

计算机

乘大模型之风，AI 芯片元老寒武纪再度起航

大模型时代来临，加速计算需求大幅增加

以 Transformer 结构为主的生成式带动了加速计算的需求突破式增长，随着 Scaling law 的进一步拓展，Nvidia 数据中心的收入从 2024 财年 Q1 的 43 亿美金增加至 2025 财年一季度的 226 亿美金，同期超威电脑和 DELL 的 AI 服务器收入也大幅增加。我们预计未来针对训练和推理的 AI 算力需求有望长时间高景气，据 IDC 及 TheNextPlatform 统计，2023 年全球 AI 服务器市场规模预计将超 500 亿美元，2027 年预计将超 946 亿美元。在这一场科技盛宴中，Nvidia 凭借前瞻的布局、CUDA 生态的壁垒、从 P100 到未来 R100 芯片的快速迭代以及完善的集群产品占据了主要的市场地位。

美国限制叠加对 AI 算力的需求，自主可控的 AI 芯片公司有望重新占领部分国内市场，寒武纪凭借优秀的软硬件能力有望脱颖而出

美国从 2022 年开始逐步加强对国产 AI 芯片的限制，从将国内 AI 芯片企业列入实体清单到限制 Nvidia 的芯片出口，国内企业获取海外算力的难度日渐加大，同时从 2024 年开始，运营商和地方政府不断加强对算力基础设施的建设力度，政策上也在不断扶持自主算力，在此背景下针对自主可控的 AI 算力的需求呼之欲出。参考海外市场，在这一广阔的市场机遇中，市场份额有望相对集中但难以一家独大，除去英伟达，我们看到 Google 和 AMD 都在努力提升自身芯片的能力。

寒武纪是 AI 芯片的元老，自 2016 年成立至今，从 AI 芯片授权走向自研云端 AI 芯片并形成了完整的软硬件产品与生态

寒武纪成立于 2016 年，创始人兼董事长陈天石博士在中科院时就参与到 AI 芯片的学术前沿研究并发表多篇论文，Google 在随后发表的论文中也参考了陈天石博士的研究成果。公司从芯片 IP 授权起步，逐步设计并自研了包括边缘产品线、云端产品线在内的多款芯片，并从底层自建了完善的软件体系，包括底层 Runtime、构建算子的编程语言 BANG 和较为完备的算子库。目前，智能计算集群系统业务和云端芯片构成了主要的收入来源，2023 年公司实现营业收入 7.09 亿元。作为领先的国产 AI 芯片企业，寒武纪拥有较强的产品研发能力和较高的市占率，在 AI 算力需求高速增长、自主可控的 AI 算力需求不断加速、算力基础设施有望加速建设的背景下，寒武纪有望成为国内自主 AI 算力供应商中的主要参与者。

建议关注：

自主 AI 算力的国内企业：寒武纪

风险提示：AI 算力景气度下降的风险、公司产品供应端的风险、AI 芯片竞争加剧的风险、客户集中度较高的风险

证券研究报告

2024 年 07 月 07 日

投资评级

行业评级

强于大市(维持评级)

上次评级

强于大市

作者

缪欣君

分析师

SAC 执业证书编号：S1110517080003
miaoxinjun@tfzq.com

刘鉴

联系人

liujianb@tfzq.com

行业走势图



资料来源：聚源数据

相关报告

- 《计算机-行业点评:央国企信创有望加速,信创迎来 2.0 阶段》2024-06-30
- 《计算机-行业专题研究:鸿蒙: AI 与 OS 深度融合,开启 OS Agent 新时代》2024-06-23
- 《计算机-行业点评:鸿蒙: 华为 2024 年“一号工程”,不止万物互联更是“AI+OS”》2024-06-17

内容目录

1. 大模型时代来临，加速计算需求突破式增长.....	5
1.1. 生成式 AI 带动训练和推理的 AI 芯片需求快速上升.....	5
1.1.1. Nvidia 数据中心与 AI 服务器厂商收入突破式增长.....	6
1.1.2. 互联网公司资本支出因 AI 逐步增加.....	7
1.2. Nvidia 软硬件体系生态紧耦合，占据 AI 芯片主要份额.....	7
1.2.1. 从 P100 到 R100，英伟达 8 年期间多次迭代 AI 芯片.....	7
1.2.2. CUDA 生态，强大的生态造就高壁垒.....	11
2. 自主可控带来国产 AI 芯片历史机遇，寒武纪有望脱颖而出.....	11
2.1. 国内 AI 芯片与智算中心急速增长.....	12
2.2. 美国芯片禁令政策，自主算力政策扶持.....	13
2.3. 对比海外，AI 芯片市场份额集中但难以一家独大.....	15
3. AI 芯片元老，从 IP 到芯片.....	15
3.1. 从 IP 到自研芯片，寒武纪为 AI 芯片而生.....	15
3.2. 核心团队出自中科院，董事长长期耕耘 AI 芯片，引领产业方向.....	16
4. 寒武纪技术能力优秀，具备 AI 芯片完整软硬件技术栈.....	19
4.1. 通用智能芯片系统复杂，寒武纪具备软硬件全栈技术.....	19
4.2. 产品体系从 IP 到端侧再到云端，软硬件深度耦合.....	20
4.2.1. 硬件覆盖 IP 和云边端一体化产品，单芯片最大算力高达 256TOPS.....	21
4.2.2. Cambricon Neuware：从底层自建适配硬件的软件生态.....	23
4.2.3. 多芯互联组网提升并行效率，实现多卡和分布式任务高效执行.....	24
4.3. 智能计算集群贡献主要收入，研发上更加聚焦，费用率逐步下降.....	24
4.4. 寒武纪在软/硬件生态综合能力强，研发针对大模型实现多点突破.....	27
5. 风险提示.....	28

图表目录

图 1：Transformer 对 AI 芯片需求快速上升.....	5
图 2：AI 高速增长下算力供给与需求的差距在加大.....	5
图 3：2020-2027 全球 AI 服务器市场规模（百万美元）.....	6
图 4：2022-2026 年全球 AI 服务器预估出货量（千台）.....	6
图 5：FY22Q2-FY25Q1 英伟达数据中心收入变化（亿美元）.....	6
图 6：FY23Q1-FY24Q3 超微电脑服务器与存储系统营收及 YOY.....	6
图 7：FY23-FY25Q1 戴尔 AI 服务器营收及 YOY.....	6
图 8：22Q2-24Q1 微软、谷歌、Meta、亚马逊资本支出对比.....	7
图 9：8 年间英伟达 GPU 实现 1053 倍性能增长.....	7
图 10：NVIDIA@NVLink™演示图.....	8
图 11：NVLink 性能的迭代历程.....	9
图 12：NVSwitch 的规格迭代对比.....	10

图 13: 2007-2021 年全球超级计算机 500 强中 IB 及其他网络供应商技术占比.....	10
图 14: NVIDIA Spectrum-X 较传统以太网 AI 性能提升 1.6 倍.....	11
图 15: 全栈优化的 NVIDIA Spectrum-X.....	11
图 16: CUDA 发展历程的持续演进与优化.....	11
图 17: 2023-2027 年中国加速计算服务器市场预测.....	12
图 18: 2022-2027 年中国 AI 服务器工作负载预测.....	12
图 19: Nvidia/Google 等 AI 加速芯片市场的出货量.....	15
图 20: AI 加速芯片在数据中心市场的份额快速提升.....	15
图 21: 中科寒武纪发展历程.....	16
图 22: 股权结构图 (截至 2024 年一季度末).....	16
图 23: Google TPU 论文引用寒武纪相关工作的专门段落.....	17
图 24: FY2019-FY2023 中科寒武纪研发人员数量及同比.....	18
图 25: 2023 年中科寒武纪研发人员学历构成.....	18
图 26: FY2020-FY2023 中科寒武纪研发费用 (亿元) 及同比.....	18
图 27: FY2019-FY2023 中科寒武纪人均研发费用 (万元) 及同比.....	18
图 28: 中科寒武纪主要产品线.....	20
图 29: 寒武纪基础软件系统平台.....	21
图 30: 寒武纪 Cambricon Neuware 架构层级关系.....	23
图 31: 寒武纪训练软件平台.....	23
图 32: 寒武纪全新推理加速引擎 MagicMind 结构.....	23
图 31: MLU370-X8 中整合了双芯片四芯粒思元 370.....	24
图 32: FY2017-FY2024Q1 中科寒武纪收入 (亿元) 及同比.....	25
图 33: FY2021-FY2023 中科寒武纪分业务收入 (亿元).....	25
图 34: FY2017-FY2023 中科寒武纪毛利率.....	25
图 35: FY2017-FY2023 中科寒武纪净利润 (亿元).....	25
图 36: FY2019-FY2023 中科寒武纪期间费用率.....	25
图 37: FY2019-FY2023 中科寒武纪研发费用 (亿元) 及同比.....	25
图 38: FY2017-FY2023 中科寒武纪经营净现金流 (亿元) 及同比.....	26
图 39: FY2017-FY2023 中科寒武纪货币资金 (亿元) 及同比.....	26
图 40: FY2017-FY2023 中科寒武纪存货 (亿元) 及同比增长.....	26
图 41: FY2017-FY2023 中科寒武纪预付款项 (亿元) 及同比增长.....	26
图 42: 公司应对大模型挑战实现的多点核心技术突破.....	28
表 1: P100、V100、A100、H100、GB200 规格参数对比.....	8
表 2: 2024 年运营商与部分各地智算中心的招标公告.....	13
表 3: 2022 年以来美国历次针对算力的禁令.....	13
表 4: 中央到地方政府出台针对算力基础设施的支持政策.....	14
表 5: 公司董事简介.....	16
表 6: 陈天石博士及相关团队早期相关工作.....	17
表 7: 2023 年寒武纪推出的股权激励计划.....	19

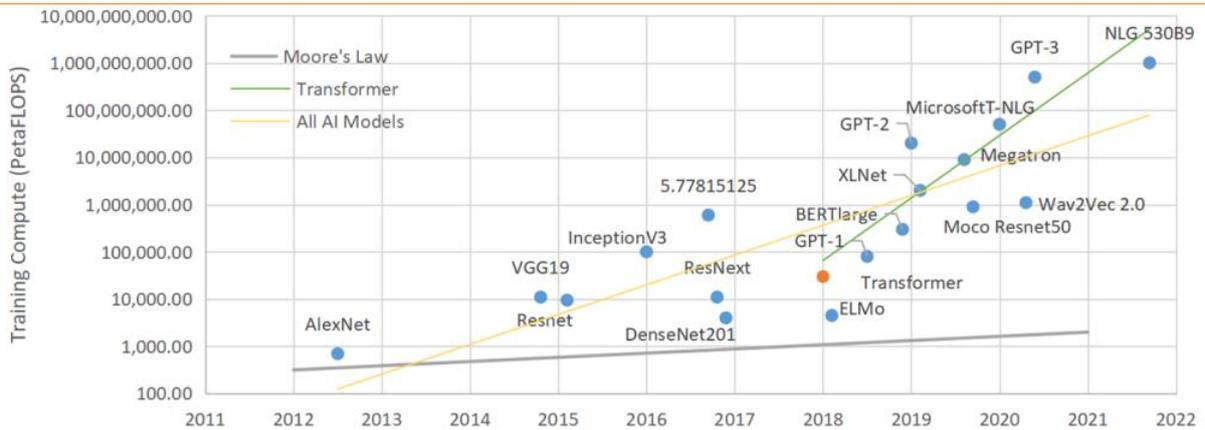
表 8：中科寒武纪核心技术框架结构	19
表 9：寒武纪终端智能处理器 IP	21
表 10：寒武纪边缘产品	22
表 11：寒武纪云端产品性能参数	22
表 12：寒武纪训练整机产品	23
表 15：寒武纪智能芯片技术及其先进性	27

1. 大模型时代来临，加速计算需求突破式增长

1.1. 生成式 AI 带动训练和推理的 AI 芯片需求快速上升

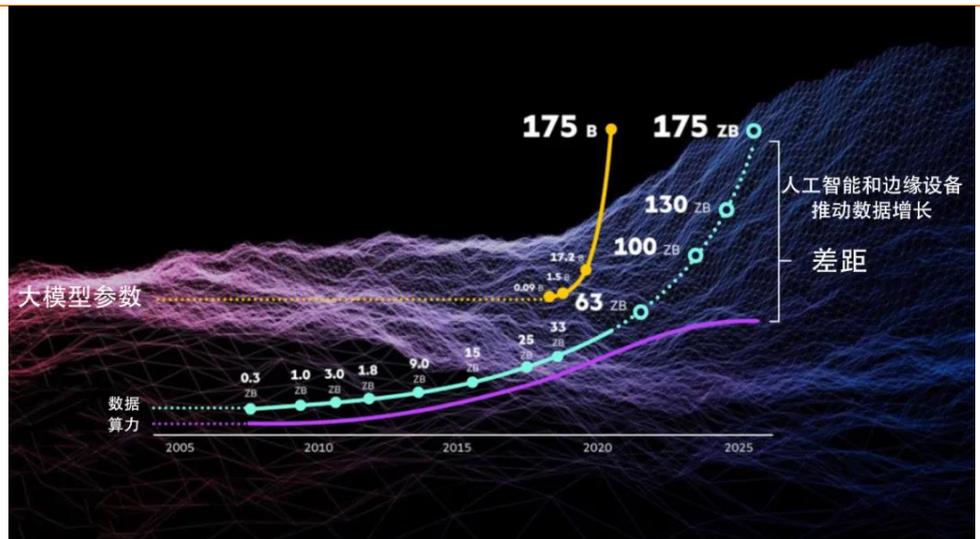
Transformer 对 AI 芯片需求快速上升,算力缺口持续扩大。GPT 和 Sora 等基于 Transformer 架构的生成式模型遵循规模效应 (Scaling Law)。Scaling Law 由 OpenAI 于 2020 年提出,他们认为基于 Transformer 的自然语言和其他模态的模型,其模型性能与模型参数量存在幂律关系。而模型参数量也与实施大模型训练及部署硬件的计算存储能力直接相关。OpenAI 公司在 2018 年对 AI 计算的趋势与硬件发展趋势做了分析,基于 Transformer 的 AI 模型的算力需求增速远大于传统的 CNN 模型,并远超依据摩尔定律的演进速度,硬件算力的增长速度与大模型所需的算力水平之间的差距越来越大。

图 1: Transformer 对 AI 芯片需求快速上升



资料来源: 摩崖芯公众号, 天风证券研究所

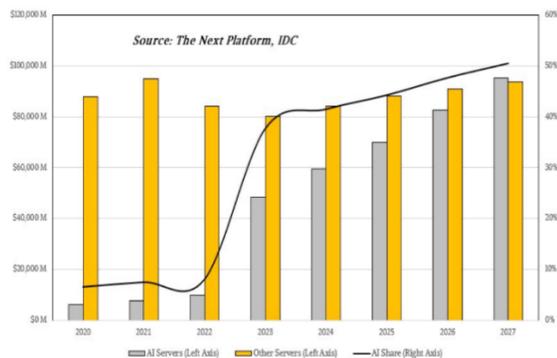
图 2: AI 高速增长下算力供给与需求的差距在加大



资料来源: 天翼智库公众号、天风证券研究所

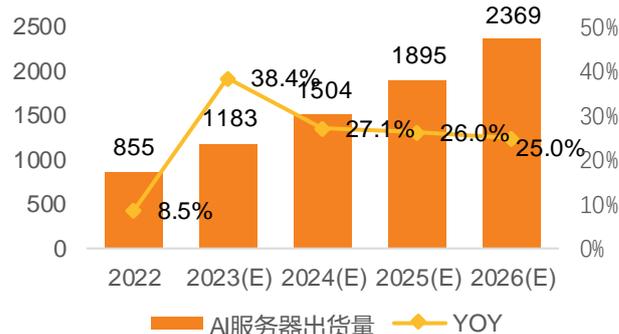
全球 AI 服务器出货加速, 预计 26 年 AI 服务器占比达 15%。据 IDC 及 TheNextPlatform 统计, 2023 年全球 AI 服务器市场规模预计将超 500 亿美元, 2027 年预计将超 946 亿美元, 2023 年-2027 年 CAGR 达 17.41%。出货量方面, 根据 TrendForce 数据, 2023 年 AI 服务器出货量约 120 万台, 占据服务器总出货量的近 9%, 年增长达 38.4%。预计 2026 年, AI 服务器出货量为 237 万台, 占比达 15%, 复合年增长率预计保持 25%。

图 3：2020-2027 全球 AI 服务器市场规模 (百万美元)



资料来源：TheNextplatform、IDC，天风证券研究所

图 4：2022-2026 年全球 AI 服务器预估出货量 (千台)



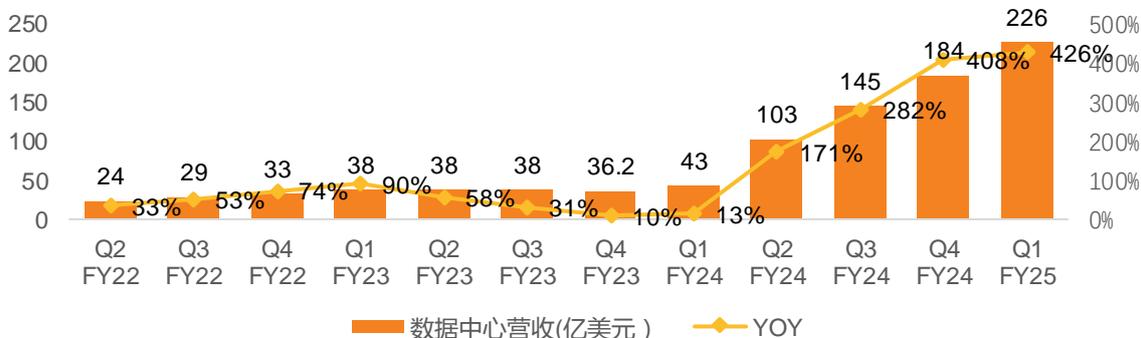
注：估算基础包含搭载 AI 训练、推论用的 GPU、FPGA、ASIC 等加速晶片

资料来源：TrendForce、半导体产业纵横公众号，天风证券研究所

1.1.1. Nvidia 数据中心与 AI 服务器厂商收入突破性增长

自 FY24Q2 起，英伟达数据中心收入加速增长，从 FY24Q1 的 43 亿美元增长至 FY24Q2 的 103 亿美元，FY24Q2 同比增长率达 171%，主要系生成式 AI 竞争已经展开，各大云服务提供商发布了大型 NVIDIA H100 AI 基础架构，以及领先的企业 IT 系统和软件提供商宣布与英伟达合作，将其 AI 技术引入各行各业。此后，在 FY24Q3、FY24Q4 和 FY25Q1 的同比增长率分别达到 282%、408% 和 426%，使其在 FY25Q1 的收入达到 226 亿美元。同时，超微电脑 (Supermicro) 服务器与存储系统的营收显著增长，从 FY24Q1 的 19.67 亿美元激增至 Q3FY24 的 36.99 亿美元，同比增长率高达 217.82%。此外，戴尔的 AI 服务器收入在 FY24 达到 16.20 亿美元，同比增长率达 625.48%，并在 FY25Q1 单季度营收高达 17 亿美元，进一步表明了 AI 服务器市场的强劲需求。

图 5：FY22Q2-FY25Q1 英伟达数据中心收入变化 (亿美元)



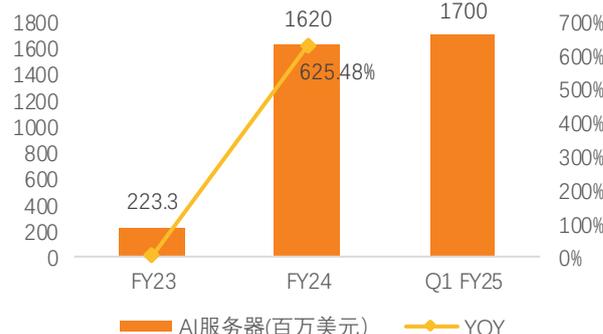
资料来源：Wind，天风证券研究所

图 6：FY23Q1-FY24Q3 超微电脑服务器与存储系统营收及 YOY



资料来源：VisibleAlpha，天风证券研究所

图 7：FY23-FY25Q1 戴尔 AI 服务器营收及 YOY

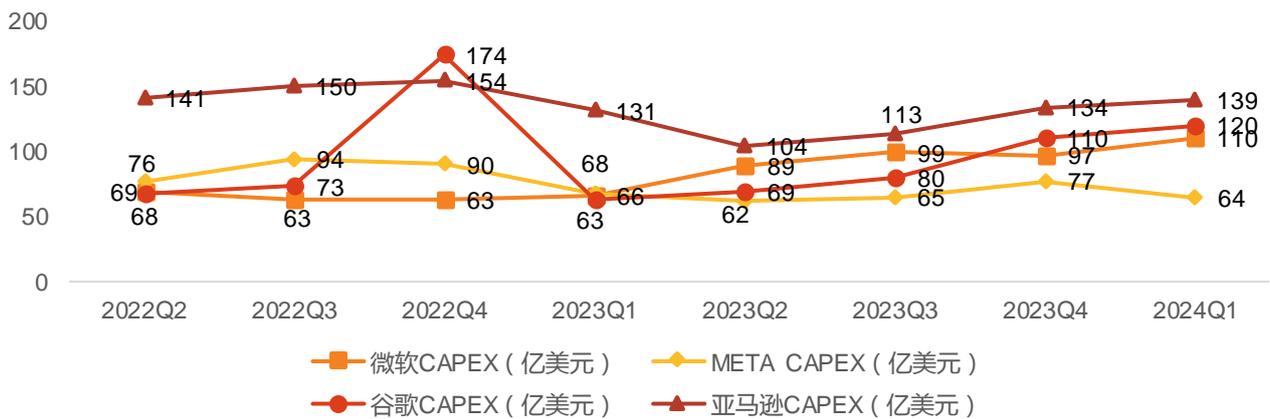


资料来源：VisibleAlpha，天风证券研究所

1.1.2. 互联网公司资本支出因 AI 逐步增加

自 2023 年 Q2 起，以微软、谷歌、亚马逊、Meta 为首的头部互联网公司资本支出整体呈上升趋势，多数投向 AI 及云服务基础设施。其中微软、谷歌资本支大幅增长，创下新高，微软资本支出由 2022 年 Q2 的 69 亿美元增长至 2024 年 Q1 的 110 亿美元，近两年 CAGR 达 26.26%，主要用于通过融资租赁收购的资产来满足云和 AI 产品的需求；谷歌资本支出由 2022 年 Q2 的 68 亿美元增长至 2024 年 Q1 的 120 亿美元，近两年 CAGR 达 32.84%，主要系其在云计算和 AI 模型算力需求上的大力投资，以保持技术基础设施领域的领先地位；2024Q1 亚马逊资本支出为 139 亿美元，位居四大头部互联网公司第一，且公司宣布计划在 2024 年大幅增加资本支出，以推进 AWS 云服务和生成式 AI 技术。

图 8：22Q2-24Q1 微软、谷歌、Meta、亚马逊资本支出对比



资料来源：Wind, 天风证券研究所

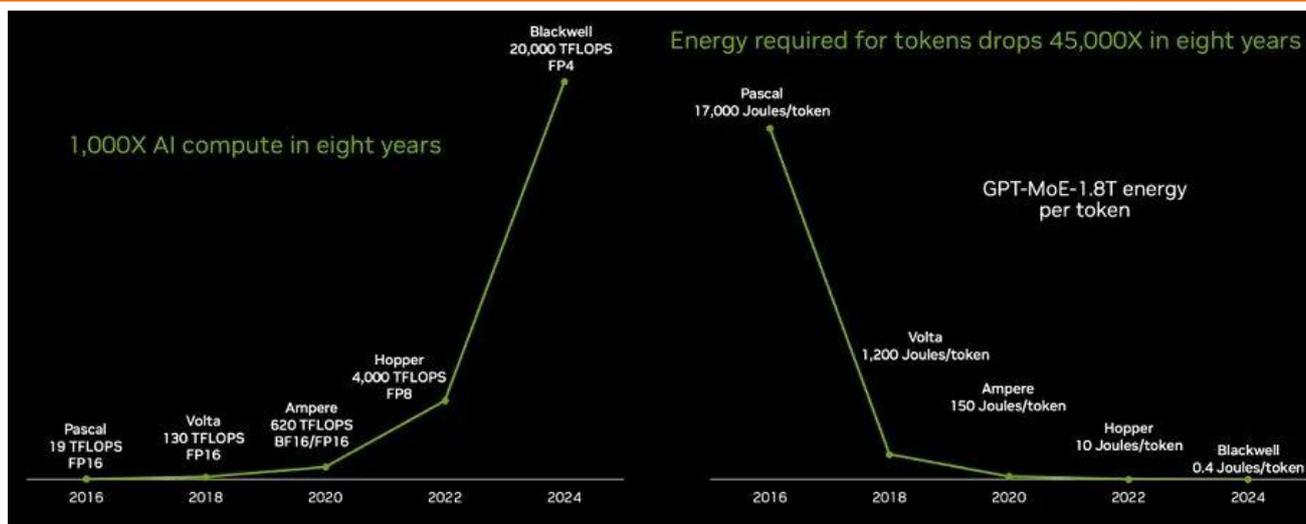
1.2. Nvidia 软硬件体系生态紧耦合，占据 AI 芯片主要份额

1.2.1. 从 P100 到 R100，英伟达 8 年期间多次迭代 AI 芯片

P100 到 B100，英伟达 GPU 性能实现超千倍性能增长。英伟达数据中心产业从 PascalP100 到 BlackwellB100 历经 8 年演进，GPU 性能实现了 1053 倍的增长。性能增长部分得益于浮点精度的降低，从 PascalP100、VoltaV100 和 AmpereA100GPU 的 FP16 性能下降了四倍，转变为 BlackwellB100 所使用的 FP4 格式。但对于大型语言模型训练，精度的降低影响较小，主要得益于数据格式的优化、软件处理的改进以及硬件中应用的复杂数学技术。

B100 发布投产，预计 2024 年 H2 发货。2024 年 3 月，英伟达发布了 BlackwellGPU 架构平台，6 月 2 日，英伟达 CEO 黄仁勋在 COMPUTEX2024 宣布，英伟达 Blackwell 芯片现已开始投产。预计 BlackwellB100 将在 24 年 H2 开始发货，并持续供应至 2025 年，或将成为英伟达 24-25 年营收重要驱动。

图 9：8 年间英伟达 GPU 实现 1053 倍性能增长



资料来源：半导体行业观察公众号，天风证券研究所

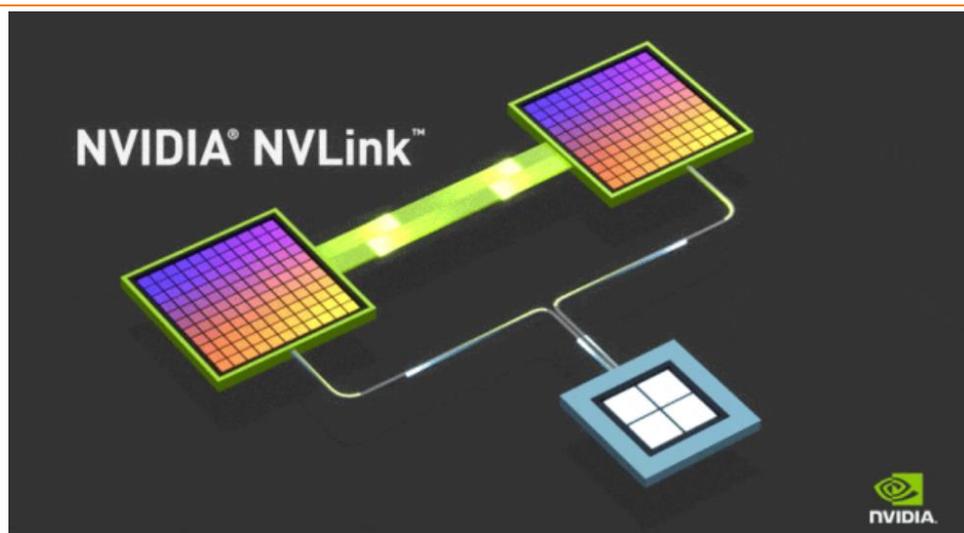
表 1：P100、V100、A100、H100、GB200 规格参数对比

	NVLink 版本 P100	NVLink 版本 V100	A10080GBSXM	H100SXM	GB200Blackwell
FP32	10.6TFLOPS	15.7TFLOPS	19.5TFLOPS	60TFLOPS	180TFLOPS
FP64	5.3TFLOPS	7.8TFLOPS	9.7TFLOPS	30TFLOPS	90TFLOPS
GPU 显存带宽	732GB/s	900GB/s	2039GB/s	3TB/s	16TB/s
最大功耗		300w	400w	700w	

资料来源：英伟达官网，天风证券研究所

多年积累卡间互联技术为集群搭建打造全面解决方案。NVLink 是世界首项高速 GPU 互连技术，与传统的 PCIe 系统解决方案相比，能为多 GPU 系统提供更快速的替代方案。NVLink 技术通过连接两块 NVIDIA 显卡，能够实现显存和性能扩展，从而满足最大视觉计算工作负载的需求。

图 10：NVIDIA@NVLink™ 演示图

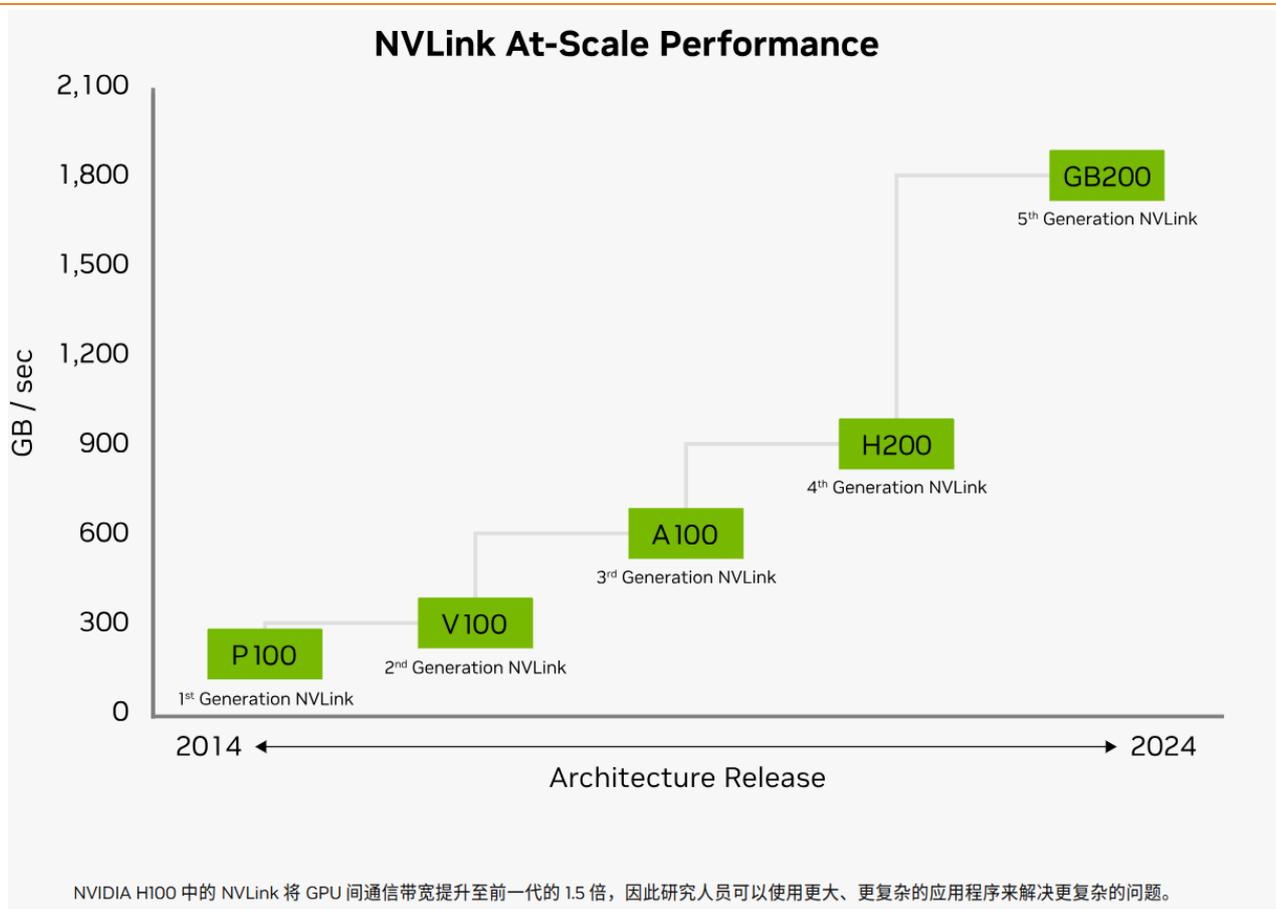


资料来源：英伟达官网，天风证券研究所

Nvlink 经历了 5 次迭代，卡间带宽增长多倍。2018 年，NVLink 首次亮相被用于连接两台超级计算机——Summit 和 Sierra 的 GPU 和 CPU，成为了高性能计算领域的焦点。2020 年，第三代 NVLink 将每个 GPU 的最大带宽翻倍提高至 600GB/s，每个 Nvidia A100 Tensor Core GPU 中都有十几条互连通道。如今一个 NVIDIA H100 Tensor Core GPU 中包含 18 条第四

代 NVLink 互联通道，这项技术已承担了一项新的战略任务——帮助打造全球领先的 CPU 和加速器。**第五代 NVLink 大幅度提高了大型多 GPU 系统的可扩展性。**单个 NVIDIA Blackwell Tensor Core GPU 支持多达 18 个 NVLink100GB/s 连接，总带宽可达 1.8TB/s，比上一代产品提高了两倍，是 PCIe5.0 带宽的 14 倍之多。

图 11: NVLink 性能的迭代历程



资料来源：英伟达官网，天风证券研究所

英伟达在服务器互联上推出 NVSwitch，为组网提供卓越的性能。通过连接多个 NVLink，NVSwitch 在单个节点内和节点之间以全 NVLink 速度提供 All-to-All GPU 通信，**凭借 NVLink 与 NVSwitch 的结合，英伟达赢得了 MLPerf1.1，这是第一个全行业范围内的 AI 基准测试。**前三代 NVSwitch 可在一个 NVLink 域内直连多达 8 个 GPU 的数量，最大带宽可达 900GB/s。

最新的 NVLinkSwitch 是首款 ASIC 芯片技术机架级交换机，能够在无阻塞计算结构中支持多达 576 个全连接 GPU，以惊人的 1800GB/s 的速度互连每对 GPU。它还支持完整的多对多通信。**NVLink 和 NVLinkSwitch 是整个 NVIDIA 数据中心解决方案的必要基础模组**，该解决方案融合了来自 NVIDIA AI Enterprise 软件套件和 NVIDIA NGC 目录中的硬件、网络、软件、库以及优化的 AI 模型和应用。

图 12: NVSwitch 的规格迭代对比

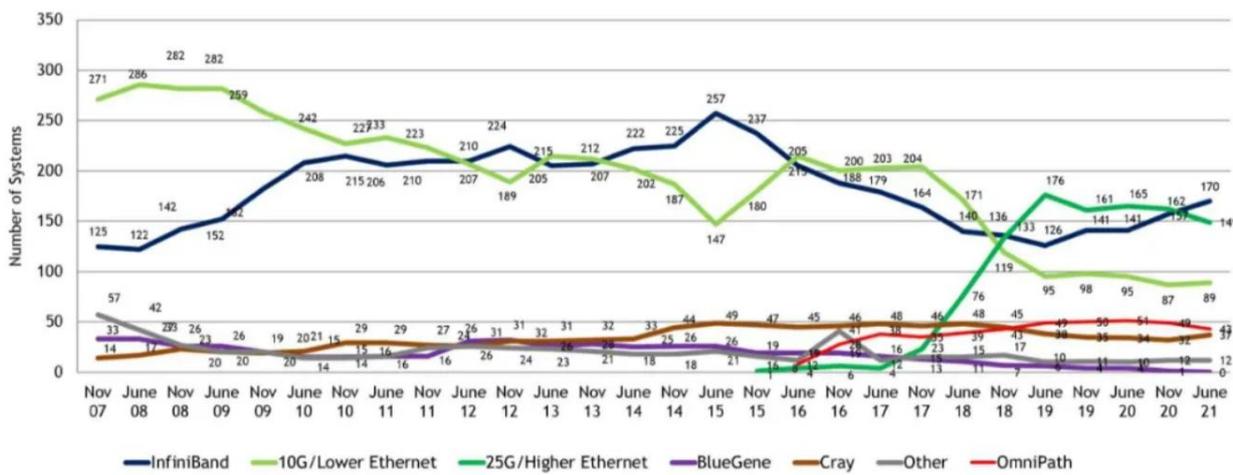
规格				
	第一代	第二代	第三代	NVLink Switch
一个 NVLink 域内直连 GPU 的数量	最多 8 个	最多 8 个	最多 8 个	最多 576 个
NVSwitch GPU 之间带宽	300GB/s	600GB/s	900GB/s	1,800GB/s
聚合总带宽	2.4TB/s	4.8TB/s	7.2TB/s	1PB/s
NVIDIA 架构支持	NVIDIA Volta™ 架构	NVIDIA Ampere 架构	NVIDIA Hopper™ 架构	NVIDIA Blackwell 架构

初步规格, 可能会有所变更

资料来源: 英伟达官网, 天风证券研究所

InfiniBand 和 Spectrum 解决方案为组网通信提供卓越优势。InfiniBand 是一种高性能通信技术协议, 起源于 21 世纪初, 旨在解决传统 PCI 总线在数据传输中的瓶颈问题。2000 年, InfiniBand 架构规范 1.0 版本发布, 引入 RDMA 协议, 提供低延迟和高带宽的 I/O 性能。Mellanox 公司是 InfiniBand 技术的关键推动者, 自 2001 年推出首款产品后, 逐渐成为市场领导者。尽管面临英特尔和微软的退出, InfiniBand 仍成功应用于高性能计算和存储设备连接。2004 年成立的 OFA 组织与 IBTA 协同推进 InfiniBand 标准。到 2009 年, InfiniBand 在 TOP500 超级计算机中广泛应用。2015 年, InfiniBand 在超级计算机内部连接技术上首次超过以太网。2019 年, Nvidia 以 69 亿美元收购 Mellanox, 进一步巩固了其在高性能计算领域的领导地位。

图 13: 2007-2021 年全球超级计算机 500 强中 IB 及其他网络供应商技术占比

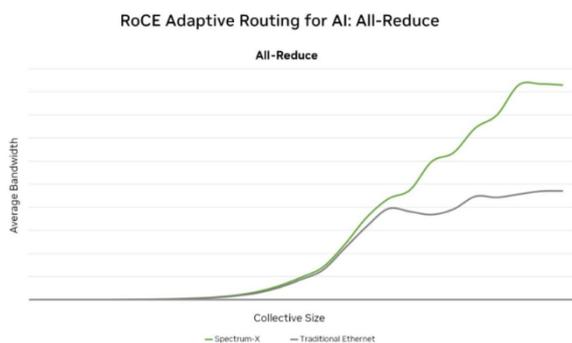


资料来源: 鲜枣课堂公众号, 天风证券研究所

2022 年 3 月 22 日, 英伟达发布新一代的以太网平台 Spectrum-4, 将为大规模数据中心基础设施提供所需的超高网络性能和强大安全性。作为全球首个 400Gbps 端到端网络平台, NVIDIA Spectrum-4 的交换吞吐量比前几代产品高出 4 倍, 达到 51.2Tbps。2023 年 5 月 29 日, 英伟达宣布推出 NVIDIA Spectrum-X 网络平台, 是业界第一款专为 AI 打造的以太网网络平台, 凭借 NVIDIA Spectrum-4 以太网交换机与 NVIDIA BlueField-3 SuperNIC 紧密耦合, 将 AI 性能和能效提升至传统以太网的 1.6 倍, 可为 AI、机器学习、自然语言处理以及各种行业应用带来卓越性能, 在英伟达的所有软、硬件端到端平台上得到了全面优化

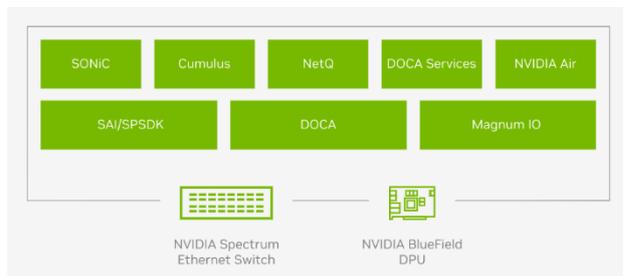
和验证。

图 14: NVIDIA Spectrum-X 较传统以太网 AI 性能提升 1.6 倍



资料来源: 英伟达官网, 天风证券研究所

图 15: 全栈优化的 NVIDIA Spectrum-X

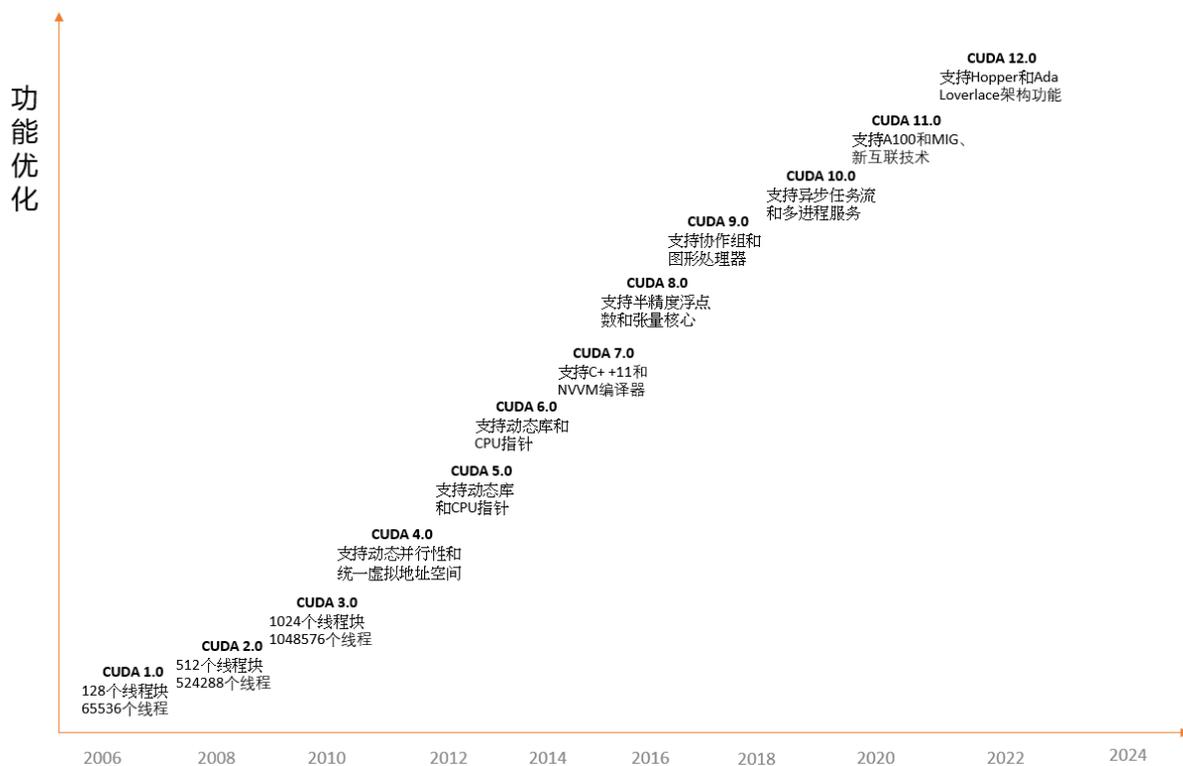


资料来源: 英伟达官网, 天风证券研究所

1.2.2. CUDA 生态, 强大的生态造就高壁垒

CUDA 生态在 AI 时代造就高壁垒。2007 年 CUDA1.0 诞生, 这一版本首次向开发者开放了 GPU 的通用计算潜能, 为开发者打开了进入 GPU 编程领域的大门; 随后在 2008 年, CUDA2.0 的推出加入了对双精度浮点运算的功能, 对于科学研究和工程模拟等高精度计算任务具有重大意义; 2010 年, CUDA3.0 进一步增强了 GPU 处理并行任务的能力, 为执行更为复杂的计算工作提供了强有力的支持; 2012 年, CUDA5.0 的问世带来了动态并行性的特性, GPU 能够在运行时自我复制内核从而显著提高了程序的适应性和执行效率。这些版本不仅促进了 CUDA 自身的技术演进, 同时也在 GPU 并行计算的历史上占据了重要的位置。

图 16: CUDA 发展历程的持续演进与优化



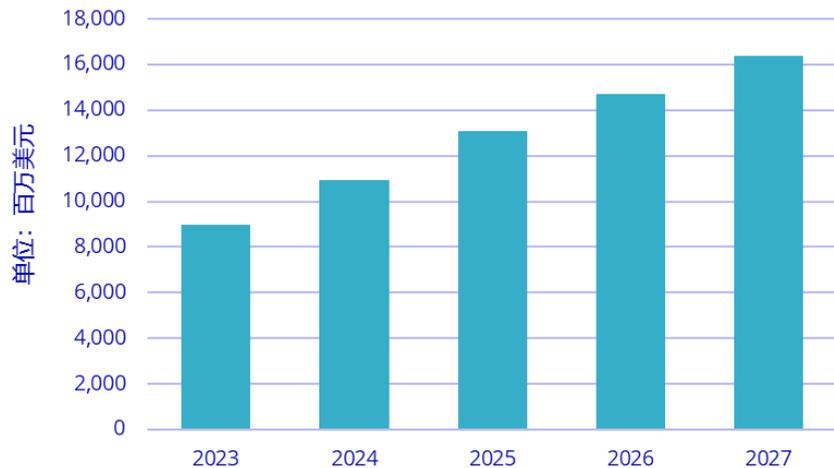
资料来源: 中嘉建信资本公众号, 天风证券研究所

2. 自主可控带来国产 AI 芯片历史机遇, 寒武纪有望脱颖而出

2.1. 国内 AI 芯片与智算中心急速增长

生成式 AI 对中国 AI 市场的发展带来了明显的拉动作用，丰富的应用场景和对技术创新迭代的热忱，让中国市场对于 AI 服务器的关注度和需求量均明显增长。根据 IDC 的数据，2023 年上半年中国加速服务器市场规模达到 31 亿美元，同比 2022 年上半年增长 54%，且 IDC 预测，到 2027 年中国加速服务器市场规模将达到 164 亿美元。随着生成式 AI 任务的不断增加，市场对于高性能和高能效的 AI 服务器需求将持续增长，未来的 AI 服务器将注重提高计算能力和处理效率，适应更复杂、更大规模的 AI 应用。

图 17：2023-2027 年中国加速计算服务器市场预测

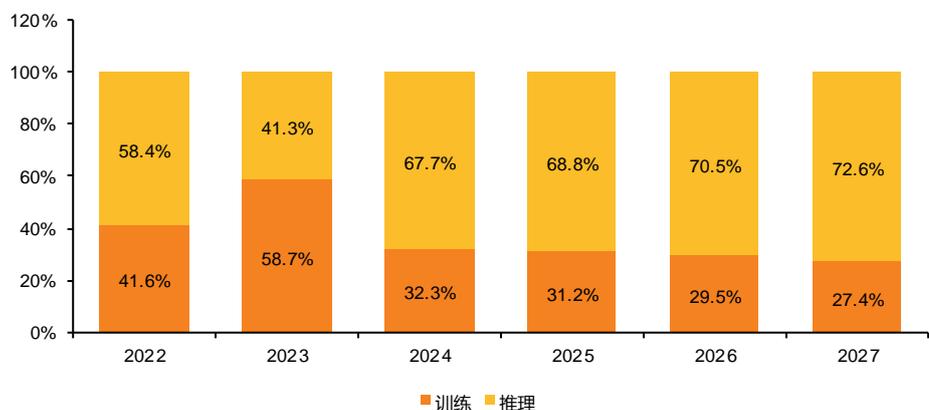


资料来源：IDC，天风证券研究所

NVIDIA 的财务数据进一步验证 2023 年中国 AI 芯片的市场规模高速增长。英伟达在 2024 财年(截止 2024 年 1 月 28 日)的数据中心收入中大约 14%来自于中国，23 财年则有 19%来自于中国，而其数据中心在 2024 财年的营业收入约为 475 亿美元，我们计算出英伟达在 2024 财年（接近 2023 年自然年）向中国销售了 66.5 亿美元的数据中心芯片，而 2023 财年中国区数据中心的收入为 28.5 亿美金，同比增长约 134%。

大模型兴起促进训练市场增长，且未来预计推理占比将攀升，到 2027 年推理占比将达 72.6%。从工作负载来看，2023 年，大模型的兴起推动了训练服务器的增长速度，IDC 数据显示，中国 2023 上半年训练工作负载的服务器占比达到 49.4%，预计全年的占比将达到 58.7%，但随着训练模型的完善与成熟，模型和应用产品逐步进入投产模式，处理推理工作负载的 AI 服务器占比将随之攀升。IDC 预计，到 2027 年，用于推理的工作负载将达到 72.6%。

图 18：2022-2027 年中国 AI 服务器工作负载预测



资料来源：IDC、浪潮信息，天风证券研究所

2024 年以来，政府与运营商也在积极部署构建算力基础设施。首先三大运营商陆续发布大额招标，中国移动 2024-2025 年新型智算中心集采项目计划采购 7994 台人工智能服务器及配套产品；2024 年中国联通发布《2024 年中国联通人工智能服务器集中采购项目资格预审公告》，计划采购人工智能服务器 2503 台。中国电信 AI 算力服务器(2023-2024 年)集中采购规模预计达到 4175 台，项目总额有望超 80 亿元。国内各地方政府的智算中心招标也逐步开启，包括佛山、山东省、襄阳市、宜昌、武汉等地方均开始了招标，其中不乏有投资额超过 10 亿的大项目规划。

表 2：2024 年运营商与部分各地智算中心的招标公告

招标方	招标内容
中国移动	启动 2024 年至 2025 年新型智算中心集采，总规模达到 8054 台，其中包含 7994 台人工智能服务器及配套产品，60 台白盒交换机
中国联通	2024 年中国联通人工智能服务器集中采购项目将采购人工智能服务器合计 2503 台，关键组网设备 RoCE 交换机合计 688 台
中国电信	中国电信 AI 算力服务器（2023-2024 年）集中采购项目，分别为训练型风冷服务器（I 系列）、训练型液冷服务器（I 系列）、训练型风冷服务器（G 系列）、训练型液冷服务器（G 系列）。预估采购规模为 4175 台。
佛山市人工智能算力公共服务平台项目	佛山市人工智能算力公共服务平台作为支撑人工智能算力的重要新型基础设施，分两期建设具备 1000P 人工智能算力的智算中心，投资总额 6.8 亿元。
山东云中心省会经济圈区域分中心数据机房—“德智未来”智算中心项目	该项目建设山东云中心省会经济圈区域分中心数据机房，算力投运后租赁运营（已实际承租）的算力值不少于 430PFlops@FP16+80PFlops@FP16(非稀疏状态下)。总投资约为 2.2 亿元。
汉阳区大模型智算中心项目	建设汉阳区大模型智算中心；建设 50PFlops@FP16 大模型人工智能算力，总投资 1.6 亿元。
襄阳市东津新区智算中心及大模型建设项目（一期）	智算中心总体规划 500P@FP16 算力，一期建设 100P@FP16 算力，总投资 4.9 亿元。
湖北宜昌点军算力枢纽中心项目	项目分两期建设，一期建设算力运营平台、运营中心、调度及交易中心，238P 智算中心，400 个机架及机房配套设施；二期建设人工智能孵化创新中心、算力调度及交易中心、融合算力适配中心、640P 智算中心、9600 个机架及机房配套设施，总投资 10 亿元。
光谷智算中心	湖北省广播电视信息网络股份有限公司发布《关于与华鑫信共同投资建设光谷智算中心的公告》，计划基建投资规模为 25 亿元人民币，部署算力规模不少于 25000P。

资料来源：C114 通信网公众号、云头条公众号、通信产业网公众号等，天风证券研究所

2.2. 美国芯片禁令政策，自主算力政策扶持

美国从 2022 年以来逐步收紧对中国的 AI 算力供应，行动措施包括将 AI 芯片厂商列入“实体清单”和禁止向中国出口高性能芯片。最早在 2022 年 8 月 26 日，美国要求英伟达将需要获得许可才能向中国出口任何基于最新架构的 A100 GPU 或 DGX/HGX 等系统，随后 2022 年 12 月 15 日，美国商务部将包括寒武纪在内的 36 家企业列入实体清单。2023 年 10 月 17 日，美国更新出口管制标准，要求先进芯片性能超过特定阈值，即需要申请出口许可，限制条款已于 10 月 23 日生效。2024 年 1 月美国商务部开始要求美国 IaaS 提供商验证外国客户的身份，并制定措施防止中国使用美国 IaaS 产品进行 AI 训练。我们认为针对中国的 AI 芯片禁令可能拖慢中国在这次生成式 AI 的步伐，但同时也给予了自主 AI 芯片供应商广阔的市场机遇。

表 3：2022 年以来美国历次针对算力的禁令

美国算力禁令政策	主要内容	时间
美国商务部下属的工业与安全局	要求美国 IaaS 提供商验证外国客户的身份，并制定措施防止中国使用	2024 年 1

(BIS) 公布《采取额外措施应对与重大恶意网络行为相关的国家紧急状态》意见稿	美国 IaaS 产品进行 AI 训练。若有证据表明 IaaS 账户被用于恶意网络活动，商务部长可以对其施加条件或禁令。	月 29 日
美国商务部发布声明称 BIS 将于 2024 年 1 月开始对美国半导体供应链展开调查	重点关注美国关键行业对中国制造的传统芯片的使用与采购情况，以评估其半导体供应链对中国芯片的依赖程度。	2024 年 1 月 18 日
美国商务部 BIS 公布半导体出口管制新规	更新对包括中国在内的武器禁运国家的先进计算半导体和半导体制造设备、及支持超算应用和最终用途的出口管制，并将中国境内更多相关实体列入实体清单。	2023 年 10 月 17 日
美国《联邦公报》官网发布了美国商务部工业和安全局对「实体清单」的增补和修订	美国商务部计划于 12 月 19 日正式将包括长江存储、寒武纪、上海微电子等 36 家中企列入实体清单。限制中国企业购买半导体和芯片制造设备。	2022 年 12 月 15 号
英伟达将需要获得许可才能向中国和俄罗斯出口任何基于最新架构的 A100 GPU 或 DGX/HGX 等系统	DGX 或任何其他采用 A100 或 H100 集成电路和 A100X 的系统的出口需要美国政府许可，未来任何峰值性能能达到等于或高于与 A100 大致相当的阈值的 NVIDIA 集成电路的出口仍受美国政府限制。	2022 年 8 月 26 日

资料来源：金杜律师事务所官网、中国证券网等，天风证券研究所

与此同时，国内顶层不断强调 AI 产业发展的重要性，从中央到地方不断发布支持算力基础设施的政策文件。《2024 年政府工作报告》提出，在人工智能方面，深化大数据、人工智能等研发应用，开展“人工智能+”行动，打造具有国际竞争力的数字产业集群；算力建设方面，适度超前建设数字基础设施，加快形成全国一体化算力体系，随后北京、上海、深圳等地也陆续发布指导性文件，对当地的算力基础设施建设进行规划。

表 4：中央到地方政府出台针对算力基础设施的支持政策

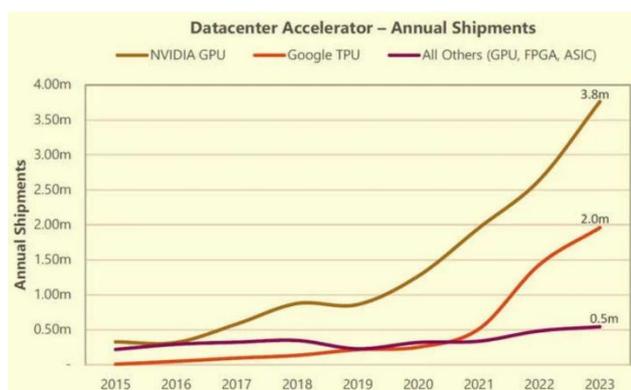
政策文件	主要内容
2024 年政府工作报告	深化大数据、人工智能等研发应用，开展“人工智能+”行动，打造具有国际竞争力的数字产业集群，适度超前建设数字基础设施，加快形成全国一体化算力体系，培育算力产业生态。
国资委中央企业人工智能专题推进会	国资央企需加快推动人工智能发展，着力打造人工智能产业集群，带头实现人工智能赋能传统行业的经济形态，加快建设一批智能算力中心。并开展 AI+ 专项行动，构建一批产业多模态优质数据集，打造从基础设施、算法工具、智能平台到解决方案的大模型赋能产业生态。
深圳市算力基础设施高质量发展行动计划	到 2025 年，深圳市将实现数据中心规模达 50 万标准机架，算力达到 14EFLOPS (FP32) 及以上，存储总量达到 90EB，并建设不少于 20 个边缘计算中心。计划强调多元算力供给体系的建设，包括超算中心、智算中心及边缘计算中心的互联互通，推动绿色低碳技术应用，提升存储和计算能力，支持全市高质量发展。
上海市新型基础设施建设贴息管理指导意见	进一步放大财政资金杠杆作用，鼓励和引导社会资金加大投入；鼓励合作银行建立总规模达到 1000 亿元以上的上海市新型基础设施建设优惠利率信贷资金，以重点支持新网络、新算力、新数据、新设施和新终端五个领域。重点对以企事业单位自筹资金为主、能形成显著规模和经济社会影响力的建设项目，按有关规定进行贴息支持，同时将对相关产业链带动作用明显的数据中心建设项目纳入贴息范围。
北京市算力基础设施建设实施方案（2024-2027 年）	方案提出加快建设智能算力基础设施，提高自主可控软硬件技术水平，推动区域算力协同发展，并强调市场主导、政府引导的建设运营机制。目标是在 2025 年实现智算资源供给集群化、设施建设自主化、能力赋能精准化、中心运营绿色化和生态发展体系化，力争到 2027 年形成国际影响力的智算产业高地。

资料来源：中国政府网、北京市人民政府官网、国务院国资委官网等，天风证券研究所

2.3. 对比海外，AI 芯片市场份额集中但难以一家独大

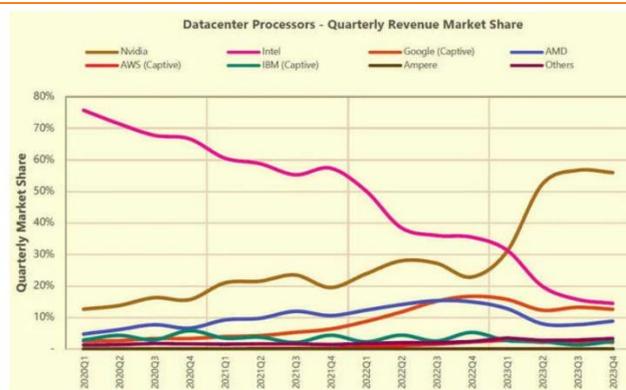
在海外市场，英伟达占据芯片市场主导地位，但 Google 与 AMD 等公司依然有望在 AI 浪潮中占据一定市场份额。根据 TechInsights 的数据，2023 年，英伟达的 GPU 出货量达到了 3.8 百万颗，位居第二名的是 Google 的 TPU，其出货量达到 200 万颗，其他厂商的出货量约为 50 万颗左右。自 2015 年推出自研的 TPU 以来，Google 在 2023 年自用的 TPU 芯片量已经突破了 200 万颗大关，就出货量来看，Google 在 AI 加速器市场的份额约 31.7%，此外，AMD 的数据中心业务部门在 2024 财年 Q1 营收 23.37 亿美元，同比增长 80%，远高于去年四季度的 38% 的同比增幅，这反应了其第四代 EPYC CPU 以及 InstinctGPU（主要 MI300 系列 AI 芯片）出货的快速增长，在 24 财年 Q1 的财报会议上，AMD 今年的 AI 芯片目标收入将提高到 40 亿美金，MI300 系列 GPU 市场表现良好，已锁定微软、Meta 和 Oracle 的采购订单。我们认为 AI 加速芯片在数据中心芯片的份额快速提升，Nvidia 依靠完善的软硬件生态成为全球 AI 芯片的领导者，但仍然有包括 Google 和 AMD 等厂商有望占据一定市场份额。

图 19: Nvidia/Google 等 AI 加速芯片市场的出货量



资料来源：TechInsights、The Register，天风证券研究所

图 20: AI 加速芯片在数据中心市场的份额快速提升



资料来源：TechInsights、The Register，天风证券研究所

3. AI 芯片元老，从 IP 到芯片

3.1. 从 IP 到自研芯片，寒武纪为 AI 芯片而生

中科寒武纪是 AI 芯片的元老级公司，成立于 2016 年 3 月。同年公司推出终端智能处理器产品，包括寒武纪 1A、寒武纪 1H 和寒武纪 1M 系列智能处理器，以 IP 形式进行销售，快速实现了技术的产业化输出，其中寒武纪智能处理器 IP 产品已集成于超过 1 亿台智能手机及其他智能终端设备中。

2018 年公司开始自研芯片，推出思元 100 云端智能芯片，2019 年推出思元 270 云端智能芯片及加速卡。2019 年 11 月，公司推出思元 220 边缘智能芯片及加速卡，实现了终端、云端、边缘端产品的完整布局。

2020 年公司上市，同年公司推出思元 290 训练芯片和玄思 1000 智能加速器，补足人工智能训练产品线，标志着寒武纪已初步建立“云边端一体、软硬件协同、训练推理融合”的新生态。2021 年，公司发布训推一体的思元 370 智能芯片及加速卡，同年设立子公司行歌科技，开展智能驾驶芯片的研发和产品化。

2022 年，公司基于思元 370 推出智能加速卡 MLU370-X8/M8 和训练整机玄思 1001 智能加速器。年末公司及部分子公司被美国商务部列入“实体清单”。2023 年，公司持续重点迭代优化新一代智能处理器微架构和指令集，针对自然语言处理大模型、视频图像生成大模型以及推荐系统大模型的训练推理等场景进行重点优化，将在编程灵活性、易用性、性能、功耗、面积等方面提升产品竞争力。例如，与智象未来达成了算力产品的合作以及视觉多模态大模型的深度适配，在语言大模型领域与百川智能等头部客户进行了大模型适配，并获得了产品兼容性认证，产品性能达到国际主流产品水平。

图 21：中科寒武纪发展历程

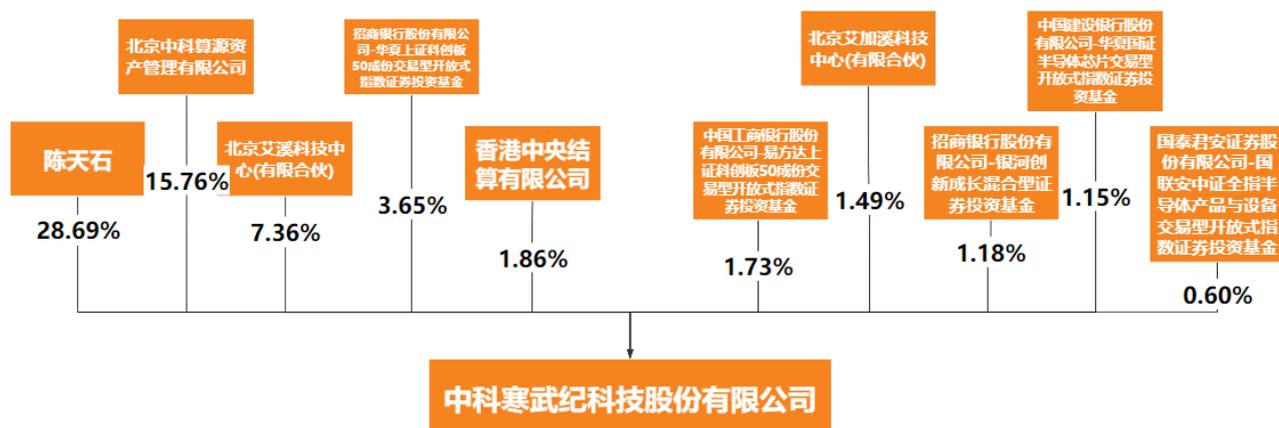


资料来源：公司年报，公司招股书，天风证券研究所

3.2. 核心团队出自中科院，董事长长期耕耘 AI 芯片，引领产业方向

公司董事长陈天石持股 28.69%，为公司最大股东。公司董事包括陈天石、刘新宇、刘少礼、王在和叶湙尹。其中刘少礼和王在兼任副总经理，分别毕业于中科院计算所和中国科学技术大学。叶湙尹兼任副总经理、财务负责人和董事会秘书。

图 22：股权结构图（截至 2024 年一季度末）



资料来源：Wind，天风证券研究所

表 5：公司董事简介

姓名	职务	简历
陈天石	董事长、 总经理	陈天石，男，出生于 1985 年，中国科学技术大学计算机软件与理论专业博士学历。中国国籍，无境外永久居留权。2010 年 7 月至 2019 年 9 月就职于中科院计算所（2018 年 4 月办理离岗创业），历任助理研究员、副研究员及硕士生导师、研究员及博士生导师。2016 年 3 月创立公司，现任公司董事长、总经理
刘新宇	董事	刘新宇，男，出生于 1972 年，哈尔滨工业大学博士学历。中国国籍，无境外永久居留权。2002 年 7 月至 2004 年 7 月，在中科院计算所博士后工作站从事研究工作；2004 年 5 月至 2006 年 2 月，历任中科院计算所技术发展处主管、分所管理办主任；2006 年 3 月至今，任苏州中科集成电路设计中心有限公司董事长、总经理；2008 年 8 月至今，历任中科院计算所技术发展处副处长、技术发展处处长、技术发展中心主任；2016 年 11 月至 2020 年 6 月，任北京中科算源资产管理有限公司总经理；现任公司董事
刘少礼	董事、 副总经理	刘少礼，男，出生于 1987 年，中科院计算所计算机系统结构博士学历。中国国籍，无境外永久居留权。2014 年至 2019 年就职于中科院计算所并任副研究员（2018 年 4 月办理离岗创

王在	董事、 副总经理	业)。2016 年作为公司创始团队成员加入公司，现任公司董事、副总经理。 王在，男，出生于 1984 年，中国科学技术大学计算机应用技术博士学历。中国国籍，无境外永久居留权。2011 年至 2015 年就职于郑州商品交易所并任核心交易系统工程师，2015 年至 2016 年就职于中原银行并任信息科技部电子银行系统主管，2016 年至 2018 年就职于中科院计算所从事科研工作。2016 年作为公司创始团队成员加入公司，现任公司董事、副总经理
叶湔尹	董事、 副总经理、 财务负责人、 董事会秘书	叶湔尹，女，出生于 1988 年，北京大学西方经济学硕士学历。中国国籍，无境外永久居留权。2012 年至 2016 年就职于中国高新投资集团公司并任投资经理、高级投资经理，2016 年至 2019 年就职于国投创业投资管理有限公司并任投资副总裁。2019 年加入公司，现任公司董事、副总经理、财务负责人及董事会秘书

资料来源：公司年报，天风证券研究所

董事长陈天石多年从事人工智能领域基础科研工作，理论功底坚实，研发经验丰富。陈天石博士毕业于中国科学技术大学计算机软件与理论专业，曾在中科院计算所担任研究员、博士生导师，在人工智能和处理器芯片等相关领域从事基础科研工作十余年，积累了坚实的理论功底和丰富的研发经验，创办并领导公司在智能芯片方向快速跻身全球初创公司前列。陈天石博士是寒武纪系列商用智能芯片的奠基人之一，负责把控公司整体的技术方向、业务进程以及战略发展方向，并牵头开展学术研究和产业化工作。

陈天石博士是 AI 芯片的开创者之一，在 14 年就提出产业进展中成果得到同行认可。陈天石等研究人员在 2013 年提出了全球首个深度学习处理器，围绕 DianNao 的研究《Dianna0: A Small-Footprint High-Throughput Accelerator for Ubiquitous Machine-Learning》获得 ASPLOS 在 2014 年的最佳论文且是亚洲首份获得计算机硬件顶会的最佳论文。接着在 2014 年、2015 年、2016 年，寒武纪又通过论文宣布了对全球首个多核深度学习处理器、摄像头上的智能识别 IP、全球首个神经网络指令集的研究。Google 发布的 TPU 论文，全文共引用了寒武纪团队成员前期发表的 6 篇论文（表 2 第 2~第 7 篇论文），并有专门的段落回顾寒武纪团队这一系列工作，并且在提到 DaDianNao/PuDianNao/ShiDianNao 时还专门用英文注释这几个名字的含义（Big computer, general computer, vision computer），对其前期工作显示了相当的尊重；在世界范围内，很少有其他工作能受到 Google 同行如此程度的关注，足以看出陈天石博士在 AI 芯片中的前瞻性和贡献度。

图 23：Google TPU 论文引用寒武纪相关工作的专门段落

The more recent DianNao family of NN architectures minimizes memory accesses both on the chip and to external DRAM by having efficient architectural support for the memory access patterns that appear in NN applications [Keu16]

13

[Che16a]. All use 16-bit integer operations and all designs dove down to layout, but no chips were fabricated. The original DianNao uses an array of 64 16-bit integer multiply-accumulate units with 44 KB of on-chip memory and is estimated to be 3 mm² (65 nm), to run at 1 GHz, and to consume 0.5W [Che14a]. Most of this energy went to DRAM accesses for weights, so one successor DaDianNao (“big computer”) includes eDRAM to keep 36 MiB of weights on chip [Che14b]. The goal was to have enough memory in a multichip system to avoid external DRAM accesses. The follow-on PuDianNao (“general computer”) is aimed at more traditional machine learning algorithms beyond DNNs, such as support vector machines [Liu15]. Another offshoot is ShiDianNao (“vision computer”) aimed at CNNs, which avoids DRAM accesses by connecting the accelerator directly to the sensor [Du15].

资料来源：Norman P. Jouppi, Cliff Young 等《In-Datacenter Performance Analysis of a Tensor Processing Unit》，天风证券研究所

表 6：陈天石博士及相关团队早期相关工作

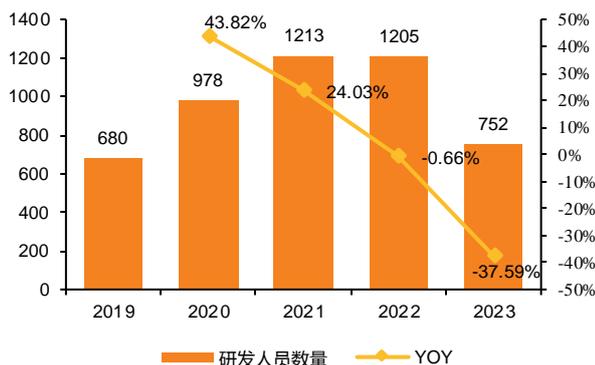
发表时间	论文名称
------	------

2012	BenchNN: On the Broad Potential Application Scope of Hardware Neural Network Accelerators
2014	Dianna: A small-foot print high-throughput accelerator for ubiquitous machine-learning. Proceedings of ASPLOS
2014	Dadianna: A machine-learning supercomputer.
2015	ShiDianNao: shifting vision processing closer to the sensor
2015	Pudianna: A polyvalent machine learning accelerator
2016	DianNaoFamily: Energy-Efficient Hardware Accelerators for Machine Learning
2016	Cambricon: An instruction set architecture for neural networks

资料来源：21 世纪经济报道公众号，《In-Datacenter Performance Analysis of a Tensor Processing Unit》，天风证券研究所

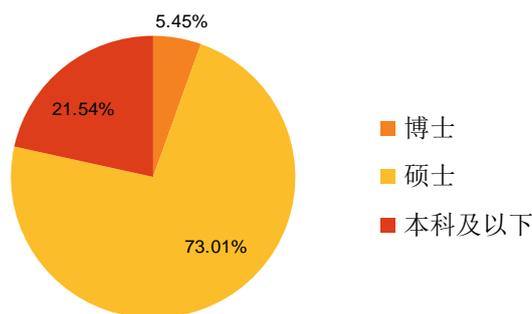
公司技术团队配备齐全，如今更加聚焦于云端 AI 芯片领域。公司在技术研发、供应链、产品销售等方面均建立了成熟团队，核心骨干均有多年从业经验。公司核心研发人员大多拥有计算机、微电子等相关专业的学历背景，毕业于国内外著名的高校和研究所；多名骨干成员拥有知名半导体公司多年的工作经历。数量上，2019-2021 年公司研发人员数量持续提升，2023 年有所下降主要系业务战略规划调整，但我们认为公司目前更加聚焦在云端 AI 芯片领域。从占比来看，研发团队中本科及以上学历占比 21.54%，73.01%的研发人员为硕士学历，5.45%为博士学历。2023 年公司研发费用由 2022 年的 15.23 亿元缩减至 11.18 亿元，但人均研发费用由 2022 年的 126.39 万元提升至 148.67 万元，保持了 2020-2022 年的增长趋势。

图 24：FY2019-FY2023 中科寒武纪研发人员数量及同比



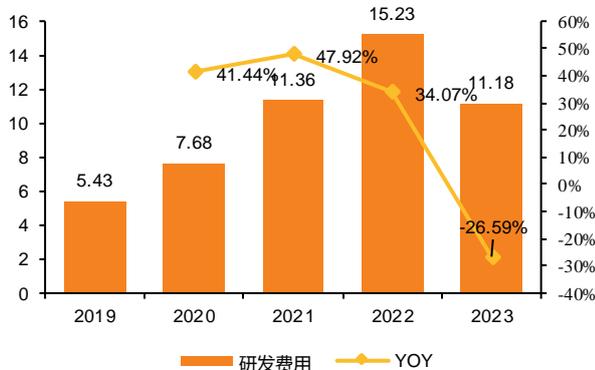
资料来源：公司年报，天风证券研究所

图 25：2023 年中科寒武纪研发人员学历构成



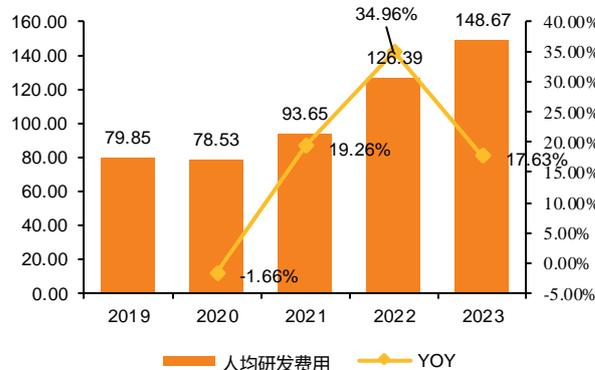
资料来源：公司年报，天风证券研究所

图 26：FY2020-FY2023 中科寒武纪研发费用（亿元）及同比



资料来源：Wind，天风证券研究所

图 27：FY2019-FY2023 中科寒武纪人均研发费用（万元）及同比



资料来源：Wind，天风证券研究所

股权激励目标进一步彰显了公司的业绩增长信心。2023 年 11 月 18 日，公司发布了《中科寒武纪科技股份有限公司 2023 年限制性股票激励计划（草案）》，股权激励针对收入提出了要求，目标值上，在 2024 年营业收入不低于 11 亿元，第二个归属期上 24-25 年累计收入不低于 26 亿元，第三个归属期 24-26 年累计收入不低于 46 亿元。结合公司 2023 年

收入仅 7.09 亿元，此次股权激励彰显了公司未来充足的业绩增长信心。

表 7：2023 年寒武纪推出的股权激励计划

归属期	目标值	触发值
	公司层面归属系数 100%	公司层面归属系数 80%
第一个归属期	2024 年营业收入值不低于 11 亿元	2024 年营业收入值不低于 8.8 亿元
第二个归属期	2024-2025 年累计营业收入值不低于 26 亿元	2024-2025 年累计营业收入值不低于 20.8 亿元
第三个归属期	2024-2026 年累计营业收入值不低于 46 亿元	2024-2026 年累计营业收入值不低于 36.8 亿元

资料来源：公司公告，天风证券研究所

4. 寒武纪技术能力优秀，具备 AI 芯片完整软硬件技术栈

4.1. 通用智能芯片系统复杂，寒武纪具备软硬件全栈技术

通用智能芯片技术复杂、覆盖面广，而寒武纪是业内少数系统掌握相关技术的企业之一。通用型智能芯片及其基础系统软件的研发需要全面掌握核心芯片与系统软件的大量关键技术，技术难度大、涉及方向广，是一个极端复杂的系统工程，其中处理器微架构与指令集两大类技术属于最底层的核心技术。寒武纪在智能芯片领域掌握了智能处理器微架构、智能处理器指令集、SoC 芯片设计、处理器芯片功能验证、先进工艺物理设计、芯片封装设计与量产测试、硬件系统设计等七大类核心技术；在基础系统软件技术领域掌握了编程框架适配与优化、智能芯片编程语言、智能芯片编译器、智能芯片高性能数学库、智能芯片虚拟化软件、智能芯片核心驱动、云边端一体化开发环境等七大类核心技术。

表 8：中科寒武纪核心技术框架结构

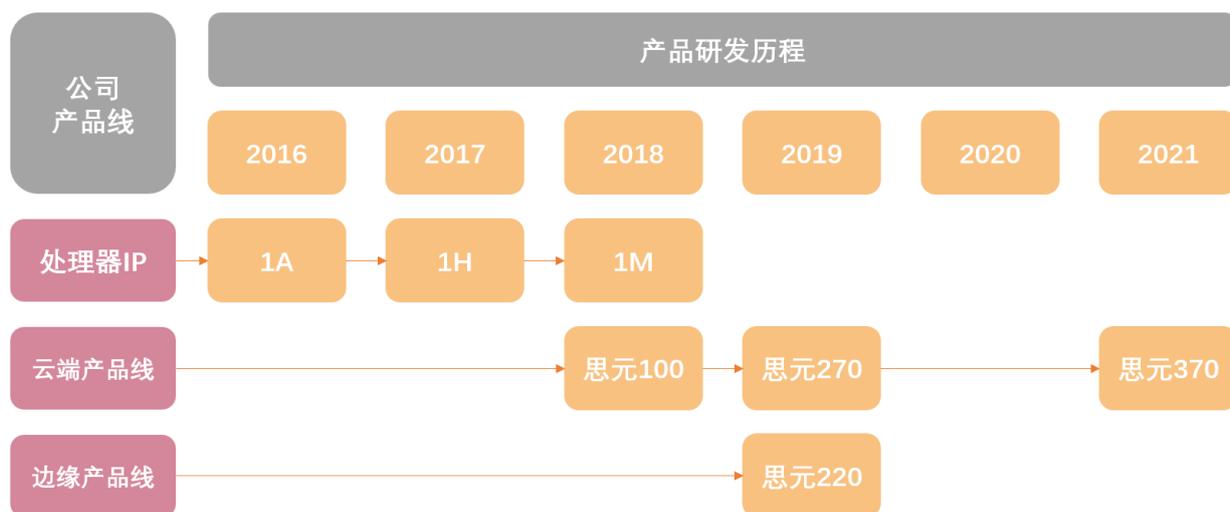
分类	名称	含义
基础系统软件技术	编程框架适配与优化	支持 Pytorch 和 TensorFlow 等框架，降低代码迁移成本，提升应用开发速度
	智能芯片编程语言	公司研发的 BANG 语言，通用性好、可扩展，用于芯片拓展应用、迭代算法
	智能芯片编译器	将 BANG 语言编译为机器码，自动优化挖掘芯片性能潜力
	智能芯片高性能数学库	开发者能够以调用数学库的形式实现常用运算，并获得较好性能
	智能芯片虚拟化软件	将物理芯片虚拟化，供多个虚拟机或容器同时使用，提升数据中心智能芯片利用率，便于数据中心管理 IT 资产
	智能芯片核心驱动	保证智能芯片在操作系统中高效运行的底层基础组件
	云边端一体化开发环境	用于应用开发、功能调试和性能调优的软件工具链，可大幅提升人工智能应用在不同硬件平台的开发效率和部署速度，使云边端异构硬件资源统一管理、调度和协同成为可能
智能芯片技术	智能处理器微架构	公司在云、边、端的智能芯片和智能处理器核均基于自研处理器架构研制
	智能处理器指令集	公司软硬件均构建与自研的 MLU 指令集基础之上
	SoC 芯片设计	公司掌握 SoC 设计的一系列关键技术，支撑了云端大型 SoC 芯片和边缘端中性 SoC 芯片的研发
	处理器芯片功能验证	公司拥有成熟先进的处理器和 SoC 芯片功能验证平台，确保了智能处理器和 SoC 芯片逻辑设计按时高质量交付，有效保障多款芯片产品的一次性流片成功
	先进工艺物理设计	公司掌握 7nm 等先进工艺下开展复杂芯片物理设计的一系列关键技术，成功应用于多款芯片的物理设计之中
	芯片封装设计与量产测试	用于公司云端、边缘端不同品类芯片产品的封装设计与量产测试过程
	硬件系统设计	用于解决高速信号完整性、大功率供电下的电源完整性、大型芯片散热、机箱模块化等关键问题，支撑公司加速卡、整机和集群等多样化的产品形态

资料来源：公司年报，天风证券研究所

4.2. 产品体系从 IP 到端侧再到云端，软硬件深度耦合

从 IP 授权到端侧芯片再到云端芯片，伴随 AI 浪潮快速变迁。2016 年，公司推出的首款智能处理器 IP 寒武纪 1A，随后于 2017 年、2018 年分别推出了升级版寒武纪 1H 和寒武纪 1M，该等处理器 IP 已累计应用于过亿台终端设备中。2018 年和 2019 年，公司相继发布云端推理芯片思元 100 和思元 270，目前已成功部署在包括联想、浪潮在内多个厂商的服务器中。2019 年 11 月，随着边缘 AI 系列产品思元 220 芯片及加速卡产品的发布，标志着公司在云、边、端实现了全方位覆盖，形成了完整的智能芯片产品群。2021 年，公司发布基于第四代智能处理器微架构 MLUarch03 的训推一体思元 370 智能芯片及加速卡。思元 370 是寒武纪首款采用 Chiplet 技术的人工智能芯片，其算力是寒武纪第二代云端推理产品思元 270 算力的 2 倍。2022 年，公司基于思元 370 云端智能芯片，推出了新款智能加速卡 MLU370-X8/M8、训练整机玄思 1001 智能加速器(MLU-X1001)。2023 年，公司的新一代智能处理器微架构和指令集正在研发中。

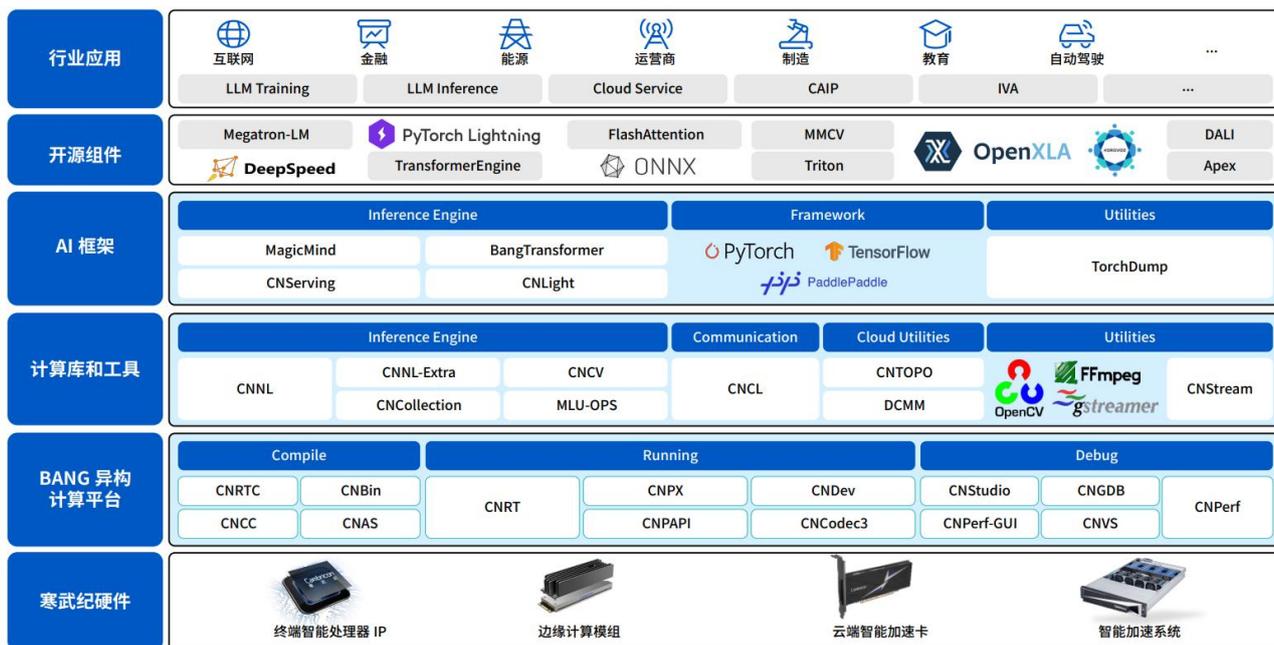
图 28：中科寒武纪主要产品线



资料来源：公司年报，天风证券研究所

软件栈方面，寒武纪拥有全面的基础软件系统平台，专门针对其云、边、端的智能处理器产品打造的软件开发平台，采用端云一体、训推一体架构，可同时支持寒武纪云、边、端的全系列产品。寒武纪基础软件系统平台包括 BANG 异构计算平台、计算库和工具、AI 框架和开源组件。其中，BANG 异构计算平台由 3 部分构成，包括编译、运行和调试。计算库和工具包括 4 个部分，分别是推理引擎、通信、云工具集和实用工具。在 AI 框架方面，寒武纪提供了 MagicMind 等推理引擎，支持 Pytorch、TensorFlow 框架，同时提供 TorchDump 作为网络精度调试工具。此外，寒武纪可在一定程度上兼容 DALI、Triton 等开源组件。寒武纪基础软件栈可为包括互联网、金融、能源、运营商在内的行业服务提供有效支撑。

图 29：寒武纪基础软件系统平台



寒武纪基础软件系统平台是寒武纪专门针对其云、边、端的智能处理器产品打造的软件开发平台，采用端云一体、训推一体架构，可同时支持寒武纪云、边、端的全系列产品。

资料来源：寒武纪 2023 年社会责任报告，天风证券研究所

4.2.1. 硬件覆盖 IP 和云边端一体化产品，单芯片最大算力高达 256TOPS

(1) 寒武纪初代处理器 IP 陆续推出 1A、1H 和 1M，累计出货量过亿。2016 年底，公司推出了全球首款商用终端智能处理器 IP 产品寒武纪 1A。此后，公司推出 1H、1M 处理器 IP。其中，Cambricon-1M 系列作为寒武纪第三代架构，具备了更优性能、更低功耗和更强的完备性，混合支持 fp32/fp16/int32/int16/int8/int4 位宽，增加了压缩解压缩模块。在上代产品的基础上，可支持个性化人工智能应用，也可使用于多路视频实时处理和自动驾驶等领域。该系列高、中、低产品分别是 Cambricon-1M-4K、Cambricon-1M-2K 和 Cambricon-1M-1K，在 1GHz 主频下，进行 8 位定点人工智能运算的峰值速度分别为 8/4/2Tops，进行 16 位定点人工智能运算的峰值速度分别为 4/2/1Tops，进行 32 位定点人工智能运算的峰值速度分别为 1/0.5/0.25Tops。

表 9：寒武纪终端智能处理器 IP

架构	版本	推出时间	峰值速度			
			INT8	INT16	INT32	FP16
Cambricon-1A		2016				0.5TOPS
Cambricon-1H	1H16	2017	1TOPS			0.5TOPS
Cambricon-1H	1H8	2017	1TOPS			
Cambricon-1H	1H8mini	2017	0.5TOPS			
Cambricon-1M	1M-4K	2018		4TOPS	1TOPS	
Cambricon-1M	1M-2K	2018	4TOPS	2TOPS	0.5TOPS	
Cambricon-1M	1M-1K	2018	2TOPS	1TOPS	0.25TOPS	

资料来源：公司官网，天风证券研究所

(2) 寒武纪边缘产品 MLU220 发布于 2019，2023 年累计销量过百万。边缘计算是近年来兴起的一种新型计算范式，在终端和云端之间的设备上配备适度的计算能力，一方面可有效弥补终端设备计算能力不足的劣势，另一方面可缓解云计算场景下数据隐私、带宽与延时等潜在问题。边缘计算范式和人工智能技术的结合将推动智能制造、智能零售、智能

教育、智能家居、智能电网等众多领域的高速发展。公司的边缘智能芯片为思元 220，MLU220 是一款专门用于边缘计算应用场景的 AI 加速产品，产品集成 4 核 ARM CORTEXA55，LPDDR4x 内存及丰富的外围接口。用户既可以使用 MLU220 作为 AI 加速协处理器，也可以使用其实现 SoC 方案。该产品和相应的 M.2 加速卡于 2019 年 11 月发布。2020 年底 MLU220 实现规模化出货，发布至 2023 年，累计销量过百万。

表 10：寒武纪边缘产品

芯片	产品	推出时间	AI 性能	CPU	内存
思元 220	MLU220-SOM 智能模组	2019	16TOPS (INT8)	ARMA55*4 1.5GHz	8GB LPDDR4 × , 64bits, 3733MHz
思元 220	MLU220-M.2 边缘端人工智能加速卡	2019	8TOPS (INT8)		LPDDR4x64bits

资料来源：公司年报、公司官网，天风证券研究所

(3) 云端产品线覆盖芯片、加速卡和整机，第三代加速卡已与主流互联网厂商开展深入适配。云端产品线目前包括云端智能芯片、加速卡及训练整机。其中，云端智能芯片及加速卡是云服务器、数据中心等进行人工智能处理的核心器件，其主要作用是云云计算和数据中心场景下的人工智能应用程序提供高计算密度、高能效的硬件计算资源，支撑该类场景下复杂度和数据吞吐量高速增长的人工智能处理任务。公司的训练整机是由公司自研云端智能芯片及加速卡提供核心计算能力，且整机亦由公司自研的训练服务器产品。**思元 370 是寒武纪第三代云端产品**，最大算力高达 256TOPS(INT8)，是寒武纪第二代云端推理产品思元 270 算力的 2 倍。同时，思元 370 芯片支持 LPDDR5 内存，内存带宽是思元 270 的 3 倍，可在板卡有限的功耗范围内给人工智能芯片分配更多的能源，输出更高的算力。思元 370 智能芯片采用了先进的 Chiplet 芯粒技术，支持芯粒间的灵活组合，仅用单次流片就达成了多款智能加速卡产品的商用。公司已推出 3 款加速卡：MLU370-S4、MLU370-X4、MLU370-X8，已与国内主流互联网厂商开展深入的应用适配。

表 11：寒武纪云端产品性能参数

芯片	智能加速卡	推出时间	制程	峰值性能						
				INT4	INT8	INT16	CINT32	FP16	BF16	FP32
思元 100		2018	16nm		32TOPS			16TOPS		
思元 270	MLU270-F4	2019	16nm	256TOPS	128TOPS	64TOPS				
思元 270	MLU270-S4	2019	16nm	256TOPS	128TOPS	64TOPS				
思元 290	MLU290-M5	2020	7nm		512TOPS	256TOPS	64TOPS			
思元 370	MLU370-S4/S8	2021	7nm		192TOPS	96TOPS		72TFLOPS	72TFLOPS	18TFLOPS
思元 370	MLU370-X4	2021	7nm		256TOPS	128TOPS		96TFLOPS	96TFLOPS	24TFLOPS
思元 370	MLU370-X8	2021	7nm		256TOPS	128TOPS		96TFLOPS	96TFLOPS	24TFLOPS

资料来源：公司官网，天风证券研究所

(4) 结合自研软硬件形成智能计算集群系统业务，服务 AI 数据中心。公司智能计算集群系统业务是将公司自研的加速卡或训练整机产品与合作伙伴提供的服务器设备、网络设备与存储设备结合，并配备公司的集群管理软件组成的数据中心集群，其核心算力来源是公司自研的云端智能芯片。智能计算集群主要聚焦人工智能技术在数据中心的应用，为人工智能应用部署技术能力相对较弱的客户提供软硬件整体解决方案，以科学地配置和管理集群的软硬件、提升运行效率。

表 12：寒武纪训练整机产品

芯片	型号	自适应精度算力		
		INT8	INT16	CINT32
思元 290	玄思 1000	2.05PetaOPS	1PetaOPS	256TOPS

资料来源：公司官网，天风证券研究所

4.2.2. Cambricon Neuware：从底层自建适配硬件的软件生态

Cambricon Neuware 是寒武纪专门针对其云、边、端得智能处理器产品打造的软件开发平台，采用云边端一体、训推一体架构，可同时支持寒武纪云、边、端的全系列产品，其包括行业应用层、AI 框架层、寒武纪开发工具层、寒武纪硬件层。

Cambricon Neuware 全面支持各类主流编程框架（如 TensorFlow, Caffe,Caffe2,MXNet 和 ONNX 等）。NeuWare 还提供了包括应用开发、功能调试、性能调优等在内得一些列工具。其中应用开发工具包括机器学习库、运行时库、编译器、模型重训练工具和特定领域(如视频分析领域) SDK 等；功能调试工具可以满足编程框架、函数库等不同层次得调试需求；性能调优工具包括性能剖析工具和系统监控工具等。

图 30：寒武纪 Cambricon Neuware 架构层级关系



资料来源：寒武纪官网，天风证券研究所

寒武纪训练软件平台支持基于主流开源框架原生分布式通信方式，同时也支持 Horovod 开源分布式通信框架，可实现单卡到集群得分布式训练任务。平台支持多种网络拓扑组织方式，并完整支持数据并行、模型并行和混合并行的训练方法。

训练软件平台支持丰富的图形图像、语音、推荐以及 NLP 训练任务。通过底层算子库 CNCL 和通信库 CNCL，在实际训练业务中达到业界领先的硬件计算效率和通信效率。同时提供模型快速迁移方法，帮助用户快速完成现有业务模型的迁移。

MagicMind 是寒武纪全新打造的推理加速引擎，也是业界首个基于 MLIR 图编译技术达到商业化部署能力的推理引擎。借助 MagicMind，用户仅需投入极少的开发成本，即可将推理业务部署到寒武纪全系列产品上，并获得颇具竞争力的性能。

图 31：寒武纪训练软件平台

图 32：寒武纪全新推理加速引擎 MagicMind 结构



资料来源：寒武纪官网，天风证券研究所

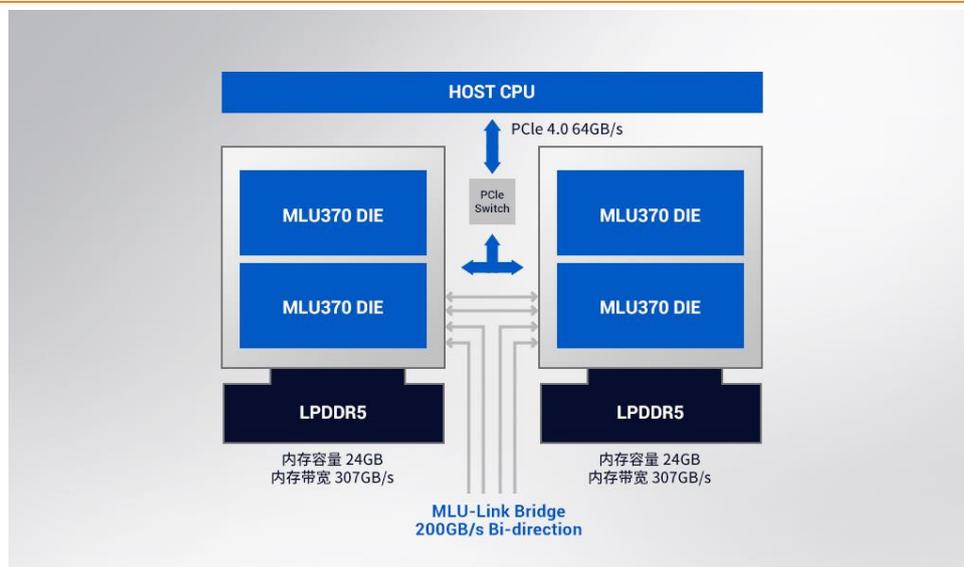


资料来源：寒武纪官网，天风证券研究所

4.2.3. 多芯互联组网提升并行效率，实现多卡和分布式任务高效执行

多芯互联提供跨芯片直接通讯能力，提高并行效率，MLU370-X8 带宽达到 PCIe4.0 的 3.1 倍。MLU-Link 多芯互联技术，搭载于寒武纪思元 370 芯片，为每颗芯片提供 200GB/s 的额外跨芯片直接通讯能力。在思元 370 应对多卡多芯并行任务时，提供更高效率的并行效率。例如，MLU370-X8 智能加速卡支持 MLU-Link™多芯互联技术，提供卡内及卡间互联功能。寒武纪为多卡系统专门设计了 MLU-Link 桥接卡，可实现 4 张加速卡为一组的 8 颗思元 370 芯片全互联，每张加速卡可获得 200GB/s 的通讯吞吐性能，带宽为 PCIe 4.0 的 3.1 倍，可高效执行多芯多卡训练和分布式推理任务。

图 33：MLU370-X8 中整合了双芯片四芯粒思元 370



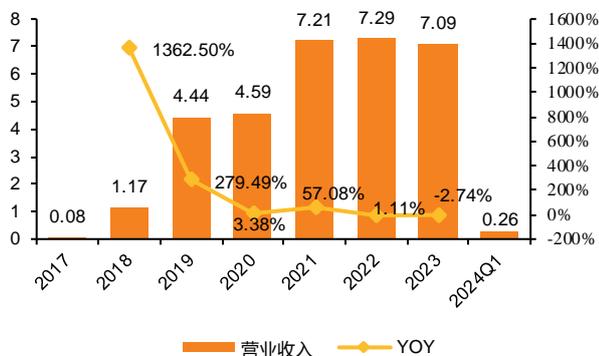
资料来源：公司官网，天风证券研究所

4.3. 智能计算集群贡献主要收入，研发上更加聚焦，费用率逐步下降

2023 年实现营业收入 7.09 亿元，其中智能计算集群系统业务收入为 6.05 亿元。2018 年公司收入增长至 1.17 亿元，2019、2020 年，公司收入达到 4.4、4.59 亿元，2021-2023 年，公司收入维持在 7 亿元以上，2023 年公司凭借人工智能芯片产品的核心优势，拓展算力基础设施项目，深耕行业客户，实现营业收入 7.09 亿元，同比减少 2.74%。分业务来看，2021-2023 年，公司收入主要来自智能计算系统、云端智能芯片及加速卡。2023 年，基于智能计算集群系统业务的竞争优势和前期落地项目的良好经验，公司积极参与台州、沈阳两地的算力基础设施建设项目，以公司的训练整机产品作为核心计算设备，集成配套的软硬件，

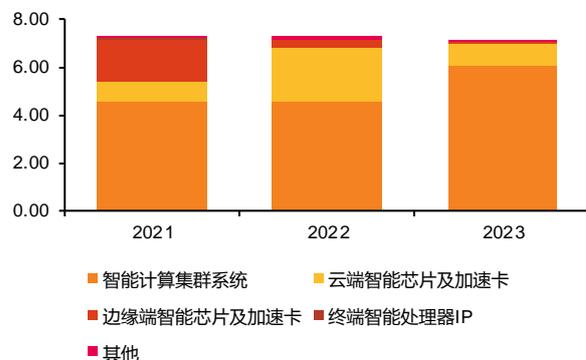
最终形成智能计算集群系统交付给客户，共实现收入 6.05 亿元；依托于智能芯片产品的技术领先优势，公司智能芯片及加速卡持续在互联网、运营商、金融、能源等多个重点行业持续落地，并在业界前沿的大模型领域以及搜索、广告推荐等领域取得了长足的进步，2023 年，公司的智能芯片及加速卡实现收入 1.01 亿元。

图 34: FY2017-FY2024Q1 中科寒武纪收入 (亿元) 及同比



资料来源: Wind, 天风证券研究所

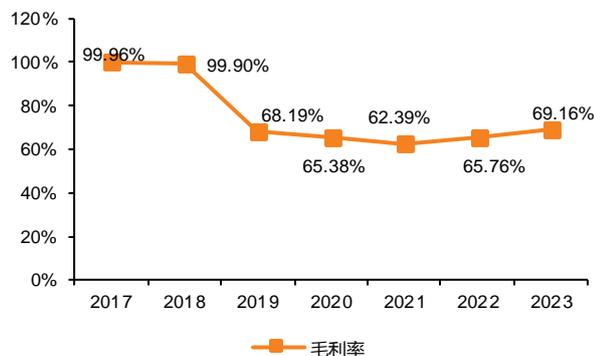
图 35: FY2021-FY2023 中科寒武纪分业务收入 (亿元)



资料来源: Wind, 天风证券研究所

2023 年亏损收窄 32.47%，战略重点集中于技术创新。2023 年公司毛利率为 69.16%，较上年同期提升 3.4pct；实现归属于上市公司股东的净利润-8.48 亿元，较上年同期亏损收窄 4.08 亿元，亏损收窄 32.47%。公司尚未实现盈利，主要系行业特性及公司战略规划两方面原因导致：芯片行业是一项高投入、长周期的行业，持续的研发投入对于企业的发展至关重要。只有通过不断的研发投入，才能推动技术创新和产品升级，提升核心竞争力，从而在激烈的市场竞争中取得优势地位。尽管受到“实体清单”的影响，寒武纪始终把技术创新作为公司的战略重点，持续进行研发投入，以打造优质的产品而易用的软件生态平台，确保在中国人工智能芯片领域的领先地位。

图 36: FY2017-FY2023 中科寒武纪毛利率



资料来源: Wind, 天风证券研究所

图 37: FY2017-FY2023 中科寒武纪净利润 (亿元)

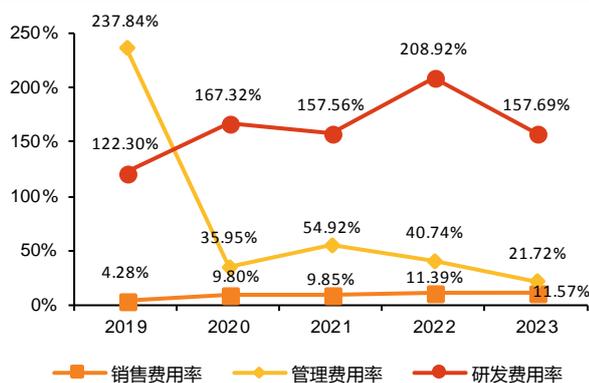


资料来源: Wind, 天风证券研究所

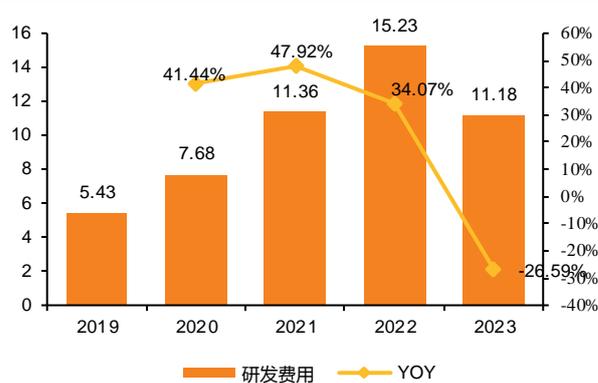
2023 年公司研发和管理费用率均同比下降，进入“实体清单”后暂停部分研发项目缩减费用率。期间费用率方面，公司销售费用率基本维持不变，管理费用率从 2022 年的 40.74% 减少至 2023 年 21.72%。公司按归属期分摊的股份支付费用较上年同期减少，从而使 2023 年管理费用较上年同期有所减少。就研发费用而言，2019-2022 年，公司研发投入持续提升，由 5.43 亿元提升至 15.23 亿元。2023 年，公司研发费用为 11.18 亿元，占营业收入的比例为 158%，较上年同期减少近 51 个百分点。2022 年 12 月 15 日，美国商务部工业和安全局 (BIS) 以国家安全和外交利益为由，将公司及部分子公司列入“实体清单”，受此影响，公司调整战略，陆续暂停部分预期毛利率较低的研发项目。

图 38: FY2019-FY2023 中科寒武纪期间费用率

图 39: FY2019-FY2023 中科寒武纪研发费用 (亿元) 及同比



资料来源: Wind, 天风证券研究所

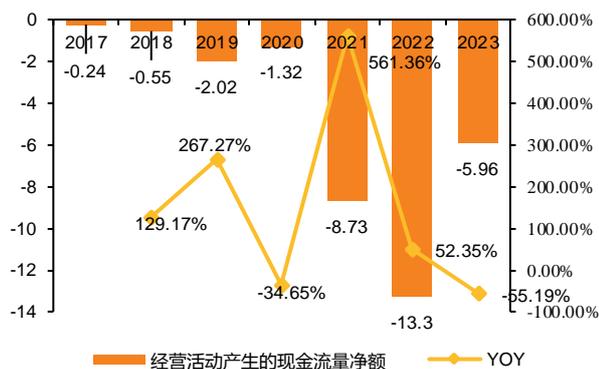


资料来源: Wind, 天风证券研究所

公司经营净现金流有显著正向增量，2023 年现金流情况良好。2023 年经营活动产生的现金流量净额变动主要系本期销售回款较上年同期增加 3.76 亿元，以及采购支出减少 2.92 亿元所致。公司目前现金流状况良好，截至 2023 年底，货币资金为 39.54 亿元，公司的现金流状况可以在未来一段时间内为公司研发投入及日常运营提供有效支撑。公司将持续拓展市场份额、加速场景落地、聚焦技术创新、持续构建生态和品牌，提升公司的核心竞争力。

图 40: FY2017-FY2023 中科寒武纪经营净现金流 (亿元) 及同比

图 41: FY2017-FY2023 中科寒武纪货币资金 (亿元) 及同比



资料来源: Wind, 天风证券研究所

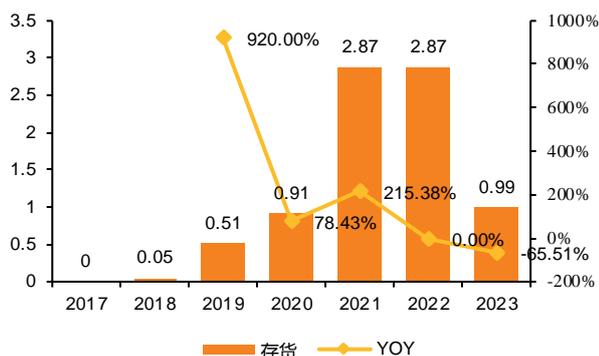


资料来源: Wind, 天风证券研究所

2023 年公司存货为 0.99 亿元，预付款项为 1.48 亿元。2023 年，公司存货为 0.99 亿元，同比减少 65.51%，主要系存货账面余额较上年期末减少，累计计提的存货跌价准备增加所致；公司预付款项为 1.48 亿元，同比增加 1750%。主要系上年预付款项余额较少。

图 42: FY2017-FY2023 中科寒武纪存货 (亿元) 及同比增长

图 43: FY2017-FY2023 中科寒武纪预付款项 (亿元) 及同比增长



资料来源：Wind，天风证券研究所



资料来源：Wind，天风证券研究所

4.4. 寒武纪在软/硬件生态综合能力强，研发针对大模型实现多点突破

寒武纪自主研发多项智能芯片技术，覆盖云边端和推理训练产品。公司能为云端、边缘端、终端提供全品类系列化智能芯片和处理器产品，是同时具备人工智能推理和训练智能芯片产品的企业，也是国内少数具有先进集成电路工艺（如 7nm）下复杂芯片设计经验的企业之一。在智能芯片技术领域，寒武纪掌握智能处理器微架构、智能处理器指令集、SoC 芯片设计、处理器芯片功能验证、先进工艺物理设计、芯片封装设计与量产测试和硬件系统设计技术均为自主研发，且成熟稳定。累计已经取得专利 829 项，其中境外专利 259 项，2023 年内增加 207 项。

寒武纪自研系统软件打破平台壁垒，兼具灵活性和可扩展性。公司能为自研云端、边缘端、终端全系列智能芯片与处理器产品提供统一的平台级基础系统软件和编程接口，公司自研的基础系统软件平台彻底打破了云边端之间的开发壁垒，兼具灵活性和可扩展性的优势，仅需简单移植即可让同一人工智能应用程序便捷高效地运行在公司云边端系列化芯片/处理器产品之上。在系统软件技术领域，公司掌握编程框架适配与优化、智能芯片编程语言、智能芯片编译器、智能芯片数学库、智能芯片虚拟化软件、智能芯片核心驱动、云边端一体化开发环境七大类核心技术，技术均来自自主研发，且成熟稳定。累计已经取得专利 281 项，其中境外专利 81 项，2023 年内增加 83 项。

表 13：寒武纪智能芯片技术及其先进性

序号	技术大类名称	在业务、产品中的贡献和应用	专利或其他技术保护措施	成熟程度	技术来源
1	智能处理器微架构	新一代智能处理器微架构正在研发中。公司是国内外在该技术方向积累最深厚的企业之一。公司在云端、边缘端、终端三条产品线的所有智能芯片和智能处理器核均基于自研处理器架构研制。	已取得专利 500 项（其中境外专利 175 项），2023 年内增加 137 项。	成熟稳定	自主研发
2	智能处理器指令集	指令集是处理器芯片生态的基石。公司是国际上最早开展智能处理器指令集研发的少数几家企业之一。新一代商用智能处理器指令集正在研发中。公司已形成了体系完整、功能完备、高度灵活的智能芯片指令集专利群。公司在云端、边缘端、终端三条产品线的所有智能芯片和智能处理器核以及基础系统软件均构建于自研的 MLU 指令集基础之上。	已取得专利 194 项（其中境外专利 53 项），2023 年内增加 36 项。	成熟稳定	自主研发
3	SoC 芯片设计	公司已掌握复杂 SoC 设计的一系列关键技术，有力支撑了云端大型 SoC 芯片（思元 100、思元 270、思元 370 和思元 590）的量产。	已取得专利 65 项（其中境外专利 25 项），2023 年内增加 13 项。	成熟稳定	自主研发

	元 290) 和边缘端中型 SoC 芯片 (思元 220) 的研发。	利 20 项), 2023 年内增加 21 项。		
4	处理器芯片功能验证 公司拥有成熟先进的处理器和 SoC 芯片功能验证平台, 确保了智能处理器和 SoC 芯片逻辑设计按时高质量交付, 有效保障了多款芯片产品的一次性流片成功。	已取得专利 14 项, 2023 年内增加 1 项。	成熟稳定	自主研发
5	先进工艺物理设计 公司已掌握 7nm 等先进工艺下开展复杂芯片物理设计的一系列关键技术, 已将其成功应用于思元 100、思元 220、思元 270、思元 290、思元 370 等多款芯片的物理设计中。	非专利技术。	成熟稳定	自主研发
6	芯片封装设计与量产测试 应用于公司云端、边缘端不同品类芯片产品的封装设计与量产测试过程, 有效支撑了公司处理器芯片的研发。	已取得专利 5 项, 2023 年内增加 2 项。	成熟稳定	自主研发
7	硬件系统设计 有效解决了高速信号完整性、大功率供电下的电源完整性、大型芯片散热、机箱模块化等关键问题, 支撑公司基于自研芯片研发模组/智能加速卡、整机、集群等多样化的产品形态。	已取得专利 51 项 (其中境外专利 11 项), 2023 年内增加 10 项。	成熟稳定	自主研发

资料来源: 公司年报, 天风证券研究所

2023 年研发成果领先, 对大模型挑战实现的多点核心技术突破。寒武纪持续进行大模型的优化和适配工作, 在芯片能力、核心 IP 能力、芯片间互联能力、并行计算能力、存储优化能力和通信优化能力均实现技术突破。2023 年, 公司先后开发了 BangTransformer 大模型分布式推理加速引擎, CNDeepspeed/Megatron-LM 大模型分布式训练加速引擎, 大模型推理与训练性能预估工具、分析工具以及精度分析工具等大模型相关软件栈, 同时基于思元 370 等系列产品, 与 LLaMa 系列、BLOOM 系列、GLM 系列、StableDiffusion、T5, 以及国内的百川、千象等大模型进行了广泛的适配和验证, 性能优异, 并可支持自然语言处理、视觉、多模态等多种场景的大模型应用。

图 44: 公司应对大模型挑战实现的多点核心技术突破



资料来源: 寒武纪 2023 年社会责任报告, 天风证券研究所

5. 风险提示

(1) AI 算力景气度下降的风险

算力支出与下游应用息息相关, 若 AI 应用需要更长期才能突破, 则算力支出的高景气可能不可持续

(2) 公司产品供应端的风险

寒武纪收到美国制裁, 同时采用 Fabless 的商业模式, 在供应端可能受到国际政治风险

(3) AI 芯片竞争加剧的风险

AI 芯片领域有较多参与者，未来市场竞争可能加剧

(4) 客户集中度较高的风险

寒武纪现阶段客户集中度较高，单个客户可能对收入利润产生较大影响，若某下游大客户订单开展不及预期有可能影响公司的收入与现金流

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，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将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。

一般声明

除非另有规定，本报告中的所有材料版权均属天风证券股份有限公司（已获中国证监会许可的证券投资咨询业务资格）及其附属机构（以下统称“天风证券”）。未经天风证券事先书面授权，不得以任何方式修改、发送或者复制本报告及其所包含的材料、内容。所有本报告中使用的商标、服务标识及标记均为天风证券的商标、服务标识及标记。

本报告是机密的，仅供我们的客户使用，天风证券不因收件人收到本报告而视其为天风证券的客户。本报告中的信息均来源于我们认为可靠的已公开资料，但天风证券对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，天风证券及/或其关联人员均不承担任何法律责任。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，天风证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。天风证券的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。天风证券没有将此意见及建议向报告所有接收者进行更新的义务。天风证券的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

特别声明

在法律许可的情况下，天风证券可能会持有本报告中提及公司所发行的证券并进行交易，也可能为这些公司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。因此，投资者应当考虑到天风证券及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突，投资者请勿将本报告视为投资或其他决定的唯一参考依据。

投资评级声明

类别	说明	评级	体系
股票投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	买入	预期股价相对收益 20%以上
		增持	预期股价相对收益 10%-20%
		持有	预期股价相对收益 -10%-10%
		卖出	预期股价相对收益 -10%以下
行业投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	强于大市	预期行业指数涨幅 5%以上
		中性	预期行业指数涨幅 -5%-5%
		弱于大市	预期行业指数涨幅 -5%以下

天风证券研究

北京	海口	上海	深圳
北京市西城区德胜国际中心 B 座 11 层	海南省海口市美兰区国兴大道 3 号互联网金融大厦 A 栋 23 层 2301 房	上海市虹口区北外滩国际客运中心 6 号楼 4 层	深圳市福田区益田路 5033 号平安金融中心 71 楼
邮编：100088	邮编：570102	邮编：200086	邮编：518000
邮箱：research@tfzq.com	电话：(0898)-65365390 邮箱：research@tfzq.com	电话：(8621)-65055515 传真：(8621)-61069806 邮箱：research@tfzq.com	电话：(86755)-23915663 传真：(86755)-82571995 邮箱：research@tfzq.com