

计算机行业深度报告

国产芯片与超节点放量在即

增持（维持）

2026 年 08 月 29 日

证券分析师 王紫敬

执业证书：S0600521080005

021-60199781

wangzj@dwzq.com.cn

证券分析师 宋鑫宇

执业证书：S0600526050001

songxy@dwzq.com.cn

投资要点

■ **需求与政策共振，国产替代进入放量兑现期。**据 IDC 与浪潮《2025 年中国人工智能算力发展评估报告》，2023 至 2028 年我国智能算力规模的年复合增长率预计达 46.2%，2026 年规模有望达到约 1460.3 EFLOPS。同时需求重心由训练转向推理，IDC 预测到 2027 年推理将占智能算力需求的 70%以上，正是超节点大显身手的主战场。政策端，算力网被列为国家六张网之一，十五五期间新增直接投资约 4 万亿元，提供长周期强确定性需求。管制端，美国出口管制转向看总部归属，客观上加速国产算力的全栈自主可控。据 IDC，2025 年国产 AI 加速卡份额首次突破四成、约 41%，叠加运营商与国资智算中心集采，国产替代已从送样验证走向规模出货。

■ **芯片厂商梯队分明，软件栈从可用走向好用。**国产 AI 芯片已形成一超（华为昇腾）、多强（寒武纪、平头哥、昆仑芯、海光信息）与新势力（摩尔线程、沐曦股份、壁仞科技、天数智芯）的清晰梯队：2025 年昇腾出货约 81.2 万张居国产第一，寒武纪上半年营收 59.96 亿元、净利 23.11 亿元进入兑现期，四小龙密集资本化。供给卡点集中在制程、HBM、先进封装，预计 2027 年起缓解。更关键的是软件栈加速成熟：DeepSeek-V4 实现九种国产芯片 Day0 适配、华为 CANN 全面开源。智谱 GLM-5.3-Flash 率先在大规模流量场景中依托国产芯片集群对外提供服务，其单 token 成本已与主流英伟达 GPU 相当。国产芯片正从可用走向好用、性价比越来越高。

■ **范式跃迁，超节点是国产算力的最优解。**单卡受制于先进制程与 HBM，国产转向以高速互联把成百上千颗芯片耦合为一台超级计算机，以系统级能力弥补单卡差距。华为 CloudMatrix 384 可提供约 300 PFLOPS 稠密算力，多项系统指标对标乃至超越英伟达 GB200 NVL72。Atlas 950 SuperPoD 最多可支持 8192 张昇腾 950DT，计划于 2026 年第四季度批量上市。中科曙光曙光 8000 建成国内首个全国产十万卡超集群。需求端亦有实证：中国移动集中采购 6208 张 AI 加速卡、折合约 776 套计算节点，昇腾 384 超节点商用超 750 套。市场往往高估国产单卡追赶英伟达的短期难度，却低估其在系统层面的赢面与兑现速度——这正是国产算力被低估之处。

■ **价值向系统环节扩散，交换与整机弹性最大。**超节点把价值从单颗芯片向交换、整机、连接、存储、散热等环节扩散。高速交换用量随超节点普及成倍提升，直接利好锐捷、新华三等国产交换与网络龙头。超节点本体交付放量，带动浪潮、华勤等整机与 ODM 量价齐升。并延伸至 HBM、先进封装、光模块与 Scale-up 标准之争。随着华为、中科曙光曙光、中兴通讯、新华三、浪潮信息订单自 2026 年下半年集中交付，数千亿级市场进入兑现期，建议沿计算核心、交换/网络、整机/ODM 三条主线布局。

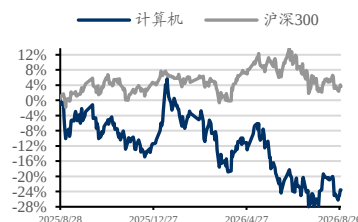
■ 相关标的

国产芯片：寒武纪、海光信息、沐曦股份、摩尔线程等。

国产超节点：紫光股份、中科曙光、浪潮信息、华勤技术、锐捷网络、中兴通讯等

■ **风险提示：**先进制程与产能瓶颈。HBM 与先进封装瓶颈。软件栈与生态发展不及预期。需求与交付不及预期风险。互联/交换芯片与连接瓶颈。外部环境风险。

行业走势



相关研究

《世界模型：从“生成世界”到“控制世界”，物理 AI 打开工业软件与自动化重估空间》

2026-08-27

《液冷：下一代 AIDC 基础设施，Rubin 带来行业确定性》

2026-07-22

内容目录

1. 需求与约束：算力需求增长、算力网建设与出口管制	4
1.1. 智能算力指数级增长，推理成为主体，算力网加速成网	4
1.2. 出口管制持续升级，卡脖子倒逼全栈自主可控	4
1.3. 需求高增下的国产算力放量	4
2. 国产 AI 芯片全景：厂商梯队分明	5
2.1. 竞争格局：一超、多强与新势力	5
2.2. 出货与份额：国产化率突破四成	5
2.3. 龙头拆解：昇腾/寒武纪/海光信息	6
2.4. GPU 新势力：从追赶到资本化	6
2.5. 供给卡点与全面自主可控：晶圆制造、HBM、先进封装	7
2.6. 大模型加速国产软件栈建设：迈过最后一公里	7
3. 范式跃迁：从单卡追赶到超节点系统竞争	8
3.1. 三堵墙：显存墙、通信墙、能效墙	8
3.2. 全球范式：NVIDIA GB200 NVL72	8
3.3. 超节点的定义与三大核心技术	9
3.4. 系统竞争如何重构价值分配	9
4. 国产超节点全景：放量元年	9
4.1. 华为 — 国产超节点引领者	10
4.2. 群雄逐鹿：中兴通讯/新华三/浪潮信息/中科曙光/超聚变	11
4.3. 曙光 8000 与海光 DCU：全国产十万卡超集群标杆	11
4.4. 开放超节点与 Scale-up 标准之争	12
4.5. 超节点产业链与受益环节	13
5. 风险提示	14

图表目录

图 1: 昇腾 950 超节点 (Atlas 950 SuperPoD) 真机首次公开亮相..... 10

图 2: 中科曙光 ScaleX40 超节点 12

表 1: 国产主要 AI 芯片厂商梳理 5

表 2: 2025 年中国 AI 加速卡出货结构 6

表 3: 国产主要超节点厂商产品..... 11

表 4: 主要 Scale-up 互联标准对比..... 13

1. 需求与约束：算力需求增长、算力网建设与出口管制

1.1. 智能算力指数级增长，推理成为主体，算力网加速成网

大模型规模化落地推动智能算力进入指数级增长通道。先进计算与算力已成为数字经济的核心底座。据 IDC 与浪潮《2025 年中国人工智能算力发展评估报告》，2023—2028 年中国智能算力规模五年复合增长率预计达 46.2%；2026 年将达约 1460.3 EFLOPS。

算力需求重心由训练转向推理，恰是超节点的用武之地。IDC 预测，到 2027 年推理将占智能算力需求的 70%以上。高并发、低时延、高性价比、MoE 亲和的推理系统成为主战场，与超节点存算强化、网络强化、MoE 亲和的设计取向高度契合。

算力网被列为国家六张网之一，上升为十五五新型基础设施的战略部署。2026 年 4 月，中央政治局会议首次明确提出加强六张网建设——水网、新型电网、算力网、新一代通信网、城市地下管网、物流网。其中算力网建设十五五期间新增直接投资约 4 万亿元。算力网由行业概念上升为国家基础设施底座，为国产算力提供长周期、强确定性的需求支撑。

算力网建设对国产算力是需求放大器。一体化算力网把分散的智算资源并网调度、提升利用率，既扩大了对国产 AI 芯片与超节点的采购需求，也强化了自主可控的底层要求——算力作为新型基础设施与国家竞争力，其供应链安全成为刚性约束。据国家数据局，2026 年国家将加快构建全国一体化算力网，推动算力一张网、协同调度、市场化资源。

1.2. 出口管制持续升级，卡脖子倒逼全栈自主可控

美国对华先进 AI 芯片管制自 2022 年多轮加码，2026 年由看收货地转向看总部归属。据国际电子商情报道，2026 年 5 月 31 日美国商务部新规明确，对总部位于中国的实体（即便采购主体在境外）出口美国先进 AI 芯片须提前申领许可证，堵住绕道漏洞。

外部约束客观上加速了国产算力的全栈自主可控进程。英伟达 CEO 黄仁勋亦公开表示已基本将中国市场让给中国公司。管制之下，中国算力被推向全栈自主可控的系统工程——从芯片设计、制程代工、HBM、先进封装，到互联协议、交换芯片、光模块、整机与软件栈，任何一环被卡都可能影响整体供给；而随着软件栈成熟与产品持续迭代，国产芯片正从能用走向好用、性价比越来越高，国产替代的确定性稳步抬升。

1.3. 需求高增下的国产算力放量

国产替代已由送样验证迈入规模出货、业绩兑现阶段。从总量看，国产 AI 芯片可及市场（TAM）有望随智能算力高增与国产化率提升而持续扩张。据 IDC，2025 年，中国市场 AI 加速卡总出货量约 400 万张，其中本土厂商合计出货约 165 万张，国产份额首次突破四成、达约 41%；英伟达以约 220 万张（约 55%）保持第一但份额被侵蚀。国

产化率快速提升，叠加运营商与国资智算中心集采，标志国产替代进入深水区。

超节点市场空间有望达数千亿级。市场普遍将国产超节点视为数千亿级新蓝海，2026 年为放量元年，华勤技术、浪潮信息、紫光股份、中科曙光等相关订单有望自 2026 年下半年进入集中交付。

2. 国产 AI 芯片全景：厂商梯队分明

2.1. 竞争格局：一超、多强与新势力

国产 AI 芯片已形成多层次格局。一超为华为昇腾；多强含寒武纪、阿里平头哥、百度昆仑芯、海光信息。另有摩尔线程、沐曦股份、壁仞科技、燧原科技、天数智芯等新兴势力。路线互补：ASIC 专用（昇腾/寒武纪）、GPGPU 通用并存。

表1：国产主要 AI 芯片厂商梳理

厂商	代表产品	路线/特点	公开进展
华为昇腾	昇腾 910B/910C、950PR	达芬奇架构+全栈+超节点	据华为，950PR 为系列首款自研 HBM 芯片，全系超节点适配 DeepSeek-V4
寒武纪	思元 590/690(MLU)	专用 AI 加速(ASIC)	2026H1 营收 59.96 亿元(+108%)、归母净利润 23.11 亿元(+123%)、毛利率约 55%
海光信息	深算 DCU+ 海光 CPU	类 CUDA 生态+x86 协同	CPU+DCU 完整方案落地高端场景
阿里平头哥	含光/PPU	云侧自用	据 IDC，国产出货份额居前
百度昆仑芯	昆仑芯 P800	自研 XPU+百度智能云	据 IDC，份额居国产前列
摩尔线程	MTT 系列	全功能 GPU	2025.12 科创板上市
沐曦股份/壁仞科技/燧原科技/天数智芯	曦云/壁砺/云燧/天垓、智铠	GPGPU	集体资本化

数据来源：华为官网、IDC、财联社、证券时报、与非网、东吴证券研究所

2.2. 出货与份额：国产化率突破四成

2025 年国产份额向头部集中。据 IDC，2025 年中国厂商合计出货约 165 万张，市场份额达到 41%。其中，华为昇腾出货量国产第一、整体第二，出货量约 81.2 万张（占比约 20%）；阿里平头哥约 26.5 万张（占比约 6.6%）；百度昆仑芯与寒武纪各约 11.6 万张并列国产第三；海光信息约 8.25 万张。昇腾、寒武纪、海光信息等构成国产算力基本盘。

表2：2025 年中国 AI 加速卡出货结构

厂商/阵营	出货(万张)	份额	说明
英伟达	220	55.0%	仍居第一，份额被侵蚀
华为昇腾	81.2	20.3%	国产第一、整体第二
阿里平头哥	26.5	6.6%	国产第二
百度昆仑芯	11.6	2.9%	并列第三
寒武纪	11.6	2.9%	增速黑马
海光信息	8.25	2.1%	CPU+DCU 差异化
其他	40.9	10.2%	

数据来源：IDC、观察者网、与非网、东吴证券研究所

2.3. 龙头拆解：昇腾/寒武纪/海光信息

华为昇腾——全栈生态 + 系统能力最强。昇腾以芯片—CANN—昇思—Atlas—超节点/集群全栈见长。2026 年华为推出 950PR（系列首款自研 HBM 芯片），并公布昇腾 950/960/970 三年路线图。其最大差异化不在单卡，而在把系统做大的工程能力。

寒武纪——专用 AI 加速的增速黑马，业绩进入兑现期。寒武纪 2026H1 营收 59.96 亿元、同比增长 108.13%，归母净利润 23.11 亿元、同比增长 122.61%，扣非净利润 21.66 亿元、同比增长 137.30%，毛利率约 55.25%；Q2 营收 31.11 亿元、净利润 12.98 亿元。公司在 CSP 客户规模化训练场景取得拓展；据其股权激励考核，2026 年营收目标不低于 135 亿元、2026—2028 年累计营收不低于 1000 亿元。

海光信息——CPU+DCU 协同的差异化路线，生态友好度较高。海光信息以 C86 系列 CPU 叠加深算 DCU 形成完整国产算力方案，DCU 采用类 CUDA 生态、迁移成本相对可控，在石油勘探、气象、金融等高端场景实现替代。

2.4. GPU 新势力：从追赶到资本化

摩尔线程——国产 GPU 第一股，营收高增、亏损收窄。摩尔线程 2026 上半年营收 17.36 亿元、同比增长 147.42%（已超 2025 全年 15.06 亿元），归母净亏损约 1156 万元（同比大幅收窄 95.7%），研发投入 7.69 亿元、占营收约 44.3%；据证券时报，2026 年 8 月公司启动 H 股赴港上市。

沐曦股份——曦云 C600 量产在即，C700 对标 H100。其首款全国产通用 GPU 曦云 C600 性能介于 A100 与 H100 之间、规划 2026 年上半年量产，下一代 C700 对标 H100、预计 2026 年下半年流片。

壁仞科技——港股 GPU 第一股，18C 最大募资。壁仞科技 H 股于 2026 年 1 月 2 日上市，发行价 19.60 港元、募资约 55.83 亿港元（港交所 18C 章实施以来最大募资）；募集净额约 85% 用于研发，旗舰 BR20X 计划 2026 年上市。

天数智芯——GPGPU 训推一体，港股上市，快速拥抱大模型生态。天数智芯（Iluvatar CoreX）于 2026 年 1 月 8 日登陆港交所。公司 2021 年推出国内首款通用 GPU 天垓并量产，形成天垓（训练）+智铠（推理）产品线，天垓 Gen4 规划 2026 年二季度推出、三季度量产。天数智芯曾一天内完成与 DeepSeek-R1 的适配，并在 DeepSeek-V4 发布时位列 Day0 适配的国产芯片之一，模型—芯片协同响应迅速。

2.5. 供给卡点与全面自主可控：晶圆制造、HBM、先进封装

国产 AI 芯片放量的核心约束，集中在晶圆制造、HBM、先进封装三大供给卡点，当前供给仍显不足。着眼后续，全面自主可控的供给体系正加速补位，支撑国产芯片持续出货放量。综合看，2026 年国产算力或呈需求增长、供给受限的紧平衡，晶圆制造/HBM/封装是核心堵点；随长鑫 HBM3 上量、先进封装扩产与先进制程产能释放，2027 年起供给瓶颈有望逐步缓解，为国产芯片出货放量提供全栈自主可控的产能支撑。

2.6. 大模型加速国产软件栈建设：迈过最后一公里

软件栈与生态是国产芯片可用性的最后一公里，而国产大模型的发展正倒逼软件栈快速成熟。芯片可用不仅取决于硬件算力，更取决于算子库、编译器、训练/推理框架与迁移工具链的完备度。以 DeepSeek 为代表的国产大模型开源与规模化落地，正推动模型—芯片协同适配，成为国产软件栈成熟的最强牵引。

DeepSeek-V4 引领多芯片 Day0 适配，国产软件栈实现从 0 到 1 的规模化验证。2026 年 4 月 24 日，DeepSeek 发布 DeepSeek-V4 系列（Pro 1.6T 与 Flash 284B）；众智 FlagOS 社区第一时间完成多芯片适配，通过统一开源系统软件栈 FlagOS 支持海光、沐曦股份、华为昇腾、摩尔线程、昆仑芯、平头哥、天数智芯、清微智能等 9 种 AI 芯片 Day0 适配、快速迁移稳定运行；DeepSeek-V4 还完成了对昇腾 950PR、天数智芯、寒武纪等主流国产芯片的全栈深度适配。

大模型还在反向加速算子开发，让国产芯片从能跑走向好用。FlagOS 核心算子库 FlagGems（基于 Triton 语言）已针对 DeepSeek-V4 推理链路的 MoE 专家调度、Attention、RMSNorm 等核心算子深度适配优化，并通过 C++ Wrapper 支持英伟达、摩尔线程、沐曦股份、天数智芯、寒武纪等多家 AI 芯片。大模型驱动的算子与编译栈迭代，正把迁移难、调优难的国产芯片痛点快速填平，显著提升好用度。

昇腾 CANN 全面开源，软件栈主导权正演变为关键竞争壁垒。华为在 2026 鲲鹏昇

腾开发者大会上进一步推进全栈开源：昇腾 CANN 已上线 50 余个代码仓、800 余个算子，兼容 PyTorch、vLLM、Triton 等主流框架，灵衢协议及五大软件组件全部开源；叠加国产框架昇思 MindSpore、飞桨 PaddlePaddle 的适配完善，国产软件栈的迁移门槛显著下降。软件栈成熟是国产芯片好用、规模可用的关键一环，也是大模型时代国产算力生态繁荣的底座。

国产大模型规模化运行于国产算力之上，进一步印证国产芯片可用度与性价比的同步提升。2026 年 8 月 26 日，智谱上线并开源 GLM-5 系列首个原生多模态模型 GLM-5.3-Flash。据智谱官方，GLM-5.3-Flash 率先在大规模流量场景中依托国产芯片集群对外提供服务；借助自研高带宽互联网络，其端到端服务性能较同一硬件的初始基线提升约 3 倍，单 token 成本已与主流英伟达 GPU 相当；而其模型能力在 Artificial Analysis 综合智能指数取得 57 分，编程表现与 Claude Opus 4.8 相当、价格约为其 1/40。此前 GLM-5 开源即获华为昇腾、摩尔线程、海光、寒武纪、昆仑芯、天数智芯、燧原等主流国产芯片 Day0 适配。国产模型与国产算力的闭环规模落地，标志国产芯片正从可用走向好用，性价比亦持续向英伟达看齐。

3. 范式跃迁：从单卡追赶到超节点系统竞争

国产算力的竞争范式，正从单卡性能追赶升级为以超节点为单位的系统级竞争。在出口管制收紧、单颗芯片受制于先进制程与 HBM 的背景下，通过高速互联把成百上千颗国产 NPU/GPU 紧耦合为一台超级计算机（超节点），以系统级的互联带宽、显存池化与在网计算弥补单卡差距，是国产算力以量补质、以系统补单点的现实解。

3.1. 三堵墙：显存墙、通信墙、能效墙

万亿参数 MoE 大模型对算力系统提出三重挑战。一是显存墙，单卡显存难以承载超大模型参数与 KV Cache；二是通信墙，专家并行/张量并行下高频 All-to-All 通信使卡间带宽与时延成为瓶颈；三是能效墙，规模扩张后功耗与散热约束凸显。解决之道是在机柜/多机柜范围内以超高带宽、低时延的 Scale-up 互联把大量加速卡紧耦合为一台逻辑超级计算机。

3.2. 全球范式：NVIDIA GB200 NVL72

英伟达以 NVLink/NVSwitch 构建的 GB200 NVL72 确立了 Scale-up 全球范式。72 颗 GPU 通过 NVLink 全互联为一个机柜级超节点，成为大模型训练与推理的事实标准之一。国产阵营的超节点，本质是在无法使用 NVLink 生态的约束下，以自研高速总线或开放以太协议重建等价甚至更大规模的 Scale-up 域；理解 NVL72 是理解国产超节点对标逻辑的基准。

3.3. 超节点的定义与三大核心技术

超节点：通过高带宽、低时延的 Scale-up 互联把多颗加速器紧耦合、对外呈现为单台虚拟计算机的 AI 算力架构。三大核心技术为：

- 高速互联（Scale-up 网络）：以自研总线（如华为 UB/灵衢）或开放以太网（RoCE/RDMA）实现卡间 TB/s 级带宽、微秒级时延、极低带宽衰减。
- 统一内存编址与显存池化：把分散 HBM 聚合为统一寻址的大显存池，突破单卡显存墙，天然亲和 MoE 与长序列推理。
- 在网计算：把 AllReduce 等集合通信卸载到交换/网络侧，减少数据搬运、提升有效算力。

3.4. 系统竞争如何重构价值分配

超节点把竞争维度从单卡 FLOPS 抬升到系统级有效算力/能效/可靠性。这对拥有全栈与供应链协同能力的厂商更有利，也为互联、光/铜连接、存储、封装、液冷等配套环节打开了远超单卡的价值量空间。

这也构成与市场认知的差异：市场易高估单卡对标英伟达的短期难度，而低估系统级可用算力的兑现速度与产业链广度。以 CloudMatrix 384 为例，其 384 颗昇腾 910C 可提供约 300 PFLOPS BF16 稠密算力，多项系统指标对标甚至超越 GB200 NVL72——即便单卡制程受限，也可在系统层面提供全球领先的可用算力规模，系统赢面是国产算力被低估的关键变量。

4. 国产超节点全景：放量元年

2026 年是国产超节点放量兑现元年，产业逻辑由概念切向产品—验证—交付。华为、中兴通讯、新华三、浪潮信息、中科曙光、超聚变等已在 WAIC 2026 集中亮相超节点方案，需求侧（运营商/云厂/国资智算中心）与供给侧（本体+零部件）双向验证，产业进入业绩兑现期。

国产超节点市场空间有望达数千亿级，2026 年是放量兑现元年。国产超节点的增长由两方面驱动：一是国产 AI 芯片出货持续放量，且用于构建超节点架构的比例逐年提升；二是单节点系统价值量随互联/交换、HBM、先进封装、液冷等环节升级而大幅抬升——仅高速交换一项，超节点普及即带动交换机需求成倍增长。随着华为、中科曙光、中兴通讯、新华三、浪潮信息等订单自 2026 年下半年集中交付，数千亿级市场进入兑现期。

需求侧的验证同样清晰，国产超节点已进入真订单、真落地阶段。运营商集采落地——中国移动 2026—2027 年 AI 超节点设备集采 6208 张加速卡、折合约 776 套计算节

点；大模型侧，昇腾全系超节点已适配 DeepSeek-V4 系列，硅基流动基于 CloudMatrix 384 上线的 DeepSeek-R1 推理服务性能比肩 H100；商用规模上，昇腾 384 超节点已商用落地超 750 套。供需两端共振，印证 2026 放量兑现元年。

4.1. 华为 — 国产超节点引领者

华为是国产超节点的引领者。CloudMatrix 384 于 2025 年 4 月发布，集成 384 颗昇腾 910C NPU 与 192 颗鲲鹏 CPU，通过 UB 协议实现点对点全对等互联，节点间带宽衰减低于 3%、时延增加低于 1 微秒；整系统提供约 300 PFLOPS BF16 稠密算力，据官方口径约为 GB200 NVL72 的近 2 倍，总显存约 3.6 倍、总内存带宽约 2.1 倍，具备 MoE 亲和、网络强化、存储强化、长稳可靠、日推夜训、开箱即用六大特性。

2026 年，昇腾超节点迎来关键进展：华为在 WAIC 2026 首次公开昇腾 Atlas 950 SuperPoD 超节点真机，支持最多 8192 张昇腾 950DT 计算卡，卡规模约为上一代 Atlas 900 的 20 多倍；据华为官网，其训练性能较 Atlas 900 提升约 17 倍、达 4.91M TPS，支持 FP4 后推理性能提升约 26.5 倍、达 19.6M TPS，计划 2026 年第四季度批量上市。华为在 2026 年世界人工智能大会上披露：昇腾 384 超节点已累计规模商用超过 750 套。

图1：昇腾 950 超节点（Atlas 950 SuperPoD）真机首次公开亮相



数据来源：华为官网、东吴证券研究所

华为推进全栈开源，把系统优势与生态开放结合。昇腾 CANN 已全面开源开放，灵衢协议（UnifiedBus/UB）及五大软件组件全部开源；灵衢 UB 采用 UB-Mesh 递归光直连拓扑替代传统 PCIe 组网，实现柜内与跨机柜 NPU 全互联。战略含义在于：即便单卡工艺受限，也可通过更大的超节点/超集群+自研互联+自研 HBM+全栈软件+开源生态，在系统层面提供全球领先的可用算力规模，把制程劣势转化为系统优势。

从系统指标看，国产超节点已在规模、显存与互联等维度对标乃至局部超越英伟达 GB200 NVL72。国产路线以更大 Scale-up 域+自研/开放互联+液冷弥补单卡差距；但需客观看待，单卡制程、HBM 与生态成熟度仍是短板，系统对标不等于单卡对标。

4.2. 群雄逐鹿：中兴通讯/新华三/浪潮信息/中科曙光/超聚变

WAIC 2026 上，国产超节点百花齐放。放量元年下主流厂商几乎同时展示超节点方案，且出现开放混插趋势——以中兴 OEX 为代表的开放架构支持多家国产 GPU 混插，有望缓解单一供应约束、加速规模落地，也使互联/交换/连接等中立环节确定性更高。

表3：国产主要超节点厂商产品

厂商	代表方案	关键特征
华为	CloudMatrix 384 / Atlas 950/960	UB 全对等互联，规模与系统能力最强
中兴通讯	OEX 超节点	三零线缆开放架构，支持壁仞科技、沐曦股份等国产 GPU 混插
新华三	UniPoD S80000	面向万亿参数大模型的超节点方案（WAIC 2026 展示）
浪潮信息	万卡级液冷超节点	出厂预调试，交付周期缩短约 60%
中科曙光	曙光 8000 超集群	十万卡级超集群落成
超聚变	整机柜超节点	单柜集成 128 个 AI 加速卡，第五代原生液冷，PUE 低至~1.06

数据来源：WAIC 2026、东吴证券研究所

4.3. 曙光 8000 与海光 DCU：全国产十万卡超集群标杆

中科曙光的曙光 8000 是国内首个全国产十万卡 AI 超集群，标志国产算力从万卡迈入十万卡规模化部署。2026 年，曙光 8000 建成投用，并接入国家超算互联网（落地郑州核心节点）；上线首周即满载运行；系统全链路自研、走超智融合路线，海光信息等国产芯片提供底层支撑。

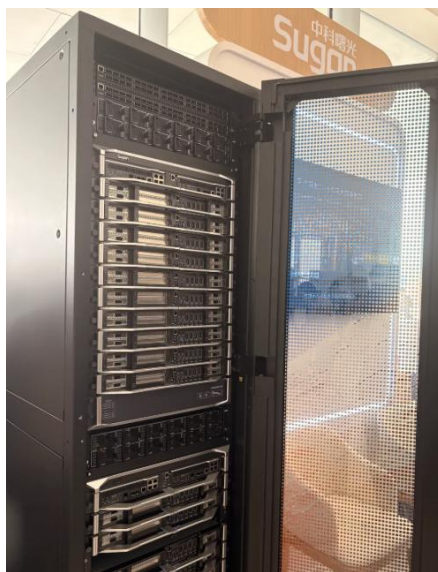
曙光超节点以 ScaleX+ 海光 DCU 构建自主互联体系，走开放系统硬件 + 相变浸没液冷路线。曙光 8000 采用自研原生 RDMA 高速网络 scaleFabric 与海光芯片级互联 HSL 协同，形成芯片设计（海光信息）+系统集成（中科曙光）的完整技术链条；散热

采用浸没式相变液冷，存储采用自研 ParaStor。

曙光超节点亦向中小规模、易部署延伸，推出世界首款无线缆箱式超节点 ScaleX40。2026 年 3 月，中科曙光发布 ScaleX40，采用正交无线缆一级互联架构，单机箱集成 40 张 GPU、FP8 总算力超 28 PFLOPS，配 5TB HBM 显存与 80TB/s 访存带宽、系统可靠性达 99.99%；相比传统 8 卡方案价格持平，训练与推理性能最高分别提升约 120%与 330%。其标准 19 英寸箱式设计仅 16U、可直接入柜，部署周期由数月级缩短至数小时级，精准适配智能体发展下的中小规模训练与推理，与十万卡曙光 8000 形成旗舰+标配的完整梯队。

在昇腾 UB 体系之外，曙光系提供了一条海光 DCU + HSL/以太的国产超节点技术路径。海光信息亦在与合作伙伴共建国产芯片开放大生态。中科曙光和海光信息以超算工程能力见长，面向科研、国家级超算与超智融合场景，与华为昇腾形成差异化并存，共同夯实国产超节点的多元供给格局。

图2：中科曙光 ScaleX40 超节点



数据来源：财联社、东吴证券研究所

4.4. 开放超节点与 Scale-up 标准之争

Scale-up 互联标准是超节点的操作系统级底座。全球看，英伟达以私有 NVLink 构筑壁垒，海外八大厂商于 2024 年组建 UALink 联盟对抗；国内则形成多套标准并进：中国移动主导 OISA、腾讯主导 ETH-X、阿里主导 ALS（Alink），以及华为 UB（灵衢）总线体系；海光 HSL、字节 Ethlink 等走以太/RDMA 路线，复用 800G/1.6T 交换与 RoCEv2

生态以控成本。

表4: 主要 Scale-up 互联标准对比

标准/总线	主导方	路线	定位
NVLink/NVSwitch	英伟达	私有高速总线	全球范式, 封闭
UALink	海外八大厂商	开放	对抗 NVLink
UB(灵衢)	华为	自研高速总线	昇腾超节点底座, 全对等互联
OISA	中国移动	开放	运营商主导
ETH-X	腾讯	以太	复用以太生态
ALS(Alink)	阿里	开放	云厂自用+开放
HSL / Ethlink	海光信息 / 字节	以太/RDMA	复用 800G/1.6T 交换、控成本

数据来源: ODCC、东吴证券研究所

标准之争的产业意义: 无论自研总线还是开放以太路线胜出, 高速互联与连接都是超节点确定性最高的增量环节。自研总线(华为 UB)系统最优、协同最强; 开放以太路线复用成熟以太与光/铜供应链, 利好交换芯片、光模块、铜缆、连接器等中立环节, 且天然支持多 GPU 混插。

4.5. 超节点产业链与受益环节

超节点的投资含义, 是把价值从单颗芯片向交换—整机—连接—存储—散热等系统环节扩散, 其中交换与整机弹性最大。据此可沿三条线索关注: 其一, 计算核心——国产 NPU/GPU 及其自研 HBM、先进封装配套; 其二, 也是弹性最高的环节——交换芯片、交换机与 AI 网络, 超节点的 Scale-up 内互联叠加集群 Scale-out, 使高速交换用量较传统集群大幅提升, 直接利好国产交换芯片与数据中心交换机龙头; 其三, 服务器与整机/ODM——超节点本体(整机柜)交付放量, 带动服务器整机、系统集成与 ODM 量

价齐升，并延伸至液冷、电源、PCB 与算力软件栈。

其中，**互联/交换芯片与高速接口 IP 是超节点特有的国产化攻坚环节**。国产高速接口 IP 正啃下最硬的骨头、突破 HBM PHY 等关键 IP；随超节点走向以太开放，开放以太路线（OISA/ETH-X/ALS）复用 800G/1.6T 交换与 RoCEv2 生态，利好国产交换芯片、光模块/CPO、高速铜缆与连接器等中立环节，且在多种技术路线下均能受益、确定性较高。软件栈（CANN/FlagOS）成熟则决定整体好用、可规模用。

超节点对高速交换的需求呈数量级放大，是产业链弹性最大的环节，直接利好国产交换/网络龙头。Scale-up 内互联叠加集群 Scale-out，交换机用量成倍提升。国产数据中心交换机龙头锐捷网络，以及新华三，均有望直接受益于 800G/1.6T 交换与 AI 网络方案的放量。

超节点本体交付放量，带动服务器整机与 ODM 环节量价齐升，浪潮、新华三、华勤等本体厂商充分受益。浪潮信息已推出元脑超节点服务器与万卡级液冷超节点；新华三推出 UniPoD S80000 超节点；华勤技术是少数兼具计算节点+网络节点全栈设计能力的 ODM。综合看，超节点最增量、弹性高的两个环节——网络/交换（锐捷、新华三）与整机/ODM（华勤、浪潮、新华三）——受益最为直接，叠加开放以太路线对国产中立环节的利好，构成放量元年的核心弹性。

5. 风险提示

先进制程与产能瓶颈：国产先进制程良率/产能爬坡不及预期，制约高端芯片供给。

HBM 与先进封装瓶颈：HBM/CoWoS 类封装产能不及预期，可能直接制约国产芯片出货上量。

软件栈与生态瓶颈：国产软件栈成熟度、框架/算子适配与迁移工具链完善不及预期。

需求与交付不及预期：互联网/运营商/智算中心国产芯片及超节点需求不及预期风险。

互联/交换芯片与连接瓶颈：高速 SerDes、接口 IP(HBM PHY)、交换芯片等国产化不及预期，制约超节点规模化。

外部环境风险：出口管制进一步升级、供应链扰动、地缘政治风险。

免责声明

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

本研究报告仅供符合中国证监会《证券期货投资者适当性管理办法》规定的专业投资者使用,且接收人须为本公司认定并建立服务关系的专业机构投资者使用,本公司不会因接收人收到本报告而视其为客户。在任何情况下,本报告中的信息或所表述的意见并不构成对任何人的投资建议,本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。本报告不面向普通投资者发布或传播。任何非专业机构投资者收到本报告后,应立即退回并销毁,不得依据本报告内容做出任何投资决策。

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

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

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

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期(A 股市场基准为沪深 300 指数,香港市场基准为恒生指数,美国市场基准为标普 500 指数,新三板基准指数为三板成指(针对协议转让标的)或三板做市指数(针对做市转让标的),北交所基准指数为北证 50 指数),具体如下:

公司投资评级:

买入: 预期未来 6 个月个股涨跌幅相对基准在 15%以上;

增持: 预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间;

中性: 预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间;

减持: 预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间;

卖出: 预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级:

增持: 预期未来 6 个月内,行业指数相对强于基准 5%以上;

中性: 预期未来 6 个月内,行业指数相对基准-5%与 5%;

减持: 预期未来 6 个月内,行业指数相对弱于基准 5%以上。

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

东吴证券研究所
苏州工业园区星阳街 5 号

邮政编码: 215021

传真: (0512) 62938527

公司网址: <http://www.dwzq.com.cn>