品竞争力，以及消费者对该品类的接受度。对于闪极宣称的“这是中国 AI 眼镜的首批量产产品”中也能感受到 AI 眼镜的高生产难度，因为这是无法复刻当下成熟的手机供应链，AI 眼镜的精细程度和材料体系都完全不一样，每个环节都可能成为卡点。

图表190: 闪极 AI 拍拍镜硬件配置



资料来源: 硬氪, 国盛证券研究所

互联网大厂、手机品牌跑步进场，AI 眼镜未来已来。但即使面对这些困难，不管是闪极这样的小厂，还是 Meta、字节等互联网大厂，乃至三星、小米等手机厂商，都在跑步进场。三星电子已于今年 11 月上旬敲定 AI 智能眼镜项目方案，产品计划于 2025Q3 上市发售，首批预计销量达 50 万台。小米也在计划推出新一代 AI 眼镜，预计于 2025Q2 发布，雷军预期销量在 30 万台以上。我们认为，在一个年出货量超 15 亿副的传统眼镜市场，如果 AI 眼镜能够逐步取代智能手机成为新一代移动终端，不管是硬件端的入口还是软件端的数据，都将成为决定相关企业生死存亡的胜负手，而在当下，这一变化正在悄然发生。

6.5 AI 耳机: 字节豆包发布 Ola Friend, 补全 AI 布局拼图

回顾国内 AI 耳机进展:

2024 年 4 月 11 日，漫步者 AI 算法通过网信备案并挂网公示。本次完成备案的“漫步者摘要提取算法”是基于预训练大语言模型架构的聊天型语言模型基础上进行精调和使用的模型，后续或将应用于公司的 AI 耳机新品。

2024 年 5 月 15 日，科大讯飞旗下品牌未来智能发布了讯飞会议耳机 Pro2 和 iFLYBUDS2 两款新耳机，讯飞会议耳机 Pro2 支持 VIAIMAI 进化大模型，可一键生成会议摘要总结、待办事项，支持 AI 会议助理有问必答、32 国语言同传翻译，支持闪录音功能，支持免费录音转文字等 AI 功能，iFLYBUDS2 具备会议摘要总结、多场景录音撰写多国语音同声传译、快速生成会议摘要和待办事项等 AI 功能。

2024 年 9 月 6 日，时空壶在德国柏林国际消费电子展览会 IFA 召开媒体见面会，正式发布时空壶新一代 W 系列旗舰翻译耳机产品 W4 Pro。在 W4 Pro 中，时空壶的双向同传技术实现了重大突破，其次，其搭载了时空壶最新的 Hybridcomm 3.0 超级沟通技术，翻译延迟低至仅一句话，让沟通瞬间响应，毫无迟滞之感，同时将基础翻译准确率由 95% 提升至 96%，并且在低音量状态下说话准确度提升 15%。

2024 年 10 月 10 日，字节跳动豆包发布 AI 智能体耳机 Ola Friend，是一款开放式耳机，单耳重量 6.6 克，为同类最轻，采用了恒玄科技的 2700 系列芯片。可接入豆包大模型，

并与豆包 App 深度结合，售价 1199 元。用户戴上耳机后，无需打开手机，只需喊出关键词“豆包豆包”，便能唤起豆包进行对话，后者能够在信息查询、旅游出行、英语学习及情感交流等场景为用户提供帮助。

图表191: 讯飞录音降噪会议耳机 Pro2



资料来源: 讯飞商城, 国盛证券研究所

图表192: Ola Friend 智能体耳机

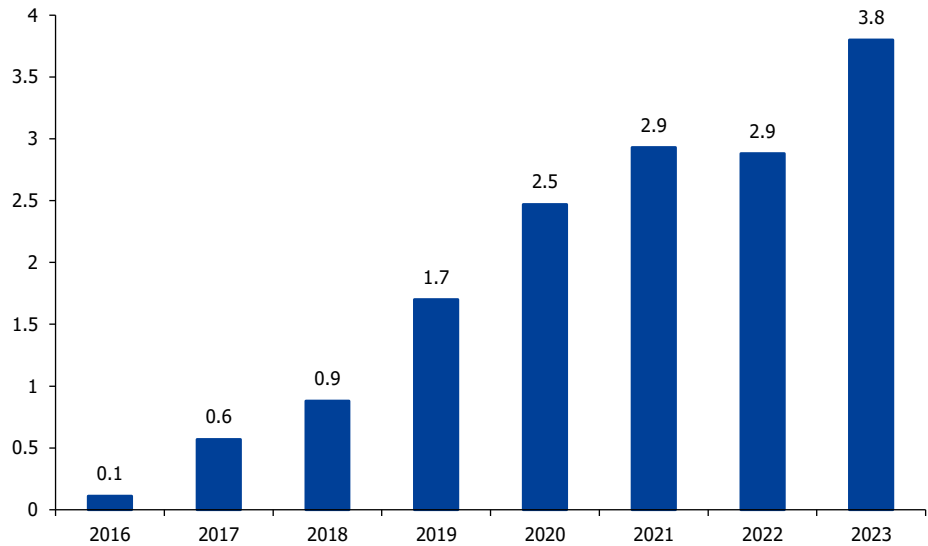


资料来源: Ola Friend, 国盛证券研究所

字节跳动全额收购 Oladance，加码 AI 硬件布局。字节跳动已成功收购开放式耳机品牌 Oladance，完成 100% 控股，标志着字节跳动在 AI 硬件领域的进一步加码及拓展。Oladance 成立于 2019 年，由前 BOSE 高管李浩乾与多名 BOSE 工程师共同创立，公司为首家推出 OWS 开放式耳机概念及相关专利技术而闻名，并在 2021 年推出了全开放式 OWS 系列耳机，锚定了中国厂商在 OWS 开放式耳机方面的领先地位。根据 36 氪，此次交易金额为 3 至 5 亿元人民币之间。通过这次收购，字节跳动不仅获得了 Oladance 的核心技术，同时也吸纳了一支由前 Bose 工程师组成的经验丰富的团队，补足了自身在音频技术方面的不足并推动 Oladance 从可穿戴设备出发向生成式 AI 服务平台的转型，进一步丰富字节自身的 AI 生态系统，通过此次收购，字节有望在前景潜力十足的开放式耳机市场中占据有利地位。

TWS 耳机市场增长势头强劲，AI 耳机进一步提振需求。根据 IDC 的数据，2024 年上半年国内蓝牙耳机市场出货量达到 5540 万台，同比增长 20.8%。其中，真无线耳机市场出货 3508 万台，同比增长 5.6%；开放式耳机市场快速增长，2024 年上半年出货 1184 万台，同比增长 303.6%。根据 FutureSource，全球 TWS 耳机出货量由 2018 年的 0.46 亿部，增加到 2023 年的 3.08 亿部，年均复合增长率为 46.27%。

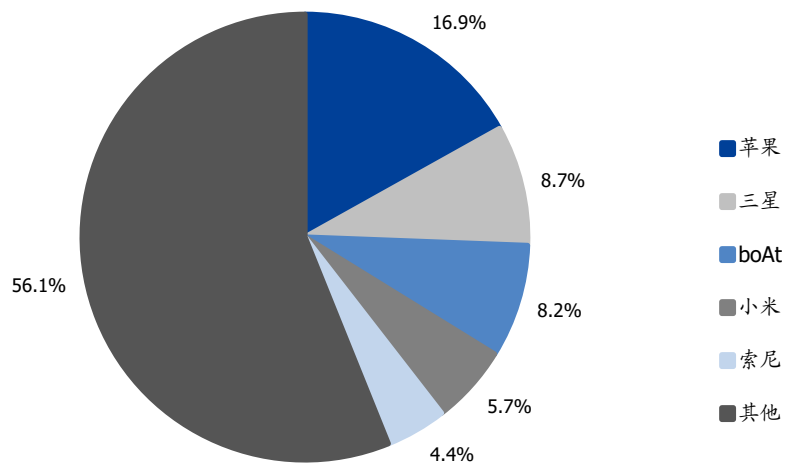
图表193: 全球 TWS 耳机出货量 (亿部)



资料来源: FutureSource, 与非网 eefocus, 国盛证券研究所

TWS 耳机市场集中度低, 品牌效应并不显著。根据 Canalsy 数据, 2024 年第三季度全球智能耳机出货量为 1.26 亿部, 同比增长 14.6%。从地区市场来看, 印度市场在推动亚太地区和全球增长方面发挥关键作用, 与 2023 年第三季度相比, 其出货量增长了 51%。从竞争格局来看, 24Q3 除 Apple 以 17% 的市场份额占据较高份额以外, 其他供应商市场份额均处于 10% 以下, 证明 TWS 用户对大部分品牌的粘性相对更低, 反而对产品的性能、功能以及设计 (开放式、入耳式、半入耳式) 等其他因素更加敏感, 市场竞争更加激烈, 这促使各厂商更加重视产品的功能迭代, 以寻找成为行业或细分领域龙头的契机。随着近年 AI 耳机的陆续推出, 全球智能耳机市场或将迎来大洗牌, 以中国厂商为首的各专业品牌或将有望获得更多的市场份额。

图表194: 24Q3 全球智能耳机市场格局

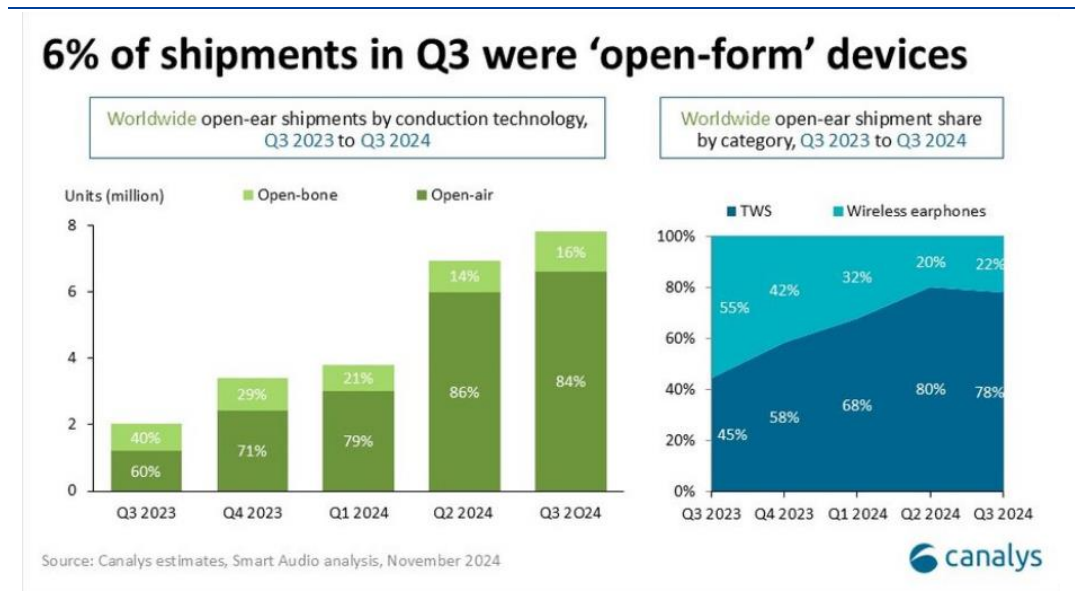


资料来源: Canalsy, 国盛证券研究所

开放式耳机出货量飞速增长, 助力中国厂商加快抢占全球智能耳机市场份额。根据 Canalsy 数据, 2024 年第三季度, 全球开放式耳机出货量占全球耳机总出货量的 6%, 开放式耳机势头迅猛与其对人耳来说更加健康的设计与市场健康和健身场景的需求的趋势兴起不谋而合, 这一转变凸显了开放式设备在消费者越来越重视产品是否拥有适合

户外及运动等生活方式的音频解决方案的大背景下日益增长。值得一提的是，目前中国大陆仍然是开放式产品出货量的领导者，中国厂商出货量占全球整体份额的59%，中国厂商有望以开放式耳机为切入点进一步蚕食全球智能耳机市场份额。

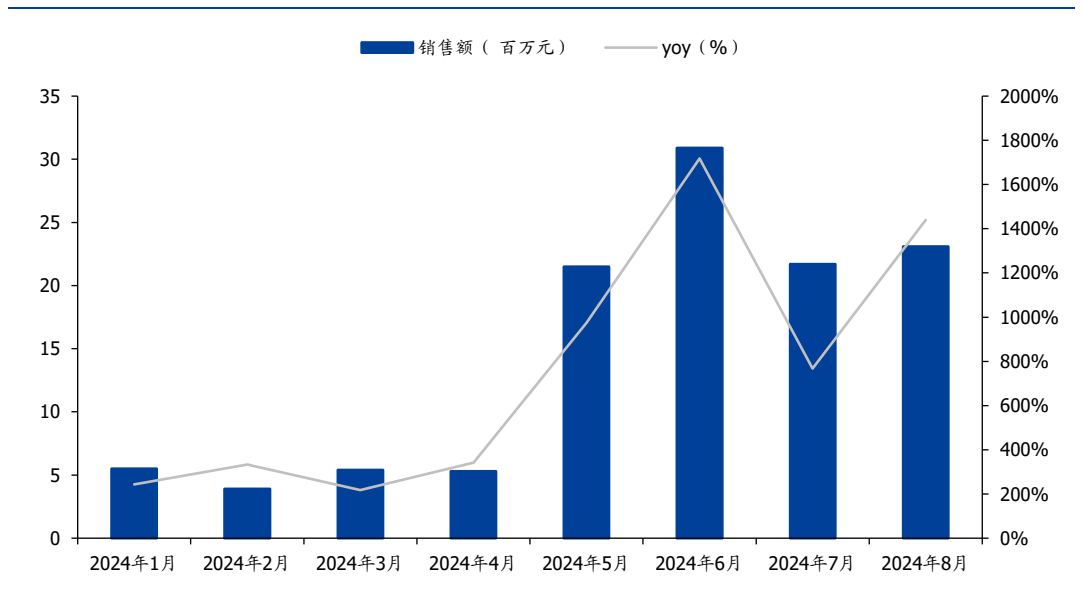
图表195: 全球开放式耳机出货量 (百万部)



资料来源: Canlys, 国盛证券研究所

2024年中国AI耳机出货量激增，AI耳机正式进入发展快车道。随着AI逐步进入大众视野，中国厂商将AI应用于主动降噪、空间音频等技术推出了一系列优质产品，尤其在OWS等细分赛道上所推出的主打AI功能的OWS耳机拥有亮眼的表现。根据洛图科技，2024年8月，具有AI功能的耳机销额占比已达1.4%，相较于上年同期销量同比增长763%，销售额同比增长1447%，宣布AI耳机在2024年下半年正式进入发展新纪元。

图表196: 截止2024年8月中国AI耳机2024年月度销售额



资料来源: 洛图科技, Ofweek, 国盛证券研究所

AI功能不断完善，AI赋能领域场景百花齐放。随着近年通用大模型的数量不断增加、性能的不断的迭代，各类解决方案层出不穷，各家厂商进一步探寻AI可赋能的应用场景。如在商务办公场景，科大讯飞便在近年推出并持续更iFLYBUD系列耳机，旗下产品可结合其自身软件的会议助手，以辅助用户在工作方面的使用体验作为切入点，率先打

开了 AI 赋能商务办公的市场赛道，此后如塞那、飞利浦以及 IKKO 等厂商均以商务办公为应用场景陆续推出旗下 AI 耳机产品。截至目前，AI 耳机的应用场景也在进一步的拓宽，如塞那在 24 年 3 月所推出的新款 OWS 耳机可与内置软件结合，实现 AI 绘画、AI 对话等多项娱乐功能，使应用场景进入轻松娱乐、日常陪伴等更加生活化的场景；Cleer 在 24 年 4 月所推出的开放式 AI 耳机 Cleer ARC 3 音弧则内置了 AI 运动算法，在商务办公场景外拓展了户外运动的应用场景，可帮助用户实时监控运动生理数据，为其提供更精准更科学的运动指导，打造用户私人的 AI 健身教练。同时它还集成了出门问问的语音控制技术，升级了 AI 语音控制，使用户无需提示词唤醒语音助手，使用“下一首”、“接听电话”等快捷口令即可进行相应操作，实现更自然直接的智能交互。我们认为随着厂商对 AI 功能的进一步投入，以及更多科技公司意识到 AI 耳机在未来的市场潜力以及作为 AI 硬件对自身产品矩阵的重要性，AI 耳机的应用场景将随着新玩家的进入而进一步扩充，并有望渗透至未来人们生活中的所有日常场景之中。

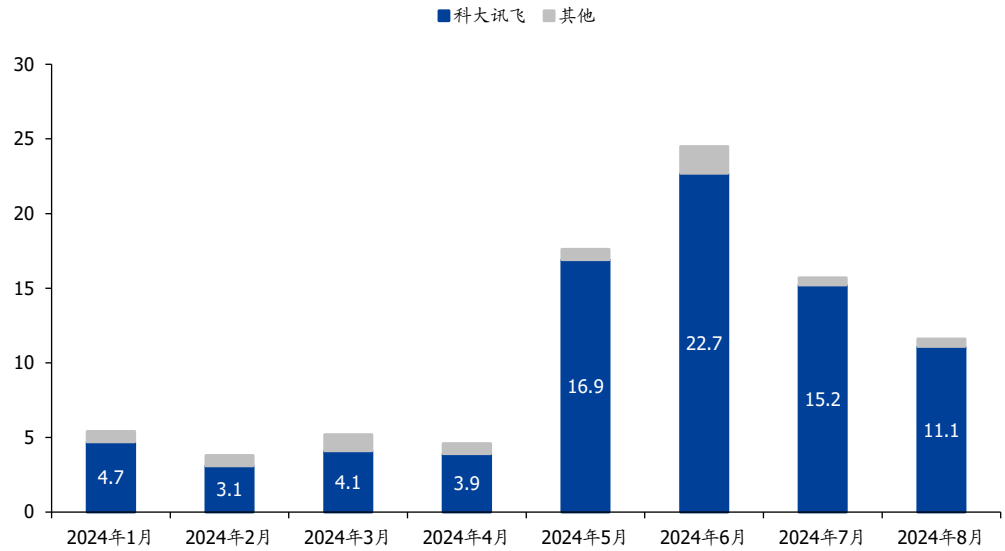
图表197: Cleer 私人 AI 运动教练



资料来源: Cleer, 国盛证券研究所

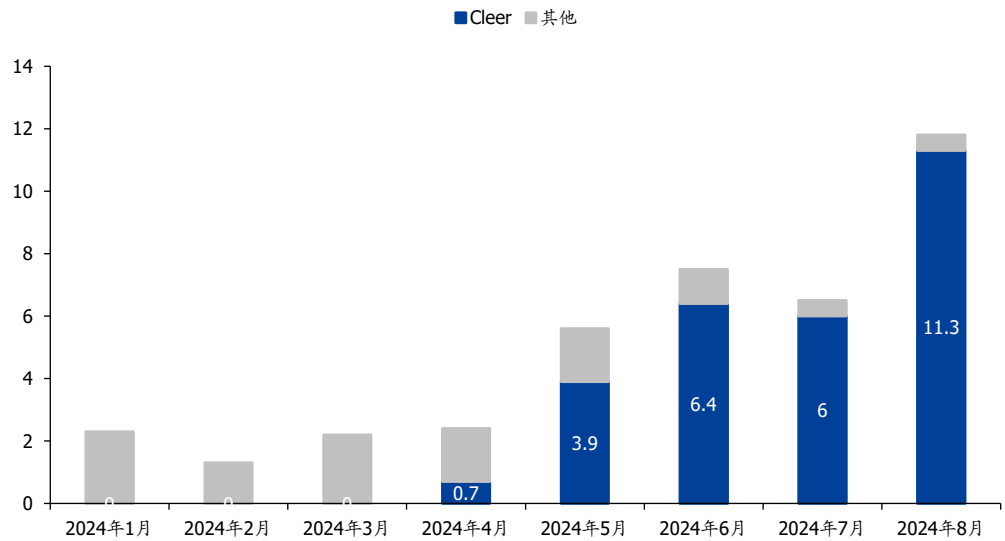
运动类 AI 耳机市场潜力十足，销售额有望进一步攀升。具体从应用场景来拆分中国 AI 耳机销售额，我们可以看到以针对商务及娱乐等应用场景的 AI 耳机销售额在 24 年 6 月以 2450 万元达到顶峰后短暂下滑，但整体销售水平依然高于去年水平，其中科大讯飞占绝大部分市场份额；在运动场景方面，Cleer ARC 3 在 4 月发布后迅速打开市场，销售在 24 年 8 月达到 1130 万元，弥补了商务娱乐 AI 耳机在 8 月份的短暂疲软，而 Cleer ARC 3 成为爆款也预示以运动作为户外场景的 AI 赋能开放式耳机拥有强大的商业潜力以及潜在市场规模。

图表198: 截止 2024 年 8 月中国商务+娱乐 AI 耳机 2024 年月度销售额 (百万元)



资料来源: 洛图科技, Ofweek, 国盛证券研究所

图表199: 截止 2024 年 8 月中国运动 AI 耳机 2024 年月度销售额 (百万元)



资料来源: 洛图科技, Ofweek, 国盛证券研究所

6.6 CES 2025 以 AI 为主旋律展开

➤ AI PC: 巨头领衔, 性能再进化。

英伟达: 持续为 AIPC 提供动力支持。新一代 GeForce RTX 50 系列采用 NVIDIA Blackwell 架构、第五代 Tensor Cores 和第四代 RT Cores。在 AI 渲染领域, 包括神经网络着色器、数字人技术、几何图形和光照等方面均有升级。此外, 英伟达还直接进入了 AIPC 赛道, 推出小型超级计算机 Project Digits, 搭载全新 GB10 超级芯片, 是全球最小的可运行 200B 参数模型的 AI 超级计算机。Project Digits 搭载的 GB10 超级芯片采用 Blackwell GPU,

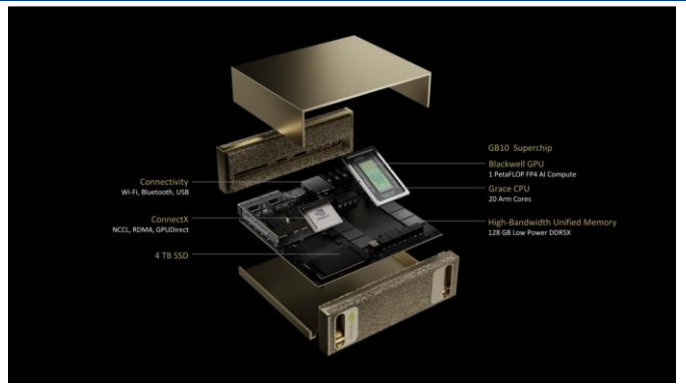
联发科参与设计，通过 NVLink-C2C 片间互连与一颗高性能 Grace CPU 互联，该 CPU 采用 ARM 架构且拥有 20 个高效节能内核，让更多开发者能在桌面端使用大模型。

图表200: 英伟达发布 GeForce RTX™ 50 GPU



资料来源: NVIDIA 英伟达, 国盛证券研究所

图表201: 英伟达 GB10 Grace Blackwell Superchip



资料来源: 英伟达, 国盛证券研究所

英特尔: 发布全新酷睿 Ultra 处理器(第二代), 涵盖全新英特尔酷睿 Ultra 200HX 和 200H 系列移动处理器、基于英特尔 vPro 平台的全新英特尔酷睿 Ultra 200V 系列移动处理器、全新英特尔酷睿 Ultra 200U 系列移动处理器、扩展的英特尔酷睿 Ultra 200S 系列台式机处理器、为边缘计算打造的全新英特尔酷睿 Ultra 处理器和英特尔酷睿处理器。酷睿 Ultra 200H 系列相比上代, 单线程和多线程性能分别提升约 17%和 20%, 核显性能提升约 22%, AI 算力高达 99 TOPS。

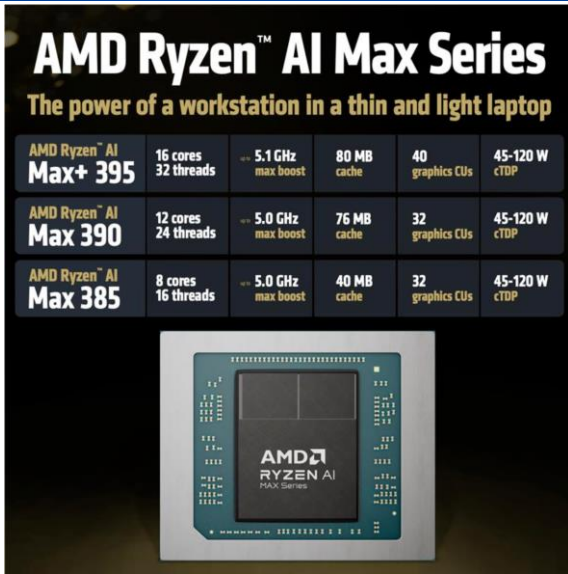
图表202: 英特尔发布全新酷睿 Ultra 200HX 和 200H 系列移动处理器



资料来源: PChome 电脑之家, 国盛证券研究所

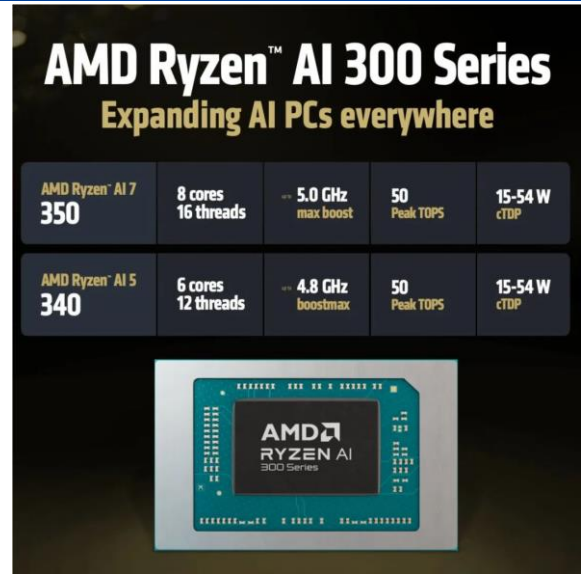
AMD: 宣布推出全新锐龙 AIMax 系列处理器, 满足高端轻薄笔记本电脑对高性能计算的需求; 同时推出基于“Zen5”架构的全新锐龙 AI300 系列处理器, 丰富了产品阵容。为延续 AMD “Zen 4” 架构的优势, 还推出了适用于日常办公的锐龙 200 系列处理器, 并扩展商用 AIPC 产品线, 将 AMDPRO 技术集成到锐龙 AIMax、锐龙 AI300 和锐龙 AI200 系列处理器中。得益于集成高达 50 TOPS AI 处理能力的 NPU, 锐龙 AIMax 系列处理器成为下一代 AI PC 的强大动力。

图表203: AMD 锐龙 AI Max 和锐龙 AI Max PRO 系列处理器



资料来源: AMD 中国, 国盛证券研究所

图表204: AMD 锐龙 AI 300 和锐龙 AI 300 PRO 系列处理器



资料来源: AMD 中国, 国盛证券研究所

高通: 推出全新系统级芯片 Snapdragon X, 瞄准中端 PC 市场。这款芯片采用 4 纳米制程工艺, 具备长续航和强大性能, 可支持运行最新人工智能软件, 搭载该芯片的 PC 售价低至 600 美元。

在芯片厂商的支持下, 联想、惠普、戴尔等电脑厂商纷纷加大在 AIPC 领域的研发和生产投入。在 CES 上, AIPC 终端厂商纷纷发布新品。联想发布了全球首款卷轴屏 ThinkBook Plus Gen 6 AI PC、YOGA Air 14 Aura AI、ThinkPad X9、搭载本地 AI 的平板电脑、掌机以及搭配英伟达 RTX5090 显卡的最新游戏笔记本等一系列新品; 戴尔则推出专为个人和专业计算打造的全新 AIPC 产品组合, 戴尔、戴尔 Pro 和戴尔 Pro Max; 华硕展示了一款被称为“最轻的 Copilot + PC”的产品。

图表205: 联想发布卷轴屏 ThinkBook Plus Gen 6 AI PC



资料来源: 联想, 国盛证券研究所

➤ AI眼镜：新品井喷，开启“百镜大战”。

根据 VR 陀螺统计，CES 现场展示的 AI 以及 AR 眼镜（同系列的 SKU 算一个）数量达到 47 款。AI 眼镜分为纯音频 AI 眼镜、音频+拍照 AI 眼镜、带显示 AI 眼镜三类，其中音频+拍照 AI 眼镜数量达到 16 款（单款眼镜有多个 SKU）。

图表206: CES 展出 AR+AI 眼镜

CES 2025 AI、AR眼镜统计					
品牌	产品	摄像头	光学	重量	其他卖点
1MORE万魔	G70	500万像素		/	280mAh电池、运动监测、磁吸充电、双计算核心、Q3上市
GetD格多维	A音频眼镜			/	售价30美金、支持ChatGPT交互
	AI拍摄眼镜	1200万像素		/	恒玄BES2800芯片、IPX4防水、计划售价200美金、Q2上市
Gyges Labs&Moody	Halliday		DigiWindow微投影仪	35g	系统集成AI助手
INAIR	INAIR 2 Pro空间套装		Micro-OLED+Birdbath	/	NAIR2Pro眼镜支持双目 3840*1080分辨率，120Hz高刷、电致变x色
	INAIR AI 空间计算机		Micro-OLED+Birdbath	/	增加外接键盘，AR轻办公
RayNeo雷鸟创新	雷鸟V3 AI拍摄眼镜	✓		39g	通义独家定制大模型、AR1旗舰级芯片、定价1799元
	雷鸟 Air 3		Micro-OLED+Birdbath	76g	观影眼镜、游戏眼镜
	雷鸟X3 Pro	✓	Micro-LED+衍射光波导	/	0.36cc体积、Q2上市
Rokid	Rokid Glasses	✓	Micro-LED+衍射光波导	49g	联合暴龙设计、多模态 AI 大模型
	Rokid AR Lite		Micro-OLED+Birdbath	75g	游戏、大屏观影、轻办公
	Rokid AR Studio		Micro-OLED+Birdbath	/	空间计算套装
Soliddd	SolidddVision	✓		/	针对黄斑变性的低视力辅助
星纪魅族	StarV Air2		Micro-LED+衍射光波导	44g	与Captify的合作，旨在为听障人士服务、提供语音转写
ThinkAR	AiLens		Micro-LED树脂光波导	37g	FoV达30°、搭载AmbigApollo4处理器
Vuzix	Ultralite智能眼镜参考设计	800万像素	光波导	78.3g	30°FOV、AR1平台、Android12
XREAL	XREAL Air 2 Ultra	✓	Micro-OLED+Birdbath	80g	XR观影眼镜、娱乐眼镜
	XREAL One系列		Micro-OLED+Birdbath	84g	X1芯片、3DoF悬停体验、Pro版光学为混合波导
大朋VR	DPVR AI Glasses系列	✓		/	配置具有较大差异化，计划在今年1、2月发布
	DPVR AI Glasses系列	✓		/	配置具有较大差异化，计划在今年1、2月发布
歌尔	AR参考设计Mulan 2	/	Micro-LED+全息波导	36g	超薄碳纤维框架和超轻钛合金铰链
	Wood 2	✓		48g	超轻材料前框、超小型SIP模组等创新技术
雷神科技	雷神 AURA 智能眼镜 AR 轻享版		Micro-OLED+Birdbath	79g	0-600度的屈光调节、600尼特入眼亮度
	雷神 AURA 智能眼镜 AR+AI 版	4800万像素	Micro-OLED+阵列光波导	/	2025年3月上市
	雷神 AURA 智能眼镜 AI探索版	✓		/	高通骁龙AR1平台，将支持1080P防抖视频录制
李未可	Meta Lens View	1200万像素		37g	4阵列麦克风
联想	Legion Glasses 2		Micro OLED+Birdbath	65g	FOV43.5°，最大入眼亮度可达800尼特
莫界科技	AI+AR眼镜参考设计	✓	树脂衍射光波导	50g	支持指环、微手势、触碰、语音等多模态交互
	AI+AR眼镜参考设计	✓	树脂衍射光波导	38g	半前框结构造型
奇景光电	AR眼镜参考设计				支持眼动追踪
闪极	loomos AI眼镜	✓			支持4K照片和1080P视频、GPT-4o驱动
亿境虚拟	AI眼镜全栈式解决方案	✓			AR1芯片、多AI支持

资料来源：VR 陀螺，国盛证券研究所

Halliday Glasses: 由 Gyges Lab 发布，重约 35 克，续航时间为 12 小时（待机时间 100 小时）。Halliday 没有采用波导镜片，而是在右侧镜框无缝集成一块超小屏幕 DigiWindow，将信息投射到用户的视野内。DigiWindow 技术是 Halliday 独有的智能眼镜投影方法，确保隐私的同时，能为佩戴者观众提供 3.5 英寸显示屏的视野，且不会阻挡视线。此外，该眼镜搭载了麦克风，同样具备 AI 助手功能，以实现提词器、翻译文本等显示功能，售价约为 399 美元至 499 美元，将在 2025 年第一季度正式发货。

图表207: Halliday 在镜框部分塞入 Digi Window 微型显示器



资料来源: Halliday, 国盛证券研究所

XREAL: XREAL One 系列在 CES 2025 迎来线下首次亮相, 自研 X1 芯片加持, XREAL One 本体实现了原生 3 DoF, 即无需借助任何外部硬/软件, 眼镜连接所有设备 (包括 MacBook、iPad、iPhone、Switch、安卓设备等), 就可带来超稳 3 DoF 体验 (无论头部如何旋转或晃动, 眼镜中显示的画面都稳定地悬停在空间中的同一位置)。

图表208: XREAL One 产品亮点



资料来源: XREAL, 国盛证券研究所

Vuzix: 发布面向原始设备制造商的最新 **Ultralite** 智能眼镜参考设计。Ultralite Pro 采用高通 Snapdragon AR1 Gen 1 平台, 采用全彩 Avegant LCoS 投影机+双目波导光学方案, 集成了声控命令、板载摄像头和麦克风阵列; Ultralite Audio OEM 平台基于 Vuzix Ultralite Z100 型号, 这是其音频迭代版本, 集成了双扬声器和降噪麦克风设置, 可增强音频和语音交互, 该平台配备了单色 microLED 投影仪, 电池续航时间可延长至两天, 可提供人工智能驱动的功能, 包括语音助手和实时语言翻译。

ThinkAR: 软银集团(SoftBank)旗下 ThinkAR 首次亮相了全新大众消费级 AI+AR 眼镜 AiLens, 集成了通话、音乐、翻译、导航和智能助手 AR 眼镜提升连接性, 适合通过免提技术增强日常互动。

莫界: 带来两款具有行业开创性的 **AI+AR 眼镜参考设计**。1) 全球首款采用树脂衍射光波导实现双目彩色异显的 **AI+AR 眼镜参考设计**, 该方案整机重量约 50 克, 在保持极致轻盈, 为用户实现全天候舒适佩戴的前提下, 配有摄像头及 Wi-Fi 模块, 支持指环、微手势、触碰、语音等多模态交互, 实现全场景创新的 **AI+AR 智能体验**; 2) 全球首款半框造型 **AI+AR 眼镜参考设计**, 重量约 38 克, 同样配有摄像头及 Wi-Fi 模块, 支持指环、微手势、触碰、语音等多模态交互, 从而开创出独具个性与科技美学的行业标杆案例。

图表209: ThinkAR 首次亮相 AI+AR 眼镜 AiLens



资料来源: ThinkAR, 国盛证券研究所

图表210: 莫界轻量型全彩 AR+AI 眼镜

莫界极致轻量型全彩AR+AI眼镜参考设计方案
荣获2025CES创新奖



资料来源: 多知, 国盛证券研究所

➤ **智能驾驶: AI 加码, 智能化赋能汽车。**

英伟达: 自动驾驶汽车平台 NVIDIA DRIVE AGX™ Hyperion 已通过两家业内权威的汽车功能安全和网络安全认证评估机构的行业安全评估, 是业内首个也是唯一一个端到端的自动驾驶平台, 集成了 DRIVE AGX™系统级芯片、参考板设计、NVIDIA DriveOS 汽车操作系统、传感器套件以及主动安全和 L2+驾驶堆栈, 为汽车制造商提供了一套完整的自动驾驶解决方案。梅赛德斯-奔驰、捷豹路虎、沃尔沃汽车等汽车安全领域的先锋已经开始部署 DRIVE Hyperion™平台, 以提升其自动驾驶车辆的性能和安全性。

英伟达为乘用车和商用车设计的最新一代 DRIVE Hyperion, 将采用基于 NVIDIA Blackwell 架构构建的高性能系统级芯片 DRIVE AGX Thor, 预计将于今年上半年上市。此外, 英伟达还宣布丰田、Aurora 和大陆集团加入其合作伙伴行列, 丰田将与其合作开发下一代自动驾驶汽车, 利用其先进的汽车芯片和操作系统, 进一步提升自动驾驶车辆的性能和安全性; Aurora 和大陆集团将与其合作大规模部署无人驾驶卡车。

图表211: NVIDIA DriveOS



资料来源：英伟达，国盛证券研究所

禾赛科技：展示了面向 L4 自动驾驶、ADAS 真值系统开发、港口物流自动化以及机器人与工业市场的新一代旗舰级 360° 远距激光雷达 OT128。这款激光雷达已与包括文远知行、西井科技、Embotech、EasyMile 在内的 90 多家国内外客户达成合作，并已开始量产交付。

尼康：展示了与三菱扶桑卡车公司共同打造的创新车载摄像系统，采用了一种兼具长焦和广角镜头功能的光学镜头系统，使用户能够同时看到远处的情况和检查周边环境。这种设计不仅提高了驾驶员的视野范围，还通过减少视差，提高了图像识别的准确性。该车载摄像系统还配备了先进的 AI 技术，能够准确识别路标、车辆等远距离信息，进一步提升驾驶的安全性。

大陆集团：首次公开了其“Emotional Cockpit”（情感驾驶舱），这一创新设计将车辆与驾驶员的连接性提升到了新的高度。该驾驶舱结合了施华洛世奇的独特设计元素和最新的显示技术，支持 AI 助手和互动部件功能，以提升用户的感性体验。此外，还展示了其隐形生物识别传感显示器，该技术由大陆集团与生物识别解决方案提供商 trinamiX 共同开发。使用安装在仪表盘显示器后面的摄像头和激光投影仪来跟踪车内人员的生命体征，支持各种安全和舒适功能。例如，3D 距离映射功能可优化安全气囊和其他约束装置部署，并检查安全带是否正确系好；检测驾驶员的心率等生命体征，以识别驾驶员在座位上的紧张情况或发生的医疗紧急情况。

宝马：首次亮相了其新世代超感智能座舱 BMW 全景 iDrive。四项创新技术亮眼，1) “视平线全景显示”利用超近距投影技术，在前风挡玻璃上呈现直观的车辆信息，取代了传统的仪表盘；2) “3D 抬头显示”通过调整信息的形状、角度和精度，优化了驾驶者读取车况信息的方式，使驾驶过程更加专注；3) “超感智控方向盘”结合了主动触觉反馈和分层显示技术，使驾驶者在不分散注意力的情况下轻松感知和操作各种信息；4) 最后，“向心中控”是宝马首创的车内交互界面，其不规则形状的中控屏不仅缩短了方向盘到显示屏的距离，还兼顾了后排乘客的观影效果。

吉利：展示了其最新的智能驾驶系统和车联网技术。吉利的智能驾驶系统集成了先进的传感器、算法和通信技术，实现了车辆的高度自动化驾驶。该系统不仅能够在高速公路上实现自动驾驶，还能够的城市道路和复杂路况下安全行驶。这一系统通过引入全新算力平台和加速无图化部署，使得主流新能源产品也能享受高阶智驾体验。此外，吉利还

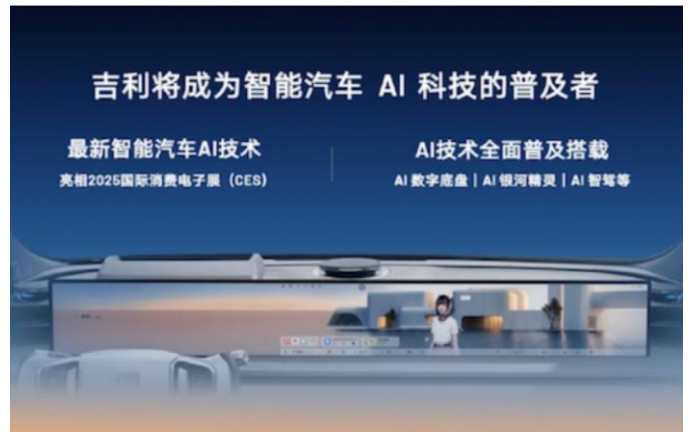
展示了其 AI 数字底盘和 AI 银河精灵等技术，这些技术的应用将进一步提升智能驾驶的性能和安全性。

图表212: 宝马推出新世代超感智能座舱 BMW 全景 iDrive



资料来源: 焉知汽车, 国盛证券研究所

图表213: 吉利展示最新智能驾驶系统



资料来源: 焉知汽车, 国盛证券研究所

相关标的

算力：中芯国际、寒武纪、海光信息、景嘉微、长电科技、通富微电

英伟达核心供应链：胜宏科技、工业富联、沪电股份、鸿腾精密科技

存储：兆易创新、普冉股份

互联：中际旭创、新易盛、天孚通信、沃尔核材、兆龙互联、立讯精密、澜起科技、龙迅股份

电源：麦格米特，欧陆通，泰嘉股份

PCB：胜宏科技、沪电股份、深南电路、景旺电子、兴森科技

消费电子：

- 消费类芯片：韦尔股份、恒玄科技、乐鑫科技、泰凌微、中科蓝讯、晶晨股份、瑞芯微、全志科技、炬芯科技等
- 智能声学硬件：歌尔股份、瑞声科技、共达电声、漫步者、国光电器、佳禾智能等
- 电池：珠海冠宇、欣旺达等
- 结构件&功能件：领益智造、长盈精密、春秋电子、莱宝高科等
- 散热：中石科技、飞荣达、思泉新材等
- FPC：鹏鼎控股、东山精密
- 光学：水晶光电、舜宇光学科技、康耐特光学、蓝特光学、宇瞳光学
- AR/VR 相关：歌尔股份、水晶光电、立讯精密、创维数字、亿道信息、兆威机电
- 代工厂&终端：华勤技术、龙旗科技、联想集团
- 智能手机：小米集团、传音控股、中兴通讯

风险提示

技术路线演进风险。以 AI 芯片为例，虽然目前 ASIC 在从推理向训练应用转变中具有潜力，但如果 GPU 技术在性能提升、能耗优化等方面取得重大突破，或者新的通用计算技术在人工智能训练场景中表现出更高的效率，可能目前的技术方案会被其他新兴技术路线所取代。

研发进展不及预期。开发先进的半导体产品需要巨额的研发投入，如果研发项目未能成功转化为可量产、有市场竞争力的产品，公司将面临研发投入无法收回的风险，对财务状况产生负面影响，甚至对相关产业链的公司造成影响。

地缘政治风险。目前美国对中国半导体展开封锁，加之特朗普上台等可能出现贸易政策收紧等风险，相关公司海外业务有可能会受到影响。

免责声明

国盛证券有限责任公司（以下简称“本公司”）具有中国证监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。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。

本报告的信息均来源于本公司认为可信的公开资料，但本公司及其研究人员对该等信息的准确性及完整性不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，可能会随时调整。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。本公司不保证本报告所含信息及资料保持在最新状态，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。

本公司力求报告内容客观、公正，但本报告所载的资料、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。投资者应当充分考虑自身特定状况，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。

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也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。

本报告版权归“国盛证券有限责任公司”所有。未经事先本公司书面授权，任何机构或个人不得对本报告进行任何形式的发布、复制。任何机构或个人如引用、刊发本报告，需注明出处为“国盛证券研究所”，且不得对本报告进行有悖原意的删节或修改。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的任何观点均精准地反映了我们对标的证券和发行人的个人看法，结论不受任何第三方的授意或影响。我们所得报酬的任何部分无论是在过去、现在及将来均不会与本报告中的具体投资建议或观点有直接或间接联系。

投资评级说明

投资建议的评级标准		评级	说明
评级标准为报告发布日后的 6 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准。	股票评级	买入	相对同期基准指数涨幅在 15%以上
		增持	相对同期基准指数涨幅在 5%~15%之间
		持有	相对同期基准指数涨幅在 -5%~+5%之间
		减持	相对同期基准指数跌幅在 5%以上
	行业评级	增持	相对同期基准指数涨幅在 10%以上
		中性	相对同期基准指数涨幅在 -10%~+10%之间
		减持	相对同期基准指数跌幅在 10%以上

国盛证券研究所

北京

地址：北京市东城区永定门西滨河路 8 号院 7 楼中海地产广场东塔 7 层

邮编：100077

邮箱：gsresearch@gszq.com

南昌

地址：南昌市红谷滩新区凤凰中大道 1115 号北京银行大厦

邮编：330038

传真：0791-86281485

邮箱：gsresearch@gszq.com

上海

地址：上海市浦东新区南洋泾路 555 号陆家嘴金融街区 22 栋

邮编：200120

电话：021-38124100

邮箱：gsresearch@gszq.com

深圳

地址：深圳市福田区福华三路 100 号鼎和大厦 24 楼

邮编：518033

邮箱：gsresearch@gszq.com