

AI景气赛道，国产全链受益

通信行业分析报告

分析师：李宏涛 S0910523030003

目录

contents

1 AI算力三要素

2 国产算力产业链受益

3 算力产业趋势和重点标的

4 风险提示

AI三要素：算力，超摩尔定律迭代

- ◆ 3月，英伟达推出Blackwell芯片，6月初，英伟达公布整体芯片计划，预计将于2025年推出Blackwell Ultra，2026年推出名为“Rubin”的下一代AI平台，该平台将采用HBM4内存。
- ◆ 海光深算二号已经发布并实现商用，深算二号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，具有全精度浮点数据和各种常见整型数据计算能力，性能相对于深算一号实现了翻倍的增长。深算三号研发进展顺利。

图表1 英伟达新一代产品序列



图表2 海光产品序列



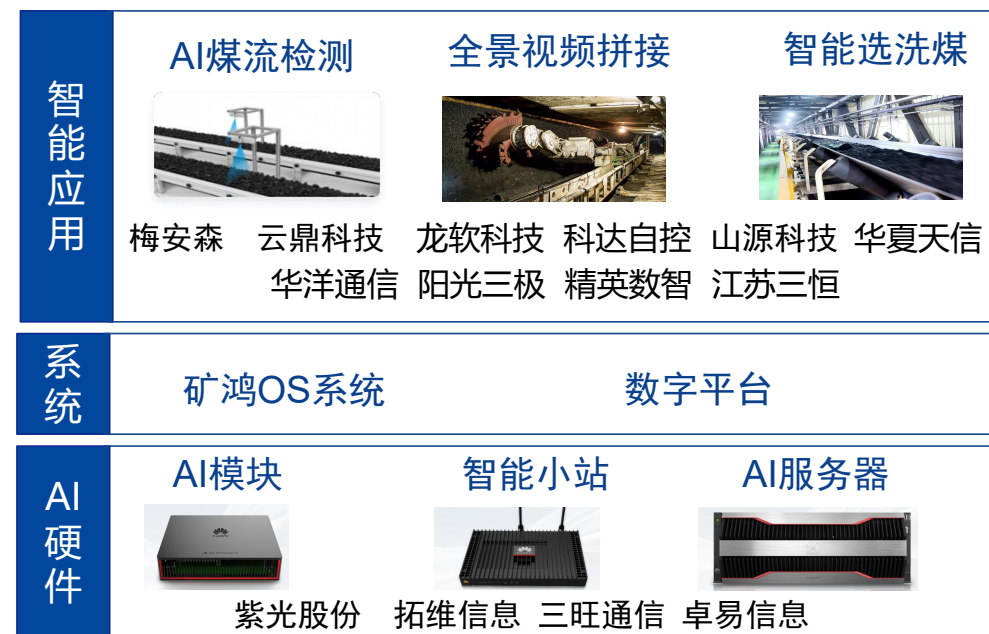
AI三要素：应用，垂直和通用选择

- ◆ 大模型趋于分化，通用和垂直两条道路。美国模型聚焦通用，国内模型聚焦垂直行业。
- ◆ 通用大模型：参数普遍达上千亿甚至万亿，需要高算力进行模型训练，适用性广，如谷歌“BERT”、OpenAI“GPT”等。
- ◆ 垂直大模型：参数普遍几十到上百亿，所需算力小于通用模型，但对数据的专有性要求高，行业客户更易接受，能够看到明显受益。如华为“盘古”大模型。

图表3 GPT性能比较

Eval Sets	GPT-4o	GPT-4T 2024-04-09	Gemini 1.0 Ultra	Gemini 1.5 Pro	Claude Opus
MMMU (%) (val)	69.1	63.1	59.4	58.5	59.4
MathVista (%) (testmini)	63.8	58.1	53.0	52.1	50.5
AI2D (%) (test)	94.2	89.4	79.5	80.3	88.1
ChartQA (%) (test)	85.7	78.1	80.8	81.3	80.8
DocVQA (%) (test)	92.8	87.2	90.9	86.5	89.3
ActivityNet (%) (test)	61.9	59.5	52.2	56.7	
EgoSchema (%) (test)	72.2	63.9	61.5	63.2	

图表4 华为矿山大模型边缘算力图谱梳理



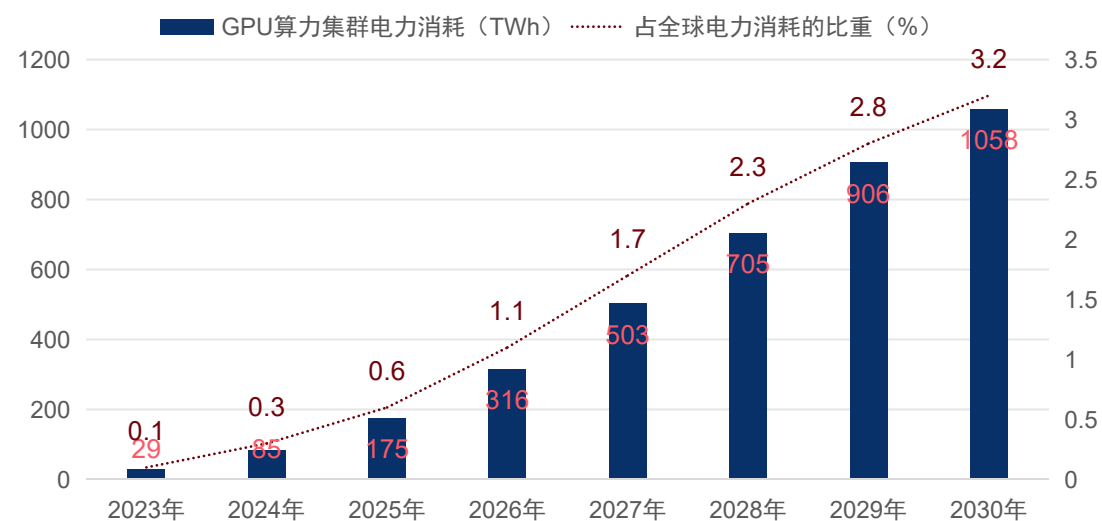
AI三要素：能耗，催生能源供应趋紧

- ◆ 2022年中国数据中心耗电量2700亿千瓦时，其中智能算力总规模已达到197EFLOPS，年耗电量约730亿度电：到2027年，全国数据中心的耗电量将达到1.1万亿千瓦时，占全社会用电量的比重将达到8.6%。
- ◆ 2024年至2030年，我国GPU算力集群年度总耗电量将增长约10倍，年复合增长率为49.8%，新增的GPU算力集群功耗远低于我国新能源装机容量的增长。

图表5 数据中心交换芯片吞吐量高速演进

	2023	2024	2025	2026	2027
智能算力年均涨幅	50%	50%	50%	50%	50%
智能算力数据中心耗电量 (亿千瓦时)	1100	1650	2475	3713	5569
通用算力及存储类数据中心年均涨幅	18%	18%	18%	18%	18%
通用算力及存储类数据中心耗电量 (亿千瓦时)	2321	2738	3231	3813	4499
数据中心总耗电量 (亿千瓦时)	3421	4388	5706	7526	10068
同比增长	27%	28%	30%	32%	34%
全社会用电量 (亿千瓦时) ⁶	92241	97775	103642	109861	116452
数据中心耗电量占比	3.7%	4.5%	5.5%	6.9%	8.6%

图表6 全球GPU算力集群电力消耗及占比的测算



目录

contents

1 AI算力三要素

2 国产算力产业链受益

3 算力产业趋势和重点标的

4 风险提示

算力产业链全景图

- ◆ 算力核心产业包括以软硬件的方式对外提供计算与服务能力，硬件主要包括AI芯片、光模块、高速铜缆、服务器、交换机、智算中心、可视化运维、液冷等。

AI芯片

- 中科曙光
- 海光信息
- 寒武纪

算力云

- 首都在线
- 优刻得
- 青云科技
- 云赛智联

服务器

- 中兴通讯
- 紫光股份
- 神州数码
- 拓维信息
- 工业富联
- 浪潮信息

光模块

- 中际旭创
- 光迅科技
- 华工科技
- 天孚通信
- 源杰科技
- 博创科技
- 新易盛

高速铜互联

- 新亚电子
- 沃尔核材
- 神宇股份
- 兆龙互连

交换机

- 紫光股份
- 菲菱科思
- 共进股份

数据中心

- 润泽科技
- 首都在线
- 数据港
- 宝信软件
- 奥飞数据
- 润建股份

可视化运维

- 恒为科技
- 亚康股份
- 新炬网络
- 亚康股份

液冷

- 英维克
- 佳力图
- 申菱环境
- 高澜股份

算力芯片：美国限制先进芯片出口中国

- ◆ 2023年10月，美政府出台新的芯片出口管制规定，阻止美国芯片制造商规避政府限制向中国出售半导体，对性能满足以下条件的芯片实施出口管制：
- ◆ (1) 总计算能力TPP达到或超过4800的芯片；(2) TPP达到或超过1600且PD达到或超过5.92的芯片；(3) $2400 \leq TPP < 4800$ ，且 $1.6 \leq PD < 5.92$ 的芯片；(4) $1600 \leq TPP$ ，且 $3.2 \leq PD < 5.92$ 的芯片。
- ◆ 受此影响，英伟达无法向我国出口高算力AI芯片。

性能参数	V100 PCIe	A100 80GB PCIe	A800 80GB PCIe	H100 80GB PCIe
FP64	7TFLOPS		9.7TFLOPS	26 TFLOPS
FP32	14TFLOPS		19.5TFLOPS	51 TFLOPS
FP16 Tensor Core			312TFLOPS	756.5 TFLOPS
GPU显存	32/16GB HBM2		80GB HBM2e	80GB
GPU显存带宽	900 GB/s		1935GB/s	2TB/s
是否被限制	否		是	是

算力芯片：国内设立安全可靠等级支持国内芯片

- ◆ 据《华尔街日报》报道，中国推动替代外国技术的重点是将美国芯片制造商从中国电信系统中剔除。电信运营商在 2027 年前逐步淘汰作为其网络核心的外国处理器，此举将打击美国芯片巨头英特尔和AMD。
- ◆ 23年12月26日，中国信息安全测评中心公布了安全可靠测评结果公告一级获证企业名单，其中在获批的18款处理器全部来自中国企业，包括鲲鹏、飞腾、龙芯、海光等芯片，英特尔和AMD均不在列。

图表7 华尔街日报

News / Technology News / China Tells Telecom Firms to ...

China Tells Telecom Firms to Phase Out Foreign Chips in Blow to Intel, AMD - WSJ

(Reuters) -Chinese officials directed the country's largest telecom carriers earlier this year to phase out foreign chips that are key to their networks by 2027, the Wall Street Journal reported on Friday, citing people familiar with the development.

The move would impact U.S. chip giants Intel and Advanced Micro Devices, according to the report. Their shares were down more than 1.5% in premarket trading.

Intel declined to comment, while AMD did not respond to a Reuters request for comment.

图表8 中央处理器（CPU）可靠测评结果

序号	产品名称	送测单位	安全可靠等级	序号	产品名称	送测单位	安全可靠等级
1	鲲鹏920	深圳市海思半导体有限公司	I级	10	飞腾FT-2000+	飞腾信息技术有限公司	I级
2	龙芯3C5000L	龙芯中科技术股份有限公司	I级	11	盘古M900	海思技术有限公司	I级
3	申威1621	无锡先进技术研究院	I级	12	飞腾腾云S2500	飞腾信息技术有限公司	I级
4	龙芯3A4000/3B4000	龙芯中科技术股份有限公司	I级	13	麒麟9006C	深圳市海思半导体有限公司	I级
5	龙芯3A5000/3B5000	龙芯中科技术股份有限公司	I级	14	海光C86-3G	海光信息技术股份有限公司	I级
6	申威SW421	无锡先进技术研究院	I级	15	麒麟990	深圳市海思半导体有限公司	I级
7	申威3231	无锡先进技术研究院	I级	16	海光2号C86	海光信息技术股份有限公司	I级
8	飞腾腾锐D2000	飞腾信息技术有限公司	I级	17	兆芯ZX-E	上海兆芯集成电路股份有限公司	I级
9	飞腾FT-2000	飞腾信息技术有限公司	I级	18	兆芯ZX-D	上海兆芯集成电路股份有限公司	I级

算力芯片：昇腾+鲲鹏，国产AI算力替代

- ◆ 昇腾芯片—AI算力：320TFLOPS，最大功率310W，基于自研华为达芬奇架构3D Cube技术，实现业界最佳AI性能与能效，架构灵活伸缩，支持云边端全栈全场景应用。
- ◆ 鲲鹏芯片—通用算力：基于ARM架构授权，由华为自主设计完成，采用开放生态模式，为合作伙伴提供主板等部件，鲲鹏产业生态稳步发展，硬件伙伴出货占比已达95%以上

图表9 华为 FusionSphere 2



- 大规模的超级计算机集
- 1024颗昇腾910 AI处理器进行协同
- ResNet-50 v1.5模型和ImageNet-1k数据集上的训练时间仅为59.8秒
- 相当于200台英伟达的H100显卡

图表10 海思 Ascend 910




Ascend 910
华为昇腾910

版本：Ascend-910

架构：

DaVinci

最大功耗：310W

工艺：N7+

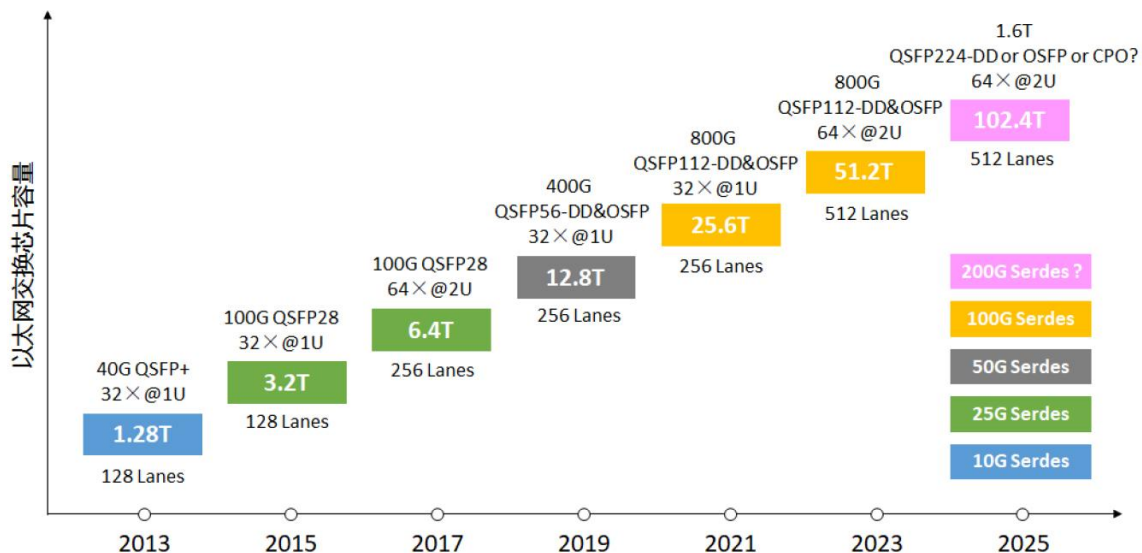
半精度算力：320 TeraFLOPS

整数精度算力：640 TeraOPS

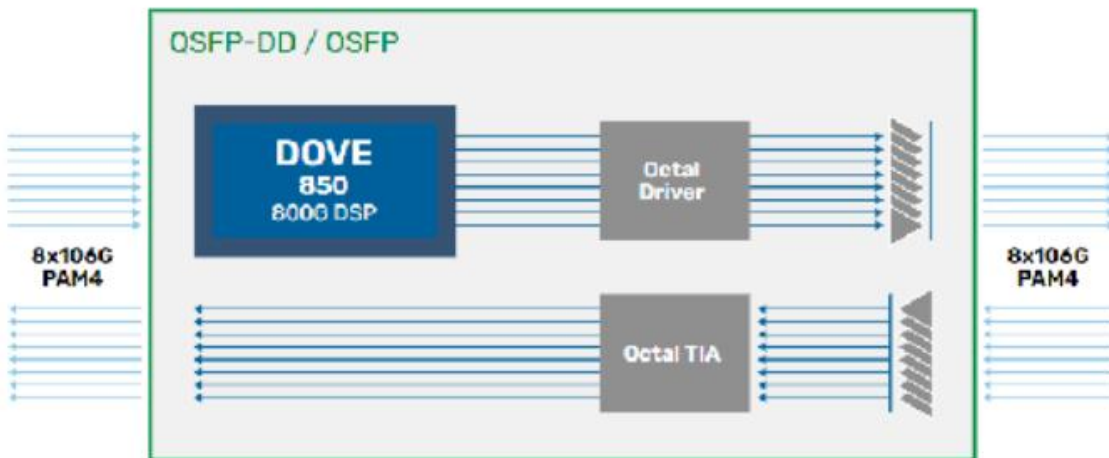
光模块：高速率、低功耗、低成本趋势明显

- ◆ 算力对互联需求倍增，带动光模块整体需求提升，随着光模块在数据中心互联中的成本占比不断攀升，高速率、低功耗、低成本成为行业主流趋势。
- ◆ 800G/1.6T将成为行业主流。数据中心交换芯片吞吐量2023年达到51.2Tb/s，2025年之后达到102.4Tb/s，800Gb/s和1.6Tb/s更高速率将成为实现高带宽数据交换的重要选择。
- ◆ CPO/LPO/LRO等新形态将是长期迭代方向。CPO/LPO/LRO差异点主要是封装模式的不同，低功耗方面，相比于可插拔光模块，LPO的功耗下降明显，与CPO的功耗接近。

图表11 数据中心交换芯片吞吐量高速演进



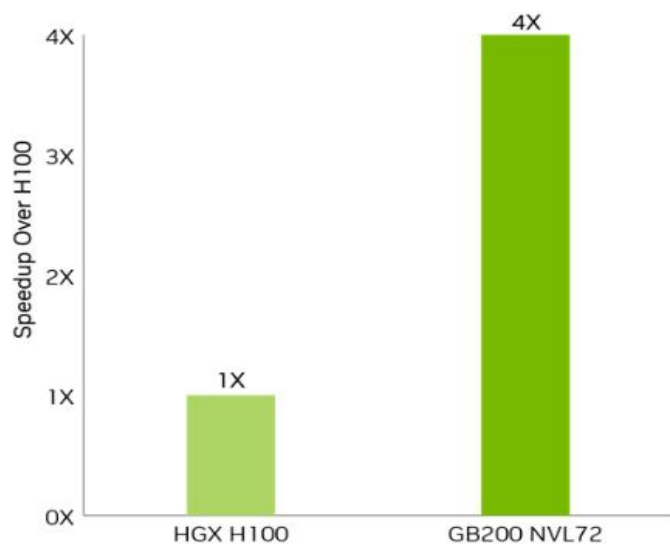
图表12 LRO内部结构



光模块：海外800G批量应用，未来将配备1.6T光模块

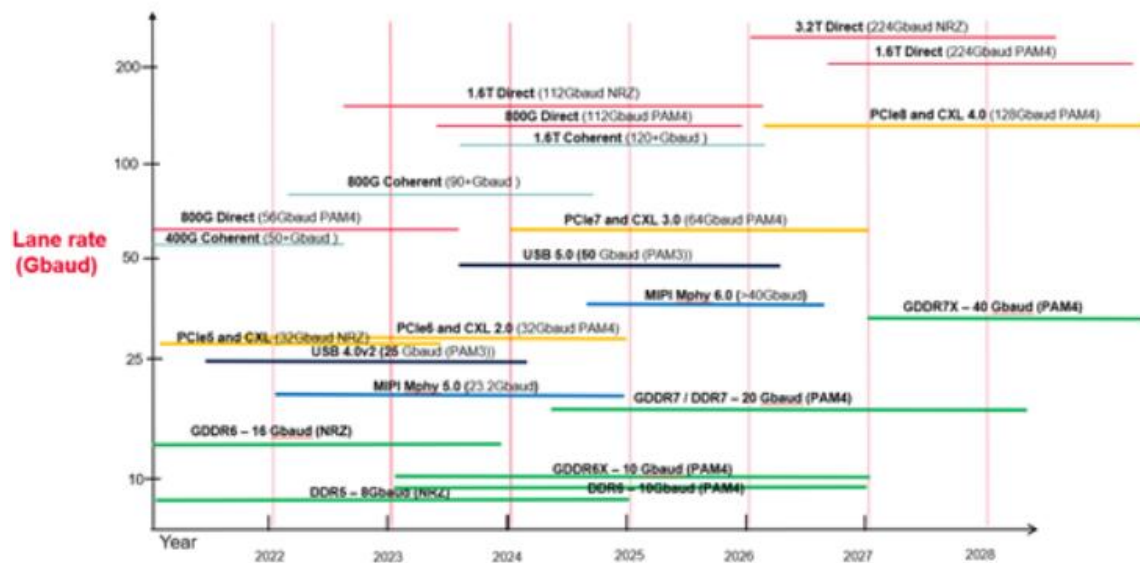
- ◆ 根据Lightcounting 预测，光模块的全球市场规模在2022-2027年或将保持以CAGR11%保持增长，2027年有望突破200亿美元，增长驱动力主要来自800G、1.6T、3.2T光模块需求，太网光模块销售的预测，将从2023年的12亿美元增至2024年的30亿美元和2025年的50亿美元。
- ◆ 根据英伟达GTC大会资料，HGX H100集群需要400G IB网络，GB200 NVL72液冷集群需求800G IB网络，1.6T光模块需求有望放量。Marvell公司市场业务将从23年的34亿美金增长到28年的111亿美金，CAGR 为27%，包括数据中心内部互联光互联，CAGR为25%；数据中心内部有源电缆DSP，CAGR为59%

图表13 GB200 NVL72 网络需求



Projected performance subject to change. 32,768 GPU scale, 4,096x HGX H100 air-cooled cluster: 400G IB network, 456x GB200 NVL72 liquid-cooled cluster: 800G IB network

图表14 高速光模块技术路线图



光模块：中国之光供应全球高端光模块

- ◆ LightCounting公布了最新版2023年全球光模块TOP10榜单，到2018年，大部分日本和美国厂商退出了这一市场，具体来看，2023年旭创科技首次排名第一，Coherent（Finisar）排名第二，中国厂商的整体表现来看，共有7家厂商入围。2023年，旭创科技（排名第1）、华为（排名第3）、光迅科技（排名第5）、海信宽带（排名第6）、新易盛（排名第7）、华工正源（排名第8）、索尔思光电（排名第9）。
- ◆ 800G/1.6T/LPO/CPO齐发力，光模块演进进入落地新阶段。从2024 OFC大会看来，1.6T光模块、硅光、LPO、CPO、相干通信、高速连接器等多项前沿创新技术在大会展现，国内企业中际旭创、新易盛、天孚通信、光迅科技、华工科技等在前沿高端领域均有布局，技术实力和量产能力全球领先。

图表15 2023年光模块全球前十大公司

Figure 1: Top 10 Transceiver Vendors in 2010, 2016, 2018 and 2023

Ranking of Top 10 Transceiver Suppliers				
2010	2016		2018	2023
Finisar	Finisar	1	Finisar	Innolight
Opnext	Hisense	2	Innolight	Coherent
Sumitomo	Accelink	3	Hisense	Huawei (HiSilicon)
Avago	Acacia	4	Accelink	Cisco (Acacia)
Source Photonics	FOIT (Avago)	5	FOIT (Avago)	Accelink
Fujitsu	Oclaro	6	Lumentum/Oclaro	Hisense
JDSU	Innolight	7	Acacia	Eoptolink
Emcore	Sumitomo	8	Intel	HGGenuine
WTD	Lumentum	9	AOI	Source Photonics
NeoPhotonics	Source Photonics	10	Sumitomo	Marvell

Source: LightCounting

图表16 OFC 2024 新产品展示



现场演示1.6T-LPO-DR8 OSFP 模块



展示业界首款4x200G LPO并宣布8x100G LPO进入量产阶段



展出适配800G/1.6T和CPO光引擎系列组件产品

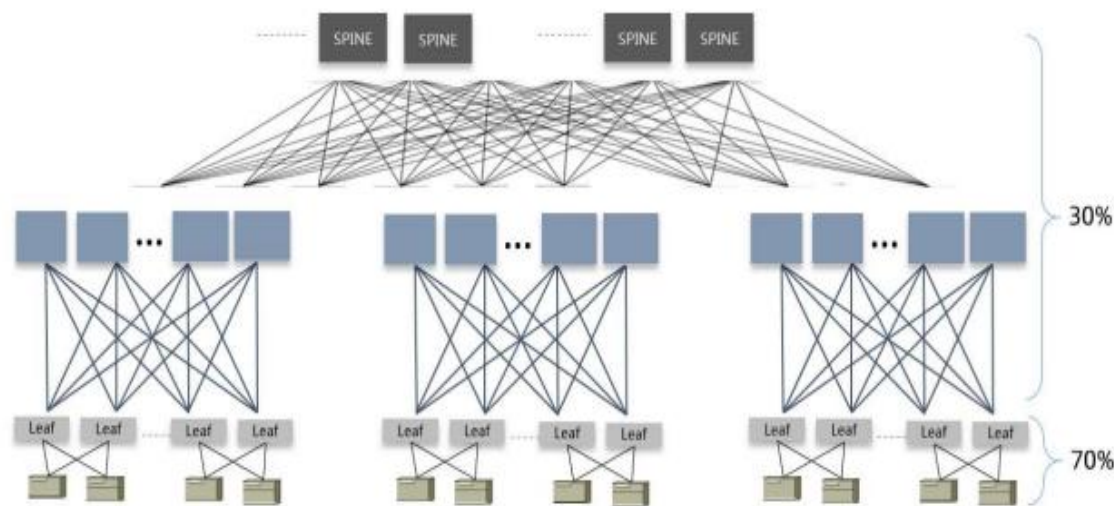
铜互连：数据中心低成本短距离连接方案

- ◆ DAC高速线缆（Direct Attach Cable）也可译为直接电缆或直连铜缆，由一根铜缆和两端的连接器模块组成，广泛应用于短距离存储区域网络、数据中心和高性能计算机连接。DAC包括无源铜缆（PCC）、有源铜缆（AAC）和有源电缆（AEC）。
- ◆ AOC有源光缆（Active Optical Cable），将2只光模块与光缆封装在一起，实际上就是带光模块的光纤跳线。AOC一般长度比较短，最多也就几百米，最大的特点是光模块和光纤做成了一体，不可分离。

图表17 AOC/DAC比较

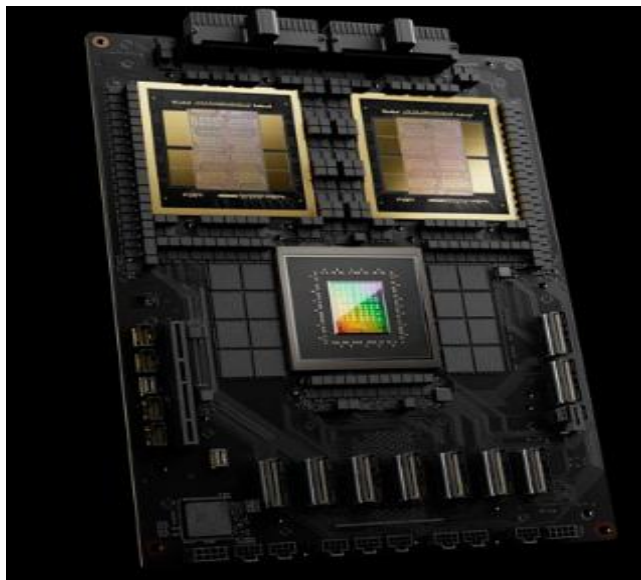
对比项	功耗	传输介质	传输距离	传输信号	价格	重量 体积	传输性能
AOC有源光缆	低	光纤	100m以内	光信号	高	重量1/4 体积1/2	误码率低
DAC高速线缆	高	铜	7m左右	低压脉冲	低	重量1 体积1	短距离低 长距离高

图表18 典型数据中心 CLOS 网络架构

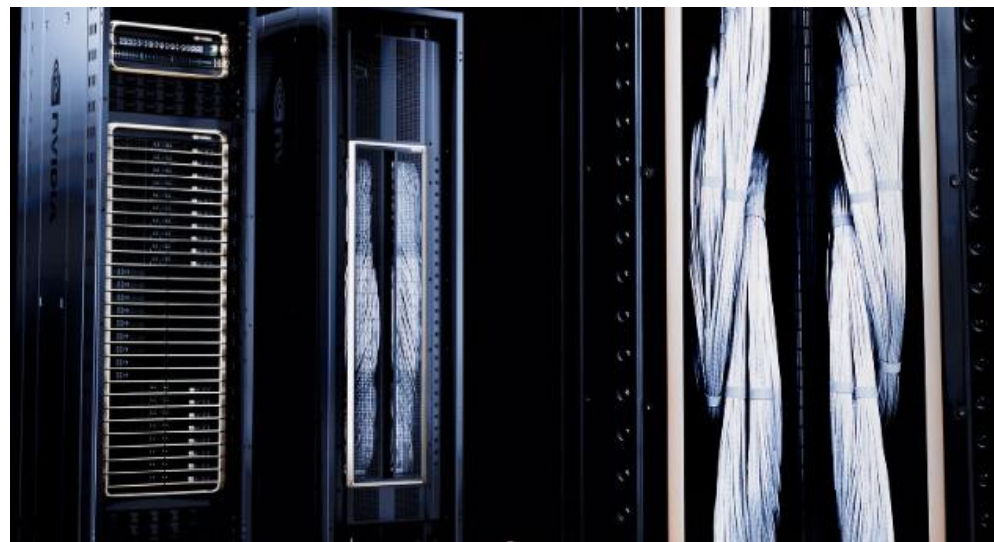


- ◆ 英伟达于 2024 年 3 月发布最新一代 Blackwell 架构 GPU，在专为 AI 和高性能计算任务设计的数据中心服务器架 NVIDIA GB200 NVL72 大机柜中采用了大量的铜连接方案，主要用于连接机柜内的 GPU 和 Switch，确保数据在不同处理单元间的迅速和准确传输。
- ◆ 单台GB200NVL72架构中利用5000根NVLink铜缆进行交换机和GPU之间的连接，总长度接近2英里，铜连接技术的显著优势主要在于其高效散热性能、成本效益以及低能耗特点，单台服务器采用铜互连方案后的价值量也更突出。

图表19 英伟达GPU内部 铜互连情况



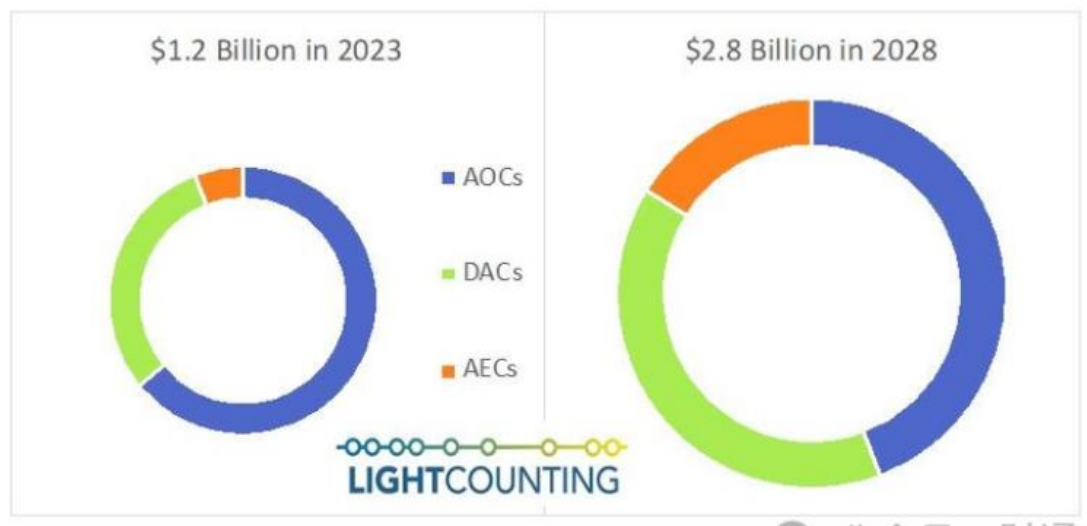
图表20 英伟达GPU外部 线缆情况



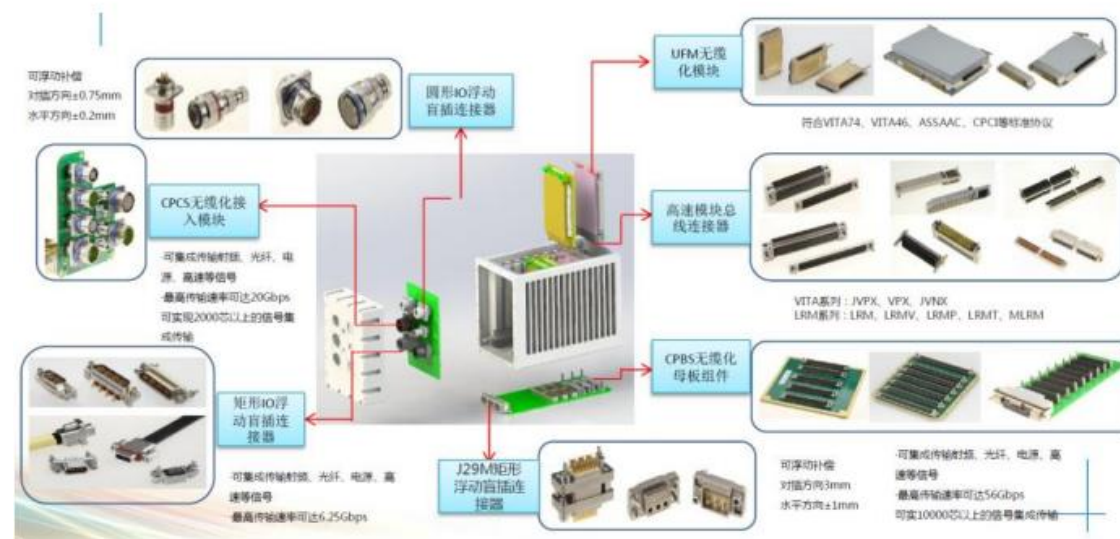
铜互连：铜互联方案获共识，打开国产高速连接空间

- ◆ 铜互联在短距通信领域有相对优势，英伟达已在新产品大量采用铜互联方案，是当下以及未来许多应用中最具成本效益的解决方案，根据Light Counting，全球无源直连电缆DAC和有源电缆AEC的市场规模将分别以25%和45%的年复合增长率增长。
- ◆ 国内高速互联厂商有华丰科技、兆龙互联、神宇股份、金信诺等。华丰科技目前取得了与国际化连接器大厂的专利授权合作，已完成正交架构的112G高速背板产品开发及小批量生产，开发并持续优化224G产品，已达到样品试制合格的状态

图表21 英伟达GPU内部 铜互连情况



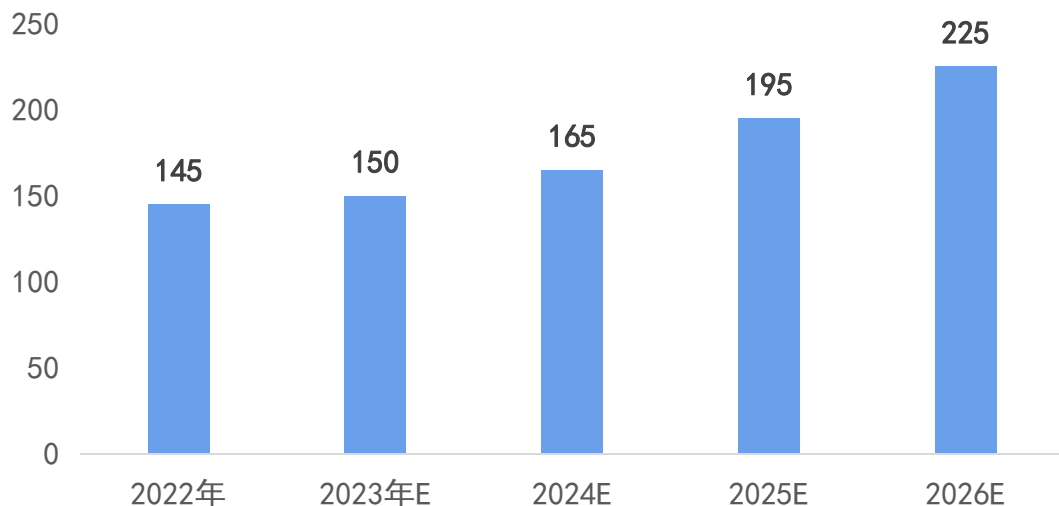
图表22 英伟达GPU内部 铜互连情况



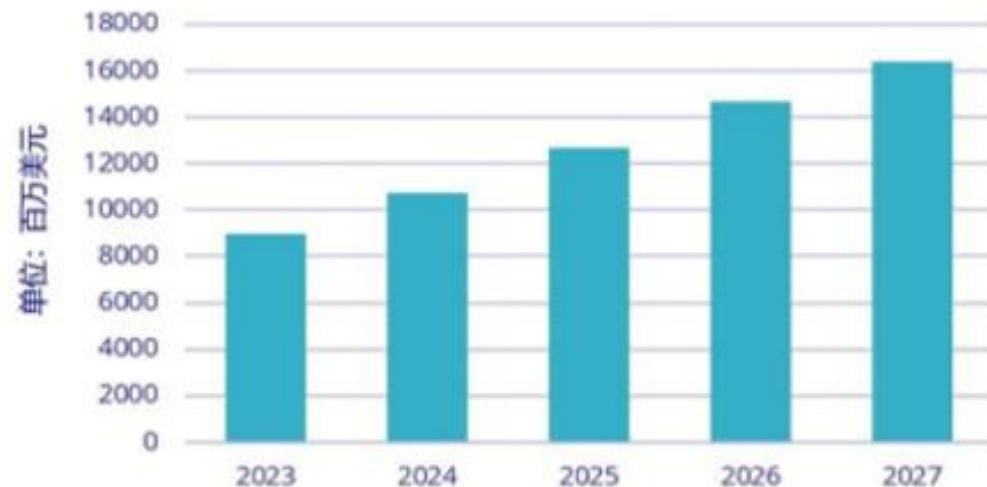
服务器：中国AI算力市场高速发展，将成为第二增长极

- ◆ 算力持续增长，AI服务器应用场景更加广泛。Trendforce预估，预估2024年全球服务器整机出货量约1,365.4万台，年增约2.05%。同时，市场仍聚焦部署AI服务器，AI服务器出货占比约12.1%。我们认为，主要受惠于北美云端数据中心业者订单带动，AI training芯片的出货量将会增长。
- ◆ 中国AI算力市场高速发展，将成为第二增长极。根据IDC预测，到2027年中国加速服务器市场规模将达到164亿美元。其中非GPU服务器市场规模将超过13%。智慧城市、智能机器人、智能家居、工业领域将成为主要应用领域。

图表23 全球AI服务器出货量预测（千台）



图表24 中国加速服务器市场预测



服务器：运营商智算集采落地，国产服务器厂商受益

- ◆ AI服务器较通用价值量提升明显。以AI服务器NF5688M6为例，该服务器采用2颗第三代Intel Xeon可扩展处理器+8颗英伟达A800 GPU的组合，每颗A800售价104000元，故该服务器芯片成本约96万元，相比于通用服务器NF5280M6，贵出近8倍。
- ◆ 运营商集采落地，国产受益。中国移动公示的2024年至2025年新型智算中心集采项目采购包1的中标结果，昆仑技术、华鲲振宇、宝德计算机、长江计算等7家中标；此外，之前据C114简单测算，中国电信集采以G系列服务器为代表的国产化服务器规模占比超过47%，接近50%。

图表25 中国电信集采情况

标包	涉及的主要评估产品品类	产品名称	数量(台)
1	I 系列服务器	主设备-训练型风冷服务器(I 系列)	2073
		配套设备-InfiniBand交换机	1129
2	I 系列服务器	主设备-训练型液冷服务器(I 系列)	125
		配套设备-InfiniBand交换机	53
3	G系列服务器	训练型风冷服务器(G系列)	1048
4	G系列服务器	训练型液冷服务器(G系列)	929

图表26 中国移动智算中心集采情况

中国移动2024-2025年新型智算中心集采			
标包	中标人	投标报价 (亿元, 不含税)	中标份额
标包1 特定场景 AI训练服务器	第1名河南昆仑技术有限公司	190.97	21.05%
	第2名四川华鲲振宇智能科技有限责任公司	190.97	17.54%
	第3名宝德计算机系统股份有限公司	190.97	15.79%
	第4名百信信息技术有限公司	191.36	14.04%
	第5名武汉长江计算科技有限公司	191.16	12.28%
	第6名神州鲲泰(厦门)信息技术有限公司	191.16	10.53%
	第7名湖南湘江鲲鹏信息科技有限公司	190.58	8.77%

- ◆ 交换机相当于一台特殊的计算机，由硬件和软件组成，负责构建网络进而实现所有设备的互联互通，是算力网络底座的主要组成部分。
- ◆ 数据中心交换机的功能：1、网络数据流量分发；2、处理大带宽数据。
- ◆ 数据中心交换机按照位置分类可分为接入层交换机、汇聚层交换机和核心层交换机。
- ◆ 数据中心交换机按照形态分类可分为盒式交换机（多位于接入层和汇聚层）和机框式交换机（多位于核心层）。

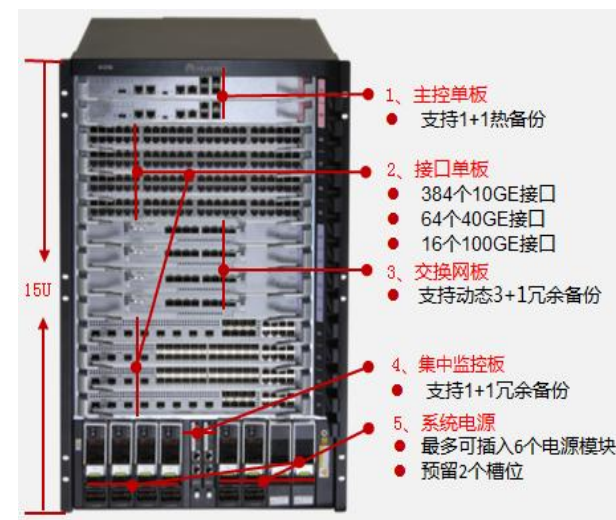
图表27 按位置分类交换机特点

位置	特点
接入层	<ul style="list-style-type: none">· 端口数量较少· 支持高速连接· 数据传输低延迟
汇聚层	<ul style="list-style-type: none">· 承载大量的流量· 连接方式灵活· 支持上行、下行链路连接
核心层	<ul style="list-style-type: none">· 高性能和高可用性· 大容量的交换能力· 高速的转发能力

图表28 按形态分类交换机示意图



盒式交换机

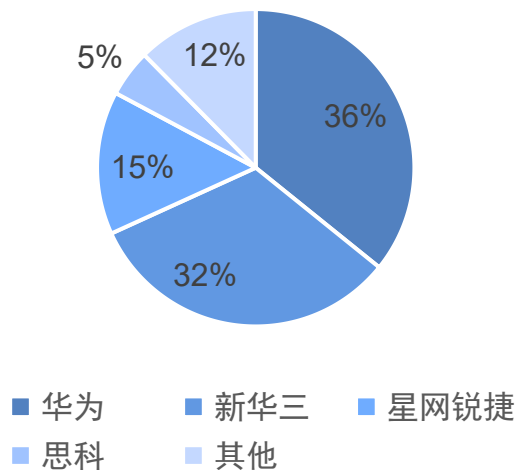


框式交换机

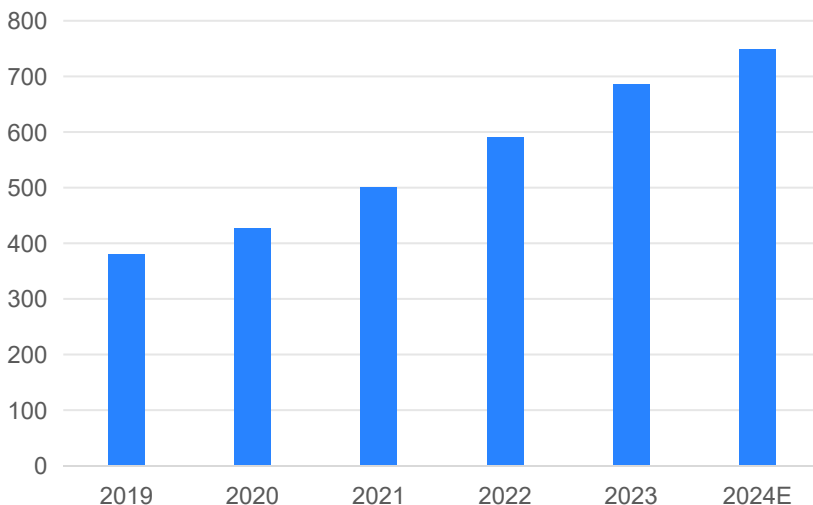
交换机：国产厂商快速崛起，800G高端交换机已突破

- ◆ 2022年中国交换机行业市场规模为591亿元，同比增长17.96%，预计2023年市场规模达到685亿元。
- ◆ 国产设备商快速崛起。2022年中国市场数据中心交换机前三厂商为华为、新华三和锐捷网络。从份额占比来看当前国内交换机市场双龙头格局特征显著，华为占比36%、新华三占比32%
- ◆ 我国主要厂商突破技术瓶颈，实现快速发展。新华三旗下S10500、S12500等系列交换机，多采用CLOS架构，提供Seerblade高性能AI计算模块，目前已形成了完整的数据中心产品体系。华为主推CloudEngine系列数据中心交换机，系列基于华为新一代的VRP8操作系统，2023Q1该系列以34.6%的份额排中国数据中心交换机市场第一，其中16800-X为全球首款800GE数据中心核心交换机。

图表29 2023年中国交换机各公司市场份额



图表30 2019-2024年中国交换机市场规模趋势（单位：亿元）



图表31 华为800G交换机

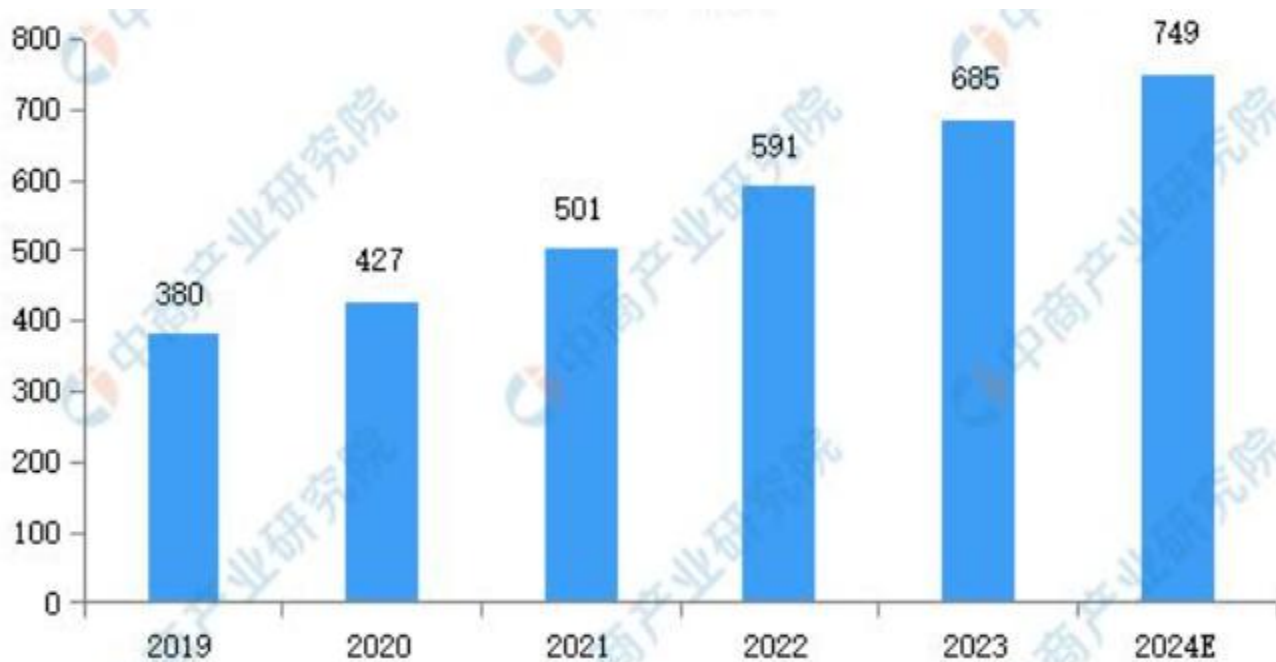


- ◆ CloudEngine 16800-X支持288*800GE端口，3.5微秒跨板转发时延，支持三网融合，总运营成本可降低36%。

交换机：400G放量元年，国产交换机厂商受益

- ◆ 据中商产业研究院分析师预测，2024年中国交换机市场规模将增至749亿元。我们认为，24年交换机市场放量的主要原因是数据中心类交换机增加所致，随着三大运营商、互联网智算中心建设加速，AI算力服务器将带动400G交换机的放量。
- ◆ 国内能够批量供应高端交换机的厂商有紫光股份、菲菱科思、共进股份等。目前菲菱科思已具备100G/400G数据中心主流的接口速率、高带宽、大容量交换机的硬件开发能力，以及2.0T/8.0T数据中心交换机形成量产交付，12.8T等其他数据中心交换机在稳步小批量验证中。

图表32 2019-2024年中国交换机市场规模预测趋势图（亿元）



图表33 2021年单机柜功率增长速度



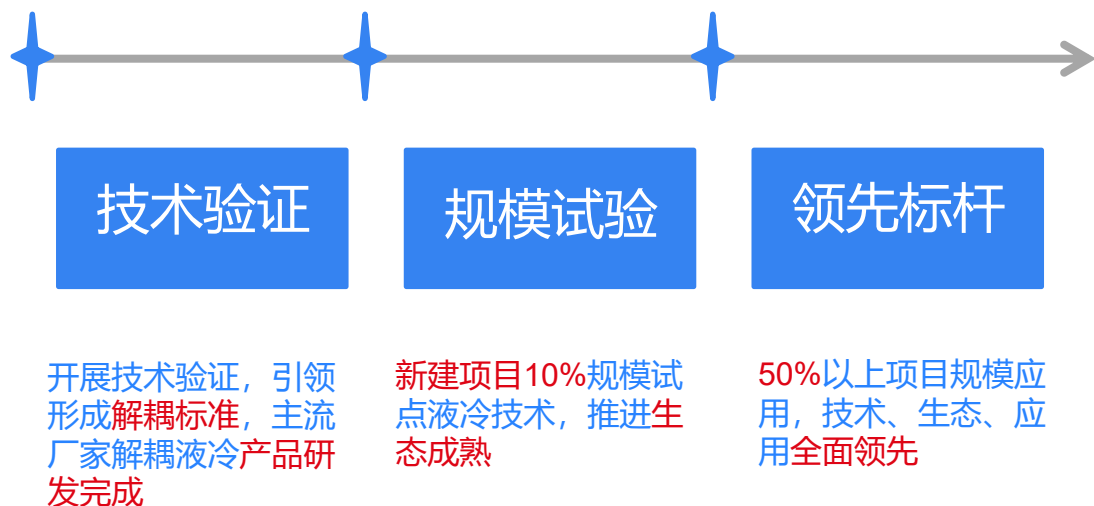
◆ 12.8T数据中心交换机

业务端口	32个400GE QSFP-DD+2个25GE SFP28
交换容量	12800Gbps交换容量，二/三层线速转发
交换芯片	云合智网 Clounix CLX850128

液冷：液冷应用逐步渗透，高端算力配液冷成必须

- ◆ 电信运营商提出三年规划，倒逼液冷渗透加速。国内三大运营商联合发布《电信运营商液冷技术白皮书》，2024 年开展规模测试，新建项目10%规模试点液冷技术。2025 年开展规模应用，2025 年及以后 50%以上项目规模应用液冷技术。
- ◆ 英伟达HGX B200单芯片功率达1000瓦；GB200 NVL72集群采用液冷方式，将多个由 GB200 驱动的系统整合到一个液冷机架中，每个采用液冷技术的机架配备36个NVIDIA GB200 Grace Blackwell Superchips。

图表34 电信运营商液冷技术白皮书



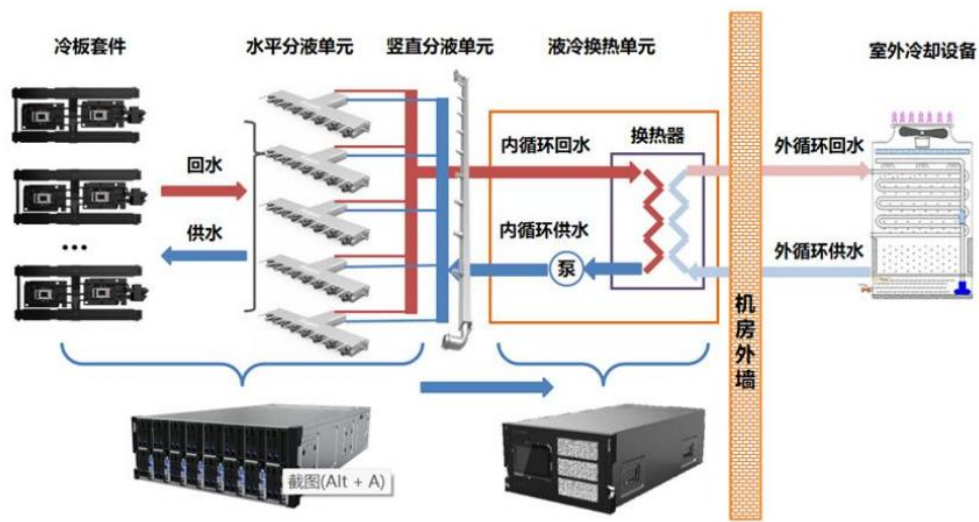
图表35 2021年单机柜功率增长速度

	HGX B200	HGX B100
TF32 Tensor Core	2.2 petaFLOPS	1.8 petaFLOPS
FP32	80 teraFLOPS	60 teraFLOPS
FP64 Tensor Core	40 teraFLOPS	30 teraFLOPS
FP64	40 teraFLOPS	30 teraFLOPS
GPU memory Bandwidth	Up to 192 GB HBM3e Up to 8 TB/s	
Max thermal design power (TDP)	1,000W	700W
Interconnect	NVLink: 1.8TB/s PCIe Gen6: 256GB/s	NVLink: 1.8TB/s PCIe Gen6: 256GB/s
Server options	NVIDIA HGX B200 partner and NVIDIA-Certified Systems with 8 GPUs	NVIDIA HGX B100 partner and NVIDIA-Certified Systems with 8 GPUs

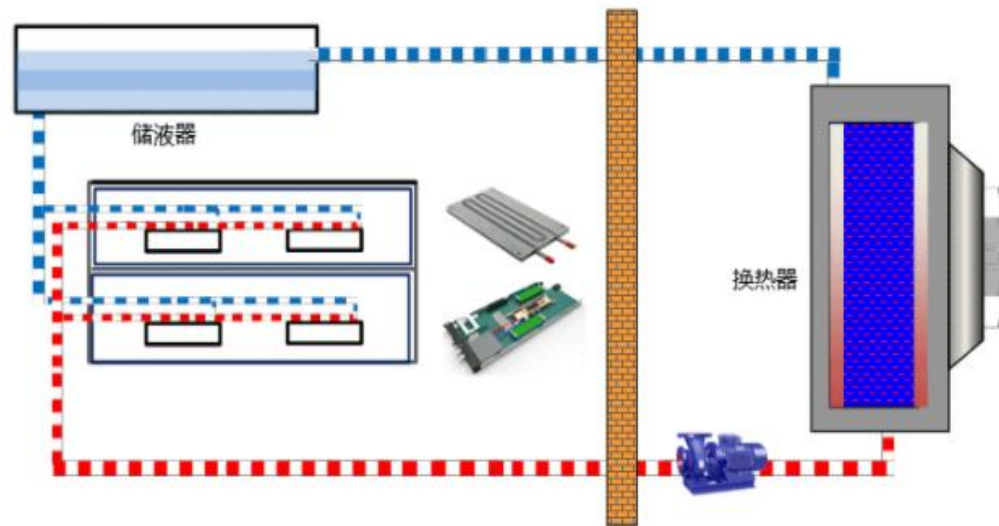
液冷：数据中心以冷板式为主，包含一二次设备

- ◆ 液冷：以液体为冷却介质，将设备发热元件产生的热量直接带走的一种冷却方式。
- ◆ 数据中心应用以冷板式为主，浸没式逐步增加：目前国内液冷数据中心技术主要为冷板式，其次是浸没式，目前冷板式液冷是主流形式。
- ◆ 液冷结构包括一次、二次设备：一次侧：包含冷却塔、一次侧管网、一次侧冷却液。二次侧：包含CDU、液冷机柜、二次侧管网和二次侧冷却液。主要通过冷却液温度的升降实现热量转移；两个部分通过 CDU中的板式换热器发生间壁式换热；

图表36 液冷整体示意图



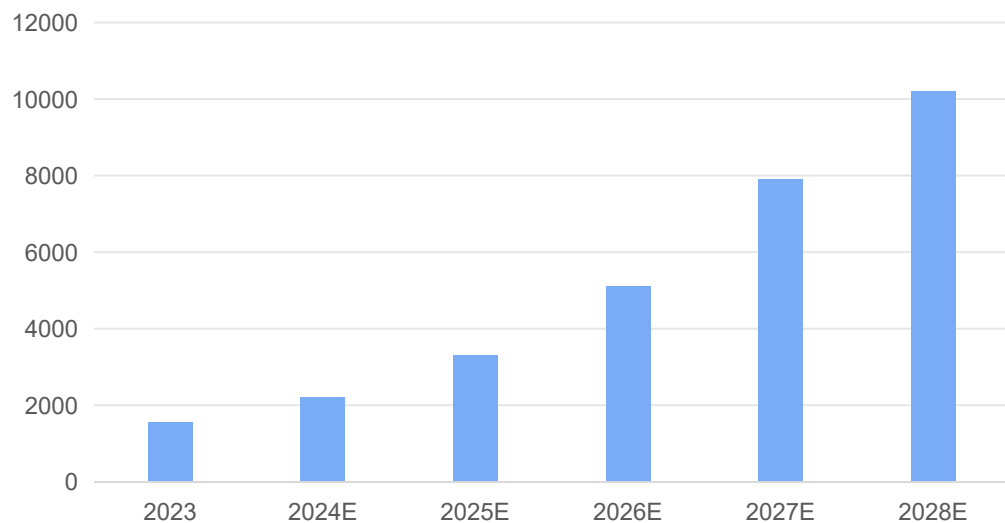
图表37 冷板式液冷示意图



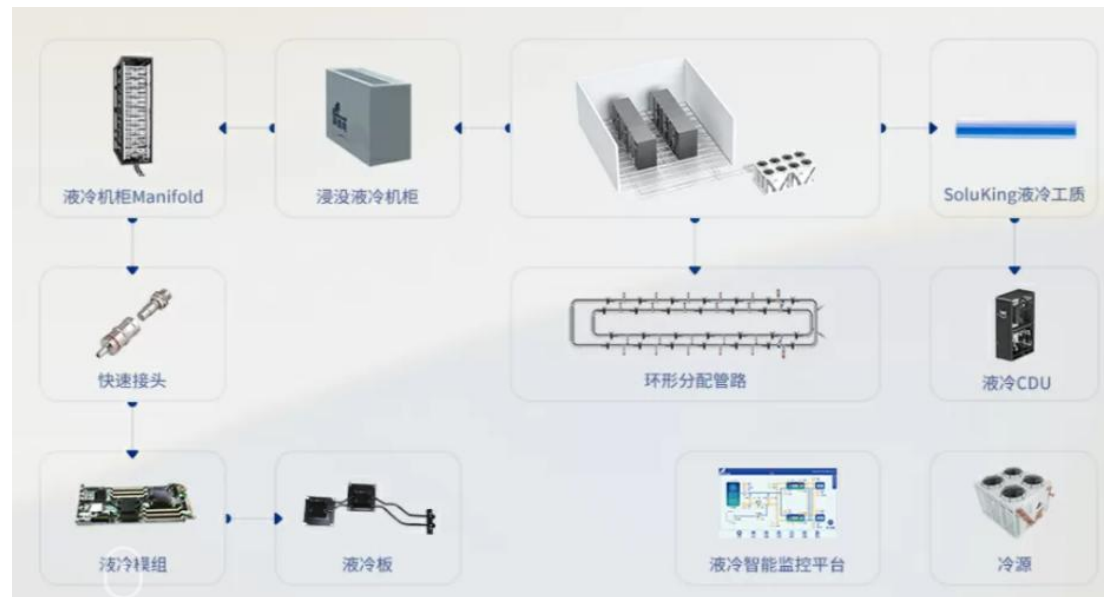
液冷：液冷市场将迎高复合增长，国产厂商布局全面

- ◆ 中国液冷服务器市场在2023年继续保持快速增长。2023全年中国液冷服务器市场规模达到15.5亿美元，与2022年相比增长52.6%，其中95%以上均采用冷板式液冷解决方案。IDC预计，2023-2028年，中国液冷服务器市场年复合增长率将达到45.8%，2028年市场规模将达到102亿美元。
- ◆ 国内专业液冷温控供应商主要有英维克、申菱环境、佳力图、高澜股份、同飞股份等，其中英维克机柜温控收入占比41%，截止2024年3月公司在液冷链条的累计交付已达900MW。合作模式为绑定服务器厂商如超聚变、宁畅、新华三等推出定制化液冷解决方案。

图表38 中国液冷服务器市场规模及预测，2023-2028（单位：百万美元）



图表39 全链条液冷解决方案



目录

contents

1 AI算力三要素

2 国产算力产业链受益

3 算力产业趋势和重点标的

4 风险提示

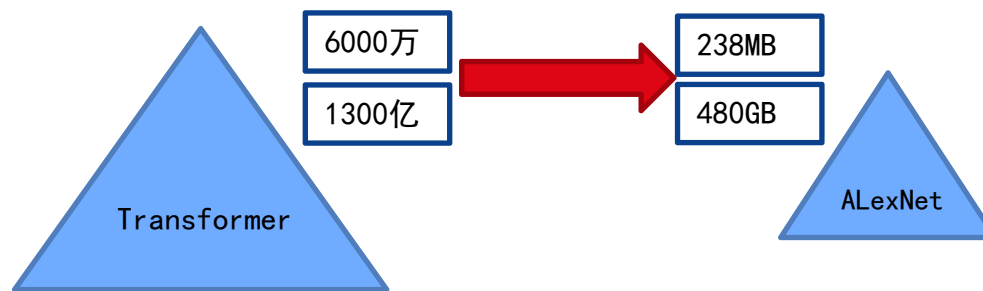
- ◆ 边缘终端嵌入大模型。高通在骁龙芯片上成功推理 stable diffusion 模型，华为推出基于手机算力的“智能搜图”功能。
- ◆ 大模型转小模型是必须之路。基于大模型开发的小模型针对性更强，可以迅速地应用到各行各业，给用户提供服务，常用的方法有知识蒸馏、剪枝、压缩、量化 模型裁剪。

图表40 特斯拉自动驾驶大脑FSD拆解



- ◆ 目前自动驾驶L3/L4瓶颈不在AI计算，而在储存带宽。以特斯拉FSD例子所示，每秒可加载5.12次权重模型。即使算力可达10万TOPs，每秒运算次数不会超过6次。

图表41 大模型转小模型例子



- ◆ AlexNet用于视觉识别神经网络，Transformer用于视觉应用模型。
- ◆ AlexNet在卷积层的参数量为34944，全连接层到卷积层参数量为37752832；Transformer在Embedding层参数量为31782912。
- ◆ 经计算可得6000万大模型可转换为238MB小模型，1300亿大模型可转换为480GB小模型。


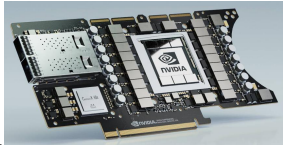

产业趋势二：算力/带宽/内存提升硬件支撑

- ◆ 目前面向行业的模型已实现边缘嵌入，如图像识别、人脸识别、简单交互，但这些算法都是基于CV算法、CNN算法，模型参数上限有限，非生成式transform架构；
- ◆ 未来大模型趋势将由作坊式向工业式转变，降低门槛。在医药研发、卫星遥感、灾害评估、自然生态监测等场景、医疗、金融、工业、教育等多个领域发挥作用。
- ◆ 目前实现大模型的运行需要具备三个要素：1、算力；2、带宽；3、内存。其中硬件是急需解决的瓶颈，在AI算力逐渐突破的背景下，所需硬件耗费将越来越多，难以实现对带宽和内存的支撑。

图表42 不同参数的大模型需求

模型参数	所需算力 (PFLOPS/day)	所需内存	所需带宽	所需硬件 (A100 GPU)
10亿	21	1GB	5.12GB/s	108张
30亿	63	3GB	15.36GB/s	325张
100亿	209	10GB	51.2GB/s	1083张
1750亿	3650	175GB	896GB/s	18953张

图表43 相关硬件价格

	骁龙8 Gen 2 CPU	1134元/颗
	英伟达A100 GPU	15万元/个
	美光GDDR6X 16G 内存	216元/颗

- ◆ 算力互联提供国内顶级的高性能AI算力支持。算力互联启动新一轮自营AI算力资源池大规模集群扩容，包括H800 AI算力服务，采用ACMCube™一体化算力基础设施解决方案，核心计算节点HGX单主板八个H800 80G GPU NVSwitch，使用Nvme高速本地数据盘和全新的PCIE5.0 IB网络，由多种最新技术构建了目前最为强大的AI算力节点。
- ◆ 市场空间巨大，产业布局加速。2022年全球数据中心互连（DCI）技术市场规模达到54.02亿美元。中国移动2023年建设了长达5000KM的400G OTN线路测试；华工科技发布400GZR/ZR+/ZR+Pro系列相干光模块。

图表44 国家“东数西算”工程驱动算力网络发展



图表45 算力互联AI/图形计算节点标准型价格表（部分）

资源名称	节点技术规格	标准单价 (元/节点天)
I-H800-80N	8*H800 80G Nvlink/lce 96c 2.1GHz/2TB DDR5/15.36TB NVME/800Gb	¥ 3,166.67
I-H800-80N	8*H800 80G Nvlink/lce 80c 2.0GHz/2TB DDR5/15.36TB NVME/800Gb	¥ 3,166.67
I-A800-80N	8*A800 80G Nvlink/lce 64c 2.6GHz/2TB DDR4/16TB NVME/400Gb	¥ 2,933.34
I-A100-80N	8*A100 80G Nvlink/lce 72c 2.7GHz/2TB DDR4/16TB NVME/400Gb	¥ 3,168.00
I-A100-40N	8*A100 40G Nvlink/lce 64c 3.0GHz/1TB DDR4/6.4T NVME/400Gb	¥ 2,496.00
I-A100-40D	8*A100 40G Nvlink DGX/lce 64c 2.6GHz/1TB DDR4/8T NVME/800Gb	¥ 2,496.00
I-A100-40P	8*A100 40G Clink/lce 56c 2.6GHz/512GB DDR4/3.84TB NVME/200Gb	¥ 1,382.40
N-V100-32N	8*V100 32G Nvlink/lce 32c 2.4GHz/512GB DDR4/7.68TB NVME/200Gb	¥ 864.00
N-V100-16	2*Tesla V100 GPU/Dhyana 64c 2.0GHz/192GB DDR4/1TB SATA/10Gb	¥ 144.00

紫光股份——新一代云计算基础设施领先者

- ◆ 公司看点：紫光股份是全球新一代云计算基础设施建设和行业智慧应用服务的领先者，提供技术领先的网络、计算、存储、云计算、安全和智能终端等全栈 ICT 基础设施及服务，在国内交换机和服务器领域的市场份额名列前茅。
- ◆ 业务增速：2023年营业收入773.08亿元，同比增长4.39%；净利润36.85亿元，同比下降1.53%。2024年一季度营业收入170.06亿元，同比增长2.89%；净利润6.99亿元，同比下降9.85%。公司加快“云-网-算-存-端”布局 and 数智创新，主动把握新一轮技术发展方向，在“AI in ALL”技术战略指引下，推出多款服务器、交换机等数字基础设施。

图表46 紫光股份子公司新华三交换机



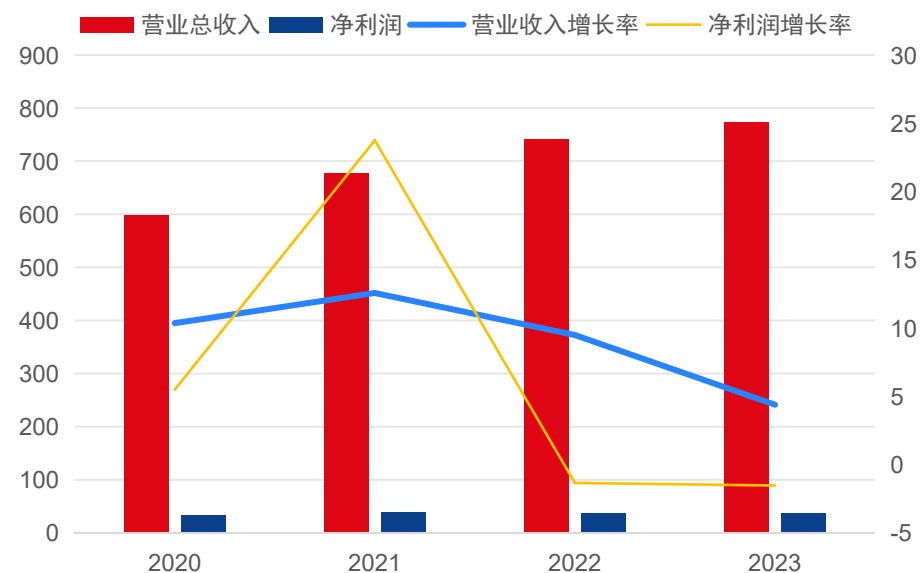
- ◆ 紫光股份旗下交换机主要分为数据中心交换机、园区网交换机、全光网络和工业和安防交换机。

图表47 紫光股份子公司新华三服务器



- ◆ 紫光股份旗下服务器主要分为机架式、刀片、塔式、高密度、边缘和关键业务服务器。
- ◆ 旗下有H3C(新华三)服务器品牌及HPE(惠普)服务器代理品牌。


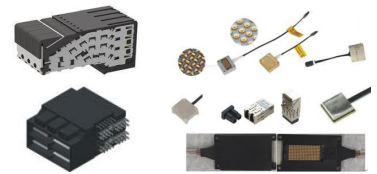

图表48 紫光股份2020-2023营业收入、净利润及增速
(单位：亿元)



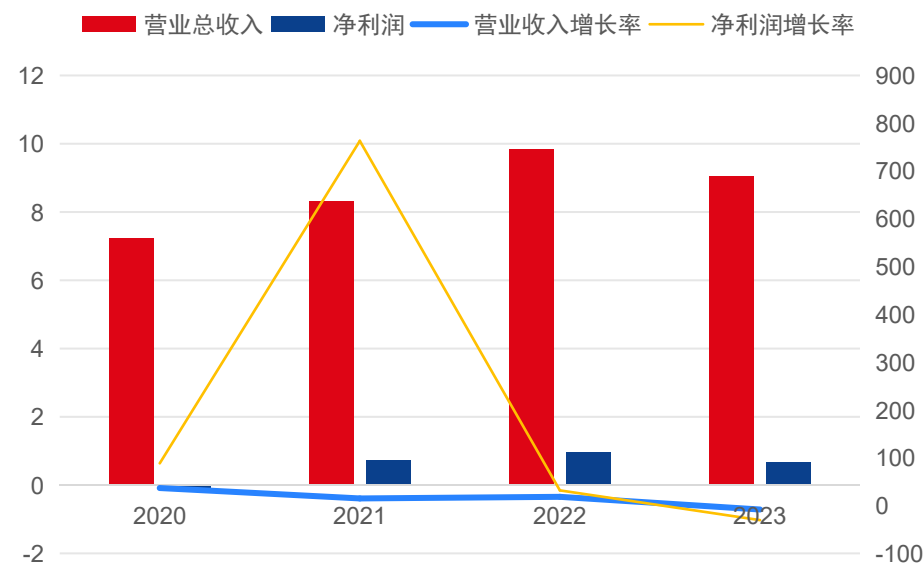
华丰科技：受益国防建设+国产替代，打开连接器成长空间

- ◆**公司看点：**全球光电连接器及互连方案提供商。研制的连接器从低频到高频、从电口到光口，从标准型到特殊定制，能适应深海环境、高压环境、高振动冲击环境、电磁波干扰环境、高低温环境、太空宇航环境等多种特殊环境。为用户提供互连解决方案一站式服务。致力于向全球防务装备、通讯装备、轨道交通装备和新能源汽车提供更专业、更可靠、更智能的互连技术和产品。
- ◆**重点客户：**华为、中兴等移动通信设备制造商，航天科工、中国电科、中国兵工等航空航天及防务集团下属单位以及上汽通用五菱、比亚迪等汽车制造厂商。
- ◆**业绩增速：**2023年营业收入9.04亿元，同比下降8.17%；净利润0.67亿元，同比下降30.36%。2024年一季度营业收入2.26亿元，同比上升26.48%；净利润877.36万元，同比下降21.26%。在连接器行业加速国产替代化的背景下，公司有望开拓成长空间，实现快速增长。

图表49：公司主流产品及应用场景

类别	主流产品	产品	产品图	应用场景
防务类连接产品	系统互连产品、防务连接器、组件	J29M矩形盲插连接器、CPBS无缆化母板组件、树型拓扑结构线缆组件		应用领域主要为航天、船舶、电子、防务装备等领域的信息系统电子设备与设备间、设备内部、模块与板卡间、印制板间的系统互连。
通讯类连接产品	高速连接器、印制板连接器、光通讯连接器、线缆组件等	P 系列、I/O连接器等、PCB 板上电源连接器、光模块等		主要用于点对点及点对多点传输接口、背板交换应用、相控阵雷达数据通信、以太网、光纤通道、Infiniband QDR 等场合
工业类连接产品	轨道交通类产品、新能源汽车类产品	轨道交通连接器、电气车钩总成、BDU/PDU 充配电系统总成等		广泛应用于高速列车、电力机车及地铁轻轨的电气控制与集成布线系统中，为不同设备或功能单元之间的电气或信号提供电连接。

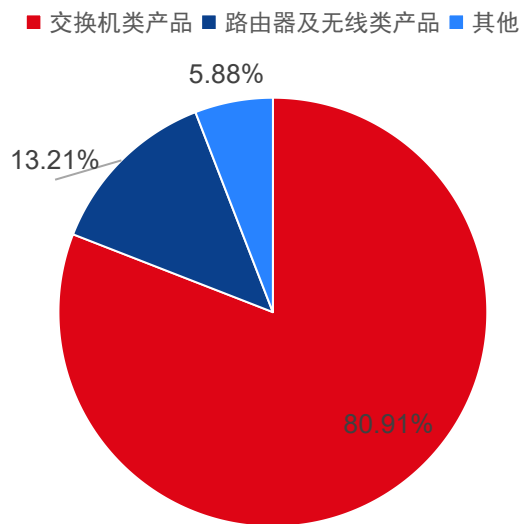
图表50：2020-2023营业收入、净利润及增速（单位：亿元）



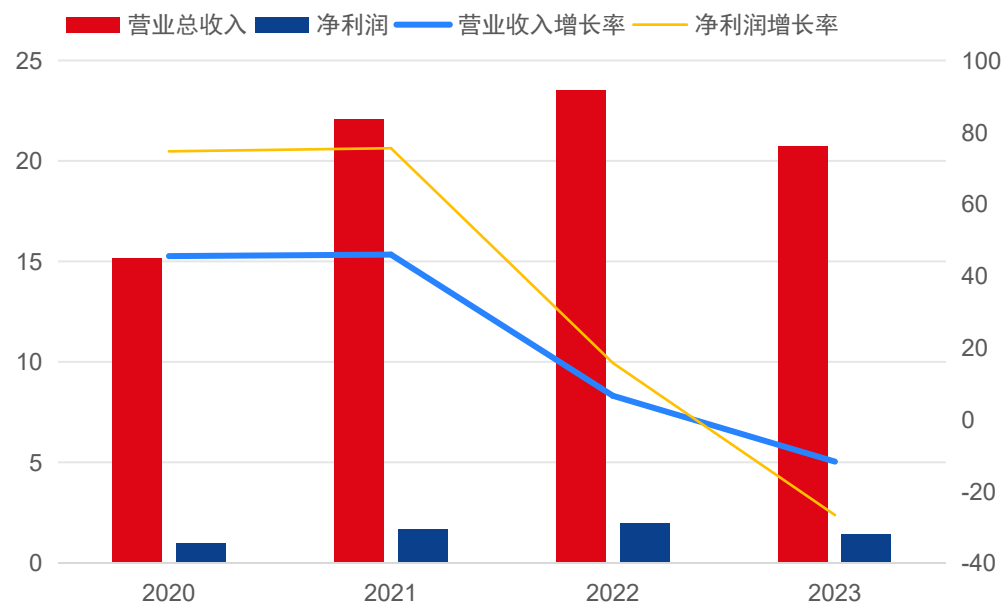
菲菱科思：中高端交换机突破，长期受益国产算力爆发

- ◆ **公司看点：**公司主要业务为网络设备的研发、生产和销售，网络设备是互联网的基础设施，属于信息化建设所需的基础架构产品。公司主要为网络设备品牌商提供交换机、企业级路由器、WiFi 无线产品、通信设备组件等产品的研发和制造服务。公司致力于为客户提供网络设备研发设计、生产制造、销售服务的一站式服务，满足客户多品种、多批次、高质量、低成本的产品制造服务需求。
- ◆ **重点客户：**新华三、S 客户、锐捷网络等知名网络设备品牌商的合作伙伴。
- ◆ **业绩增速：**2023年营业收入20.75亿元，同比下降11.80%；净利润1.44亿元，同比下降26.68%。2024年一季度实现营业收入4.07亿元，同比下降19.45%，实现净利润0.34亿元，同比下降38.45%。公司抓住数字经济快速发展以及新一代算力信息基础设施建设周期中的市场机遇，建立自己鲜明优势，推动公司竞争中稳步向前发展。

图表51：主要产品营收占比



图表52：2020-2023营业收入、净利润及增速（单位：亿元）



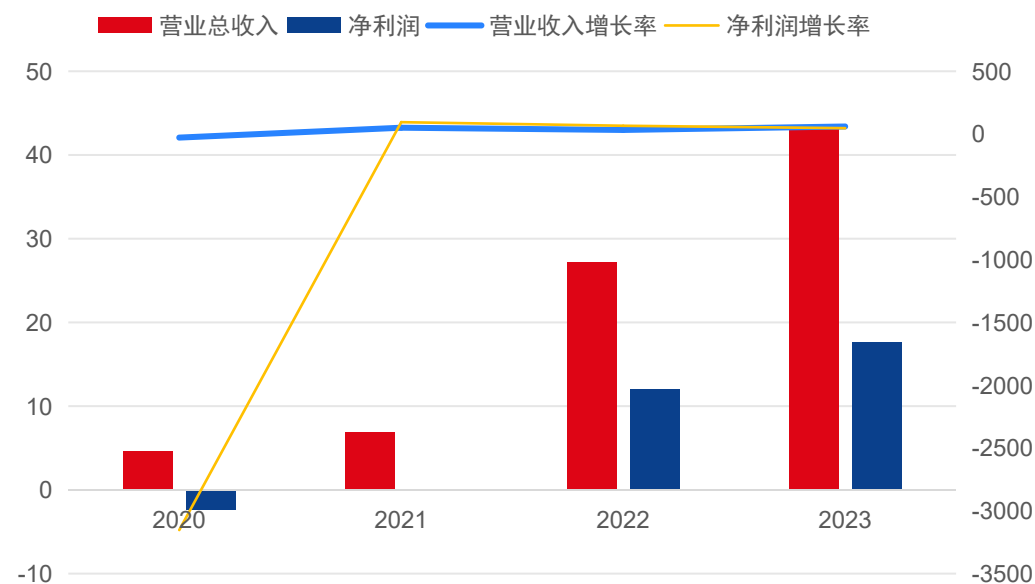
润泽科技：国内IDC龙头，盈利能力领先同行业

- ◆ **公司看点：**公司是国内IDC龙头企业，致力于面向大数据、云计算、物联网、5G 技术等行业应用需求，以算力为基础、数字技术为手段、智慧应用为示范，为各行业提供新一代数字经济产业技术、产品、服务和系统解决方案。
- ◆ **重点客户：**中国电信（第一）、中国联通（第二）、字节跳动、华为、京东、快手等。
- ◆ **业务增速：**2023年营业收入43.51亿元，同比上升60.27%；净利润17.58亿元，同比上升46.84%。2024年一季度，公司实现营业收入12.42亿元，同比增长53.65%；净利润4.75亿元，同比增长44.00%。

图表53 公司主流业务（单位：万元）

产品	总投资金额	拟使用募集资金
润泽（佛山）国际信息港A2、A3、数据中心项目	169668	169668
润泽（平湖）国际信息港A2数据中心项目	81306	76306
偿还银行借款	243800	209026
中介机构费用及相关发行费用	15000	15000
合计	509774	470000

图表54 2020-2023营业收入、净利润及增速（单位：亿元）



目录

contents

1 AI算力三要素

2 国产算力产业链受益

3 算力产业趋势和重点标的

4 风险提示

- ◆ 算力落地不及预期；应用是算力落地的关键，目前大模型开发、迭代已有一年，但终端爆品应用还未出现，大模型落地需要应用场景加持，市场在等待爆款应用的推出。
- ◆ 芯片成本过高影响落地进度。目前英伟达芯片供给虽有改善，但需求增长的速度远大于产能速度，订单执行周期过长，此外国内芯片虽需求旺盛，但产能受限。
- ◆ 智算中心建设进度不及预期。智算中心建设需要综合考虑，此外智算所需的卡以及客户都是市场关心的重点，这些资源要素在推进过程中面临着不确定性。

代码	简称	总市值 (亿元)	EPS-2022A	EPS-2023A	EPS-2024E	PE-2022A	PE-2023A	PE-2024E
300308.SZ	中际旭创	1597.77	1.54	2.80	1.30	130.54	73.51	32.28
300502.SZ	新易盛	784.39	1.79	0.97	0.46	86.81	113.95	53.34
300394.SZ	天孚通信	536.63	1.03	1.85	0.71	133.18	73.52	39.08
000977.SZ	浪潮信息	534.83	1.39	1.18	0.21	25.71	30.00	24.80
000938.SZ	紫光股份	647.24	0.75	0.74	0.15	29.99	30.78	25.60
301191.SZ	菲菱科思	53.00	4.09	2.08	0.49	27.13	36.71	29.29
603118.SH	共进股份	55.40	0.29	0.08	0.01	24.43	85.61	14.63
300913.SZ	兆龙互连	92.68	0.71	0.38	0.07	71.05	94.43	62.86
002130.SZ	沃尔核材	173.61	0.49	0.56	0.15	28.25	24.78	20.63
603912.SH	佳力图	33.48	0.12	0.07	0.00	91.73	94.67	48.55
301018.SZ	申菱环境	54.54	0.69	0.40	0.19	32.80	51.99	19.10
002837.SZ	英维克	178.01	0.64	0.61	0.11	63.51	51.75	34.44
300442.SZ	润泽科技	469.20	1.57	1.03	0.28	39.16	26.63	20.74

公司评级体系

收益评级：

- 买入 — 未来6个月的投资收益率领先沪深300指数15%以上；
- 增持 — 未来6个月的投资收益率领先沪深300指数5%至15%；
- 中性 — 未来6个月的投资收益率与沪深300指数的变动幅度相差-5%至5%；
- 减持 — 未来6个月的投资收益率落后沪深300指数5%至15%；
- 卖出 — 未来6个月的投资收益率落后沪深300指数15%以上。

风险评级：

- A — 正常风险，未来6个月投资收益率的波动小于等于沪深300指数波动；
- B — 较高风险，未来6个月投资收益率的波动大于沪深300指数波动。

行业评级体系

收益评级：

领先大市 — 未来6个月的投资收益率领先沪深300指数10%以上；

同步大市 — 未来6个月的投资收益率与沪深300指数的变动幅度相差-10%至10%；

落后大市 — 未来6个月的投资收益率落后沪深300指数10%以上；

风险评级：

A — 正常风险，未来6个月投资收益率的波动小于等于沪深300指数波动；

B — 较高风险，未来6个月投资收益率的波动大于沪深300指数波动。

分析师声明

李宏涛声明，本人具有中国证券业协会授予的证券投资咨询执业资格，勤勉尽责、诚实守信。本人对本报告的内容和观点负责，保证信息来源合法合规、研究方法专业审慎、研究观点独立公正、分析结论具有合理依据，特此声明。

本公司具备证券投资咨询业务资格的说明

华金证券股份有限公司（以下简称“本公司”）经中国证券监督管理委员会核准，取得证券投资咨询业务许可。本公司及其投资咨询人员可以为证券投资人或客户提供证券投资分析、预测或者建议等直接或间接的有偿咨询服务。发布证券研究报告，是证券投资咨询业务的一种基本形式，本公司可以对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向本公司的客户发布。

免责声明：

本报告仅供华金证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因为任何机构或个人接收到本报告而视其为本公司的当然客户。

本报告基于已公开的资料或信息撰写，但本公司不保证该等信息及资料的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映本公司于本报告发布当日的判断，本报告中的证券或投资标的价格、价值及投资带来的收入可能会波动。在不同时期，本公司可能撰写并发布与本报告所载资料、建议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，本公司将随时补充、更新和修订有关信息及资料，但不保证及时公开发布。同时，本公司有权对本报告所含信息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。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以本公司向客户发布的本报告完整版本为准。

在法律许可的情况下，本公司及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或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务，提请客户充分注意。客户不应将本报告为作出其投资决策的惟一参考因素，亦不应认为本报告可以取代客户自身的投资判断与决策。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，无论是否已经明示或暗示，本报告不能作为道义的、责任的和法律的依据或者凭证。在任何情况下，本公司亦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。

本报告版权仅为本公司所有，未经事先书面许可，任何机构和个人不得以任何形式翻版、复制、发表、转发、篡改或引用本报告的任何部分。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“华金证券股份有限公司研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。

华金证券股份有限公司对本声明条款具有惟一修改权和最终解释权。

风险提示:

报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价或询价。投资者对其投资行为负完全责任，我公司及其雇员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。

华金证券股份有限公司

办公地址:

上海市浦东新区杨高南路759号陆家嘴世纪金融广场30层

北京市朝阳区建国路108号横琴人寿大厦17层

深圳市福田区益田路6001号太平金融大厦10楼05单元

电话: 021-20655588

网址: www.huajinsec.cn