计算机

报告日期：2024 年 04 月 08 日

海外科技启示录：英伟达（1）——超级工厂是怎样炼成的

——算力行业深度报告

投资要点

- **英伟达：全球算力王者，加速计算时代的 AI 超级工厂**

公司成立自 1993 年，早期专注图形计算芯片，针对游戏机、PC 市场，2006 年推出 CUDA 进军通用计算，逐步形成 GPU+CPU+DPU 的三芯布局，应用领域也从游戏拓展至数据中心、专业可视化、自动驾驶等，目标客户涵盖 CSP、商业企业、消费者；公司上市以来，收入成长超 160 倍，市值增长超 2700 倍，在全球 AI 芯片市场的市占率最高已经达到 90%。

- **历史复盘：强自研+前瞻并购丰富产品矩阵，通用生态形成坚实护城河**

（1）2000 年前，从“狂野西部”通用图形计算起步，凭借更加高效的研发模式、绑定大客户微软胜出并定义世界首款 GPU：研发上，公司采用了“三团队-两季度”的高效研发模式实现了技术和产品的快速迭代；产品上，通过 DDR、SDR 和 TNT 三个系列实现了高中低端的全面布局；战略上，绑定大客户微软，凭借 ows 95　　操作系统的高占有率打败并收购图形计算先驱 3dfx;

（2）2000-2005 年，客户多元化，通过性能优势掌握 PC 独显龙头地位：上市初期，公司与微软合作失败并导致业绩和股价的调整，后续公司积极开拓 Intel、索尼等大客户，并通过 GeFroce 6800Ultra 优异的产品性能再次夺回 GPU 老大的地位，同时打败了竞争对手 ATI;

（3）2006-2015 年，以游戏业务筑基，培育以 CUDA 为核心的通用计算体系：在终端多元化导致 PC 出货量下降的背景下，公司聚焦高端游戏卡稳定增长态势并实现了逆势增长，2010-2015 年游戏业务收入五年 CAGR 达到 21%；2006 年公司超前推出 CUDA，从图形计算拓展至通用计算，虽然在前期经历了产品生态不完善导致的业绩和股价的调整，但是公司坚持投入研发，研发费用逐年攀升，形成了快速迭代的硬件+深度捆绑的软硬件+大量外围的二次开发者和易用的软件生态三位一体的生态飞轮。

（4）2016 年至今，CUDA 开花结果，云端数据中心业务开启新一轮成长曲线：算力需求侧经历了大数据及云服务（2016-2018）、云端办公和娱乐（2020-2021）、云端 AI 训练（2023 至今）三大阶段，公司套娃式收购了 IB 和以太网领军 Mellanox、网络安全和智能 IP 提供商 Titan IC，并形成了 GPU+CPU+DPU 三芯底座，公司凭借综合全面的生态优势在三次浪潮中逐步崛起，2023 年数据中心业务收入 475.25 亿美元，同比增长 224.36%。

- **成长之路总结：专注带来前瞻，通用诞生生态**

（1）专注计算芯片，带来前瞻战略思维：公司三十年专注计算芯片，并保持在单一赛道上超过竞争对手的持续、高强度投入，同时公司具备前瞻的战略眼光，通过收并购提前布局，最终带来了领先全球的产品迭代速度;

（2）注重技术复用性，让研发投入落到实处：公司在前期收并购布局时也曾遇到过错失移动终端的失误，但公司将前期投入复用到边缘计算领域，并带来了智能驾驶、机器人等新的业务机会，研发投入始终累积;

（3）重视通用性，引入生态共建：公司在图形计算领域做深做专后，大力扩展应用领域，并持续降低开发者的使用门槛，高通用性、低门槛带来的是生态共建飞轮，进而形成了坚实的护城河。

- **国内相关公司：逐步追赶，国产化趋势已现**

华为昇腾：国产算力龙头，性能优势明显，昇腾生态有望成为国内 AI 算力第二级;

海光信息：中科院背景，研发实力强劲，拥有国产稀缺的 x86 架构，类 CUDA 带来生态兼容性，有望部分替代英伟达在国内算力产业中的生态位;

行业评级：看好(维持)

分析师：刘雯蜀
执业证书号：s1230523020002
liuwenshu03@stocke.com.cn

分析师：李佩京
执业证书号：S1230522060001
lipeijing@stocke.com.cn

相关报告

1 《行业逐步回暖，算力确定性最强 》 2024.04.05

2 《英伟达引领 AI 行业变革，打造新时代的“苹果公司”》 2024.03.19

3 《华为合作伙伴大会：未来三年中国政企业务有望快速增长》 2024.03.17

更多免费行业报告 并购家 www.ipoipo.cn

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

更多免费行业报告 并购家 www.ipoipo.cn

寒武纪：产品栈实现云、边、端全覆盖，AI 训练推理卡同步快速发展，性能持续迭代升级；

景嘉微：景宏系列高性能计算产品有望持续突破下游客户，商业化可期。

- **风险提示**

历史成长轨迹与股价表现并不能代表未来趋势；上游供应不足的风险；技术迭代难度加大的风险；中游竞争加剧的风险；下游客户需求变化的风险；其他导致生态壁垒削弱的风险；宏观环境变化的风险；

更多免费行业报告 并购家 www.ipoipo.cn

正文目录

1 英伟达：全球算力王者，加速计算时代的 AI 超级工厂 6
1.1 上市以来收入成长超 160 倍，市值增长超 2700 倍 6
1.2 三芯片四领域，构筑全面产品矩阵 6
1.3 三十年专注的高研发投入，奠定高业绩成长基石 10
2 历史复盘：用产品定义算力发展史 12
2.1 2000 年前：从“狂野西部”通用图形计算起步，凭借更加高效的研发模式、绑定大客户微软胜出并定义世界首款 GPU 12
2.2 2000-2005：客户多元化，通过性能优势掌握 PC 独显龙头地位 14
2.3 2006-2015：以游戏业务筑基，培育以 CUDA 为核心的通用计算体系 14
2.4 2016 至今：CUDA 开花结果，云端数据中心业务开启新一轮成长曲线 18
2.4.1 2016-2019：大数据与云服务阶段，凭借 GPU 的通用性登上数据中心芯片王座 19
2.4.2 2020-2022：云端办公和娱乐阶段，收购整合打造最强数据中心异构芯片+高速互联+网络安全解决方案体系护城河 19
2.4.3 2023 至今：大模型浪潮引爆公司数据中心业务成长 21
3 巨头成长之路总结：专注带来前瞻，通用诞生生态 23
3.1 专注计算芯片，带来前瞻战略思维 23
3.2 注重技术复用性，让研发投入落到实处 25
3.3 重视通用性，引入生态共建 26
4 国内相关公司：逐步追赶，国产化趋势已现 29
4.1 华为昇腾：对标英伟达，有望成为国内第二 AI 算力生态 29
4.2 海光信息：类 CUDA 带来更好的生态兼容性 30
4.3 寒武纪：云端芯片性能持续扩展，云边终端协同覆盖 31
4.4 景嘉微：发布景宏系列高性能计算产品，商业化布局有望加速 32
5 风险提示 32

更多免费行业报告 并购家 www.ipoipo.cn

图表目录

图 1: 公司分季度收入情况（亿美元、%）.....6
图 2: 公司三芯布局产品线.....7
图 3: 公司 FY2024Q3 分业务营收情况.....7
图 4: DOCA 软件体系.....7
图 5: 英伟达数据中心产品迭代情况.....7
图 6: 英伟达数据中心 GPU 发展历程.....8
图 7: 公司数据中心业务收入情况（亿美元、%）.....8
图 8: 公司游戏业务收入情况（亿美元、%）.....9
图 9: 公司游戏显卡发展历史.....9
图 10: NVIDIA Ada Lovelace 架构能实现更拟真的光线追踪效果.....9
图 11: DLSS 利用 AI 实现性能的成倍提升.....9
图 12: 公司专业可视化业务收入情况（亿美元、%）.....10
图 13: 公司专业可视化产品发展历史.....10
图 14: 公司自动驾驶业务收入情况（亿美元、%）.....10
图 15: 公司自动驾驶产品发展历史.....10
图 16: 可比公司毛利率情况（%）.....11
图 17: 可比公司净利率情况（%）.....11
图 18: 可比公司研发费用绝对值情况（亿美元）.....11
图 19: 可比公司研发费用率情况（%）.....11
图 20: 可比公司销售及管理费用率（%）.....12
图 21: “Pre GPU”时期图形计算芯片技术发展路线图.....12
图 22: 公司发展前期营业收入情况（亿美元，%）.....14
图 23: 独立 GPU 市占率变化情况（%）.....14
图 24: 全球智能手机出货情况（亿部、%）.....15
图 25: 研发费用绝对值对比：Nvidia vs. AMD（亿美元）.....16
图 26: 英伟达数据中心平台.....17
图 27: 英伟达 GPU 架构演进图.....17
图 28: 全球 PC 出货量（亿台、%）.....18
图 29: 全球桌面显卡出货量（万块、%）.....18
图 30: 全球独立显卡市场占有率情况（%）.....18
图 31: 公司游戏业务收入情况（亿美元、%）.....18
图 32: 海外主要云厂商季度 Capex（亿美元、%）.....18
图 33: 海外主要云厂商季度 Capex 同比增速 vs 英伟达数据中心业务季度同比增速（%）.....18
图 34: 四大云服务厂商 GPU 分布（%）.....19
图 35: 英伟达在 TOP500 系统中份额（%）.....19
图 36: 未来算力生态.....20
图 37: 英伟达三芯布局产品线.....20
图 38: DOCA 软件体系.....20
图 39: 监督微调-奖励建模-强化学习过程.....21
图 40: Transformer 架构.....21
图 41: 全球数据中心 GPU 市场规模（亿美元、%）.....22
图 42: Transformer 架构示意图.....22

更多免费行业报告 并购家 www.ipoipo.cn

图 43: 可比公司研发费用绝对值情况（亿美元） 24
图 44: 可比公司研发费用率情况（%） 24
图 45: 独立显卡市场占有率情况（%） 24
图 46: 英伟达 Tegra 芯片 25
图 47: 同时搭载高通芯片与 Tegra 的小米 3 25
图 48: Tegra X1 相较前一代的性能巨大提升 26
图 49: Tegra X1 采用 Maxwell 架构 26
图 50: Tegra X1 基于深度学习可以识别现实世界的事物 26
图 51: CPU 与 GPU 架构的根本区别 27
图 52: 支持 C++ API 接入 28
图 53: 支持 Python API 接入 28
图 54: TensorRT 最大化 GPU 在 AI 推理中的性能 28
图 55: TensorRT 最小化客户总拥有成本 28
图 56: CUDA 平台部分生态伙伴 28
图 57: CUDA 生态部分伙伴 29
图 58: 昇腾芯片迭代计划 29
图 59: 华为昇腾 AI 产业生态 30
图 60: 华为行业智能化参考框架 30
图 61: 海光 DCU 组成部分 31
图 62: 海光软件栈体系 31

表 1: 公司历代 GPU 架构情况 8
表 2: Riva128 与 i740 对比情况 13
表 3: 6800Ultra、X800 参数比较 14
表 4: 2002-2013 年英伟达收购汇总 15
表 5: 大模型训练阶段对算力的需求 21
表 6: 大模型推理阶段对算力的需求 22
表 7: AI 芯片与 CPU 芯片在大模型训练中的效率和速度对比 23
表 8: Nvidia VS AMD 产品迭代情况 24
表 9: CUDA VS ROCm 25
表 10: GPU VS CPU 26
表 11: 主流 AI 芯片性能比较 30
表 12: 寒武纪产品线 32

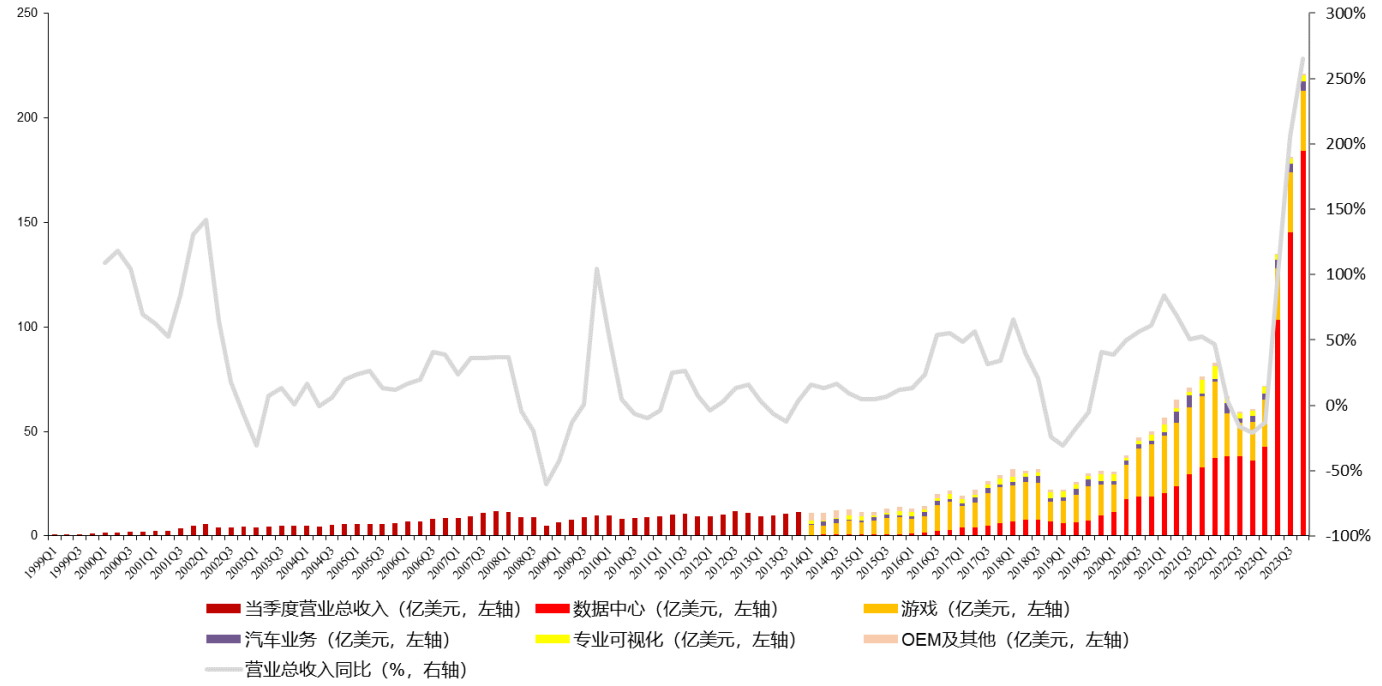
更多免费行业报告 并购家 www.ipoipo.cn

1 英伟达：全球算力王者，加速计算时代的 AI 超级工厂

1.1 上市以来收入成长超 160 倍，市值增长超 2700 倍

公司是全球领先的 GPU 算力龙头。公司于 1993 年由 Jenson Huang（黄仁勋）及来自于 Sun Microsystem 的两位工程师 Chris Malachowsky 和 Curtis Priem 创立，专注图形计算芯片的设计与研发，公司经历了起步积累、困境反转、转型升级和快速成长四个阶段，并成长为全球 AI 算力领军，根据 Gartner 预计，公司在全球 AI 芯片市场的市占率最高已经达到 90%。

图1： 公司分季度收入情况（亿美元、%）



资料来源：彭博、浙商证券研究所

1.2 三芯片四领域，构筑全面产品矩阵

基础芯片层面，公司基于 GPU 技术路线，通过自研+并购形成 GPU+CPU+DPU 的三芯布局：

（1）2000 年推出全球首款 GPU 以来不断进行迭代升级，目前 GPU 领域产品涵盖消费级、工作站级、移动级到高性能计算的多种类型，即将在 2024Q2 出货的 H200 Tensor Core GPU 基于 NVIDIA Hopper™ 架构，FP16 下算力达到 989TFLOPS，同时是首款提供 HBM3e 的 GPU，以每秒 4.8 TB 的速度提供 141GB 内存，与前身 A100 相比容量几乎翻倍，带宽增加 2.4 倍，针对 GPT-3 模型的推理性能是 A100 的 18 倍；

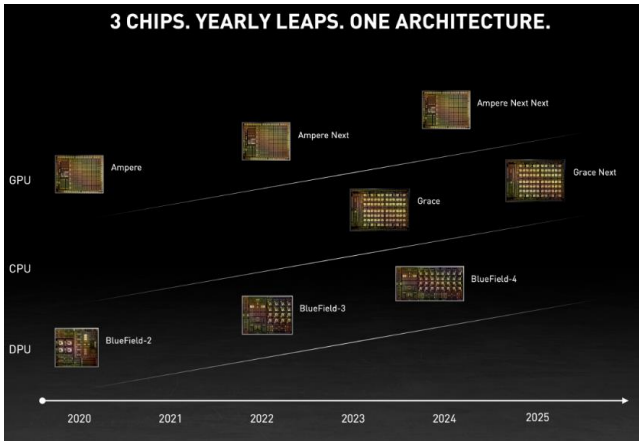
（2）2020 年收购 Mellanox 后推出的 BlueField DCU 能够有效减少 CPU 负荷，提升整体系统性能，BlueField-3 DPU 与上一代相比，具备 2 倍的网络带宽、4 倍的计算能力和几乎 5 倍的内存带宽，能够以高达 8 倍的速度运行工作负载，同时降低 TCO 并提高数据中心能效；

（3）2021 年推出的自研 Grace 系列 CPU 超级芯片基于 ARM v9 架构设计，相较于现有数据中心使用的 x86 CPU，运行微服务的速度快 2.3 倍，内存密集型数据处理性能快 2

更多免费行业报告 并购家 www.ipoipo.cn

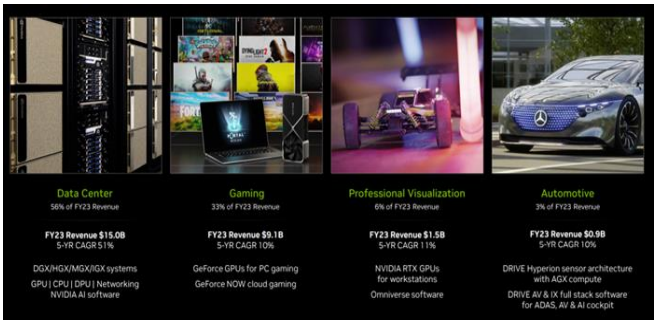
倍，在多个技术计算应用上运行流体力学计算工作时速度快 1.9 倍；为了进一步满足巨型人工智能和高性能计算工作负载的需求，公司还发布了将 Grace CPU 和 Hopper GPU 封装在一起的 Grace Hopper Superchip，以及将两个 Grace CPU 在同一款 PCB 上互联的 Grace Superchip;

图2： 公司三芯布局产品线



资料来源：SDNLAB，浙商证券研究所

图3： 公司 FY2024Q3 分业务营收情况



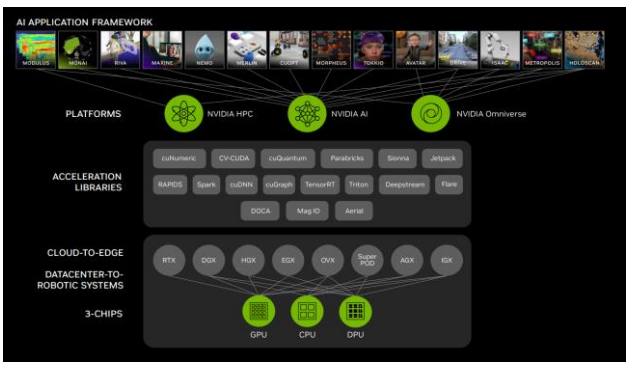
资料来源：英伟达官网、浙商证券研究所

行业客户层面，公司布局了游戏、数据中心、专业可视化、自动驾驶市场四大领域:

（1）数据中心：公司 2016 年至今以及未来长期的增长点。公司为云厂商（CSP）、企业、公共部门的数据中心、智算中心、超算中心提供基于 CPU+GPU+DPU 芯片、IB+以太网等网络设备的硬件系统，以及 AI 加速库、开发工具、应用等软硬件一体的解决方案；随着云端数据中心需求的爆发以及公司三芯布局的形成，公司数据中心业务近五年收入复合增速高达 74.56%、近三年复合增速高达 92.18%，2023 年全年实现收入 475.25 亿美元，同比+216.73%，其中 2023Q4 实现收入 184.04 亿美元，同比+409%;

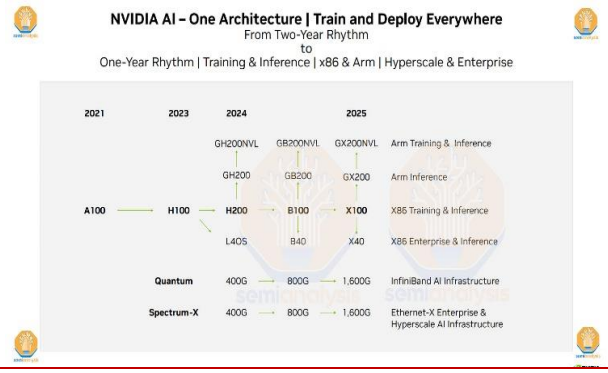
公司自 2011 年的 Tesla M2090 开始不断更新迭代数据中心产品，下一代 B100 采用 Blackwell 架构，将使用台积电的 4nm 工艺，与现有采用 Hopper 架构的 H200 系列相比，性能提升超过 100%。

图4： DOCA 软件体系



资料来源：英伟达，浙商证券研究所

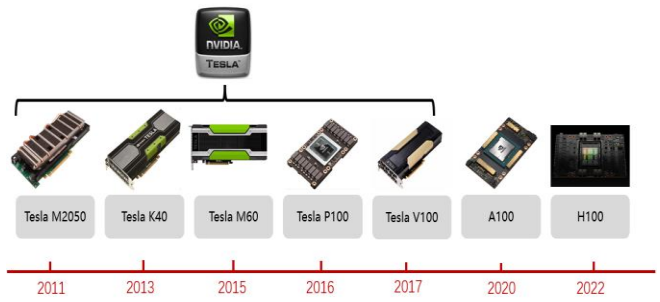
图5： 英伟达数据中心产品迭代情况



资料来源：Semianalysis、浙商证券研究所

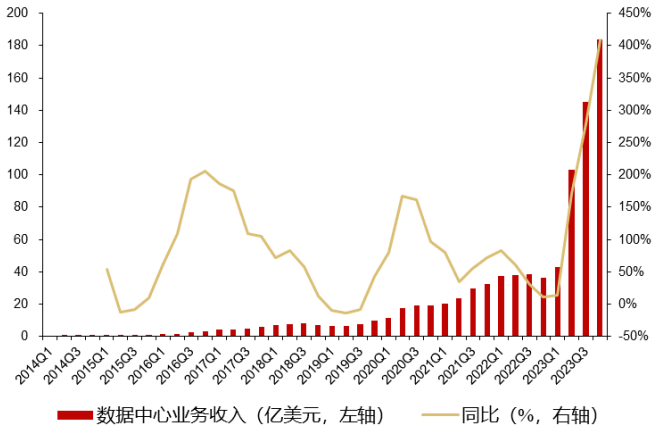
更多免费行业报告 并购家 www.ipoipo.cn

图6： 英伟达数据中心 GPU 发展历程



资料来源：CSDN、英伟达、浙商证券研究所整理

图7： 公司数据中心业务收入情况（亿美元、%）



资料来源：彭博、浙商证券研究所

表1： 公司历代 GPU 架构情况

架构代号	Tesla	Fermi	Kepler	Maxwell	Pascal	Volta	Turing	Ampere	Hopper	Blackwell
中文代号	特斯拉	费米	开普勒	麦克斯韦	帕斯卡	伏特	图灵	安培	赫柏	布莱克威尔
时间	2006	2009	2012	2014	2016	2017	2018	2020	2022	2024
制程	-	40nm	28nm	28nm	16nm	12nm	12nm	8nm	4nm	4nm
核数	128 个	16 个 SM*32CUDA Core，共计 512 个 CUDA Core	15 个 SMX*（192 个单精度+64 个双精度 CUDA core）	具有 3072 个 CUDA 核心	3840 个 CUDA 核心	5120 个 CUDA 核心，新增了 640 个张量核心	具有 4608 个 CUDA 核心、576 个张量核心	具有 6912 个 CUDA 核心、432 个张量核心	具有 18432 个 FP32 CUDA 核心、576 个张量核心	-
特点	首个通用 GPU 计算架构，采用全新的 CUDA 架构，支持使用 C 语言进行 GPU 编程，标志着 GPU 开始从专用图形处理器转变为通用数据并行处理器	引入 L1/L2 快速缓存、错误修复功能和 GPU Direct 技术	首个支持超级计算和双精度计算的 GPU 架构，计算能力比 Fermi 架构提高 3-4 倍，GPU 开始成为高性能计算的关注点	在功耗效率、计算密度上获得重大提升，计算密度是 Kepler 的两倍，标志着 GPU 的节能计算时代到来	增强了 GPU 的能效比和计算密度，功耗只有 300W，比 Maxwell 架构提高 50%以上，使 GPU 可以进入更广泛的人工智能、汽车等新兴应用市场。	AI 计算能力达到 112 TFLOPS，比 Pascal 架构提高了近 3 倍，可以大大加速人工智能和深度学习的训练与推理	新增了 Ray Tracing 核心(RT Core)，可硬件加速光线追踪运算	在人工智能、光线追踪和图形渲染等方面性能大幅跃升，功耗却只有 400W，能效比显著提高	Hopper Transformer 引擎可以做到 FP16 和 FP8 之间逐层交换，利用 NVIDIA 提供的启发算法来降低所需精度	包含 2080 亿个晶体管，采用双倍光刻极限尺寸的裸片，通过 10 TB/s 的片间互联技术连接成一块统一的 GPU。

资料来源：英伟达、腾讯云社区、芯智讯等、浙商证券研究所整理

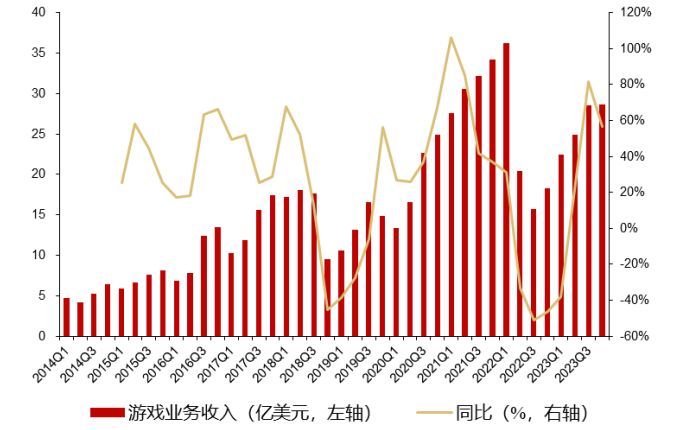
（2）游戏业务：公司的起家业务和基本盘，以先进技术引领行业发展。公司提供分别面向 PC 和笔记本的 GeForce 系列 RTX 和 GTX 显卡硬件、可在性能不足的设备上玩 PC 游戏的 GeForce NOW 云游戏服务、在电视上播放高质量流媒体的 SHIELD 服务以及用于游戏机的平台和开发服务；随着挖矿浪潮的兴起以及公司光追系列显卡的推出，公司游戏业务近五年复合增长率达到 10.91%，2023 年全年实现收入 104.82 亿美元，同比+15.61%，其中 2023Q4 实现收入 29 亿美元，同比+56%，为其他业务提供了稳定的现金流；

游戏业务见证了公司的成长史，每一代微架构的升级都带来了性能的显著提升：2018 年公司推出的首款支持实时光线追踪的 Turing 架构 GeForce RTX 2080 显卡，可以在游戏中

更多免费行业报告 并购家 www.ipoipo.cn

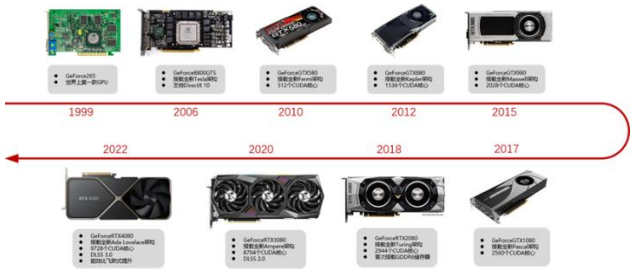
通过模拟光线的物理行为，实现电影级质量的实时渲染，引领了 3A 大作发展的方向；公司推出的 DLSS（深度学习超级采样抗锯齿）技术在不影响游戏性能的同时，能提供与 TAA 抗锯齿技术几乎相同的画质，根据快科技测试，RTX 2080 DLSS 的性能领先上一代 GTX 1080 TAA 达到了 80%。

图8： 公司游戏业务收入情况（亿美元、%）



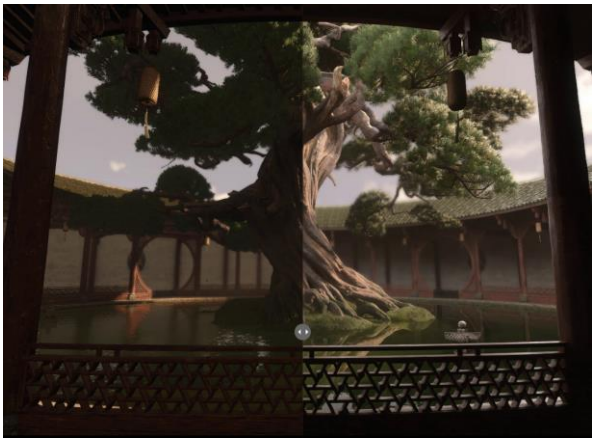
资料来源：彭博，浙商证券研究所

图9： 公司游戏显卡发展历史



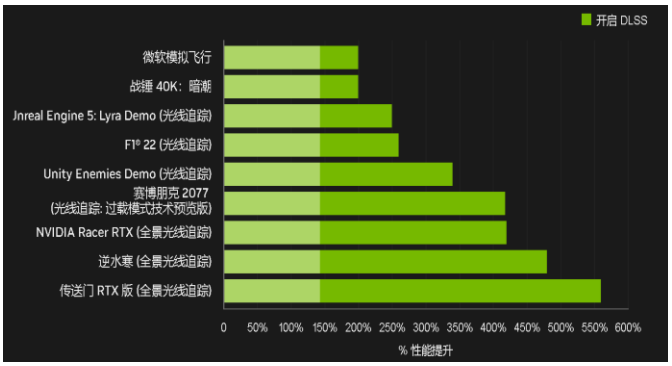
资料来源：CSDN、英伟达、中号硬核玩家等、浙商证券研究所整理

图10： NVIDIA Ada Lovelace 架构能实现更拟真的光线追踪效果



资料来源：英伟达官网，浙商证券研究所

图11： DLSS 利用 AI 实现性能的成倍提升

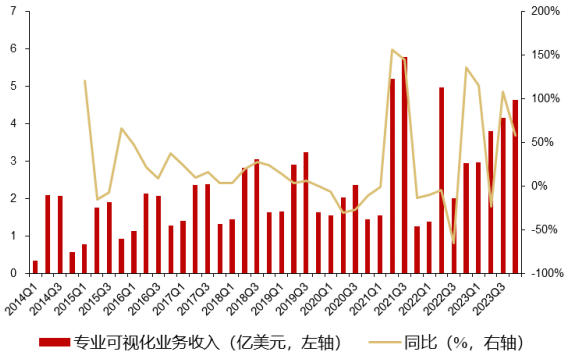


资料来源：英伟达官网，浙商证券研究所

（3）专业可视化：专业图形领域领导者。公司为独立软件供应商（ISV）合作，为在设计与制造环节与数字内容创作环节的 3D 艺术家、建筑师和产品设计师等提供从桌面到云端的 RTX 和 Quadro 解决方案；随着大模型的兴起，企业工作站也开始进行更新迭代，2023 年全年实现收入 12.72 亿美元，同比-3.56%，其中 2023Q4 实现收入 4.63 亿美元，同比+105%。

更多免费行业报告 并购家 www.ipoipo.cn

图12： 公司专业可视化业务收入情况（亿美元、%）



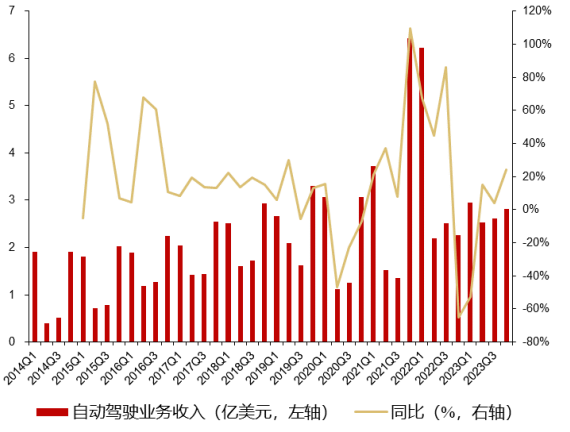
资料来源：彭博，浙商证券研究所

图13： 公司专业可视化产品发展历史

资料来源：CSDN、英伟达、浙商证券研究所整理

（4）自动驾驶：前瞻布局的中长期增长业务。公司通过 DRIVE 系列品牌，为交通运输业构建出软件定义的端到端平台及解决方案，客户可以基于该平台快速高效地开发自动驾驶产品；硬件端包括 Orin、Atlan、以及将于 2025 年投入生产的 Thor SOC，软件端包括针对车载加速计算率先推出的安全操作系统 OS，针对自动驾驶汽车开发的 DriveWorks 中间件，包含感知、地图构建和规划层的 AV 软件栈，AI 辅助驾驶平台 Chauffeur，为 AI 驾驶舱创新解决方案提供舱内感知的开放软件平台 IX，实现实时对话式 AI 的 Concierge，使用准确的真值地图和可扩展的车队来源地图来创建和更新自动驾驶汽车地图 Map 等产品；2023 年全年实现收入 10.90 亿美元，同比-17.36%，其中 2023Q4 实现收入 2.81 亿美元，同比+24.34%。

图14： 公司自动驾驶业务收入情况（亿美元、%）



资料来源：彭博，浙商证券研究所

图15： 公司自动驾驶产品发展历史

资料来源：英伟达等、浙商证券研究所整理

1.3 三十年专注的高研发投入，奠定高业绩成长基石

与 Intel、AMD 相比，CUDA 软件生态带来的毛利率、净利率优势明显。公司上市以来毛利率及净利率随行业及业务的变化经历了三个阶段：

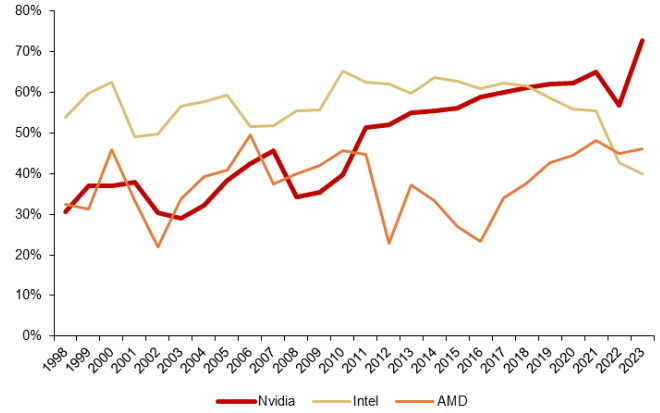
1）上市前至 2003 年，受公司与微软就 Xbox 降价问题的影响，公司毛利率从 2001 年的 37.92%下跌至 2003 年的 29.01%、利润率则从 12.92%下降至 4.08%，后续随着公司与 Intel、索尼签单，毛利率、净利率逐步回升；

更多免费行业报告 并购家 www.ipoipo.cn

2）2004-2008 年，受全球经济危机、研发 CUDA 初期的技术不成熟影响，当时的 G84/G86 核心产品出现了过热而导致花屏的“显卡门”事件，公司毛利率从 2007 年的 45.62%下跌至 2008 年的 34.29%，利润率则从 19.46%转负为-0.88%，后续随着 2009 年 Fermi 架构的 GPU 推出，毛利率、净利率实现了快速修复；

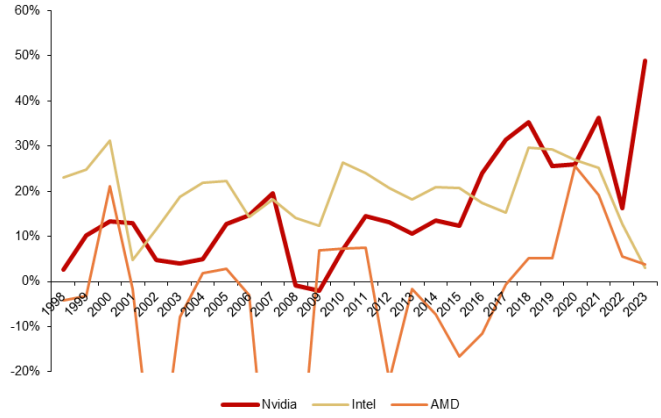
3）2009 年至今，在全球经济复苏以及 CUDA 带来的软件生态优势拉动下，与 Intel、AMD 相比，公司毛利率、利润率开始呈现明显领先的上升态势，毛利率持续上行从 2008 年的 34.29%提升至 2023 年的 72.7%、净利率也同步从-0.88%大幅提升至 48.85%。

图16： 可比公司毛利率情况（%）



资料来源：彭博，浙商证券研究所

图17： 可比公司净利率情况（%）



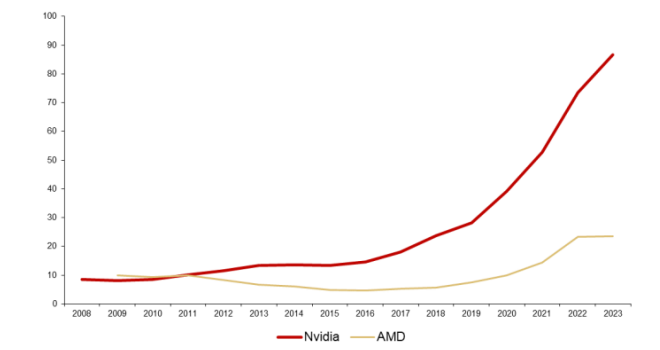
资料来源：彭博，浙商证券研究所

公司常年专注投入研发，研发费用处于高位水平。上市以来，公司通过“三团队-两季度”的研发模式以及聚焦 GPU 研发，实现了在单一领域与友商相比更高的研发投入：

1）在 1999-2005 年与 ATI 的竞争中，公司的研发费用从 1999 年的 0.32 亿美元快速提升至 2005 年的 3.52 亿美元，CAGR 达到 61.54%；

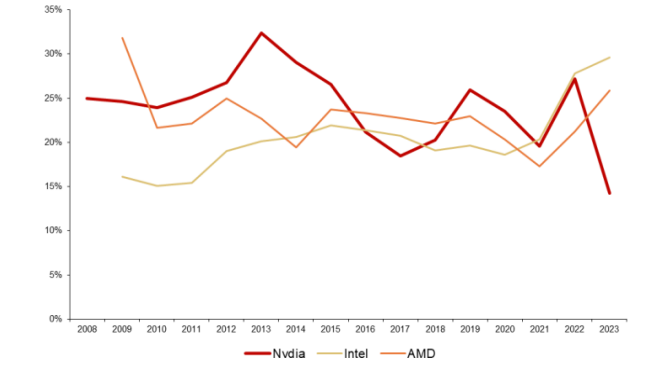
2）在 2006 至今与 AMD 的竞争中，公司的研发费用从 2005 年的 3.6 亿美元快速提升至 2023 年的 86.75 亿美元，CAGR 达到 19.34%；与 AMD 相比，2005 年公司研发投入为 AMD 的 32.73%，而 18 年后的 2023 年，公司研发投入已是 AMD 的 1.47 倍；同时由于公司聚焦 GPU 的研发而 AMD 则需要同时对 CPU 进行投入并于 Intel 展开竞争，我们估计 2023 年在 GPU 领域的研发投入倍数将大于 1.47。

图18： 可比公司研发费用绝对值情况（亿美元）



资料来源：彭博，浙商证券研究所

图19： 可比公司研发费用率情况（%）

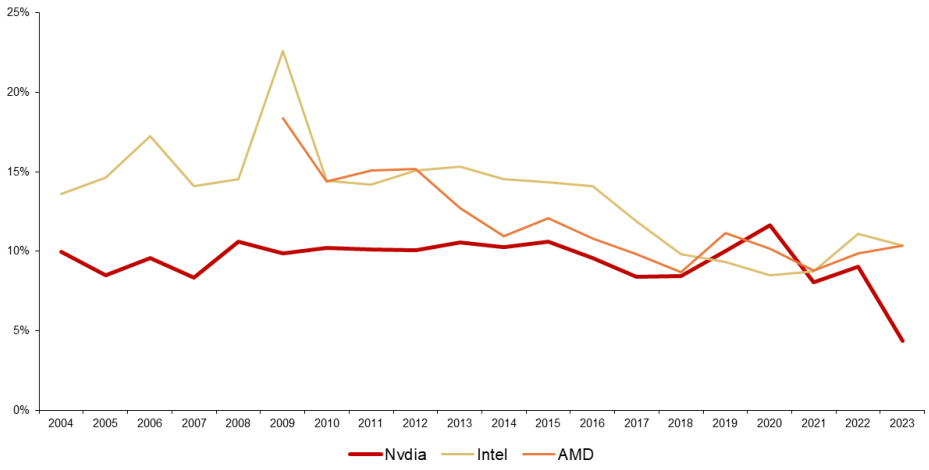


资料来源：彭博，浙商证券研究所

更多免费行业报告 并购家 www.ipoipo.cn

凭借超强的研发投入实现了强大的产品力，销售及管理费用率从常年的 10%逐步下降。公司上市以来销售费用率稳定在 10%并呈逐步下降趋势；2021 年销售及管理费用率从 10.01%提升至 11.63%，主要原因是收购 Mellanox；2023 年，由于公司数据中心业务的爆发，公司销售及管理费用率下降至 4.36%。

图20： 可比公司销售及管理费用率（%）



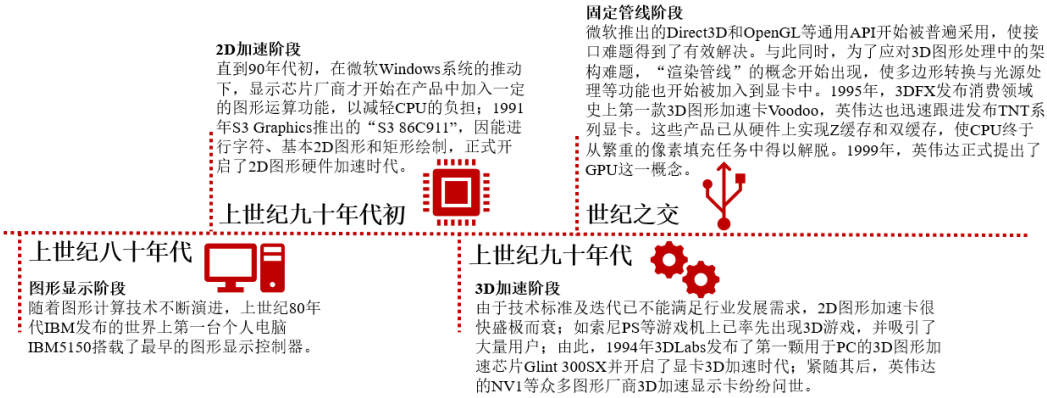
资料来源：彭博、 、浙商证券研究所

2 历史复盘：用产品定义算力发展史

2.1 2000 年前：从“狂野西部”通用图形计算起步，凭借更加高效的研发模式、绑定大客户微软胜出并定义世界首款 GPU

“Pre GPU”时期，图形计算芯片技术快速迭代，百家争鸣：上世纪 90 年代初，高性能图像主要用于图形工作站和视频游戏机中，1995 年微软推出的 ows 95 具备音视频等多媒体功能、大量 3D 渲染游戏登录 PC 平台、图形芯片集成度提升推动了 3D 图像市场的发展；因而“Pre GPU”时期的图形计算芯片技术路线经历了单纯辅助 CPU 进行图形显示、可进行 2D 加速计算、可进行 3D 加速计算、具备固定的渲染管线四个阶段，S3、ATI、AMD、英伟达、3DFX 等众多大小玩家展开激烈竞争，一度形成“百家争鸣”局面，激烈的市场竞争带来的是图形处理芯片的快速迭代和演进。

图21： “Pre GPU”时期图形计算芯片技术发展路线图



资料来源：爱集微、浙商证券研究所

更多免费行业报告 并购家 www.ipoipo.cn

英伟达成立初期专注图形计算芯片的 PC 消费市场：1993 年，Jenson Huang（黄仁勋）及来自于 Sun Microsystem 的两位工程师 Chris Malachowsky 和 Curtis Priem 认为个人电脑将会成为游戏、多媒体的主流消费设备，因而共同创立了英伟达（Nvidia），专注于图形计算芯片的设计与研发。

1994-2000 年公司完成了技术和产品积累：1994 年，公司与意法半导体首次开展战略合作，意法半导体为公司制造单芯片图形用户界面加速器；1995 年，公司发布首款产品 NV1；1997 年，公司推出 RIVA 128 系列产品，凭借高性能+低结构成本而广受市场好评，出货量超过 100 万台，在性能方面甚至优于英特尔于下一年推出的 i740，而 Intel 则逐步退出了独立显卡市场；1998 年，公司与台积电建立合作伙伴关系，自此 OEM 成为公司重要的销售模式；1999 年，公司推出了世界第一款 GPU——GeForce 256，整合了关键的硬件变换和光照（T&L）、立方环境材质贴图和顶点混合、纹理压缩和凹凸映射贴图、双重纹理四像素 256 位渲染引擎等功能，并且兼容 DirectX 和 OpenGL 两大通用 API；2000 年，公司成功收购曾在 1995 年推出消费领域史上第一款 3D 图形加速卡 Voodoo 的图形显卡先驱 3dfx。

表2： Riva128 与 i740 对比情况

	RIVA128ZX	i740
填充率（百万像素/秒）	100	66
储存器接口	128-bit	64-bit
帧缓冲总线带宽	1.6GB/s	800MB/s

资料来源：英伟达，Intel，浙商证券研究所整理

在图形市场发展初期，面对技术及标准不成熟、行业迭代速度快且竞争激烈等难题，**公司凭借“三团队-两季度”的更快速高效的研发运营模式比竞争对手更加快速地响应下游需求的变化、推出全面的产品矩阵、果断绑定大客户微软实现了份额的快速提升，从而在竞争中胜出：**

（1）研发上，公司采用了“三团队-两季度”的高效研发模式实现了技术和产品的快速迭代：图形市场产品研发周期包括短周期（6-9 个月）和长周期（12-18 个月）两类，公司则采用“三团队-两季度”的研发模式，即采用三个并行开发团队专注于第一年秋季、第二年春季、第二年秋季这三个独立的分阶段产品开发，这使得公司可以每 6 个月推出一次新产品，领先市场 1-2 个研发周期，能够更快满足下游需求的变化；

（2）产品上，公司不断丰富产品矩阵：公司在 GeForce 256 时代便通过 DDR、SDR 和 TNT 三个系列实现了高中低端的全面布局；

（3）战略上，公司绑定大客户微软：NV1 由于不兼容竞争 3dfx 的 GLIDE 3D 主流技术标准、成本高、无性能优势因而市场表现平平，公司一度陷入破产的边缘；此时公司做出了重大决定：支持当时微软刚刚推出的 Direct 3D 标准与 GLIDE 进行竞争，依靠着 ows 95 操作系统的高占有率，英伟达 Riva 128 显卡出货量逐渐上升并超越 3dfx，后续凭借 GeForce 256 扩大优势并最终收购 3dfx。

更多免费行业报告 并购家 www.ipoipo.cn

2.2 2000-2005：客户多元化，通过性能优势掌握 PC 独显龙头地位

GPU 时代初期，大客户微软引领图形硬件标准，图形显卡双雄局面形成：2001 年，微软发布了包含全新 Shader Model（优化渲染引擎模式）1.0 标准的 DirectX 8.0，由于遵循这一接口标准的 GPU 具备顶点和像素的可编程性，微软开始引领图形硬件标准，图形显卡领域呈现英伟达、ATI（后被 AMD 收购）双寡头的局面。

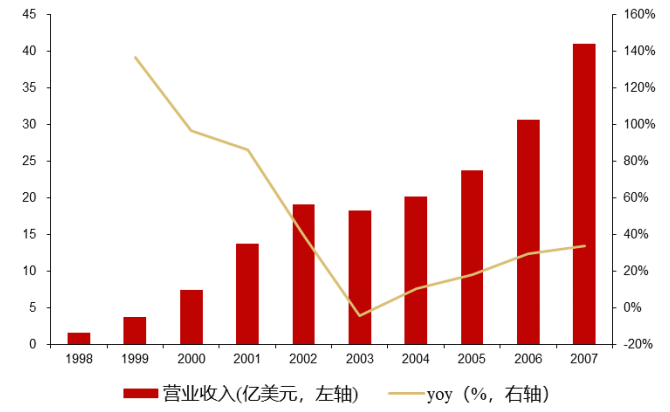
2001-2005 年，与微软合作失败后，公司积极寻求多元客户支持，并通过产品性能再度占据 PC 独显龙头地位：

（1）2000 年公司为微软首款 Xbox 游戏机提供图形处理器，但因交付价格问题产生矛盾而失去了订单（改为竞争对手 ATI 供应），这使得公司 2003 年营收减少，错过了微软 DirectX9 规格确立的重要消息，直接导致当年推出的 GeForce FX 由于兼容性问题败给 ATI 的 Radeon9700;

（2）面对困境，公司积极寻求多元客户支持：1）主动与微软和解，争取再次合作；2）和 Intel 达成了专利交叉许可协议；3）争取到为索尼 PS3 游戏机开发处理器的订单、与暴雪娱乐合作发布基于 3D 图形世界的全球大型多人在线游戏《魔兽世界》;

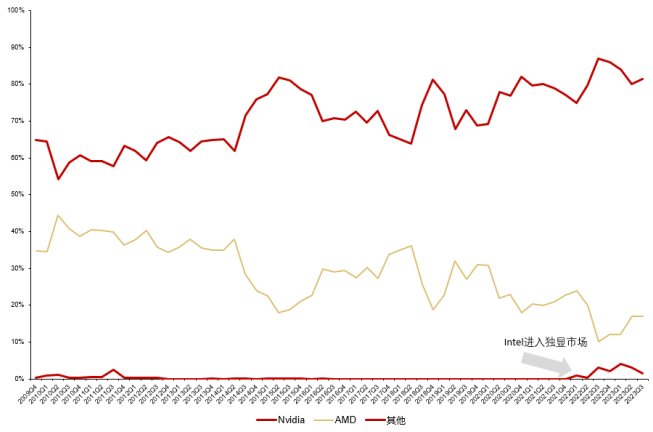
（3）持续迭代：2004 年，公司汲取以往教训推出全新的 GeFroce 6800Ultra，并凭借优异的产品性能再次夺回 GPU 老大的地位；至 2006 年，ATI 被 AMD 斥资达 54 亿美元收购，后续专向中低端市场，自此公司牢牢掌控了 GPU 高端市场并重回增长轨道;

图22：公司发展前期营业收入情况（亿美元，%）



资料来源：彭博、浙商证券研究所

图23：独立 GPU 市占率变化情况（%）



资料来源：Jon Peddie Research、浙商证券研究所

表3：6800Ultra、X800 参数比较

	芯片厂商	制造工艺	核心频率	显存频率	显存类型	显存容量	显存位宽	显存带宽	DirectX 版本
GeForce6800Ultra	NVIDIA	0.13 微米	400MHz	1100MHz	DDR3	256MB	256bit	35.2GB/s	9.0C
X800	ATI	0.11 微米	400MHz	700MHz	DDR3	256MB	256bit	22.4GB/s	9

资料来源：芯参数、浙商证券研究所整理

2.3 2006-2015：以游戏业务筑基，培育以 CUDA 为核心的通用计算体系

因时机和定位失误，错失手机终端机遇：2006-2011 年，以智能手机为代表的移动终端逐步兴起，2010 年功能与设计理念领先业界 3 年的划时代产品 iPhone 4 带来了全球智能手机渗透率的二阶导拐点，2009-2015 年智能手机渗透率从 14.38%提升至 74.08%；公司早

更多免费行业报告 并购家 www.ipoipo.cn

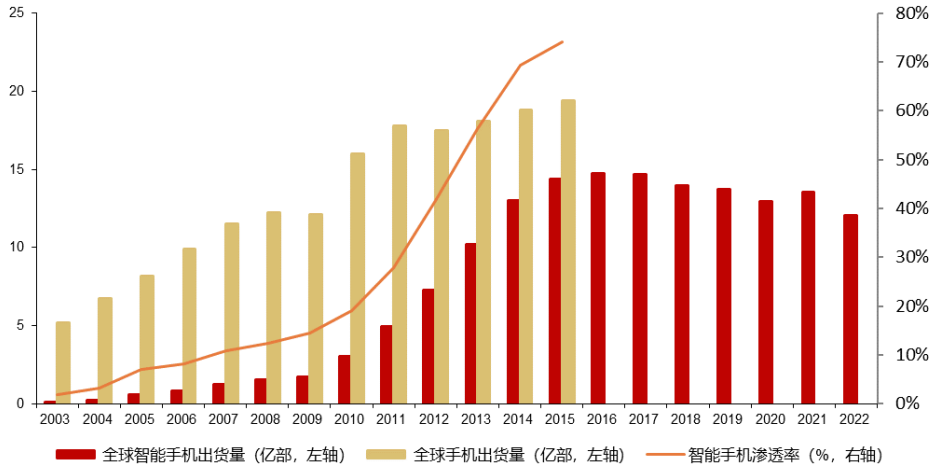
在 2003 年便开始通过收购布局移动端图像芯片，认为未来能实现通话和多媒体功能的手机将成为重要市场，此后的 2008 年公司依靠平板和游戏机的优势推出了针对移动端的 Tegra，但由于高通凭借基带技术占据主流，而 Tegra 后续芯片未能及时整合基带技术而无法及时占领市场，公司因而错失了移动时代机遇，此后公司果断放弃手机市场并将 Tegra 处理器运用在智能汽车、智慧城市和云端服务上。

表4： 2002-2013 年英伟达收购汇总

年份	收购公司	收购目的
2002	Exluna	提供设计人才，推动 CG 语言进入电影行业。
2003	MediaQ	打开快速增长的移动和手持市场领域。
2004	iReady	获得用于支持超高性能以太网络的传输技术
2005	ULIElectronics	ULI 为 ATI 提供南桥部件
2006	Hybrid Graphics	打开手持设备领域，开发图像解决方案
2006	PortalPlayer	将 GPU 和 PortalPlayer 应用处理器结合，完善手持产品线战略分布
2008	Ageia	将 PhysX 物理引擎和 GPU 集成
2011	Icera	帮助代工厂缩短产品上线时间，满足下一代移动计算需求
2013	PGI	为 HPC 系统提供关键部件

资料来源：英伟达、浙商证券研究所整理

图24： 全球智能手机出货情况（亿部、%）



资料来源： 、浙商证券研究所

超前推出 CUDA 进军 GPGPU，开始构建生态护城河：

（1）让只做 3D 渲染的 GPU 技术通用化：早期的 GPU 使用顶点着色单元和像素渲染单元两种计算资源，两种处理器数量的最佳比例是随应用的变化而变化的，因此经常出现一种处理器不够用、而另一种处理器闲置的情况，公司首席科学家 David Kirk 认为给 GPU 装备一组完全相同的、具有较强编程能力的内核，根据任务情况在顶点和片元处理任务之间动态分配可以极大程度提升 PC 的计算性能，同时将丰富的并行运算资源分享给开发者，便可具备重要的战略意义，因而公司开始投入大量研发资源。

更多免费行业报告 并购家 www.ipoipo.cn

（2）坚定方向铺长路：2006 年，公司推出了能够让 GPU 计算变得通用化的 CUDA（Compute Unified Device Architecture）编程技术，并让公司的每一颗 GPU 都支持 CUDA；2007 年，公司推出了不具备绘图能力的第一代大规模并行运算芯片 Tesla；CUDA 初期投入成本较高，并给公司带来了较大的业务压力：

1）在技术方面，芯片面积增大、散热增加、成本上升、故障率增高，直接导致后续 G84/G86 核心的产品出现了过热而导致花屏的“显卡门”事件，而公司也因此付出了近 2 亿美元的一次性支出代价来解决产品质量问题；

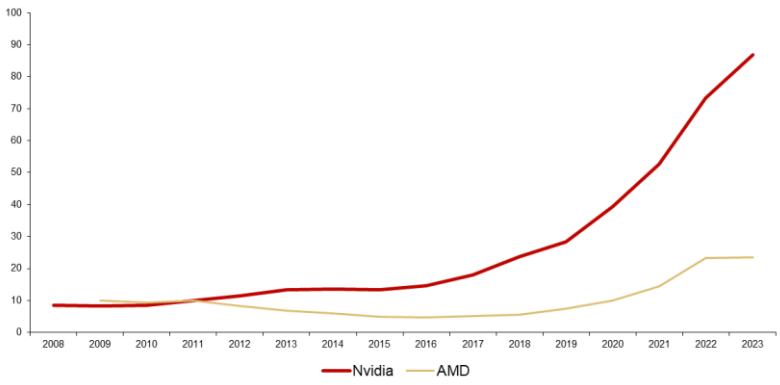
2）在研发上，保证每款产品的软件驱动都支持 CUDA，会对公司的工程师带来巨大的额外工作量；

3）在资金上，一旦项目启动，在当时每年公司预估要在核心业务关系并不紧密的 CUDA 平台上投资高达 5 亿美元，而 2006 年公司总收入 30.68 亿美金；

4）在外部环境上，2008 年 CPU 巨头 AMD 收购公司对手 ATI 并形成了 CPU 整合 GPU 的新解决方案；Intel 也终止了与英伟达的合作并在自家芯片组中集成了 3D 图形加速器；2008 年经济危机也导致了全球 PC 和独显出货量同时出现了负增长；

（3）公司在内忧外困的情况下仍然坚持投入研发，研发费用逐年攀升：2009 年公司推出 Fermi 架构的 GPU，因而经营得以快速恢复，再次夺回市场领先地位，此后公司通过制程的进步及芯片设计的优化持续迭代引领行业；根据创事记，事后来看，公司在将 GPU 转化为更通用的计算工具上投入了将近 100 亿美元；

图25： 研发费用绝对值对比：Nvidia vs. AMD（亿美元）

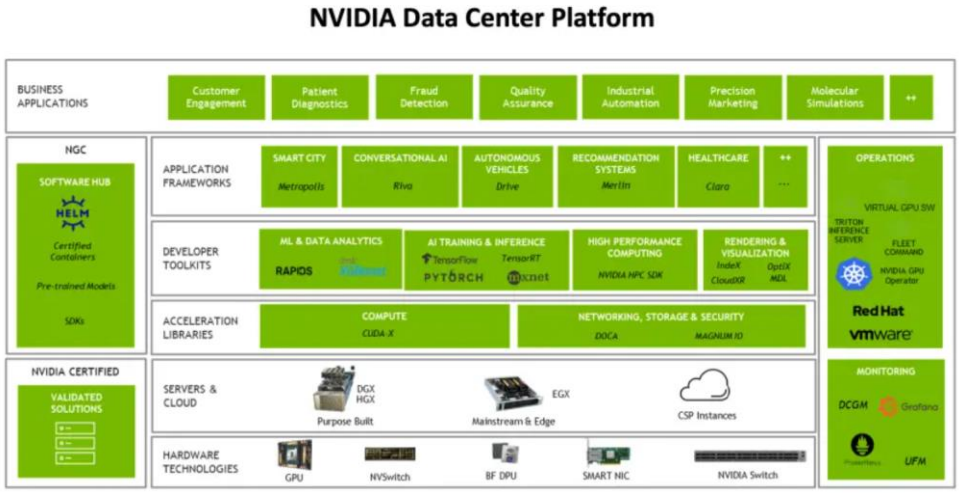


资料来源：彭博、wind、浙商证券研究所

（4）通用计算价值初现，应用领域拓展：CUDA 拓展了 GPU 的应用领域，让只做 3D 渲染的 GPU 得以从游戏（图形渲染）向外扩展至高性能计算、自动驾驶等多个领域，结合前期在游戏、移动领域的积累，公司逐步形成四大产品线：GeForce（PC、笔记本）、Quadro（工作站）、Tesla（大型高性能计算）、Tegra（移动产品）；

更多免费行业报告 并购家 www.ipoipo.cn

图26： 英伟达数据中心平台



资料来源：英伟达、浙商证券研究所

（5）2006 年以来，公司持续推进 CUDA 通用计算生态建设：

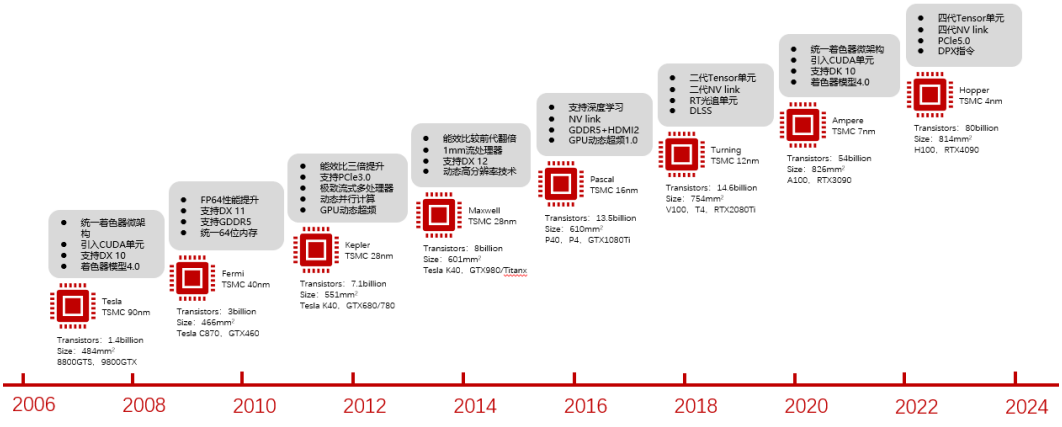
1）架构端，公司每 2 年推出一个微架构，并对四大产品线进行全面升级；

2）硬件端，公司 2019 年收购了网络芯片龙头 Mellanox，并形成了 CPU+GPU+DPU 三芯布局；

3）软件端，公司研发了大量的加速库、开发工具链，极大程度提升了易用性；

最终公司形成了快速迭代的硬件+深度捆绑的软硬件+大量外围的二次开发者和易用的软件生态三位一体的生态飞轮。

图27： 英伟达 GPU 架构演进图



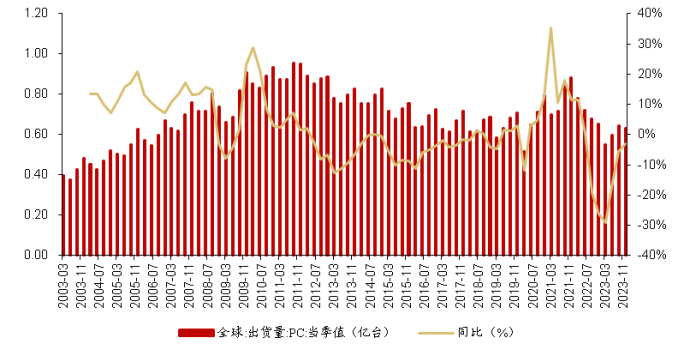
资料来源：英伟达、浙商证券研究所

终端多元化背景下，聚焦高端游戏卡稳定增长态势：2012 年，平板电脑、笔记本电脑等终端的多元化使得 PC 出货开始呈下降趋势，同时集显性价比的逐步提升挤占了独显市场空间，公司战略聚焦高端游戏卡，通过 GeForce 系列站稳脚跟；根据公司披露，2010-2015 年公司游戏业务收入五年 CAGR 达到 21%，游戏显卡出货量五年 CAGR 为 9%、ASP

更多免费行业报告 并购家 www.ipoipo.cn

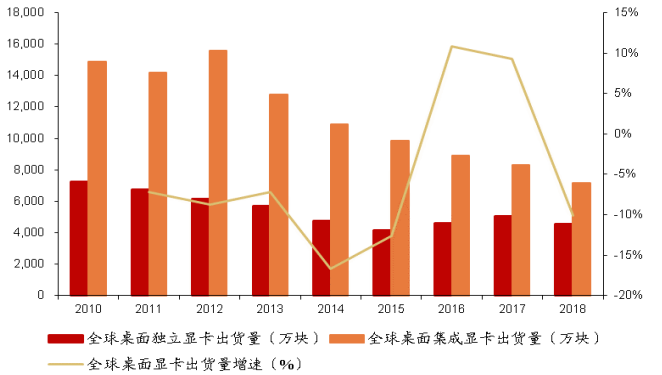
五年 CAGR 为 11%，游戏业务实现了逆势增长，并为公司创造了确定的业绩增长与健康的现金流，奠定了通用 GPU 和 AI 业务爆发的基石。

图28： 全球 PC 出货量（亿台、%）



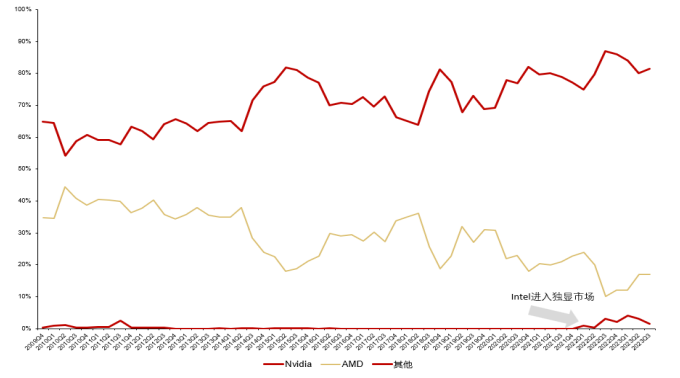
资料来源：wind、浙商证券研究所

图29： 全球桌面显卡出货量（万块、%）



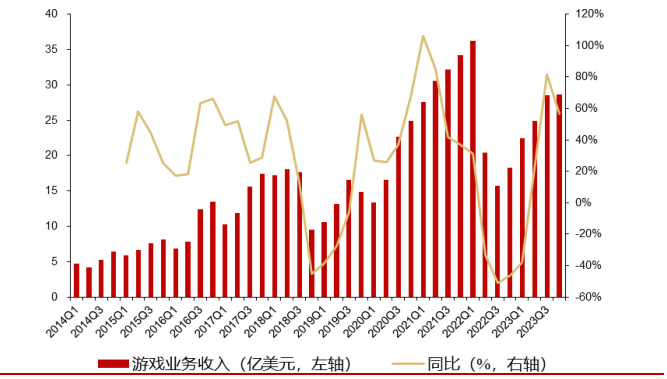
资料来源：华经产业研究院、浙商证券研究所

图30： 全球独立显卡市场占有率情况（%）



资料来源：Jon Peddie Research、浙商证券研究所

图31： 公司游戏业务收入情况（亿美元、%）

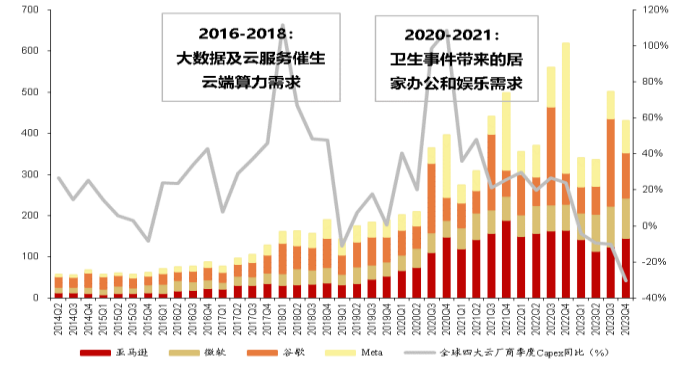


资料来源：彭博、浙商证券研究所

2.4 2016 至今：CUDA 开花结果，云端数据中心业务开启新一轮成长曲线

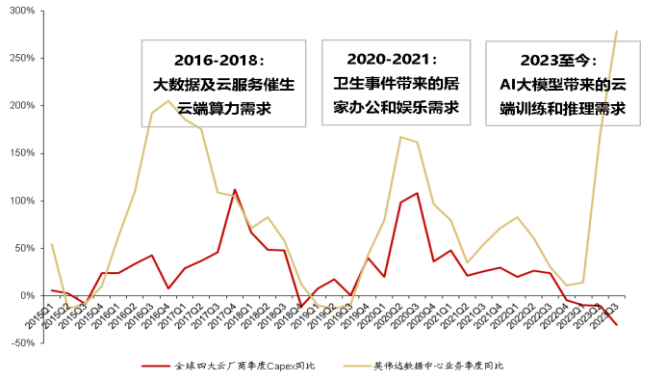
2016 年至今，算力需求侧经历了大数据及云服务（2016-2018）、云端办公和娱乐（2020-2021）、云端 AI 训练（2023 至今）三大阶段，公司借助 CUDA 成长为全球算力龙头。

图32： 海外主要云厂商季度 Capex（亿美元、%）



资料来源：彭博、wind、浙商证券研究所

图33： 海外主要云厂商季度 Capex 同比增速 vs 英伟达数据中心业务季度同比增速（%）



资料来源：彭博、wind、浙商证券研究所

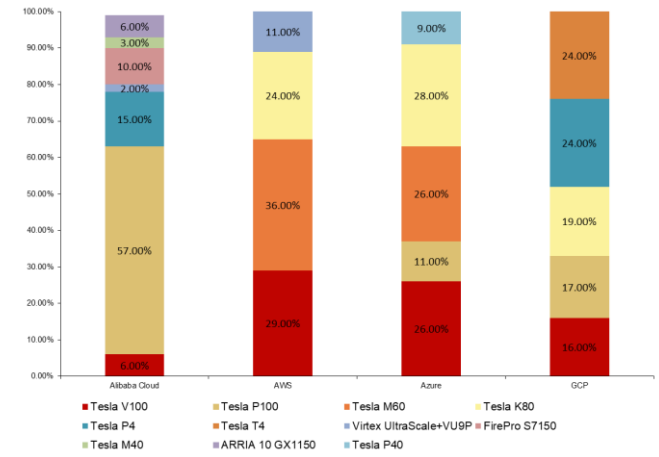
更多免费行业报告 并购家 www.ipoipo.cn

2.4.1 2016-2019：大数据与云服务阶段，凭借 GPU 的通用性登上数据中心芯片王座

大数据催生数据上云需求。2016-2018 年，大数据的发展及移动互联网流量产生的海量数据催生了数据上云的需求，同时分布式编程模式 MapReduce、分布存储和管理技术、虚拟化技术等关键技术的成熟使得云服务能给客户带来的的经济价值提升，海外云厂商纷纷自建大规模数据中心，2016-2017 年，亚马逊、微软、谷歌、Meta 四大云厂商 Capex 从 309.62 亿美元增长至 405.66 亿美元，资本支出同比增长 31%。

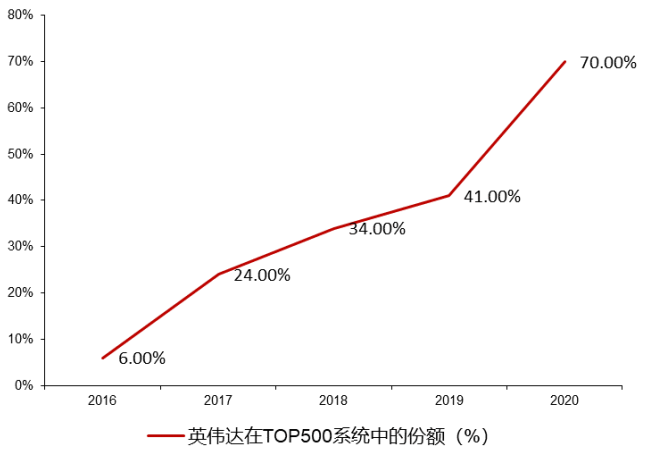
CUDA 积累的 GPU 通用能力直接带来公司数据中心业务爆发。海量数据带来的计算需求激增，公司通过开发 CUDA 将 GPU 实现了通用化，2016 年推出的 Pascal 架构的 P100 具备 3840 个 CUDA 核，在海量数据的并行运算上具备显著优势，公司凭借 Tesla 系列 V100、T4、P100、P4 等产品拔得头筹。在 Liftr 与 2019 年进行的调查中显示，公司的 Tesla 系列产品在阿里云、亚马逊 AWS、微软 Azure、谷歌 GCP 四大云厂商中的专用加速器占有率分别为 82%、89%、100%和 100%，呈现绝对龙头地位；根据海豚投研，公司在 TOP500 系统中的份额从 2016 年的 6%迅速增长至 2017 年的 24%，一年内增长了 3 倍。

图34：四大云服务厂商 GPU 分布（%）



资料来源：forbes、Liftr、浙商证券研究所

图35：英伟达在 TOP500 系统中份额（%）



资料来源：GS、海豚投研、浙商证券研究所

2.4.2 2020-2022：云端办公和娱乐阶段，收购整合打造最强数据中心异构芯片+高速互联+网络安全解决方案体系护城河

卫生事件带来的云端办公和娱乐需求驱动了云厂商的第二轮数据中心建设周期，数据处理及带宽互联是关键。2020 年，卫生事件导致全球生产经营和日常活动都受到了影响，催生了企业上云、人民消费娱乐上云的需求，亚马逊、微软、谷歌、Meta 四大云厂商合计 Capex 从 2020Q1 至 2022Q4 连续 12 个季度保持同比 20%以上的高速增长，云厂商开启了第二轮建设周期；同时，企业要求数据中心除了简单的存储以外，能够实现一定程度的数据分析，虚拟机和容器（containers）等技术的流行也使得应用程序更多放在多台服务器上进行分布式运行，这两大趋势要求未来的数据中心需要同时具备大规模数据处理能力和高带宽互连技术。

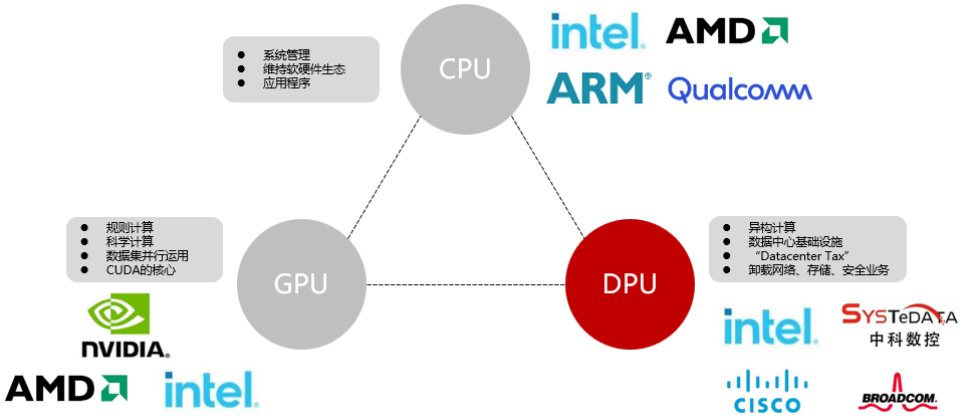
收购整合构筑数据中心上下游一体的体系化护城河。公司于 2019 年斥资 69 亿美金收购了 InfiniBand 和以太网领军 Mellanox，而 Mellanox 则在 2020 年收购了网络安全和智能 IP 提供商 Titan IC，该套娃式收购大大提升了公司云端 AI 产品体系的竞争力，将公司在单机上的生态优势成功拓展到了分布式集群中：分布式训练对于数据交互的需求非常高，而

更多免费行业报告 并购家 www.ipoipo.cn

Mellanox 的数据互联方案+英伟达的 GPU 底层接口可以成为完善的工程解决方案，Titan IC 提供的网络安全和内容智能又能够实现在硬件加速器中检测恶意入侵的网络流量并减少了 CPU 负载，最终公司构建起了包含人工智能芯片及生态+高速数据互联解决方案+网络安全加速的横跨多个领域的完整方案。

DPU 专为减少 CPU 负荷、进一步提升大规模数据中心系统效率而生。数据大爆发的时代，仍存在 CPU 处理效率低下、GPU 处理不了的负载，如网络虚拟化、硬件资源池化等基础设施层服务，DPU 可作为 CPU 的卸载引擎，通过承担网络、存储、安全等业务，提升整个计算系统的效率、降低整体系统的总体拥有成本（TCO）。

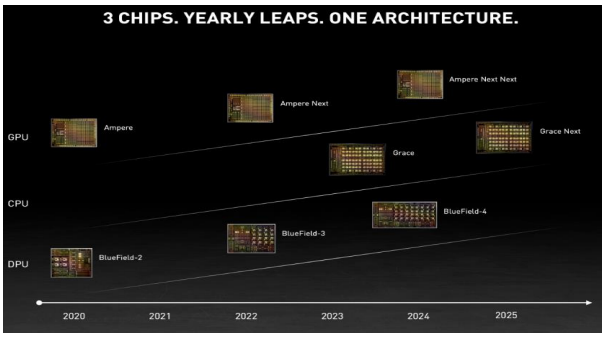
图36： 未来算力生态



资料来源：智东西、浙商证券研究所

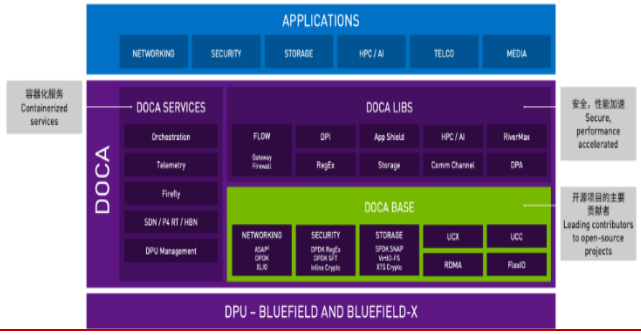
公司集成 CPU+GPU+DPU 形成三芯异构硬件布局，实现数据中心芯片体系的“降本增效”。公司在 2020 年凭借 Mellanox 原有的 ConnectX 系列高速网卡技术，推出了 DPU（数据处理器）BlueField-2，并在 2021 年推出了匹配的 DOCA（Data-Center-Infrastructrue-On-A-Chip-Architectrue，即“线上数据中心基础设施体系结构”）生态，BlueField 系列 DPU 在支持网络处理、安全和存储功能的同时，实现网络虚拟化、硬件资源池化等基础设施层服务，同时可释放高达 30%的 CPU 资源；而 DOCA 软件框架使开发者能够在 BlueField DPU 上快速创建应用程序和服务，为开发者构建软件定义、硬件加速网络、存储、安全和其他基础设施应用程序提供了一个全面的开放平台。

图37： 英伟达三芯布局产品线



资料来源：NVDIA GTC2021，浙商证券研究所

图38： DOCA 软件体系



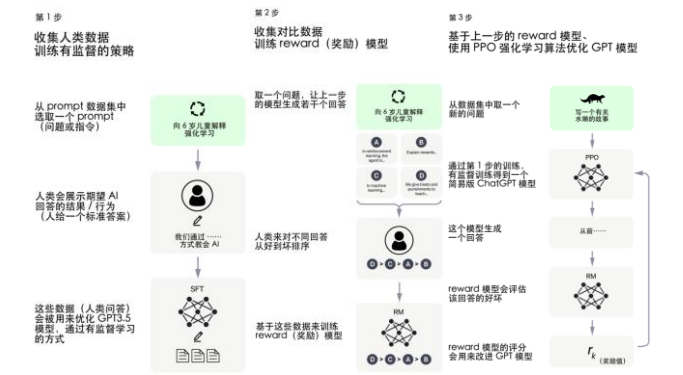
资料来源：英伟达官网、浙商证券研究所

更多免费行业报告 并购家 www.ipoipo.cn

2.4.3 2023 至今：大模型浪潮引爆公司数据中心业务成长

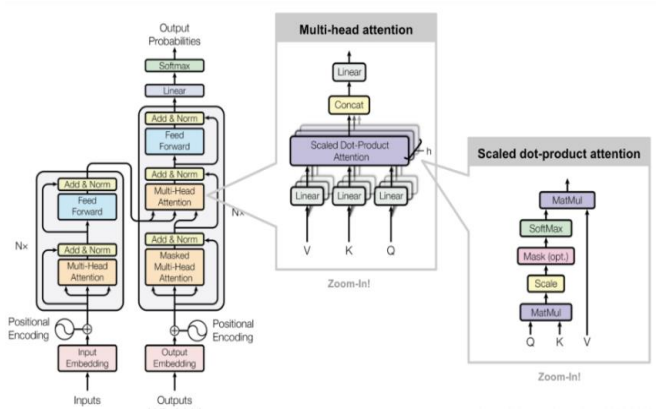
GPT 本质是基于 Transformer 架构的大模型。GPT，全称"Generative Pre-training Transformer"，最初是一个由 OpenAI 开发的自然语言处理（NLP）的模型，**通过预训练和生成技术以及 Transformer 的自注意力机制，可以理解和生成人类的自然语言**，比传统的 RNN、CNN 更快、更稳定、准确率更高、回答更富有逻辑性、并具备强大的泛化能力。

图39： 监督微调-奖励建模-强化学习过程



资料来源：画宇宙、浙商证券研究所

图40： Transformer 架构



资料来源：CSDN、浙商证券研究所

大模型对于算力的需求体现在模型训练和推理应用两个阶段：

（1）训练阶段：根据 OpenAI 的论文《Scaling Laws for Neural Language Models》（2020 年发表），训练阶段算力需求=3×前向传递操作数×模型参数数量×训练集规模，训练所需 GPU 数量=总算力需求/（每个 GPU 每秒运算能力×训练时间×有效算力比率），因此我们可以得到，单次训练 GPT-4 需要约 2.65 万张 A100。

表5： 大模型训练阶段对算力的需求

训练算力需求	GPT-3	GPT-4	SORA
平均参数数量（亿个，N）	1750	2800	100
单 Token 训练所需运算次数（TFLOPS，6N）	1.05	1.68	0.06
训练数据			5 亿图片+1000 万个视频
图片分辨率*像素数			9.72E+04
patch 量（个）			1.75E+16
压缩比例			20%
patch 到 tokens 的换算比例			1.30E-03
单次训练 Tokens 数量（亿个）	3000	130000	45689
训练步数（steps）			20
单次训练所需总算力（TFLOPS）	3.15E+11	2.184E+13	5.48E+12
单次训练所需时间（天）	90	90	90
按上述时间计算，每秒的训练算力需求（TFLOPS）	4.05E+04	2.81E+06	7.05E+05
A100 算力值（非稀疏，TFLOPS）	312	312	312
集群利用率（MFU）	34%	34%	34%
所需卡数	382	26,477	6,647

资料来源：Open AI、英伟达、智东西、新智元、CSDN、AIGC 开放社区、华尔街见闻、第一财经、浙商证券研究所

更多免费行业报告 并购家 www.ipoipo.cn

（2）推理阶段：同样根据 openAI 论文可以得到，单次 GPT-4 推理所需要的算力成本约为 0.05 美分，按照 AIPRM 统计，截至 2023 年 12 月，ChatGPT 拥有约 1.8 亿用户，平均每月产生 17 亿次网站浏览量，则平均每天访问次数为 567 万次，假设每次访问进行 10 轮推理对话，则平均每秒进行推理次数为 17/30*10/3600*10^8≈157407 次，对应 GPT-4 需要 A100 为 27.7 万张。

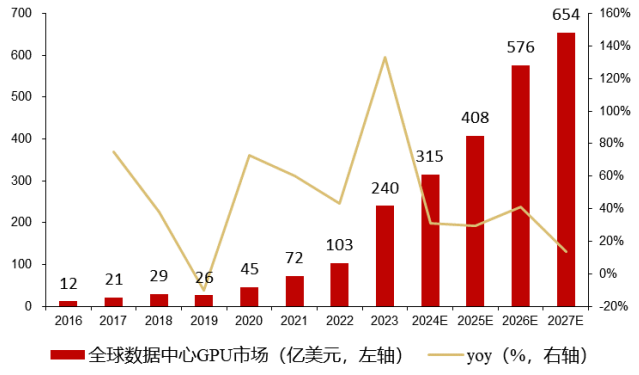
表6： 大模型推理阶段对算力的需求

推理算力需求	GPT-3	GPT-4	SORA
平均参数数量（亿个，N）	1750	2800	100
单 Token 推理所需运算次数（TFLOPS，2N）	0.35	0.56	0.02
单次推理视频长度（秒）			60
每秒帧数			30
图片分辨率*像素数			1.94E+05
patch 量（个）			3.50E+08
patch 到 tokens 的换算比例			1.30E-03
单次推理 Tokens 数量（亿个）	1.00E-05	1.00E-05	4.56E-03
推理步数（steps）			20
单次推理所需算力（TFLOPS）	350.00	560.00	182,250
假设单次推理所需时间（秒）	3	3	3
单次按上述时间计算，每秒的推理算力需求（TFLOPS）	116.67	186.67	60,750.00
A100 算力值（非稀疏，TFLOPS）	312	312	312
集群利用率（MFU）	34%	34%	34%
所需卡数	1.10	1.76	572.68

资料来源：Open AI、英伟达、智东西、新智元、CSDN、AIGC 开放社区、华尔街见闻、第一财经、浙商证券研究所

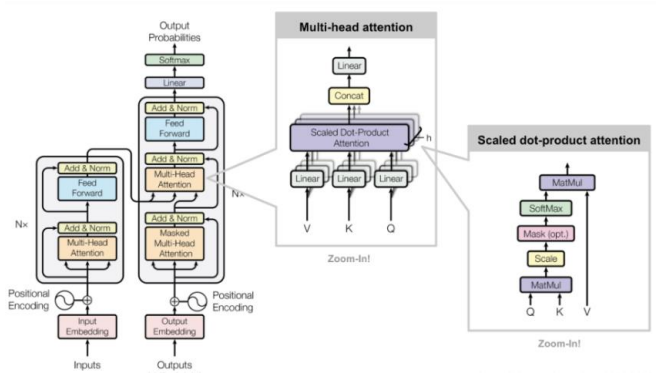
大模型引爆算力需求。根据 IDC 预计，数据中心 GPU 市场预计将从 2022 年的 103 亿美元增长至 2027 年的 654 亿美元，CAGR 达到 44.55%；AMD 报告显示，2023 年全球 AI 芯片市场规模会达 450 亿美元左右，预计 2027 年将增长到 4000 亿美元，2023 年-2027 年复合增速超过 70%。

图41： 全球数据中心 GPU 市场规模（亿美元、%）



资料来源：IDC、浙商证券研究所

图42： Transformer 架构示意图



资料来源：CSDN、浙商证券研究所

公司凭借数据中心产品和生态体系一飞冲天。公司 2023Q3-2023Q4，数据中心业务收入分别达到 145.14、184.04 亿，同比增速达到 278.66%、408.96%；截至 2024 年 3 月 27 日，市值达到 2.3 万亿美元，较 2023 年初涨幅超过 530%。

更多免费行业报告 并购家 www.ipoipo.cn

顺势而为切入云端定制 ASIC。由于算力成本高企，云计算公司纷纷开始自研芯片以部分替代英伟达产品，根据财联社，2 月 9 日消息人士透露，英伟达正在建立一个新的业务部门，专注于为云计算等公司设计定制芯片以及先进的人工智能（AI）处理器；

我们认为公司此举既能减少客户自研芯片带来的替代压力，又能为长期芯片走向降本化、定制化提前做好准备。我们认为，从 10 年维度下的长期视角来看，单位算力成本的下降是确定的趋势，根据 CSET 报告《AI Chips: What They Are and Why They Matter》，与 GPU 相比，ASIC 芯片在训练上的效率平均约为 10 倍、推理上的效率约为 100 倍，因而随着大模型的发展逐步进入成熟期，ASIC 的芯片市场规模有望逐步提升，根据研究公司 650 Group 的 Alan Weckel 的估计，数据中心定制芯片市场今年将增长到高达 100 亿美元，到 2025 年将翻一番。

表7： AI 芯片与 CPU 芯片在大模型训练中的效率和速度对比

	训练		推理		通用性	推理准确度
	效率	速度	效率	速度		
CPU	一倍基准				非常高	98%-99.7%
GPU	10-100 倍	10-1000 倍	1-10 倍	1-100 倍	高	98%-99.7%
FPGA	-	-	10-100 倍	10-100 倍	中等	95%-99%
ASIC	100-1000 倍	10-1000 倍	100-1000 倍	10-1000 倍	低	90%-98%

资料来源：CSET、浙商证券研究所

前瞻布局移动基站，剑指边缘计算。根据新浪财经、财联社，英伟达正在与电信基础设施建设者爱立信就一款包含芯片设计公司的图形处理单元（GPU）技术的无线芯片进行谈判，同时软银和英伟达将联合成立一个新的行业协会“AI-RAN 联盟”，电信巨头爱立信、诺基亚等全球约 10 家公司也将参加其中，致力于将利用移动通信基站分散 AI 处理的技术实用化。

我们认为长期看，未来 AI 算力增量需求场景将逐步由云向边缘、端侧转移，公司有望在边缘侧复制云端的成功经验。根据 TDIA 预计，2023 年底全球 5G 基站将突破 480 万个，650 Group 的 Weckle 预计电信定制芯片市场每年约为 40 亿至 50 亿美元。

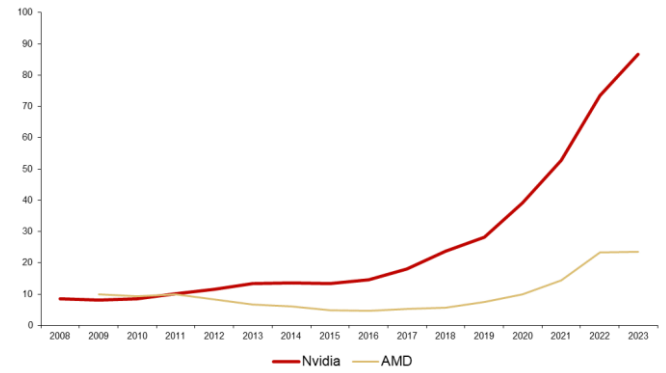
3 巨头成长之路总结：专注带来前瞻，通用诞生生态

3.1 专注计算芯片，带来前瞻战略思维

专注 GPU，保持高强度研发投入。英伟达以图形处理器起家，1999 年，英伟达发布了世界上第一个正式的 GPU——GeForce 256。随后，公司持续加大研发投入，深耕 GPU 领域，无论是游戏显卡、数据中心加速卡、自动驾驶芯片、可视化领域，英伟达不断推展产品终端用户群体的同时，始终保持 GPU 产品的研发和迭代。公司研发费用率常年保持在 20%-30%，同时随着营收规模的增长，研发人员以及研发金额也不断攀升。

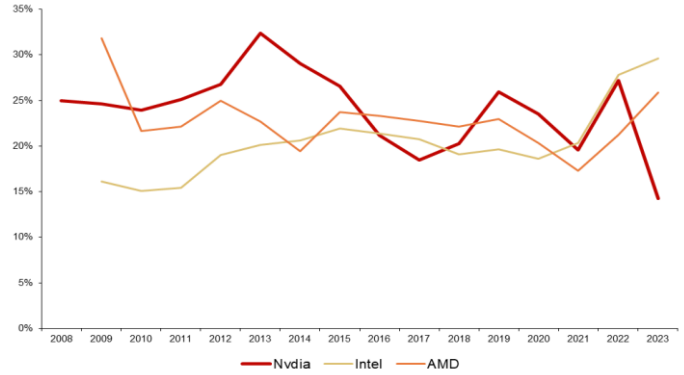
更多免费行业报告 并购家 www.ipoipo.cn

图43： 可比公司研发费用绝对值情况（亿美元）



资料来源：彭博，浙商证券研究所

图44： 可比公司研发费用率情况（%）



资料来源：彭博，浙商证券研究所

“三团队两季度”驱动创新。英伟达成立伊始，为了应对图形芯片市场激烈的竞争，采用“三团队两季度”的研发策略，将研发团队分为产品、硬件、软件三个团队，每个团队专注于自身负责领域以保证产品创新性，同时两季度研发模式保证公司每六个月迭代一次产品，领先市场研发周期的同时，充分满足下游市场需求。

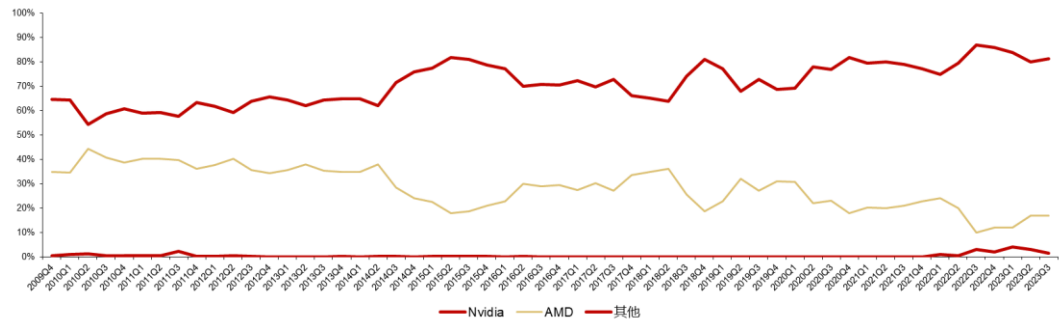
产品端：专注游戏显卡领域竞争，迎合玩家需求快速迭代产品。2010 年后，英伟达和 AMD 逐渐抢占其他公司份额，成为独立显卡领域唯二的巨头。随后在游戏显卡领域，英伟达加速产品迭代速度以及性价比的提升，逐步提升市场份额，截止 2023Q3，英伟达独立显卡市场份额达到 81.50%。**我们认为，正是公司专注于 GPU 研发，带来了游戏显卡的快速迭代以及性价比的快速提升，牢牢抓住玩家需求，最终成长为市场龙头。**

表8： Nvidia VS AMD 产品迭代情况

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Nvidia	GeForce GTX 580	GeForce GTX 680	GeForce GTX 780	GeForce GTX 980	GeForce GTX 980 Ti	GeForce GTX 1080	TITAN V	GeForce RTX 2080	GeForce RTX 2080 Super	GeForce RTX 3080	GeForce RTX 3080 Ti	GeForce RTX 4090
AMD	Radeon HD 6990	Radeon HD 7970	Radeon HD 8970	Radeon R7 250X	Radeon R9 380X	Radeon RX 470	Radeon RX 560	Radeon 550X	Radeon VII	Radeon RX 6900XT	Radeon RX 6600	Radeon RX 7900XTX

资料来源：CSDN 等、浙商证券研究所整理

图45： 独立显卡市场占有率情况（%）



资料来源：Jon Peddie Research、浙商证券研究所整理

微架构创新，从底层突破 GPU 性能与效率。公司自 2006 年自研的 Tesla 架构开始。不断更新 GPU 架构，平均两年迭代一次 GPU 架构。目前公司旗舰产品采用 Hopper 架构，采用台积电 4nm 制程，下一代游戏显卡 RTX50 系产品以及数据中心产品 B100 将采用

更多免费行业报告 并购家 www.ipoipo.cn

Blackwell 架构，将使用台积电的 4nm 工艺，与现有采用 Hopper 架构的 H200 系列相比，性能提升超过 100%。公司 GPU 架构发展历程详见图 27。

前瞻性：预见 CUDA 生态的革命性。英伟达于 2006 年开发 CUDA，从今天的眼光来看，这项决策领先了市场十年之久，对比另一家 GPU 巨头 AMD，AMD 于 2015 年为了对标 CUDA 生态开发了 ROCm，落后英伟达 9 年。

表9： CUDA VS ROCm

	CUDA	ROCm
开发时间	2006 年	2015 年
兼容性	NVIDIA GPU	AMD Radeon GPU、CPU
开源	专有平台	开源平台
生态系统	TensorFlow、PyTorch、CUDNN 等	TensorFlow、PyTorch、Mlopen 等
开发者体验	NVIDIA 提供丰富的文档和案例	缺乏详细的指导
行业采用	远超 ROMc 和 OpenCL 的份额	相对较少

资料来源：CSDN 等、浙商证券研究所整理

3.2 注重技术复用性，让研发投入落到实处

决策高效干脆，放弃手机芯片市场。2008 年公司依靠平板和游戏机的优势推出了针对移动端的 Tegra，但由于高通凭借基带技术占据主流，而 Tegra 后续芯片未能及时整合基带技术而无法及时占领市场，公司因而错失了移动时代机遇，此后公司果断放弃手机市场并将 Tegra 处理器运用在智能汽车、智慧城市和云端服务上。

图46： 英伟达 Tegra 芯片

资料来源：英伟达、浙商证券研究所

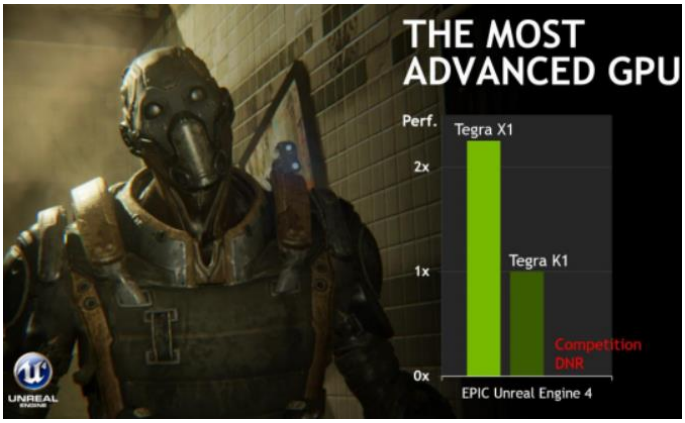
图47： 同时搭载高通芯片与 Tegra 的小米 3

资料来源：小米、浙商证券研究所

技术复用，开辟汽车芯片市场。2015 年国际消费类电子产品展览会上，英伟达发布新一代移动超级芯片 Tegra X1 处理器，该处理器在性能上是上一代产品 Tegra K1 的两倍。该处理器适配 NVIDIA DRIVE PX 汽车自动驾驶计算平台，可处理来自 12 个车载摄像头的视频，以实现环绕视觉（Surround-Vision）和自动代客泊车（Auto-Valet）等功能。虽然英伟达在移动芯片领域折戟沉沙，但其技术上极强的复用性，最终使其成功转向至汽车芯片领域。

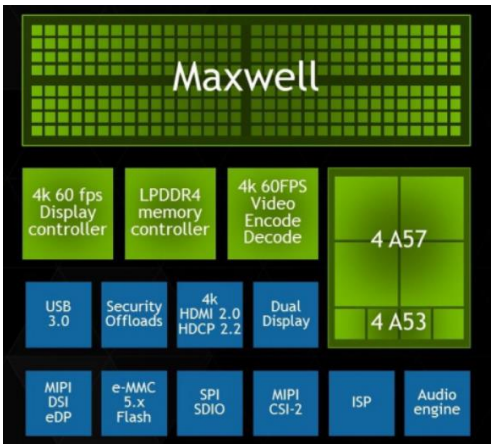
更多免费行业报告 并购家 www.ipoipo.cn

图48： Tegra X1 相较前一代的性能巨大提升



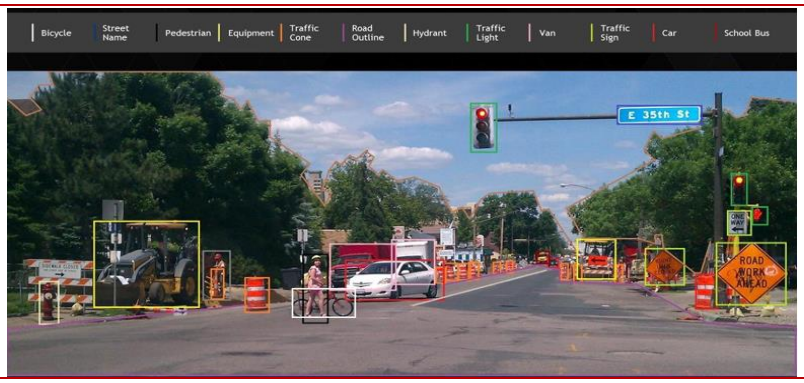
资料来源：英伟达、浙商证券研究所

图49： Tegra X1 采用 Maxwell 架构



资料来源：英伟达、浙商证券研究所

图50： Tegra X1 基于深度学习可以识别现实世界的事物



资料来源：英伟达、浙商证券研究所

3.3 重视通用性，引入生态共建

GPGPU：通用化漫漫长路。GPU 一开始只是为了图形加速，替 CPU 分担工作。其主要用于显示图像、视频的编解码与显示、游戏渲染等工作，起初并不支持编程。在 1999 年之前甚至没有 GPU 这一概念，只是称为图形加速卡。英伟达发现了 GPU 在并行计算方面的优势，并且在科学、工程和人工智能领域，许多问题都可以通过并行计算来加速解决。

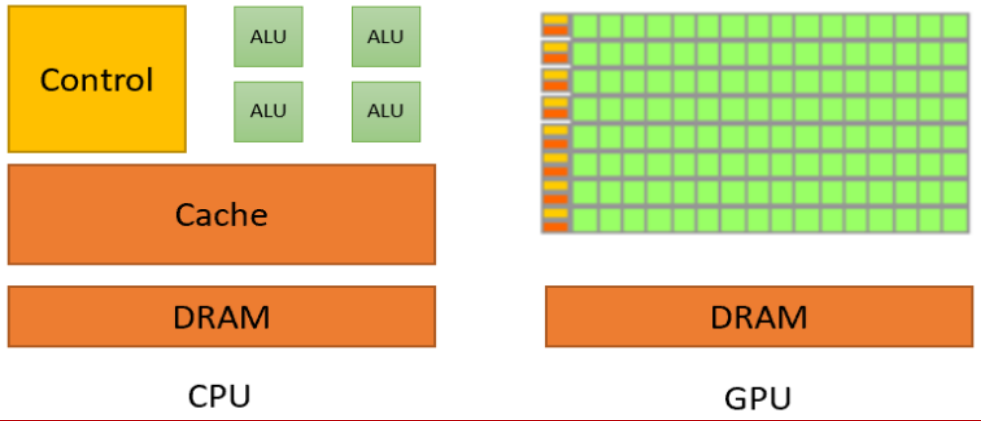
表10： GPU VS CPU

	GPU	CPU
核心数量	数百到数千个核心，更强的并行性	通常几个到十几个
核心类型	专用型	通用型
指令执行方式	并行执行	顺序执行
内存访问速度	相对较快	相对较慢
应用示例	图形渲染、深度学习、科学模拟	数据库管理、Web 服务器、桌面应用
编程模型	CUDA、OpenCL、ROMc 等专用语言	Python、C++等通用编程语言
功耗与效率	相对低功耗，适合高性能计算与深度学习	相对高功耗，适用于多种用途
并行性	高度并行	相对有限

资料来源：英伟达、浙商证券研究所整理

更多免费行业报告 并购家 www.ipoipo.cn

图51： CPU 与 GPU 架构的根本区别



资料来源：英伟达、浙商证券研究所

CUDA 出现之前，GPU 的编程面临多重挑战：

1）编程模型不足：没有一种通用的编程模型来利用 **GPU** 的计算能力。开发人员需要使用低级别的图形 API（如 OpenGL 或 Direct3D）来执行计算任务。这些 API 并不专门用于通用计算，因此编写代码变得复杂且容易出错。

2）数据传输成本高：将数据从 CPU 传输到 **GPU** 的成本很高。这涉及到数据的复制**和传输，而这些操作会降低性能。**

3）无法发挥并行性：开发人员需要手动管理 GPU 并行性，这对于复杂的计算任务来说是巨大的人力成本。

降低开发者门槛，让 GPU 真正意义上“可编程”。在经 GPU 加速的应用中，工作负载的串行部分在 CPU 上运行，且 CPU 已针对单线程性能进行优化，而应用的计算密集型部分则以并行方式在数千个 GPU 核心上运行。**使用 CUDA 时，开发者使用主流语言（如 C、C++、Fortran、Python 和 MATLAB）进行编程，并通过扩展程序以几个基本关键字的形式来表示并行性。**

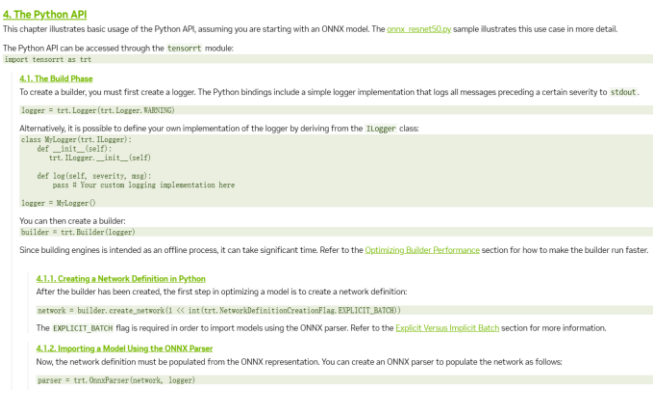
英伟达的 CUDA 工具包提供了开发 GPU 加速应用所需的一切。TensorRT 基于 CUDA 平台并行编程模型构建，使开发者能够在英伟达 GPU 产品上使用量化、层和张量融合、内核调整等技术来优化推理。TensorRT 是英伟达基于 CUDA 一款用于高性能深度学习推理的 SDK（Software Development Kit），使用门槛相对较低，可以运用 C++ 、Python API 导入和加速模型。

更多免费行业报告 并购家 www.ipoipo.cn

图52： 支持 C++ API 接入

资料来源：英伟达、浙商证券研究所

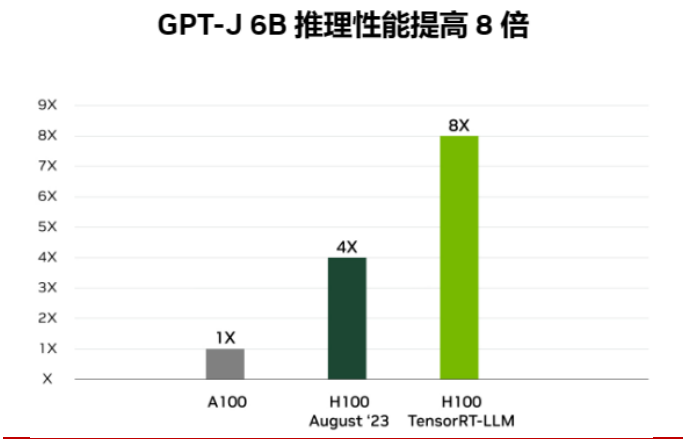
图53： 支持 Python API 接入



资料来源：英伟达、浙商证券研究所

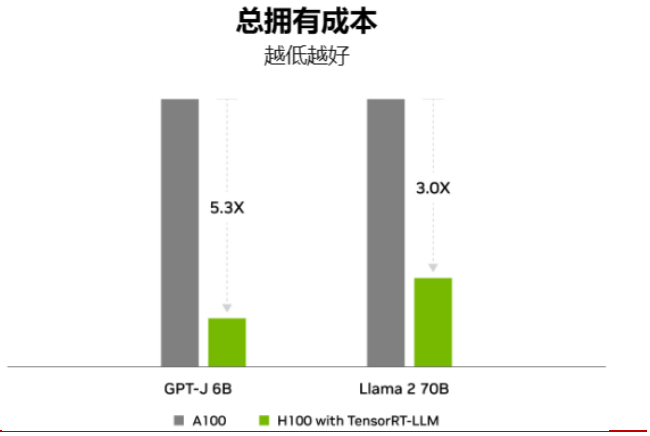
TensorRT 在低门槛开发的同时，能够最大化挖掘 GPU 性能。NVIDIA TensorRT-LLM 是一个开源库，可加速和优化 NVIDIA AI 平台上最新大型语言模型 (LLM) 的推理性能，而无需深入了解 C++ 或 CUDA。

图54： TensorRT 最大化 GPU 在 AI 推理中的性能



资料来源：英伟达、浙商证券研究所

图55： TensorRT 最小化客户总拥有成本



资料来源：英伟达、浙商证券研究所

CUDA 平台允许开发者利用英伟达的 GPU 来加速计算密集型任务。在全球范围内，许多行业领军者采用 CUDA 平台最大化其 GPU 性能，

图56： CUDA 平台部分生态伙伴

资料来源：英伟达、浙商证券研究所

我们认为，正是公司 CUDA 低门槛的特性、GPU 过硬的性能，引入了大批开发者建设 CUDA 生态社区，最终 CUDA 绑定了数百万 AI 开发者，当 CUDA 几乎与 AI 画等号

更多免费行业报告 并购家 www.ipoipo.cn

的时候，会有大量的社区力量为其助力。这就是一种良性循环：好的性能带来好的生态，好的生态会有助于更好的性能。最终帮助英伟达构建了强大的 CUDA 生态护城河。

图57： CUDA 生态部分伙伴

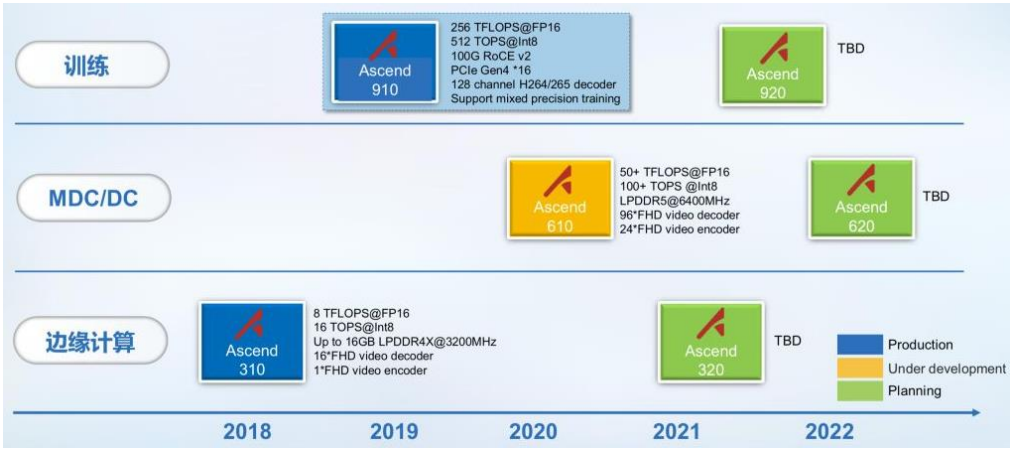
资料来源：英伟达、浙商证券研究所

4 国内相关公司：逐步追赶，国产化趋势已现

4.1 华为昇腾：对标英伟达，有望成为国内第二 AI 算力生态

昇腾处理器支持全场景。昇腾处理器是全球首个覆盖全场景 AI 芯片，基于统一的达芬奇架构，可以支持端边云不同场景的差异化算力需求，并具备从几十毫瓦 IP 到几百瓦芯片的平滑扩展，覆盖了端边云全场景部署的能力：

图58： 昇腾芯片迭代计划



资料来源：华为、新智元、199it、科大讯飞、IDC、AMD、英伟达、海光信息招股书等、浙商证券研究所

昇腾 910 训练处理器具有超高算力，FP16 下性能最高可达 320TFLOPS。昇腾 910 集成了 CPU Core、DVPP 和任务调度器(Task Scheduler)，可以减少和 Host CPU 的交互，充分发挥其高算力的优势；还集成了 HCCS、PCIe 4.0 和 ROCE v2 接口，为构建横向扩展(Scale Out)和纵向扩展 (Scale Up)系统提供了灵活高效的方法，科大讯飞创始人、董事长刘庆峰表示华为的 GPU 能力可以对标英伟达 A100;

更多免费行业报告 并购家 www.ipoipo.cn

表11： 主流 AI 芯片性能比较

		AMD	英伟达			华为昇腾		海光信息	寒武纪
		MI300X	L40S	A100 SXM	H100 SXM	昇腾 310	昇腾 910	DCU	思远 370
算力指标	FP64	47.9T	-	9.7T	34T	-	-	11.5T	-
	FP32	47.9T	183T	19.5T	67T	-	-	-	24T
	FP16	383T	362.05T	312T	989.5T	11T	320T	-	96T
	INT8	-	733T	624T	1979T	22T	640T	-	256T
内存容量		192GB	48GB	80GB	80GB	-	-	32GB	24GB
内存带宽		5.05TB/S	864GB/S	1.99TB/S	3.35TB/S	-	-	1TGB/S	307.2GB/S
功耗		600W	350W	400W	700W	8W	310W	260-350W	150W

资料来源：华为、新智元、199it、科大讯飞、IDC、AMD、英伟达、海光信息招股书等、浙商证券研究所整理

华为昇腾 AI 产业生态包括昇腾 AI 基础软硬件平台，即 Atlas 系列硬件、异构计算架构 CANN、全场景 AI 框架昇思 MindSpore、昇腾应用使能 MindX 以及一站式开发平台 ModelArts 等。基于昇腾 910 系列板卡，华为推出了 AI 训练集群 Atlas900、AI 训练服务器 Atlas800、智能小站 Atlas500、AI 推理与训练卡 Atlas300 和 AI 加速模块 Atlas200，完成了 Atlas 全系列产品布局，支持万亿参数大模型训练，同时覆盖云、边、端全场景。

华为提出了具备分层开放、体系协同、敏捷高效、安全可信等特征的，全行业通用的行业智能化参考架构。其中智能底座提供大规模 AI 算力、海量存储及并行计算框架，支撑大模型训练，提升训练效率，提供高性能的存算网协同。根据场景需求不同，提供系列化的算力能力。适应不同场景，提供系列化、分层、友好的开放能力。另外，智能底座层还包含品类多样的边缘计算设备，支撑边缘推理和数据分析等业务场景。

图59： 华为昇腾 AI 产业生态

资料来源：鲲鹏社区、昇腾社区、浙商证券研究所

图60： 华为行业智能化参考框架

资料来源：鲲鹏社区、昇腾社区、浙商证券研究所

4.2 海光信息：类 CUDA 带来更好的生态兼容性

海光 DCU 属于 GPGPU 的一种。性能上，海光深算一号 DCU 内核频率、显存位宽已逐步接近英伟达 A100，在显存容量、带宽、算力、互联性能上仍有一定的进步空间；深算二号已于 2023 年 Q3 发布，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。

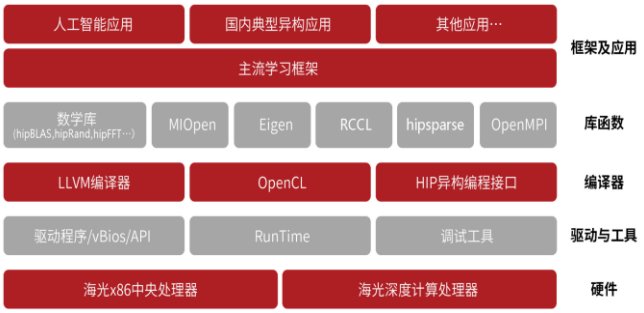
更多免费行业报告 并购家 www.ipoipo.cn

在 AIGC 持续快速发展的时代背景下，海光 DCU 能够完整支持大模型训练，实现 LLaMa、GPT、Bloom、ChatGLM、悟道、紫东太初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。

图61： 海光 DCU 组成部分

资料来源：海光招股书、海光官网、浙商证券研究所

图62： 海光软件栈体系



资料来源：海光招股书、海光官网、浙商证券研究所

DCU 产品具备完善的软件栈支持。海光 DCU 协处理器全面兼容 ROCm GPU 计算生态，由于 ROCm 和 CUDA 在生态、编程环境等方面具有高度的相似性，CUDA 用户可以以较低代价快速迁移至 ROCm 平台，因此 ROCm 也被称为“类 CUDA”。因此，海光 DCU 协处理器能够较好地适配、适应国际主流商业计算软件和人工智能软件，软硬件生态丰富，可广泛应用于大数据处理、人工智能、商业计算等计算密集类应用领域，主要部署在服务器集群或数据中心，为应用程序提供高性能、高能效比的算力，支撑高复杂度和高吞吐量的数据处理任务。

海光 DCU 具备开放式生态和统一底层硬件驱动平台，支持常见计算框架、库和编程模型层次化软件栈，适配不同 API 接口和编译器可最大限度利用已有的成熟 AI 算法和框架。

4.3 寒武纪：云端芯片性能持续扩展，云边终端协同覆盖

云端 AI 推理，高能效比国产芯片。思元 270 集成了寒武纪在处理器架构领域的一系列创新性技术，处理非稀疏人工智能模型的理论峰值性能提升至上一代思元 100 的 4 倍，达到 128TOPS(INT8)；同时兼容 INT4 和 INT16 运算，理论峰值分别达到 256TOPS 和 64TOPS；支持浮点运算和混合精度运算。思元 270 采用寒武纪 MLUv02 架构，可支持视觉、语音、自然语言处理以及传统机器学习等多样化的人工智能应用，更为视觉应用集成了充裕的视频和图像编解码硬件单元。

推理卡研发迭代，性能持续提升。思元 370 是寒武纪首款采用 chiplet（芯粒）技术的 AI 芯片，集成了 390 亿个晶体管，最大算力高达 256TOPS(INT8)，是寒武纪第二代产品思元 270 算力的 2 倍。凭借寒武纪最新智能芯片架构 MLUarch03，思元 370 实测性能表现更为优秀。思元 370 也是国内第一款公开发布支持 LPDDR5 内存的云端 AI 芯片，内存带宽是上一代产品的 3 倍，访存能效达 GDDR6 的 1.5 倍。搭载 MLU-Link™多芯互联技术，在分布式训练或推理任务中为多颗思元 370 芯片提供高效协同能力。全新升级的寒武纪基础软件平台，新增推理加速引擎 MagicMind，实现训推一体，大幅提升了开发部署的效率，降低用户的学习成本、开发成本和运营成本。

更多免费行业报告 并购家 www.ipoipo.cn

先进工艺保障芯片制造。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，并且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。

表12： 寒武纪产品线

产品线	产品类型	寒武纪主要产品	推出时间
云端产品线	云端智能芯片及加速卡	思元 100（MLU100）芯片及云端智能加速卡	2018 年
		思元 270（MLU270）芯片及云端智能加速卡	2019 年
		思元 290（MLU290）芯片及云端智能加速卡	2020 年
		思元 370（MLU370）芯片及云端智能加速卡	2021 年、2022 年
	训练整机	玄思 1000 智能加速器	2020 年
		玄思 1001 智能加速器	2022 年
边缘产品线	边缘智能芯片及加速卡	思元 220（MLU220）芯片及边缘智能加速卡	2019 年
IP 授权及软件	终端智能处理器 IP	寒武纪 1A 处理器	2016 年
		寒武纪 1H 处理器	2017 年
		寒武纪 1M 处理器	2018 年
	基础系统软件平台	寒武纪基础软件开发平台（适用于公司所有芯片与处理器产品）	持续研发升级

资料来源：寒武纪 2022 年年报、浙商证券研究所

4.4 景嘉微：发布景宏系列高性能计算产品，商业化布局有望加速

公司是国内首家成功研制国产 GPU 芯片并实现大规模工程应用的企业，掌握了包括芯片底层逻辑/物理设计、超大规模电路集成验证、模拟接口设计、GPU 驱动程序设计等关键技术，2010 年，公司自研 GPU 项目获得国家“核高基”专项支持。2014 年研发 JM5400，2018 年研发 JM7200。公司在 GPU 体系结构、图形绘制高效处理算法、高速浮点运算器设计、可复用模块设计、快速大容量存储器接口设计、低功耗设计等方面有深厚的技术积累，先后研制成功 JM5 系列、JM7 系列、JM9 系列等具有自主知识产权的高性能 GPU 芯片。

2024 年 3 月，公司披露景宏系列产品研发进展。景宏系列是公司推出的面向 AI 训练、AI 推理、科学计算等应用领域的高性能智算模块及整机产品，支持 INT8、FP16、FP32、FP64 等混合精度运算，支持全新的多卡互联技术进行算力扩展，适配国内外主流 CPU、操作系统及服务器厂商，能够支持当前主流的计算生态、深度学习框架和算法模型库，大幅缩短用户适配验证周期。

5 风险提示

1）历史成长轨迹与股价表现并不能代表未来趋势：我们在报告中对英伟达的历史业务发展以及股价情况进行了复盘分析，但由于公司所处的宏观环境、产业趋势以及公司自身业务不断变化，历史经验并不能代表未来趋势；

2）上游供应不足的风险：目前公司的主要芯片已接近光刻掩膜版尺寸极限，GPU 面积、制程均达到瓶颈，并且 cowos 产能短期内供给仍显不足，因此未来供应仍是制约公司业绩的重要因素；

更多免费行业报告 并购家 www.ipoipo.cn

3）技术迭代难度加大的风险：技术进步难度往往随着时间不断加大，如果公司核心技术和产品的迭代速度变慢，则可能会导致公司的产品竞争力减弱，进而影响未来业绩表现；

4）中游竞争加剧的风险：生成式 AI 需要海量充沛、低成本的算力供应已成为共识，AMD、Intel、华为等芯片厂商以及 CSP 云客户均在更新自研芯片，公司不可避免地会面对未来各个层面的竞争，如果竞争加剧导致公司份额丢失，则可能会影响公司未来的业绩及估值水平；

5）下游客户需求变化的风险：生成式 AI 处于快速成长期，客户需求尚不明确且变化较大，公司若无法快速满足客户需求，则可能会在未来的竞争中落败；

6）其他导致生态壁垒削弱的风险：公司核心的护城河为强大的生态体系，目前 AMD、华为等都在自建生态，可能会削弱公司的生态优势；

7）宏观环境变化的风险：公司作为全球重要标的，不可避免的会受到美国和全球宏观经济环境、政策环境等因素的影响。

更多免费行业报告 并购家 www.ipoipo.cn

股票投资评级说明

以报告日后的 6 个月内，证券相对于沪深 300 指数的涨跌幅为标准，定义如下：
1.买　入 ： 相对于沪深 300 指数表现 + 20％以上；
2.增　持 ： 相对于沪深 300 指数表现 + 10％ ~ + 20％；
3.中　性 ： 相对于沪深 300 指数表现 - 10％ ~ + 10％之间波动；
4.减　持 ： 相对于沪深 300 指数表现 - 10％以下。

行业的投资评级：

以报告日后的 6 个月内，行业指数相对于沪深 300 指数的涨跌幅为标准，定义如下：
1.看　好 ： 行业指数相对于沪深 300 指数表现 + 10%以上；
2.中　性 ： 行业指数相对于沪深 300 指数表现 - 10% ~ + 10%以上；
3.看　淡 ： 行业指数相对于沪深 300 指数表现 - 10%以下。
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取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者不应仅仅依靠投资评级来推断结论。

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。
本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。
本报告仅反映报告作者的出具日的观点和判断，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。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。
本公司的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。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、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。
本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。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。

浙商证券研究所

上海总部地址：杨高南路 729 号陆家嘴世纪金融广场 1 号楼 25 层
北京地址：北京市东城区朝阳门北大街 8 号富华大厦 E 座 4 层
深圳地址：广东省深圳市福田区广电金融中心 33 层
上海总部邮政编码：200127
上海总部电话：(8621) 80108518
上海总部传真：(8621) 80106010
浙商证券研究所：https://www.stocke.com.cn

更多免费行业报告 并购家 www.ipoipo.cn