

研究所：

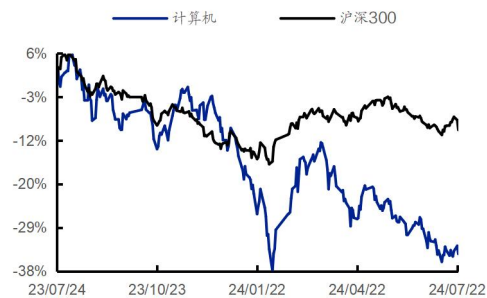
证券分析师：

刘熹 S0350523040001  
liux10@ghzq.com.cn

# 国产 AI 算力：万卡集群、多芯混合时代来临

## ——计算机“自主可控”系列报告

最近一年走势



行业相对表现

2024/07/23

表现	1M	3M	12M
计算机	-7.1%	-9.9%	-33.8%
沪深 300	-1.6%	-1.9%	-10.0%

相关报告

《计算机“智能驾驶”系列专题一：萝卜快跑订单激增，Robotaxi 商业化奇点到来（推荐）\*计算机\*刘熹》——2024-07-18

《多环节验证算力景气提速、且 2024Q3 预期乐观——AI 算力月度跟踪（202407）（推荐）\*计算机\*刘熹》——2024-07-12

《芯片散热：从风冷到液冷，AI 驱动产业革新——AI 算力“卖水人”专题系列（2）（推荐）\*计算机\*刘熹》——2024-06-25

《电力 IT 系列专题（1）：能源转型+电改深化，电力 IT 迎新机遇（推荐）\*计算机\*刘熹》——2024-06-23

《计算机事件点评：英伟达 COMPUTEX 2024：下一代 AI 平台 Rubin 将推出，拥抱 AI+机器人时代（推荐）\*计算机\*刘熹》——2024-06-04

### 投资要点：

#### ■ 美国或加剧对华出口限制，AI 芯片自主可控势在必行

据 7 月 17 日彭博社，美国拜登政府或进一步限制日本京东电子、荷兰 ASML 等全球半导体公司向中国提供先进技术。美国政府或将考虑采取最严厉贸易限制措施，扩大实施“外国直接产品规则”（FDPR）。我们认为，在国际形势持续不明朗、AI+战略升级等背景下，我国加大发展 AI 芯片自主可控或将成为必然选择。

目前，NVIDIA H20 是英伟达可在中国销售的符合美国规定的最高性能的产品，预计 5-6 月已经开始规模化出货。我们认为，在当前我国 AI 战略加速推进，以及 H20 可供货窗口期，H20 仍将有持续的需求，2024 年或呈现 H20 和国产 AI 芯片出货量同步高增的景象。

#### ■ 国产 AI 芯片：加速升级迭代，开启万卡集群时代

据中商产业研究院，2024 年中国 AI 芯片市场规模有望增长至 2302 亿元、同比+91%，其中，国产 AI 芯片市场空间广阔。厂商方面：华为海思、寒武纪、海光信息、燧原科技、摩尔线程、天数智芯等国产 AI 芯片持续迭代升级，推进万卡集群建设，阿里/百度等互联网厂加大发展自研芯片，优化卡间互联、集群建设、软件栈等能力，未来国产 AI 芯片有望实现从“能用”到“好用”。

国家发文推动智算中心建设，昇腾等国产芯片份额领先。2023 年 10 月 8 日，工信部、网信办等六部门联合印发《算力基础设施高质量发展行动计划》，提出 2025 年，算力规模超 300 EFLOPS，智算中心 50 个，智能算力占比达到 35%。截至 2023 年的 18 个国家新一代人工智能创新发展试验区中，已有 12 个城市披露采用昇腾 AI 芯片，占比高达 2/3。

#### ■ AI 服务器：积极推进多元异构，将受益于国产 AI 算力高增

AI 服务器景气度持续提升。根据 Gartner 数据，2024Q1，大中华地区服务器出货量同比+30.3%；据中商产业研究院，2024 年预计中国 AI 服务器出货量将达到 42.1 万台，景气度高增。据 IDC，2023 年，中国加速服务器下游客户中互联网占比近 60%，金融、电信等行业增速领先。

我国服务器厂商积极推进多元异构算力，如浪潮信息、紫光股份等；百度、阿里等中国互联网厂商也开始转向“多芯片混合”策略。我们认为，多元异构、多芯混合策略的推动，将有利于缓解 AI 芯片供应链问题、降低对英伟达 AI 芯片的依赖，有望进一步驱动国内 AI 服务器市场增长。

- **行业评级及投资策略：**在国际形势持续不明朗、AI+战略持续升级等背景下，我国发展 AI 芯片自主可控或将是大势所趋，国产 AI 芯片厂商加速迭代、推进万卡集群建设，或将驱动国产算力市场持续增长，维持对计算机行业“推荐”评级。
- **相关公司：**1) **AI 芯片：**寒武纪、海光信息、龙芯中科、景嘉微、英伟达、AMD、Intel；2) **服务器：**中科曙光、工业富联、浪潮信息、华勤技术、紫光股份、中国长城、协创数据、神州数码、烽火通信、拓维信息、软通动力、鸿海、纬创、广达、英业达、纬颖、超微电脑、戴尔、联想；3) **服务器液冷：**曙光数创、飞荣达、英维克、同方股份、申菱环境、高澜股份、网宿科技、奇鋆科技、双鸿、健策、VERTIV。4) **光模块：**新易盛、中际旭创、天孚通信、光迅科技、华工科技。5) **数据中心：**云赛智联、奥飞数据、光环新网、宝信软件、数据港、电科数字、世纪互联。
- **风险提示：**宏观经济影响下游需求、大模型产业发展不及预期、市场竞争加剧、中美博弈加剧、相关公司业绩不及预期等，各公司并不具备完全可比性，对标的相关资料和数据仅供参考。

## 内容目录

1、 美国或加剧对华出口限制，AI 芯片自主可控势在必行.....	5
1.1、 国际形势持续不明朗，AI 芯片自主可控势在必行.....	5
1.2、 2024 年 NVIDIA H20 或将与国产算力同步增长.....	5
2、 国产 AI 芯片：加速升级迭代，开启万卡集群时代.....	6
2.1、 国产算力加速迭代，推进 AI 万卡集群项目.....	6
2.2、 AI 芯片市场景气高增，国产化空间广阔.....	9
3、 AI 服务器：积极推进多元异构，将受益于国产 AI 算力高增.....	10
3.1、 “多元异构”或将成为服务器和 AI 用户主流选择.....	10
3.2、 服务器市场规模：通用服务器复苏，AI 服务器景气向上.....	11
3.3、 国产 AI 芯片崛起有望驱动服务器市场增长.....	13
4、 行业评级及投资策略.....	14
5、 风险提示.....	15

## 图表目录

图 1: 中国芯片市场规模预期 .....	10
图 2: 2023 年, 国产 AI 芯片出货量占比约 14% .....	10
图 3: 异构 AI 算力操作平台架构 .....	10
图 4: 中国服务器市场规模 .....	12
图 5: 中国 AI 服务器出货量 .....	12
图 6: 2022 年国内服务器厂商格局 .....	12
图 7: 2022 年国内 AI 服务器厂商格局 .....	12
图 8: 2022 年, 中国分行业 AI 服务器采购比重 .....	13
表 1: 美国对华半导体限制升级 .....	5
表 2: AI 芯片性能与美国禁令情况分析 .....	6
表 3: 国产 AI 芯片厂商介绍 .....	7
表 4: 国内主流人工智能芯片性能对比情况 .....	8
表 5: 部分使用昇腾芯片的智算中心 .....	8
表 6: 各家芯片厂商也持续推进智算中心建设 .....	9
表 7: 服务器厂商发展异构算力 .....	11
表 8: 多芯片混合的主要优势 .....	11
表 9: 2023-2024 年运营商 AI 服务器采购情况 .....	14

# 1、美国或加剧对华出口限制，AI 芯片自主可控势在必行

## 1.1、国际形势持续不明朗，AI 芯片自主可控势在必行

美国政府或进一步限制中国的先进半导体技术。7月17日彭博社报道，美国拜登政府向盟友表示，如果日本东京电(Tokyo Electron Ltd.)和荷兰 ASML 等全球半导体巨头继续向中国提供先进的半导体技术，美国政府将考虑采取最严厉的贸易限制措施，或扩大实施“外国直接产品规则”(FDPR)。

表 1：美国对华半导体限制升级

	内容
FDPR	“外国直接产品规则”：相当于对外国产品实施“长臂管辖”。即使是在美国境外生产的特定物项，如果开发或制造直接利用了哪怕是少量的美国特定受管控的软件或技术，则该物项也将受到《出口管理条例》(EAR)管辖。
过去施行情况	2022年，美国商务部工业安全局对 EAR 进行修订，新增了针对先进计算和超级计算机的 FDPR，被认为是美国政府近年来针对中国半导体领域最严厉的限制措施。此前该规则仅被用于少数被美国商务部列入“实体清单”的企业。
近期变化	7月17日彭博社报道，拜登政府正在考虑是否扩大实施“外国直接产品规则”(FDPR)。

资料来源：芯智讯，国海证券研究所

国际形势持续不明朗，AI 芯片国产化势在必行。美国 2022 年、2023 年 2 次对中国出口 AI 芯片进行了限制，分别导致英伟达高端 AI 芯片不能向中国出口。我们认为，面临国际环境持续不明朗、中美博弈加剧，英伟达高端 AI 芯片向中国的长期销售仍具备不确定性，中国市场也将不得不考虑国产化 AI 芯片的路径。

## 1.2、2024 年 NVIDIA H20 或将与国产算力同步增长

美国对中国半导体制裁加剧，随后英伟达推出 H20 等改良版 AI 芯片。2023 年 10 月 17 日，美国商务部发布针对芯片的出口禁令新规，对原有 2022 年 10 月 7 日规则进行了修改，更加严格地限制了中国购买英伟达 A100\A800\H100\H800\L40S 等高端芯片。根据 IT 之家，2023 年 11 月 9 日，英伟达开发出针对中国区的最新改良版 AI 芯片，包括 HGX H20、L20 PCIe 和 L2 PCIe。H20 采用台积电的 CoWoS 封装，理论上等于 H100 的 1/6 综合算力性能，而且增加 HBM 显存和 NVLink 互联模块以提高算力成本。

**H20 在实际使用中具备内存、生态等优势。**据 SemiAnalysis，虽然 H20 芯片在纸面性能上弱于中国厂商的国产芯片，但凭借 HBM 内存上的优势，其在使用过程中仍有一定优势。此外，英伟达具备生态优势，许多公司一开始就使用英伟达的生态系统和软件来训练自己的人工智能模型，切换基础设施也意味着额外的开支和时间。

表 2: AI 芯片性能与美国禁令情况分析

型号	Memory Capacity (GB)	Memory Bandwidth (Tbps)	Tera FLOPs	Bit length	TPP (TeraFLOPs x Bitlength)	Die size (mm <sup>2</sup> )	Performance density (TPP/Die size)	Rule 3A090. a	Rule 3A090. b
H100 SXM	80	3.4	1,979	8	15,832	814	19.4	✓	×
H20 SXM	96	4.0	296	8	2,368	814	2.9	×	×
L40S	48	0.9	733	8	5,864	608	9.6	✓	×
L40	48	0.9	362	8	2,896	608	4.8	×	✓
L20	48	0.9	239	8	1,912	608	3.1	×	×
L4	24	0.3	242	8	1,936	295	6.6	✓	×
L2	24	0.3	193	8	1,544	295	5.2	×	×
A100 SXM	40	1.6	312	16	4,992	826	6.0	✓	×
V100 SXM	16	0.9	125	16	2,000	815	2.5	×	×
RTX 4090	24	1.0	661	8	5,285	609	8.7	✓	×
RTX 4080	16	0.7	320	8	2,560	379	6.8	✓	×
AMD MI210	64	1.6	181	16	2,896	770	3.8	×	✓
AMD MI250X	128	3.2	383	16	6,128	1,540	4.0	✓	×
AMD MI300X	192	5.6	2,400	8	19,200	2,381	8.1	✓	×
Intel Gaudi2	96	2.5	700	8	5,600	826	6.8	✓	×

资料来源：IT 之家，国海证券研究所

**互联网厂商系英伟达 H20 购买主力，2024 年 H20 或与国产 AI 芯片同步增长。**据 6 月 27 日《科创板日报》，目前互联网厂商是英伟达 H20 购买主力，实际出货时间在 5、6 月份，价格对标华为昇腾 910。目前互联网厂商们购买的规模不算大，每个厂商大概为上千片 GPU。我们认为，NVIDIA H20 是 NV 目前可在中国销售的符合美国规定的最高性能的产品，预计 5-6 月已规模化出货，在当前我国 AI 战略加速推进，以及 H20 可以供货的窗口期，H20 仍将有持续的需求，2024 年或呈现 H20 和国产 AI 芯片出货量同步高增的景象。

## 2、国产 AI 芯片：加速升级迭代，开启万卡集群时代

### 2.1、国产算力加速迭代，推进 AI 万卡集群项目



1) **华为海思**: 旗下 AI 芯片为昇腾系列, 其中昇腾 910 面向训练推理场景, 昇腾 310 面向推理场景, 也研发开源 AI 框架与异构计算框架, 并且结合自身通信技术的基础, 昇腾芯片使用效率不断提升。

2) **寒武纪**: 目前自研云端芯片, 已发布思元 290、370 产品, 2022 年 9 月 WAIC 大会, 寒武纪董事长、总经理陈天石博士透露思元 590 等在研新品信息, 其采用 MLUarch05 全新架构, 提供更大内存容量和更高内存带宽, PCIe 接口也较上代实现升级。

3) **海光信息**: 持续迭代深算 DCU, 目前已推出深算二号产品, 兼容通用的“类 CUDA”环境、具备一定生态优势, 并与大股东中科曙光合作拓展集群建设能力。

我们认为, 随着海外限制持续升级, 发展 AI 芯片等产业自主可控是大势所趋。国产 AI 芯片目前产品迭代升级, 并且各家芯片公司联合服务器、交换机等厂商提升卡间互联、集群建设等能力, 软件栈架构也协同优化发展, 未来有望实现从“能用”到“好用”, 国产化比例加速提升。

表 3: 国产 AI 芯片厂商介绍

厂商	介绍
华为海思	华为旗下的半导体与器件设计公司 <b>昇腾处理器</b> : 是基于华为达芬奇架构的 NPU。2018-2019 年, 华为推出昇腾 310、910 NPU。昇腾 310 主要用于推理场景, 具有较高的能效比; 昇腾 910 芯片在 2019 年推出, 面向训练场景, 可实现业界最佳 AI 性能与能效。 <b>生态</b> : 昇思 MindSpore 开源 AI 框架、CANN 异构计算框架
寒武纪	<b>研发思元系列 AI 芯片</b> 。寒武纪提供云边端一体、软硬件协同、训练推理融合、具备统一生态的系列化智能芯片产品和平台化基础系统软件; 云端产品方面发布思元 100 芯片系列, 随后持续迭代至思元 270、思元 290 芯片和思元 370, 性能不断提升
海光信息	中科院旗下领先的高端处理器企业 <b>研发海光 CPU 与深算 DCU, 其中深算二号已经发布</b> 。海光 DCU 以 GPGPU 架构为基础, 能够支持全精度模型训练, 实现 LLaMa、GPT、Bloom 等为代表的大模型的全面应用, 与国内包括文心一言等大模型全面适配, 达到国内领先水平。 <b>兼容通用的“类 CUDA”环境</b>
天数智芯	<b>产品包括天垓与智铠两大系列</b> 。天垓系列基于通用 GPU 架构设计, 聚焦大模型训练, 天垓 100 支持 200 余种人工智能模型 (数量持续增加)。智铠系列聚焦大模型推理, 智铠 100 芯片支持 FP32、FP16、INT8 等多精度推理混合计算, 兼容 CUDA 生态
燧原科技	<b>产品包括邃思 2.0 训练芯片、邃思 2.5 推理芯片</b> : 面向数据中心, 可广泛应用于互联网、金融、教育、医疗、工业以及政务等行业。 <b>推进“燎原”生态合作计划</b> 。以全国算力一体化布局的智算中心算力网络为根基, 发展云服务、大模型、垂直模型 (ISV)、AIDC 集成部署运营运维等战略生态伙伴, 解决谁来建、谁来运与谁来用的问题, 打造从算力建设到应用的完整落地闭环。
摩尔线程	摩尔线程全功能 GPU、大模型智算加速卡 MTT S4000、AI 大模型训推一体机 MCCX D800、AI 超融合一体机 KUAE FUSION、夸娥 (KUAE) 万卡智算集群解决方案和 AIGC 应用在内的全栈 AI 产品

资料来源: 各公司官网, 各公司公众号, 海光信息公告, 寒武纪公告, 国海证券研究所

表 4: 国内主流人工智能芯片性能对比情况

型号	华为		海光	寒武纪	壁仞科技
	昇腾 310	昇腾 910	深算一号	MLU370-X4	壁砺 104P
芯片架构	HUAWEI Da Vinci	HUAWEI Da Vinci	GPGPU	Cambricon LUarch03	壁立仞架构
峰值 INT8 计算性能	16 TOPS	640 TOPS	-	256 TOPS	1024 TOPS
峰值半精度 (FP16) 性能	8 TOPS	320 TFLOPS	-	96 TFLOPS	-
显存容量	-	-	32GB HBM2	24GB LPDDR5	32GB HBM2e
最大功耗	8W	310W	350W	75 W	300 W
发布时间	2018Q4	2019Q3	2021H2	2021Q4	2022Q3

资料来源: 各公司官网, 海光信息招股书, 壁仞科技公众号, 国海证券研究所

**国家发文推动智算中心建设。**2023 年 10 月 8 日, 工信部、网信办等六部门联合印发《算力基础设施高质量发展行动计划》, 提出计算力、运载力、存储力、应用赋能等到 2025 年量化指标。到 2025 年, 计算力方面, 算力规模超过 300 EFLOPS, 智算中心 50 个, 智能算力占比达到 35%, 东西部算力平衡协调发展。

**智算中心建设提速, 昇腾处理器份额领先。**昇腾处理器性能国内领先, 成为国内智算中心的主流选择, 截至 2023 年的 18 个国家新一代人工智能创新发展试验区中, 已有 12 个城市披露采用昇腾 AI 芯片, 占比高达 2/3。此外, 海光信息、寒武纪、天数智芯、燧原科技、摩尔线程等国产 AI 芯片公司, 也与服务器厂商、地方政府合作建设智算中心, 发展自身集群建设能力。

表 5: 部分使用昇腾芯片的智算中心

超算中心	项目状态	运营时间	算力规划
天津人工智能计算中心	已运营	2023.3	300P
重庆人工智能计算中心	已运营	2023.5	400P
广州人工智能计算中心	已运营	2022.6	99P
沈阳人工智能计算中心	已运营	2022.8	300P
南京人工智能计算中心	已运营	2022.5	40P
成都智算中心	已运营	2022.5	300P
杭州人工智能计算中心	已运营	2022.5	一期 40P, 二期 100P

资料来源: 昇腾社区官网, IDC 圈, 云头条, 界面新闻, 央视新闻国海证券研究所



表 6: 各家芯片厂商也持续推进智算中心建设

公司	内容
寒武纪	智能计算集群系统: 2019 年主要系西安沣东、珠海横琴项目; 2020 年主要系江苏南京、无锡数据湖项目; 2021 年主要系江苏昆山项目; 2022 年主要系江苏南京项目 (二期); 2023 年参与台州、沈阳两地的算力基础设施建设项目。
海光信息	大股东中科曙光持续发展智算中心建设。中科曙光不断提升智算中心解决方案综合能力, 凭借分布式全闪存存储、浸没式液冷、计算服务等多项优势赋能多区域多行业数智化发展
天数智芯	天数智芯与爱特云翔合作建设千卡 GPU 算力集群, 可支持千亿以上参数大模型的全量预训练; 天数智芯参与建设的中国移动智算中心 (呼和浩特) 项目已正式上线运营。
燧原科技	目前, 燧原科技已在四川成都、湖北宜昌和甘肃庆阳展开智算中心建设及合作。其中, 宜昌点军智算中心仅用 1 年时间完成 300P 国产算力建设和交付, 并实现上线全消纳, 成为智算中心建设标杆案例。
摩尔线程	7 月 3 日, 摩尔线程宣布 AI 旗舰产品夸娥 (KUAE) 智算集群解决方案实现重大升级, 从千卡级别扩展至万卡规模, 其以全功能 GPU 为底座, 旨在打造国内领先的、能够承载万卡规模、具备万 P 级浮点运算能力的国产通用加速计算平台。 摩尔线程联合中国移动、中国联通、北京德道信科集团等分别就三个万卡集群项目进行战略签约。

资料来源: 各公司官网, 各公司公告, 国海证券研究所

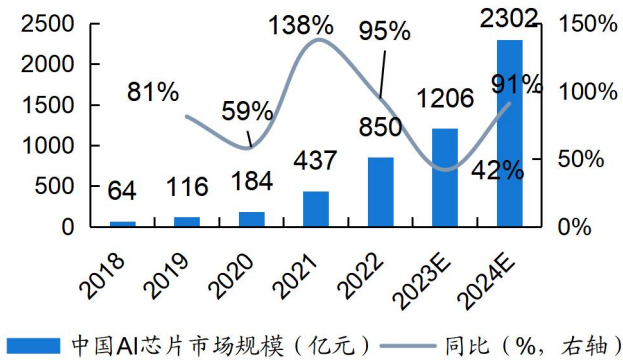
**互联网厂商也持续推动自研芯片发展。**1) **阿里平头哥:** 是阿里巴巴集团的全资半导体芯片业务主体。平头哥拥有端云一体全栈产品系列, 涵盖数据中心芯片、IoT 芯片等, 实现芯片端到端设计链路全覆盖。2) **字节:** 表示自研专用芯片, 未来将服务于抖音、西瓜视频等视频业务, 并通过火山引擎视频云开放给企业客户。3) **百度昆仑芯:** 前身为百度智能芯片及架构部, 布局 AI 加速领域, 深耕十余年, 已发布昆仑芯 1 代与 2 代产品。

## 2.2、AI 芯片市场景气高增，国产化空间广阔

**中国 AI 芯片市场规模快速增长。**据中商产业研究院, 2023 年中国 AI 芯片市场规模预计为 1206 亿元, 2024 年有望增长至 2302 亿元、同比+91%, 行业市场规模不断增长, 景气度较高。

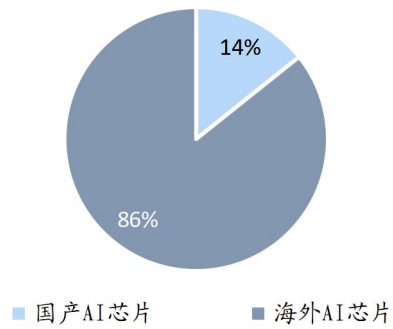
**AI 芯片国产化空间广阔。**据 IDC, 2023 年, 中国加速芯片的市场规模达到近 140 万张, 其中 GPU 卡占据 85% 的市场份额; 从品牌角度来看, 中国本土人工智能芯片品牌的出货量已超过 20 万张、占比约 14%。据 IDC, 2022 年, 中国 AI 加速卡 (公开市场) 出货量约 109 万张, 其中英伟达市场份额 85%, 华为约 10%, 百度约 2%, 寒武纪和燧原科技均为 1%。因此在国内厂商中, 华为市场份额较高。

图 1：中国芯片市场规模预期



资料来源：中商产业研究院，国海证券研究所

图 2：2023 年，国产 AI 芯片出货量占比约 14%



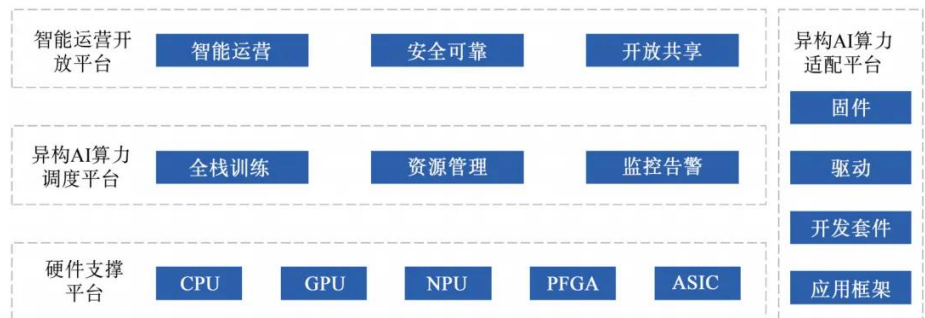
资料来源：IDC，国海证券研究所

### 3、AI 服务器：积极推进多元异构，将受益于国产 AI 算力高增

#### 3.1、“多元异构”或将成为服务器和 AI 用户主流选择

中国的服务器厂商积极推进多元异构算力。根据中国信通院报告，异构算力是指 CPU、GPU、FPGA、ASIC 等多种算力协同的处理体系，能够满足不同场景中的应用需求，实现计算效力最大化。我国主流的服务器厂商均在积极推进多元异构算力，以满足客户对多样 GPU 算力的需求。

图 3：异构 AI 算力操作平台架构



资料来源：中国信通院 CAICT 公众号

**表 7: 服务器厂商发展异构算力**

公司	内容
浪潮信息	公司作为算力系统供应商，支持多元异构算力，可适配多种架构的 GPU，与很多国内外服务器部件厂商均保持合作关系
紫光股份	2024 年 5 月，紫光股份旗下新华三集团发布全新多元异构算力平台 H3C UniServer R5330 G7 和 H3C UniServer R5500 G7，以多元异构、开放包容、灵活高效等多重优势，提供高性能、高可靠的多元算力

资料来源：金融界，新紫光集团官网，国海证券研究所

**中国互联网厂商也开始转向“多芯片混合”策略。**7月4日 Digitimes 报道，为解决人工智能（AI）芯片算力问题，中国 AI 公司正实施“多芯片混合”的策略来提高 AI 计算能力，并避免供应链安全问题。多芯片混合是将不同芯片组成一个训练集群。例如：百度表示其可组合来自不同供应商的 GPU 并将其用于 AI 训练。阿里巴巴自 2021 年以来一直致力于“一云多芯片”解决方案，且阿里云已开始转向使用基于以太网的高性能网络。

**表 8: 多芯片混合的主要优势**

公司	内容
提升训练速度	利用多个不同型号的 GPU 并行训练，来共同提高大语言模型（LLM）训练速度
更好利用内存	同时可以处理更多数据，可更好利用内存
降低使用成本	中国厂商可降低对于更昂贵的英伟达芯片的依赖，降低成本

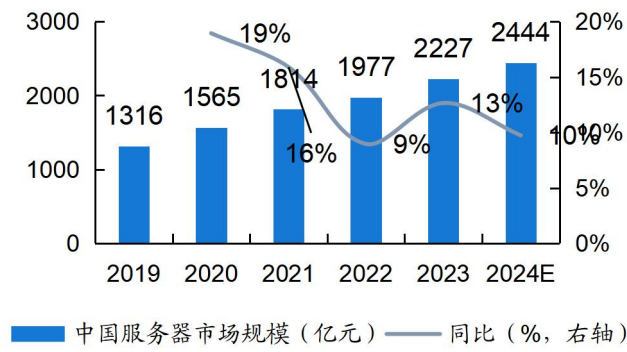
资料来源：芯智讯微信公众号，国海证券研究所

## 3.2、服务器市场规模：通用服务器复苏，AI 服务器景气向上

**2024Q1 服务器产业持续高景气。**据中商产业研究院，我国服务器市场规模由 2019 年的 1315.82 亿元增长至 2023 年的 2226.78 亿元，复合年均增长率达 11%。根据 Gartner 数据，2024Q1，全球服务器市场销售额 407.5 亿美元，同比增长 59.9%，出货量 282.0 万台，同比增长 5.9%。从地域来看，随着宏观经济的复苏，Q1 出货量上，成熟亚太地区、新兴亚太地区、拉丁美洲、大中华地区同比增速分别为 12.5%、6.5%、3.9%、30.3%。

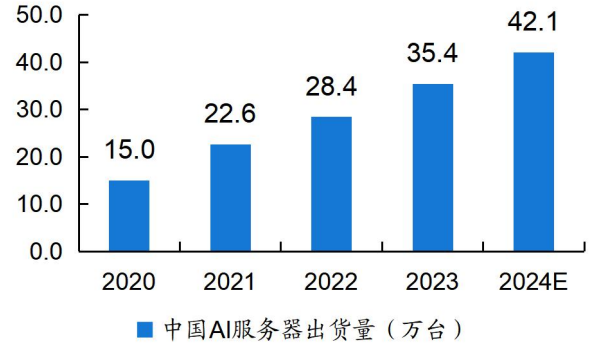
**中国 AI 服务器市场规模高速增长。**据中商产业研究院，2023 年约为 35.4 万台，2024 年预计中国 AI 服务器出货量将达到 42.1 万台。

图 4: 中国服务器市场规模



资料来源: 中商产业研究院, IDC, 国海证券研究所

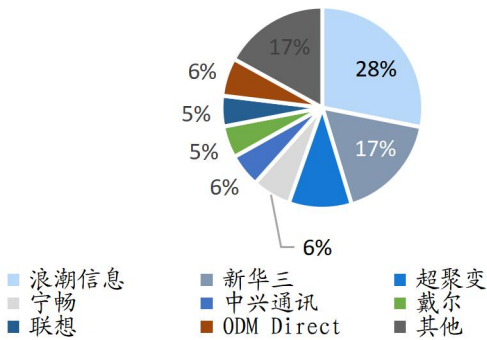
图 5: 中国 AI 服务器出货量



资料来源: 中商产业研究院, 国海证券研究所

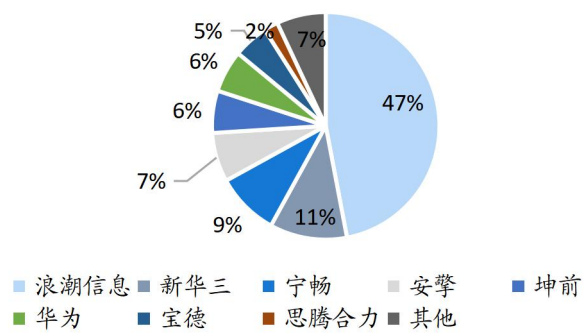
浪潮、新华三、宁畅在国内加速服务器市场份额领先。2023 年, 从厂商销售额角度看, 浪潮、新华三、宁畅位居前三, 占据了近 70% 的市场份额; 从服务器出货台数角度看浪潮、坤前、新华三位居前三名, 占有超过 50% 以上的市场份额。据华经产业研究院, 2022 年我国 AI 服务器市场 (按销售额), 浪潮以 46.6% 份额占据首位, 新华三、宁畅分别以 11% 和 9% 占据第二、第三名。

图 6: 2022 年国内服务器厂商格局



资料来源: IDC, 中商产业研究院, 国海证券研究所

图 7: 2022 年国内 AI 服务器厂商格局

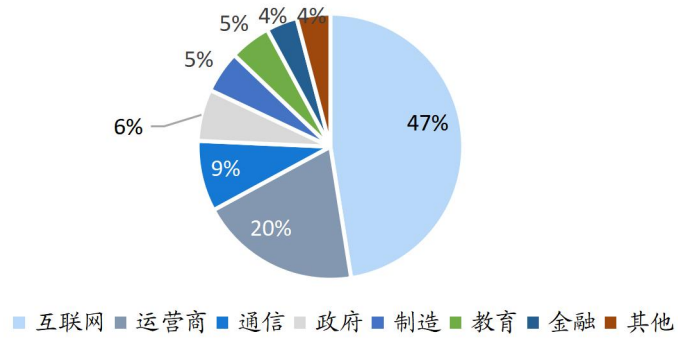


资料来源: 华经产业研究院, 国海证券研究所

加速服务器采购: 互联网采购量级大, 金融、电信等行业增速领先。据 IDC, 2023 年, 从行业的角度看, 互联网依然是最大的加速服务器采购行业, 占整体加速服务器市场近 60% 的份额, 此外金融、电信、交通和医疗健康等多数行业均有超过一倍以上的增长。

国内互联网厂商服务器需求较高。据 7 月 1 日 Trendforce, 在 CSPs 动能上, ByteDance 因新业务的需求, 拉大全年服务器采购; Alibaba 与 Tencent 也因换机周期, 纷纷于近期上修服务器整机订单。

图 8：2022 年，中国分行业 AI 服务器采购比重



资料来源：观知海内，国海证券研究所

### 3.3、国产 AI 芯片崛起有望驱动服务器市场增长

算力是我国 AI+战略的重要支撑，国产 AI 芯片崛起将加速服务器市场增长。面对发展人工智能的迫切需求以及英伟达供应链的不确定性，国内市场积极寻找算力的“第二选择”。随着国产 AI 芯片性能迭代、互联能力与集群建设能力增强，被互联网、运营商、政企客户等广泛接受，我国昇腾、海光、寒武纪等国产算力需求有望激增。目前，中国移动、中国电信均已规划针对国产算力的万卡算力集群，且在不断加大在国产 AI 算力的投入。我们认为，随着国产 AI 芯片的出货量增长，相关服务器厂商有望持续受益。

**表 9: 2023-2024 年运营商 AI 服务器采购情况**

公司	项目名称		总数量	总金额	AI 服务器数量	国产化比例
中国移动	《2023 年至 2024 年新型智算中心(试验网)采购》	标包 1-11	人工智能服务器 1204 台、高性能无损交换机 324 台(含集群管理软件 2050 端口)、集群管理服务器 2 台、RoCE 交换机 204 台、分布式文件存储 119PB、智算资源池化软件(虚拟化软件) 608 套	-	人工智能服务器 1204 台	-
		标包 12	扣卡风冷型特定场景 AI 训练服务器 106 台, 扣卡液冷型特定场景 AI 训练服务器 1144 台	24.74 亿元	人工智能服务器及配套产品 1250 台	-
	《2024 年至 2025 年新型智算中心采购招标》		7994 台人工智能服务器及配套产品、60 台白盒交换机, 总计 8054 台设备	191 亿元	训练服务器 7994 台	中标候选人: ① 昆仑技术 21.05%; ② 华坤振宇 17.54%; ③ 宝德计算机份额 15.79%; ④ 百信份额 14.04%; ⑤ 长江计算份额 12.28%; ⑥ 神州鲲泰份额 10.53%; ⑦ 湘江鲲鹏份额 9.77%
中国联通	《2024 年中国联通人工智能服务器集中采购项目》		2503 台人工智能服务器, 688 台关键组网设备 RoCE 交换机, 总计 3191 台	20.46 亿元	人工智能服务器 2503 台	昆仑、宝德、虹信和长江 4 家入围
中国电信	《AI 算力服务器 (2023-2024 年) 集中采购项目》		AI 服务器 4175 台	84.63 亿元	AI 服务器 4175 台	I 系列规模为 2198 台, G 系列规模为 1977 台。I 系列 CPU 采用 Intel 至强可扩展处理器, G 系列 CPU 采用鲲鹏处理器
	《服务器 (2024-2025 年) 集中采购项目》		服务器 15.6 万台	-	GPU 服务器 13135 台	G 系列(国产化系列)数量达 10.53 万台, 占比达 67.5%; GPU 服务器中, A 系列 6295 台, G 系列 6840 台

资料来源: 中国移动采购与招标网, C114 通信网, 中国联通采购与招标网, 中国电信阳光采购网, IT 之家, 国海证券研究所

## 4、行业评级及投资策略

在国际形势持续不明朗、AI+战略持续升级等背景下,我国发展 AI 芯片自主可控或将是顺势所趋,国产 AI 芯片厂商加速迭代、推进万卡集群建设,或将驱动国产算力市场持续增长,维持对计算机行业“推荐”评级。

相关公司:

1) **AI 芯片:** 寒武纪、海光信息、龙芯中科、景嘉微、英伟达、AMD、Intel。

2) **服务器:** 中科曙光、工业富联、浪潮信息、紫光股份、华勤技术、中国长城、协创数据、神州数码、烽火通信、拓维信息、软通动力、鸿海、纬创、广达、英业达、纬颖、超微电脑、戴尔、联想。

3) **服务器液冷:** 曙光数创、飞荣达、英维克、同方股份、申菱环境、高澜股份、网宿科技、奇鋆科技、双鸿、健策、VERTIV。



4) **光模块**: 新易盛、中际旭创、天孚通信、光迅科技、华工科技。

5) **数据中心**: 云赛智联、奥飞数据、光环新网、宝信软件、数据港、电科数字、世纪互联。

## 5、风险提示

1) **宏观经济影响下游需求**: 宏观经济环境下行, 将影响客户对信息化基础设施的采购需求;

2) **大模型产业发展不及预期**: 行业的核心驱动力是人工智能大语言模型的训练和推理对算力的需求, 如果大模型行业发展不及预期将会影响 AI 算力相关需求;

3) **市场竞争加剧**: IT 产品和服务行业是成熟且完全竞争的行业, 新进入者可能加剧整个行业的竞争态势;

4) **中美博弈加剧**: 国际形势持续不明朗, 美国不断通过“实体清单”等方式对中国企业实施打压, 若中美紧张形势进一步升级, 将可能导致中国半导体供应链供应受到影响;

5) **相关公司业绩不及预期**: 市场环境变化、公司治理情况变化、其他非主营业务经营不及预期等原因或将导致相关公司的整体业绩不及预期。

6) 各公司并不具备完全可比性, 对标的相关资料和数据仅供参考。

## 【计算机小组介绍】

刘熹，计算机行业首席分析师，上海交通大学硕士，多年计算机行业研究经验，致力于做前瞻性深度研究，挖掘产业投资机会。新浪金麒麟新锐分析师、Wind 金牌分析师团队核心成员。

## 【分析师承诺】

刘熹，本报告中的分析师均具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观的出具本报告。本报告清晰准确的反映了分析师本人的研究观点。分析师本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收取到任何形式的补偿。

## 【国海证券投资评级标准】

### 行业投资评级

推荐：行业基本面向好，行业指数领先沪深 300 指数；  
 中性：行业基本面稳定，行业指数跟随沪深 300 指数；  
 回避：行业基本面向淡，行业指数落后沪深 300 指数。

### 股票投资评级

买入：相对沪深 300 指数涨幅 20%以上；  
 增持：相对沪深 300 指数涨幅介于 10%~20%之间；  
 中性：相对沪深 300 指数涨幅介于-10%~10%之间；  
 卖出：相对沪深 300 指数跌幅 10%以上。

## 【免责声明】

本报告的风险等级定级为 R3，仅供符合国海证券股份有限公司（简称“本公司”）投资者适当性管理要求的客户（简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。客户及/或投资者应当认识到有关本报告的短信提示、电话推荐等只是研究观点的简要沟通，需以本公司的完整报告为准，本公司接受客户的后续问询。

本公司具有中国证监会许可的证券投资咨询业务资格。本报告中的信息均来源于公开资料及合法获得的相关内部外部报告资料，本公司对这些信息的准确性及完整性不作任何保证，不保证其中的信息已做最新变更，也不保证相关的建议不会发生任何变更。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。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。报告中的内容和意见仅供参考，在任何情况下，本报告中所表达的意见并不构成对所述证券买卖的出价和征价。本公司及其本公司员工对使用本报告及其内容所引发的任何直接或间接损失概不负责。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或者金融产品等服务。本公司在知晓范围内依法合规地履行披露义务。

## 【风险提示】

市场有风险，投资需谨慎。投资者不应将本报告视为作出投资决策的唯一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向本公司或其他专业人士咨询并谨慎决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。投资者务必注意，其据此做出的任何投资决策与

本公司、本公司员工或者关联机构无关。

若本公司以外的其他机构（以下简称“该机构”）发送本报告，则由该机构独自为此发送行为负责。通过此途径获得本报告的投资者应自行联系该机构以要求获悉更详细信息。本报告不构成本公司向该机构之客户提供的投资建议。

任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。本公司、本公司员工或者关联机构亦不为该机构之客户因使用本报告或报告所载内容引起的任何损失承担任何责任。

### 【郑重声明】

本报告版权归国海证券所有。未经本公司的明确书面特别授权或协议约定，除法律规定的情况外，任何人不得对本报告的任何内容进行发布、复制、编辑、改编、转载、播放、展示或以其他方式非法使用本报告的部分或者全部内容，否则均构成对本公司版权的侵害，本公司有权依法追究其法律责任。