

电子AI+系列专题报告（二） 复盘英伟达的AI发展之路

行业研究 · 行业专题

电子

投资评级：超配（维持评级）

证券分析师：胡剑

021-60893306

hujian1@guosen.com.cn

S0980521080001

证券分析师：胡慧

021-60871321

huhui2@guosen.com.cn

S0980521080002

证券分析师：周靖翔

021-60375402

zhoujingxiang@guosen.com.cn

S0980522100001

证券分析师：李梓澎

0755-81981181

lizipeng@guosen.com.cn

S0980522090001

联系人：詹浏洋

010-88005307

zhanliuyang@guosen.com.cn

- GPU是人工智能时代下满足深度学习大量计算需求的核心AI芯片。

过去五年，大型语言模型的参数规模以指数级增长；从2018年起，OpenAI开始发布生成式预训练语言模型GPT以来，GPT更新换代持续提升模型及参数规模；2022年12月，OpenAI发布基于GPT-3.5的聊天机器人模型ChatGPT，参数量达到1750亿个。ChatGPT引领全球人工智能浪潮，人工智能发展需要AI芯片作为算力支撑。据Tractica数据，全球AI芯片市场规模预计由2018年的51亿美元增长至2025年的726亿美元，对应CAGR达46.14%；据前瞻产业研究院数据，中国AI芯片市场规模预计由2019年的122亿元增长至2024年的785亿元，对应CAGR达45.11%。AI芯片中由于GPU通用型强，满足深度学习大量计算的需求，因此GPU在训练负载中具有绝对优势。以GPT-3为例，在32位的单精度浮点数数据下，训练阶段所需GPU数量为1558个，谷歌级应用推理阶段所需GPU数量为706315个。

- 英伟达开辟GPGPU加速计算格局，GPU架构演进及产品布局赋能AI时代。

英伟达(NVIDIA)成立于1993年，总部位于美国加利福尼亚州圣克拉拉，是一家人工智能计算公司。据JPR数据，4Q22英伟达独立GPU出货量占比为82%，位居市场第一。公司股价经历2016-2018年、2020-2021年、2022年9月以来三轮快速增长；其中2022年9月至今，受AI驱动下针对芯片算力需求提升，公司股价呈现大幅度反弹；截至2023年4月30日，公司市值为6854.00亿美元。回顾历史，1999年，公司发明了图形处理器，定义了现代计算机图形学；2006年，公司推出用于通用GPU(GPGPU)计算的CUDA平台。自2015年以后，随着AI浪潮迅猛推进，公司业务不断多元化，向数据中心、游戏、移动设备、汽车电子等市场发展。公司GPU产品能够并行计算的性能优势满足深度学习需求，通过对GPU架构升级不断推出新产品，其运算性能得到显著提升，广泛用于数据中心等计算密集领域。

- 构建加速计算平台、完善“三芯片”产品布局，重点发力AI及数据中心领域。

英伟达业务模式拟打造成类似于计算堆栈或神经网络，包含硬件、系统软件、平台软件和应用四层，公司结合芯片、系统和软件的全栈创新能力构建加速计算平台，并且完善针对AI加速计算及数据中心的GPU、CPU、DPU三种芯片产品结构。AI布局方面，早在生成式AI变革初期就已参与并与OpenAI、微软合作。2023年3月，英伟达在GTC大会上推出4个针对各种生成式AI应用程序进行优化的推理平台，其中发布带有双GPU NVLink的H100 NVL加速计算卡，以支持ChatGPT类大型语言模型推理。与适用于GPT-3处理的HGX A100相比，配备四对H100与双GPU NVLink的标准服务器的速度最高可达10倍。

- 计算与网络事业部收入占比提升，数据中心超过游戏成为收入主要来源。

FY23，公司收入为269.74亿美元，同比增长0.22%，主要由于数据中心及汽车领域收入的高速增长抵消了游戏及专业可视化领域所带来的收入减少；净利润为43.68亿美元，同比减少55.21%，主要由于运营费用同比增长50%所致，包括与Arm交易相关的13.5亿美元收购终止费用。公司业务部门包括计算与网络事业部和图形事业部，FY23收入占比分别为55.86%、44.14%。公司的平台及产品主要应用于数据中心、游戏、专业可视化、汽车等四大领域，FY23占比分别为55.63%、33.61%、5.72%、1.69%。分地区来看，FY23来自美国地区的收入占比最高，达30.7%。公司FY1Q24实现收入71.92亿美元（YoY -13.2%，QoQ 18.9%），FY2Q24收入指引为107.8-112.2亿美元（YoY 60.8%至67.4%，QoQ 49.9%至56.0%）。

- **产业链相关公司：**算力：英伟达、海光信息、寒武纪、全志科技；服务器：工业富联、国芯科技、环旭电子、闻泰科技、易德龙；PCB：沪电股份、胜宏科技、东山精密、鹏鼎控股；AI终端：晶晨股份、瑞芯微；先进封装：长电科技、通富微电、芯原股份；存储：深科技、江波龙、佰维存储、兆易创新。

- **风险提示：**宏观AI推广不及预期，AI投资规模低于预期，AI服务器渗透率提升低于预期，AI监管政策收紧等。

- [01] GPU与人工智能
- [02] 英伟达的AI发展之路
- [03] 英伟达业绩实现与拆解
- [04] 风险提示

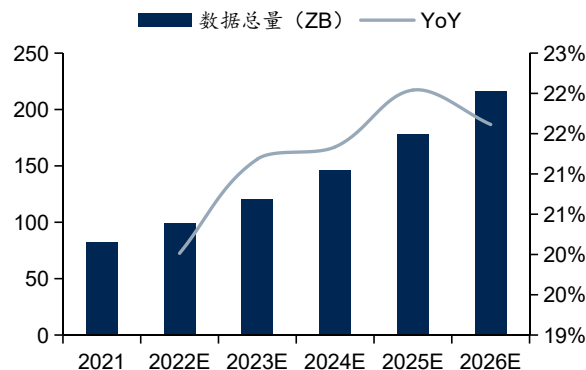
一、GPU与人工智能

全球数据量及数据中心负载量大幅上涨，AI模型参数呈现指数级增长

●全球数据总量及数据中心负载任务量大幅上涨，数据中心算力需求快速增长。随着人工智能等新技术发展，海量数据的产生及其计算处理成为数据中心发展关键。

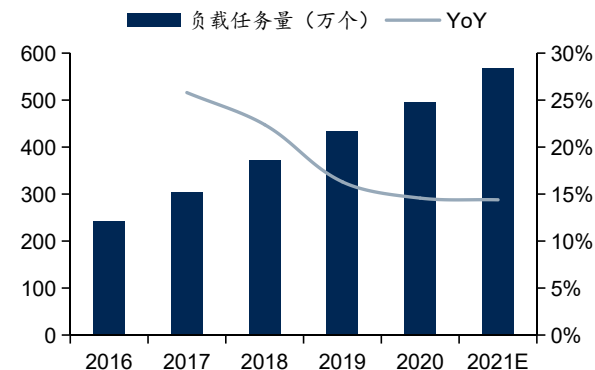
- 据IDC数据，全球数据总量预计由2021年的82.47 ZB上升至2026年的215.99 ZB，对应CAGR达21.24%。其中，大规模张量运算、矩阵运算是人工智能在计算层面的突出需求，高并行度的深度学习算法在视觉、语音和自然语言处理等领域上的广泛应用使得算力需求呈现指数级增长。
- 据Cisco数据，全球数据中心负载任务量预计由2016年的241.5万个上升至2021年的566.7万个，对应CAGR达18.60%；其中，云数据中心负载任务量CAGR预计达22%。

图：2021-2026年全球数据总量及预测



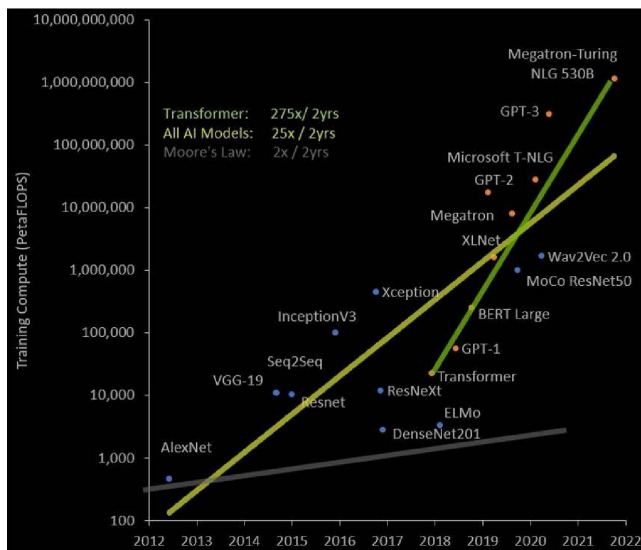
资料来源：IDC，国信证券经济研究所整理

图：2016-2021年全球数据中心负载任务量及预测



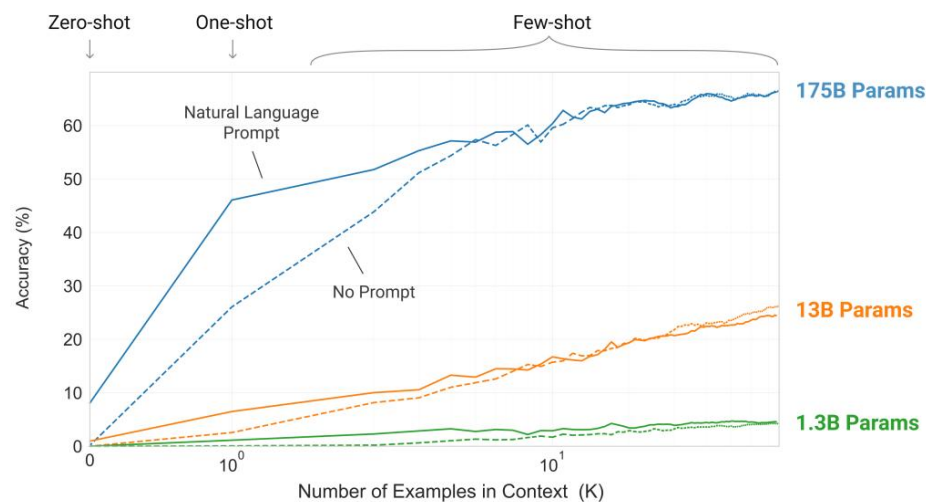
资料来源：Cisco Global Cloud Index，国信证券经济研究所整理

图：模型参数规模大幅上涨带来算力需求提升



资料来源：英伟达官网，国信证券经济研究所整理

图：简单任务中不同参数模型上下文学习性能



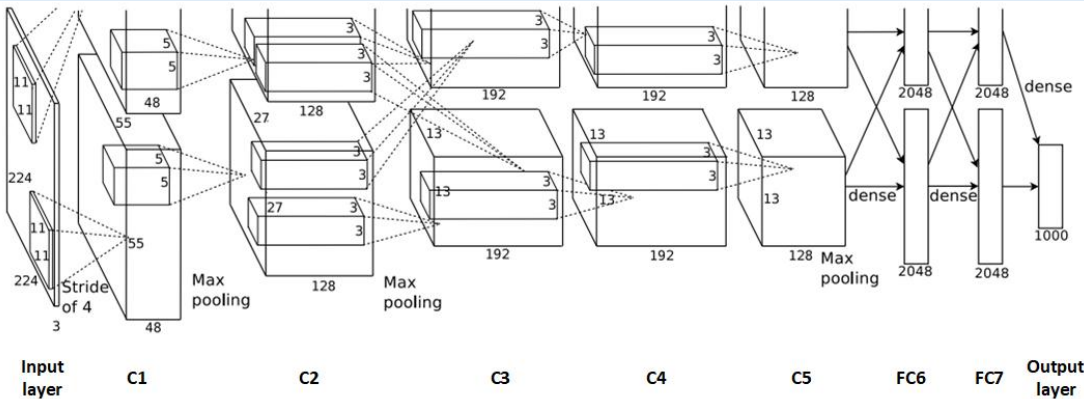
资料来源：《Language Models are Few-Shot Learners》，国信证券经济研究所整理

●以模型中的参数数量衡量，大型语言模型的参数在过去五年中以指数级增长。随着参数量和训练数据量的增大，语言模型的能力会随着参数量的指数增长而线性增长，这种现象被称为Scaling Law。但当模型的参数量大于一定程度的时候，模型能力会突然暴涨，模型会突然拥有一些突变能力（Emergent Ability），如推理能力、无标注学习能力等。例如GPT之前的大语言模型主流是深度神经网络驱动，参数在数十亿水平，而ChatGPT达到1750亿参数。

神经网络AlexNet使用GPU训练开启AI时代，AI芯片是算力的重要支撑

- 随着互联网时代对于数据量的积累，大数据背景下神经网络成为机器学习的重要方法。2012年，深度卷积神经网络AlexNet凭借在图像分类识别领域中性能的大幅提升及错误率的大幅降低，成为人工智能的标志性事件。
- 在此过程中，其训练者Alex Krizhevsky创新性地使用英伟达GPU成功训练了性能有突破性提升的深度神经网络AlexNet，从而开启了新的人工智能时代。英伟达GPU伴随着深度学习模型训练和推理所需要的大量算力成为了人工智能时代的新基础设施。

图：AlexNet网络结构示意图



资料来源：AI有道，国信证券经济研究所整理

- ChatGPT引领全球人工智能浪潮，人工智能发展需要AI芯片作为算力支撑。从2018年起，OpenAI开始发布生成式预训练语言模型GPT以来，GPT更新换代持续提升模型及参数规模，当时GPT-1参数量只有1.17亿个。
- 2020年，OpenAI发布GPT-3预训练模型，参数量为1750亿个，使用1000亿个词汇的语料库进行训练，在文本分析、机器翻译、机器写作等自然语言处理应用领域表现出色。
- 2022年12月，OpenAI发布基于GPT-3.5的聊天机器人模型ChatGPT，具有出色的文字聊天和复杂语言处理能力。ChatGPT的发布引爆AI领域，海内外科技公司纷纷宣布发布大语言模型，而用户爆发式增长对大语言模型的算力需求同样带来挑战，AI芯片成为算力提升关键。

表：GPT-1至GPT-3模型参数

版本	GPT-1	GPT-2	GPT-3
时间	2018年6月	2019年2月	2020年5月
参数量	1.17亿	15.4亿	1750亿
预训练数据量	5GB	40GB	45TB
训练方式	Pre-training+Fine-tuning	Pre-training	Pre-training
序列长度	512	1024	2048
# of Decoder Layers	12	48	96
Size of Hidden Layers	768	1600	12288

资料来源：腾讯云开发者，国信证券经济研究所整理

表：国内外科技企业发语言模型发布情况

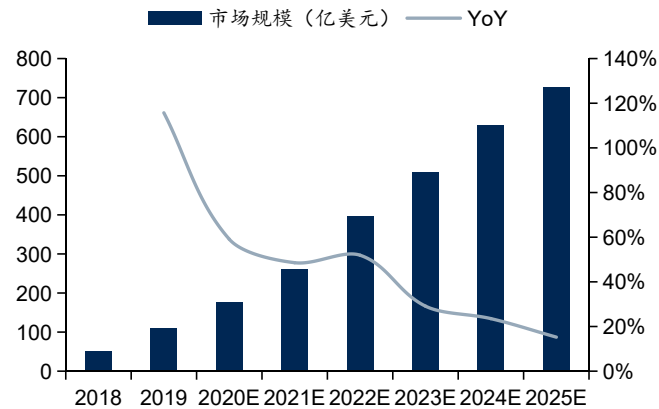
公司	产品	(拟)发布日期	阶段	链接
OpenAI	ChatGPT	2022年11月30日	开放注册	https://chat.openai.com/
Google	Bard	2月8日	公开测试	http://bard.google.com/
复旦大学	Moss	2月21日	公开测试(目前升级中)	https://moss.fastnlp.top/
澜舟科技	孟子	3月14日	已发布	https://www.langboat.com/portal/mengzi-model
百度	文心一言	3月16日	企业用户内测	https://yiyan.baidu.com/
达观数据	曹植	3月21日	可申请使用	http://www.datagrand.com/products/aigc/
清华大学	ChatGLM-6B	3月28日	已开源	https://github.com/THUDM/ChatGLM-6B
阿里巴巴	通义千问	4月7日	企业用户内测	https://tongyi.aliyun.com/
360	360智脑	4月10日	企业用户内测	http://www.360dmodel.com/
商汤科技	日日新	4月10日	即将邀请内测	https://www.sensecore.cn/
昆仑万维	天工3.5	4月17日	即将邀请内测	http://tiangong.kunlun.com/
科大讯飞	1+N认知智能大模型	5月6日	即将发布	-
网易有道	子曰	近期	即将发布	-
华为	盘古NLP模型	近期	即将发布	-
腾讯	混元助手	近期	未开放	-
京东	言犀	今年	未开放	-

资料来源：金十数据，国信证券经济研究所整理

全球及中国AI芯片市场规模不断扩展，GPU占比具有绝对优势

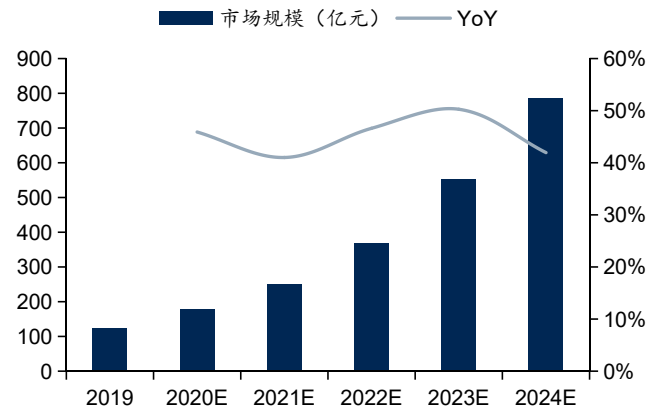
- AI芯片又称AI加速器或计算卡，是专门用于处理人工智能应用中大量计算任务的模块。随着数据海量增长、算法模型趋向复杂、处理对象异构、计算性能要求高，AI芯片能够在人工智能的算法和应用上做针对性设计，高效处理人工智能应用中日渐多样繁杂的计算任务。
- 随着人工智能技术进步及应用场景多元化，全球及中国AI芯片市场得到进一步发展。据Tractica数据，全球AI芯片市场规模预计由2018年的51亿美元增长至2025年的726亿美元，对应CAGR达46.14%。据前瞻产业研究院数据，中国AI芯片市场规模预计由2019年的122亿元增长至2024年的785亿元，对应CAGR达45.11%。

图：2018-2025年全球AI芯片市场规模及预测



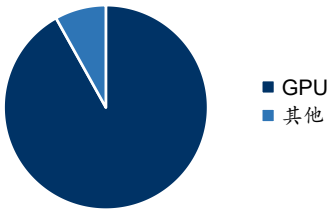
资料来源：Tractica，国信证券经济研究所整理

图：2019-2024年中国AI芯片市场规模及预测



资料来源：前瞻产业研究院，国信证券经济研究所整理

图：1H21中国AI芯片市场份额



资料来源：IDC，国信证券经济研究所整理

- 当前主流的AI芯片主要包括图形处理器（GPU）、现场可编程门阵列（FPGA）、专用集成电路（ASIC）、神经拟态芯片（NPU）等。其中，GPU、FPGA均是前期较为成熟的芯片架构，属于通用型芯片。ASIC属于为AI特定场景定制芯片。另外，中央处理器（CPU）是计算机的运算和控制核心，是信息处理、程序运行的最终执行单元，是计算机的核心组成部件。
- GPU在训练负载中具有绝对优势。据IDC数据，1H21中国AI芯片市场份额中，GPU占比高达91.9%，依然是实现数据中心加速的首选。GPU通用型较强、适合大规模并行运算，设计和制造工艺成熟，适用于高级复杂算法和通用性人工智能平台。

表：不同技术架构AI芯片比较

种类	定制化程度	可编辑性	算力	价格	优点	缺点	应用场景
GPU	通用型	不可编辑	中	高	通用型较强、适合大规模并行运算；设计和制造工艺成熟。	并行运算能力在推理段无法完全发挥。	高级复杂算法和通用性人工智能平台。
FPGA	半定制化	容易编辑	高	中	可通过编程灵活配置芯片架构适应算法迭代，平均性能较高；功耗较低；开发时间较短。	量产单价高；峰值计算能力较低；硬件编程困难。	适用于各种具体的行业。
ASIC	全定制化	难以编辑	高	低	通过算法固化实现极致的性能和能效、平均性很强；功耗很低；体积小；量产成本最低。	前期投入成本高；研发时间长；技术风险大。	当客户处在某个特殊场景，可以为其独立设计一套专业智能算法软件。

资料来源：亿欧智库，国信证券经济研究所整理

AI芯片包括GPU、FPGA、ASIC等，不同芯片定制化程度存在差异



●FPGA（Field Programmable Gate Array，现场可编程门阵列）是一种硬件可重构的集成电路芯片，通过编程定义单元配置和链接架构进行计算。FPGA具有较强的计算能力、较低的试错成本、足够的灵活性以及可编程能力，在5G通信、人工智能等具有较频繁的迭代升级周期、较大的技术不确定性的领域，是较为理想的解决方案。

●ASIC（Application Specific Integrated Circuit，专用集成电路）是一种根据产品的需求进行特定设计和制造的集成电路，能够更有针对性地进行硬件层次的优化。由于ASIC能够在特定功能上进行强化，因此具有更高的处理速度和更低的能耗。相比于其他AI芯片，ASIC设计和制造需要大量的资金、较长的研发周期和工程周期，在深度学习算法仍在快速发展的背景下存在一旦定制则难以修改的风险。

●GPU（Graphics Processing Unit，图形处理器）能够并行计算的性能优势满足深度学习需求。GPU最初承担图像计算任务，目标是提升计算机对图形、图像、视频等数据的处理性能，解决CPU在图形图像领域处理效率低的问题。由于GPU能够进行并行计算，其架构本身较为适合深度学习算法。因此，通过对GPU的优化，能够进一步满足深度学习大量计算的需求。

表：部分数据中心FPGA产品

公司名称	产品型号	发布时间	制程	等效 LUT4	分布式RAM (Kbit)	最大用户IO	DDR rate (Mbps)
安路科技	PH1A系列 PH1A180SFG676	2022	28nm	210240	3277	396	1866
紫光同创	Logos-2系列 PG2L200H	2020	28nm	239700	2528	500	1066

资料来源：安路科技、紫光同创官网，国信证券经济研究所整理

表：部分数据中心ASIC产品

公司名称	产品型号	发布时间	制程	单精度浮点运算性能 (TFLOPS)	半精度浮点运算性能 (TFLOPS)	整型定点运算性能 (TOPS)	显存 (GB)	显存带宽 (GB/s)	功耗 (W)
Google	TPU v4	2021	7nm	-	275	275	32	1200	192
平头哥	含光800	2021	12nm	-	-	825	-	-	276
寒武纪	思元370 MLU370-X8	2021	7nm	24	96	256	48	614.4	250
华为海思	昇腾910	2018	7nm	-	320	640	-	-	310

资料来源：Google、平头哥、寒武纪、华为海思官网，国信证券经济研究所整理

典型GPU架构以及GPU与CPU的异同

●典型GPU架构包括：

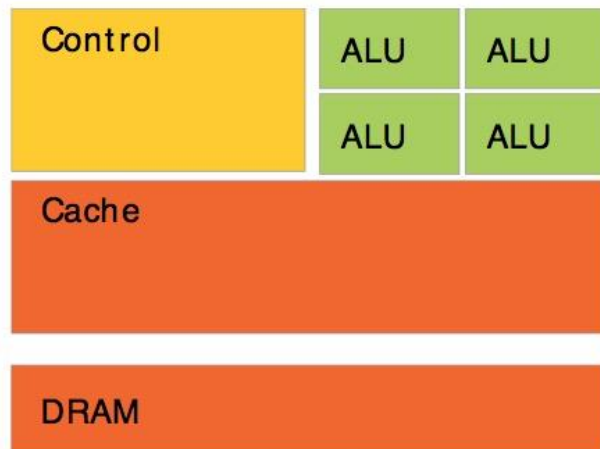
- 一个GPU由多个Processor Cluster组成
- 一个Processor Cluster由多个Streaming Multiprocessors组成
- 一个Streaming Multiprocessors里面可能包含多个Core，Streaming Multiprocessors中一定数量的Core共享一级缓存，多个Streaming Multiprocessors共享二级缓存

●CPU（Central Processing Unit，中央处理器）是计算机系统的运算和控制核心，是信息处理、程序运行的最终执行单元。CPU的结构主要包括运算器（ALU，Arithmetic and Logic Unit）、控制单元（CU，Control Unit）、寄存器（Register）、高速缓存器（Cache）和之间通讯的数据、控制及状态的总线。

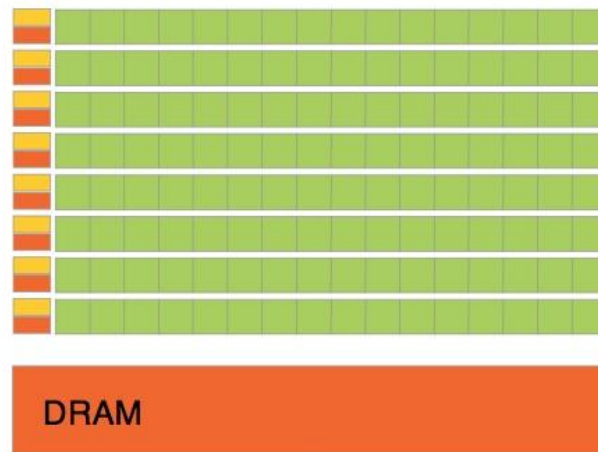
●相同点：CPU和GPU都是运算的处理器，在架构组成上都包括3个部分：运算单元ALU、控制单元Control和缓存单元Cache。

●不同点：CPU为低延时设计，擅长处理逻辑复杂、串行的计算任务。CPU需要很强的通用性来处理各种不同的数据类型，同时又要逻辑判断又会引入大量的分支跳转和中断的处理，因此CPU内部结构复杂，擅长逻辑控制和通用类型数据运算。GPU为高吞吐设计，为大规模数据并行计算任务量身定做。GPU面对的则是类型高度统一的、相互无依赖的大规模数据和相对纯净的计算环境。GPU采用数量众多的计算单元和超长流水线，擅长大规模并发运算。

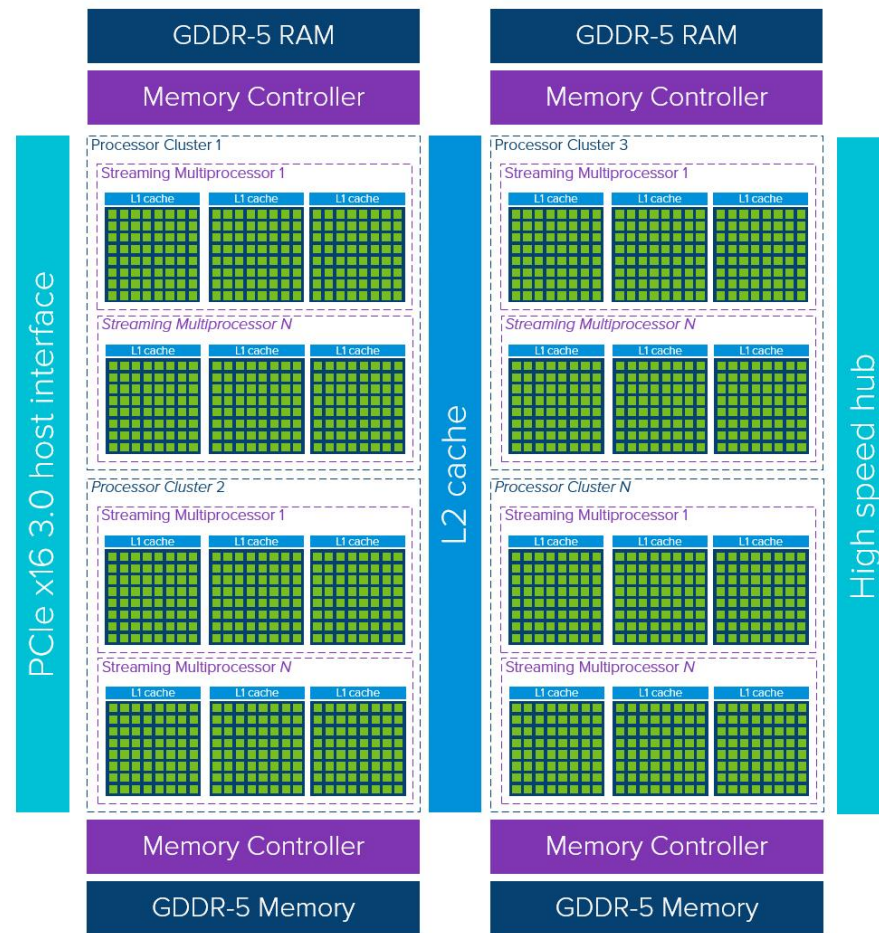
图：CPU基础架构示意图



图：GPU基础架构示意图



图：典型GPU架构示意图



AI大语言模型对于GPU算力需求测算（以GPT-3为例）

●《AI大语言模型的原理、演进及算力测算》

●**训练阶段：**考虑采用精度为32位的单精度浮点数数据进行训练和推理。以A100 PCIe芯片为例（H100 PCIe芯片同理），GPT-3训练所需运算次数为：样本token数3000亿个*6*参数量1750亿个=315*10²¹FLOPs；考虑训练时间要求在30天完成（训练时间为2592000秒），则对应GPT-3训练所需算力为121528TFLOPS；结合A100有效算力78TFLOPS，**得到所需GPU数量为1558个，对应AI服务器为195台。**

●**推理阶段：**按谷歌每日搜索量35亿次进行假设，估算得到GPT-3每日需推理token数为79330亿个，则推理所需运算次数为4760*10²¹FLOPs；考虑推理时间以每日为单位（推理时间为86400秒），则对应GPT-3推理所需算力为55*10⁶TFLOPS；结合A100有效算力78TFLOPS，**得到所需GPU数量为706315个，对应AI服务器为8.8万台。**

●根据上述结论，GPT-3新增GPU价值达到英伟达公司FY23收入的39.4%。
综合IDC及TrendForce数据估算，GPT-3新增AI服务器数量达到2021年全球AI服务器数量的65.35%。

表：训练阶段算力需求测算过程及结论

GPU相关		A100 PCIe	H100 PCIe
	Tensor Float 32(TF32)	156TFLOPS	756TFLOPS
	有效算力	78TFLOPS	378TFLOPS
	GPT-3训练所需运算次数	315*10 ²¹ FLOPs	315*10 ²¹ FLOPs
	GPT-3训练所需算力	121528TFLOPS	121528TFLOPS
	所需GPU数量	1558	322
	GPU单价	1.5万美元	3.65万美元
	对应GPU价值	2337万美元	1175.3万美元
AI服务器相关		DGX A100	DGX H100
	单个服务器对应GPU数量	8	8
	所需服务器数量	195台	40台

资料来源：英伟达官网，国信证券经济研究所整理及预测

表：推理阶段算力需求测算过程及结论

GPU相关		A100 PCIe	H100 PCIe
	Tensor Float 32(TF32)	156TFLOPS	756TFLOPS
	有效算力	78TFLOPS	378TFLOPS
	GPT-3推理所需运算次数	4760*10 ²¹ FLOPs	4760*10 ²¹ FLOPs
	GPT-3推理所需算力	55*10 ⁶ TFLOPS	55*10 ⁶ TFLOPS
	所需GPU数量	706315	145748
	GPU单价	1.5万美元	3.65万美元
	对应GPU价值	105.95亿美元	53.2亿美元
AI服务器相关		DGX A100	H100
	单个服务器对应GPU数量	8	8
	所需服务器数量	8.8万台	1.8万台

资料来源：英伟达官网，国信证券经济研究所整理及预测

二、英伟达的AI发展之路

英伟达是人工智能计算领域领导者，独立GPU出货量位居市场第一

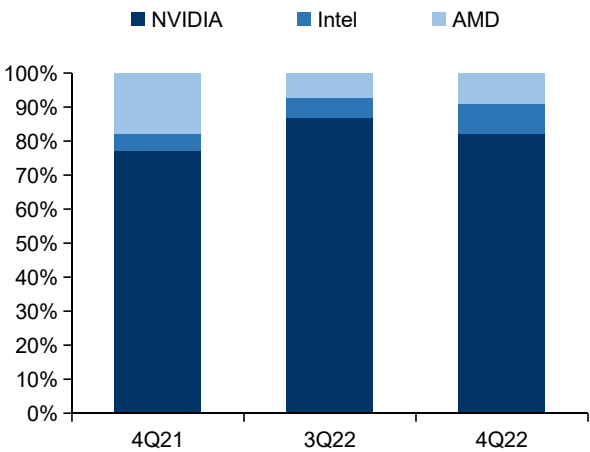
- 英伟达（NVIDIA）成立于1993年，总部位于美国加利福尼亚州圣克拉拉，是一家人工智能计算公司。公司作为加速计算的先驱，由最初专注于PC图形计算，扩展到了各类重要大型计算密集型领域。公司利用其GPU产品和架构为科学计算、人工智能（AI）、数据科学、自动驾驶汽车（AV）、机器人、元宇宙和3D互联网应用创建平台。
- 截至FY23（各财年截止日为1月最后一个周日，FY23截止日为2023年1月29日），公司在全球35个国家和地区拥有员工人数约2.62万人。截至2023年4月30日，公司市值为6854.00亿美元。

表：2022年全球半导体收入前十公司

排名	公司	2022年收入（亿美元）	市场份额	同比增速	地区
1	三星	670.6	11.3%	-10.8%	韩国
2	英特尔	608.1	10.2%	-20.6%	美国
3	高通	367.2	6.2%	25.2%	美国
4	海力士	341.0	5.7%	-7.3%	韩国
5	博通	269.6	4.5%	28.1%	美国
6	美光	268.7	4.5%	-7.4%	美国
7	AMD	237.8	4.0%	47.2%	美国
8	英伟达	210.5	3.5%	2.3%	美国
9	德州仪器	189.0	3.2%	9.9%	美国
10	联发科	185.2	3.1%	6.1%	中国台湾省

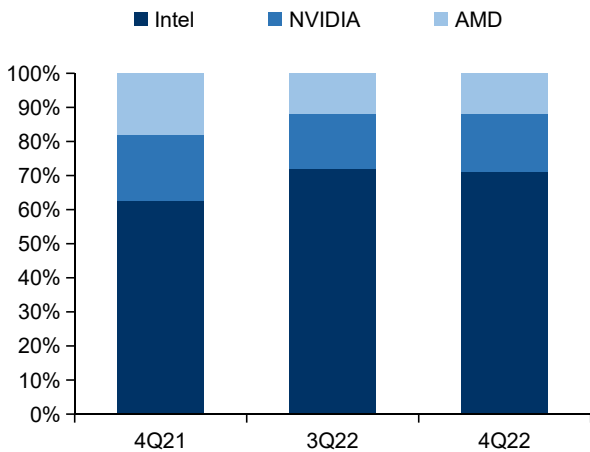
资料来源：Omdia，国信证券经济研究所整理

图：全球独立GPU市场出货量占比



资料来源：Jon Peddie Research，国信证券经济研究所整理

图：全球PC GPU市场出货量占比



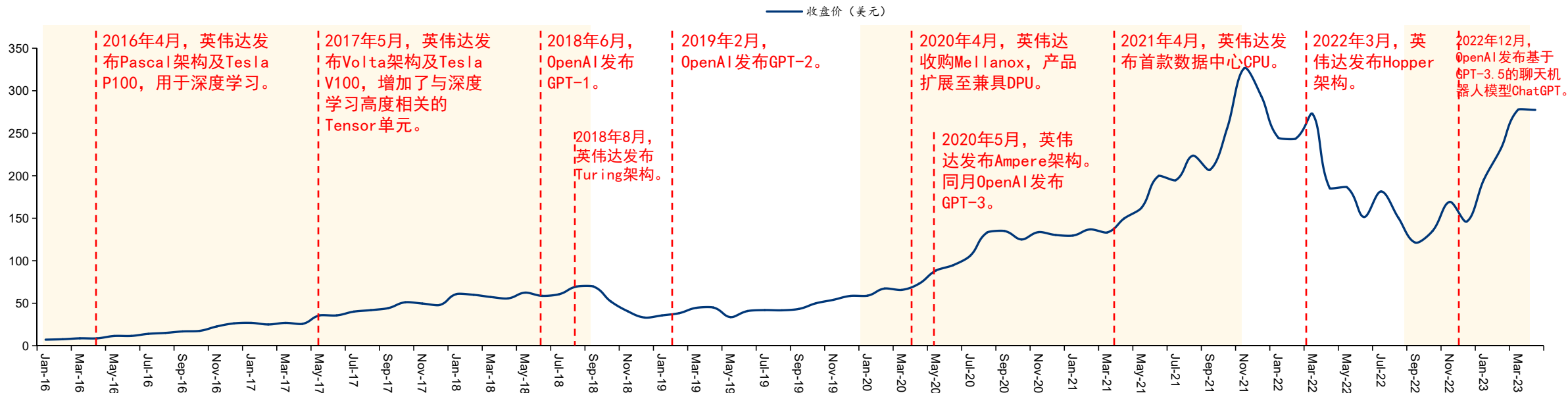
资料来源：Jon Peddie Research，国信证券经济研究所整理

- 据Omdia数据，在2022年全球半导体收入前十公司中，英伟达排名第八，市场份额约3.5%。
- 公司在全球独立GPU市场及数据中心市场加速芯片中处于领先地位。据Jon Peddie Research（JPR）数据，4Q22英伟达独立GPU出货量占比为82%，位居市场第一；PC GPU出货量占比为17%，仅次于全球最大的处理器厂商英特尔凭借其桌面端集成显卡优势占据的最大份额。

股价经历2016-2018年、2020-2021年、2022年9月以来三轮快速增长

- 2016-2018年，受益于AI及挖矿机的兴起带动算力需求增长，公司股价由2016年初的8.05美元上涨至2018年9月末的69.66美元。
- 2020-2021年，由于游戏及数据中心等领域业务需求的快速增长大幅提升公司业绩，公司股价呈现大幅增长，由2020年初的58.61美元快速上涨至2021年11月末的326.36美元。而后一段时间内，由于全球半导体产业需求衰退、库存压力等因素影响下，公司股价出现大幅度回调。
- 2022年9月至今，受AI驱动下针对芯片算力需求提升，公司股价呈现大幅度反弹，由2022年9月末的121.34美元上涨至2023年4月末的277.49美元。截至2023年4月30日，公司市值为6854.00亿美元。

图：公司股价变化情况



公司发展历史及收并购历史



●1993年，黄仁勋等人共同创立了英伟达。1994年，公司与SGS-Thomson合作，为其制造单芯片图形用户界面加速器。1995年，公司推出首款产品NV1；而1998年发布的RIVA TNT巩固了公司在开发功能强大的图形适配器方面的市场地位。此外，公司1998年与台积电签约建立战略合作关系，台积电开始协助制造公司产品。

●1999年，公司发明了图形处理器，全球首款GPU GeForce 256诞生。公司发明GPU定义了现代计算机图形学，由此走上了重塑行业的道路，确立了在该领域的领导地位。而此后，公司GPU产品出货量快速增长，2002年处理器出货量累计超1亿台，2006年累计出货超5亿台，2011年累计出货超10亿台。

- **2006年，公司推出用于通用GPU（GPGPU）计算的CUDA平台。** 软件开发者可以通过该平台使用C语言编写GPU片上程序来完成复杂的计算。从G80开始，英伟达GPU体系结构已经全面支持通用编程，GPU实际脱离图像处理的单一用途，成为了真正的通用GPU。

●自2015年以后，随着AI浪潮迅猛推进，公司业务不断多元化，向数据中心、游戏、移动设备、汽车电子等市场发展。2017年，公司专为数据中心和高性能计算打造Tesla V100 GPU，为DGX系列AI超级计算机提供支持。FY23，公司来自于数据中心业务领域的收入超过游戏领域业务，成为公司最大的收入来源。

表：公司历年收购史

年份	收购标的	标的简介
2000	3dfx	显卡技术先驱
2003	Media Q	无线领域图形和多媒体技术的领导者
2005	Uli Electronics	核心逻辑技术开发商
2006	Hybrid Graphics	手持设备嵌入式2D和3D图形软件开发商
2007	Portal Player	个人媒体播放器半导体、固件及软件供应商
2008	Mental images	视觉渲染软件公司
2008	AGEIA	游戏物理技术开发商
2013	Portland Group	加速计算开发者工具提供商
2020	Mellanox	高性能互联技术领域龙头
2022	Excelero	存储方案提供商

资料来源：公司官网，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：公司发展历史



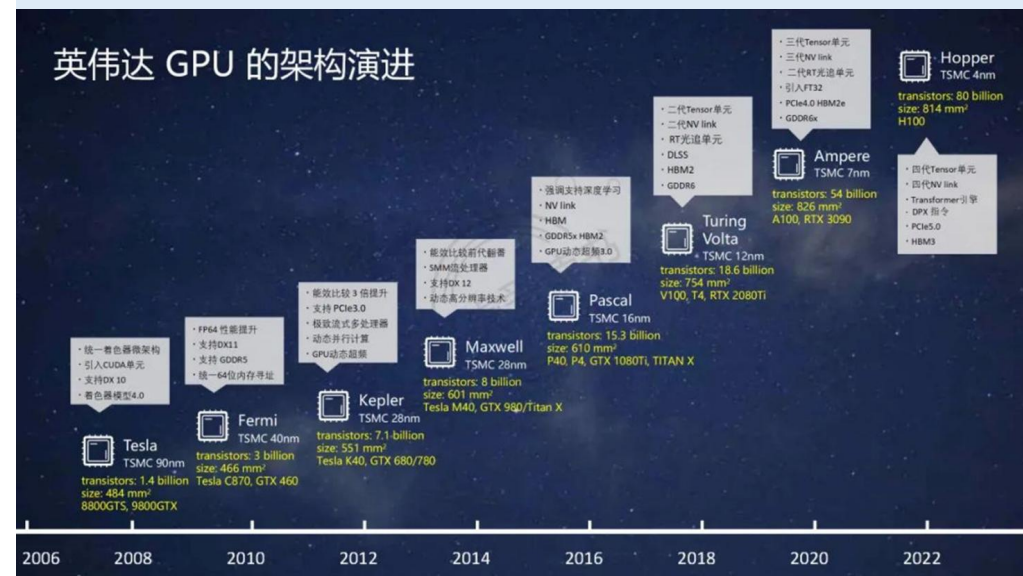
资料来源：公司官网，国信证券经济研究所整理

●**公司通过收购不断巩固业务实力，拓展业务边界。**较为代表性的重要收购事件是收购3dfx公司。3dfx是一家专门研发与生产显卡与3D芯片的公司，在二十世纪九十年代末一直是显卡芯片的领导者，但3dfx因破产最终被公司所收购。此外，公司通过收购多家图形渲染相关公司，巩固自身在传统图形计算优势领域的实力。2020年，公司收购高性能互联技术领域龙头Mellanox，产品布局从GPU扩展至兼具DPU。

GPU架构约每两年换代演进，CUDA降低GPGPU应用门槛

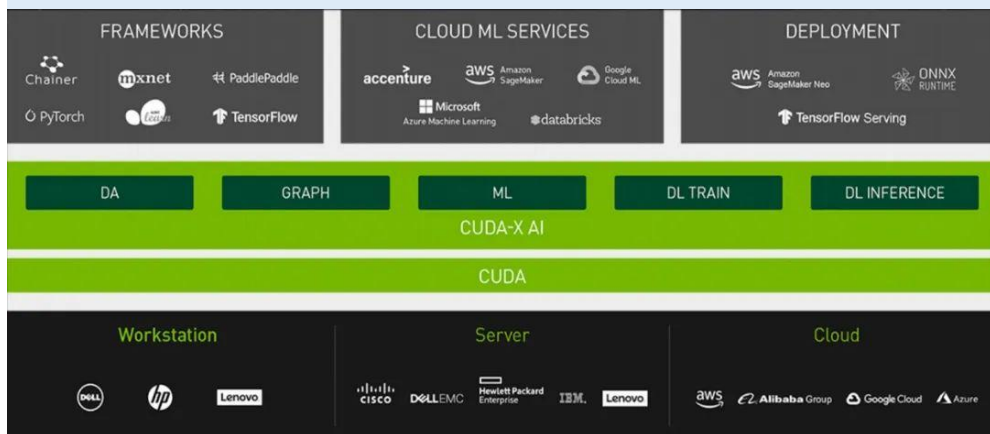
- 2006年，伴随CUDA平台推出，首款针对统一渲染图形的处理器G80诞生，采用台积电90nm制程。
- 2010年，Fermi架构推出，是第一个支持DirectX 11的GPU计算架构，采用台积电40nm制程。
- 2012年的Kepler架构是Fermi的升级版，整体架构保持一致性，采用台积电28nm制程。
- 2014年的Maxwell架构通过优化架构，提供了可观的能耗比提升。
- 2016年，Pascal架构推出，采用台积电16nm制程，支持DirectX 12标准，是首个为深度学习而设计的GPU架构。
- 2017年的Volta架构专注于提高深度学习的性能，采用台积电12nm制程，
- 2018年的Turing架构是全球首款支持实时光线追踪的GPU架构。
- 2020年，Ampere架构推出，采用台积电7nm/三星8nm制程，统一了AI训练和推理，并在光线追踪和DLSS（深度学习超级采样）方面有显著的改进。
- 2022年的Hopper架构采用台积电4nm制程，集成多达800亿个晶体管，主要面向AI及数据中心等构建。

图：英伟达GPU架构演进史



资料来源：woshipm，国信证券经济研究所整理

图：CUDA生态体系



资料来源：公司官网，国信证券经济研究所整理

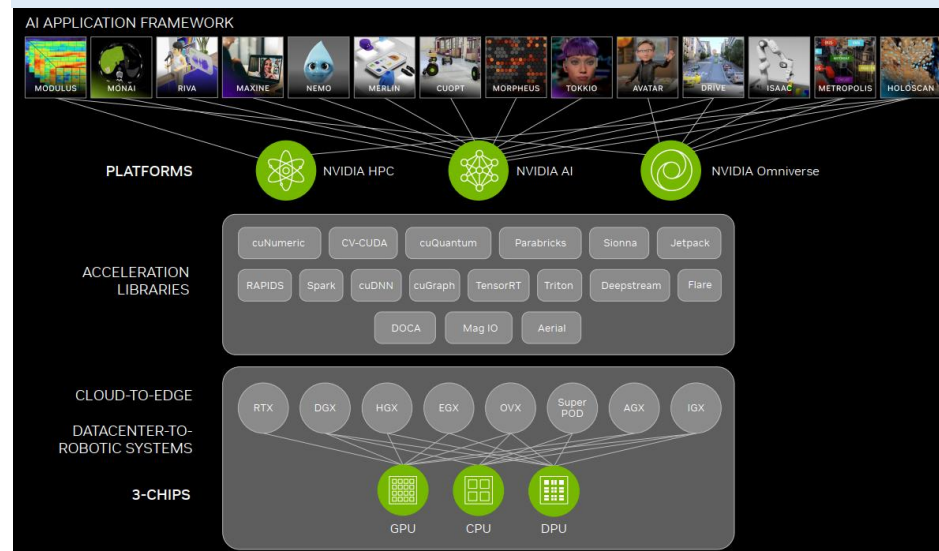
●**CUDA（Compute Unified Device Architecture，统一计算架构）**是由英伟达推出的一种集成技术，用户可以通过这个技术利用GPU进行图像处理之外的运算。CUDA也是首次可以利用GPU作为C-编译器的开发环境。

●2006年，英伟达发布了针对GPU的编程的架构CUDA，它提供了GPU编程的简易接口，可以基于CUDA编译基于GPU的应用程序，利用GPU的并行计算能力更高效的解决复杂计算难题。CUDA让GPU真正实现更广泛的通用计算，降低了GPGPU的应用门槛。CUDA与GPGPU直接推动了AI与深度学习的发展与产业革命。

构建加速计算平台，完善“三芯片”产品结构

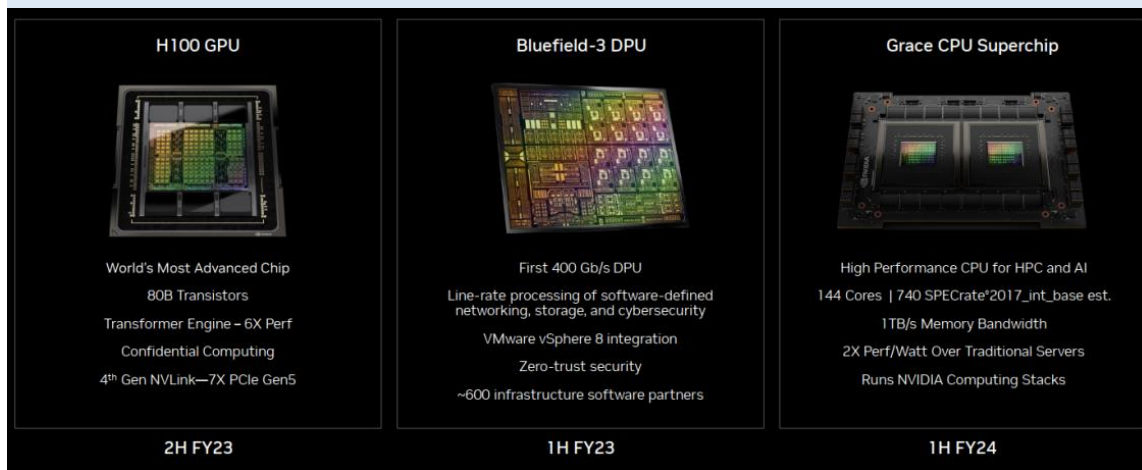
- 公司业务模式拟打造成类似于计算堆栈或神经网络，包含硬件、系统软件、平台软件和应用四层，每层都向计算机制造商、服务提供商和开发者开放，并将加速计算平台从个人电脑延伸到超级计算中心、企业数据中心、云和边缘环境中。
- 公司整体业务主要战略包括：推进NVIDIA加速计算平台；扩大公司在人工智能领域的技术和平台领导地位；扩展公司在计算机图形领域的技术和平台领导地位；推进领先的自动驾驶汽车平台等。

图：公司基于全栈创新能力的加速计算平台



资料来源：公司公告，国信证券经济研究所整理

图：公司GPU、DPU、CPU三种芯片产品结构



资料来源：公司公告，国信证券经济研究所整理

- 随着2020年公司收购收购高性能互联技术领域龙头Mellanox，产品布局从GPU扩展至兼具DPU（数据处理器）；以及2021年公司发布首款专为巨型AI和高性能计算工作负载设计的代号为“Grace”的CPU，公司成为具有GPU、CPU、DPU三种芯片的公司。

- 随着数据中心浪潮推进，公司陆续完善三种芯片产品结构。其中，集成800亿晶体管的迄今为止最先进的GPU H100已于FY2H23推出；首款400Gb/s全线速软件定义网络、存储和网络安全的DPU Bluefield-3已于FY1H23推出；为HPC（高性能计算）及AI应用而设计的Grace CPU Superchip计划于FY1H24推出。

●公司GPU产品能够并行计算的性能优势满足深度学习需求，由图形领域扩展至通用计算领域。GPU最初承担图像计算任务，目标是提升计算机对图形、图像、视频等数据的处理能力，解决CPU在图形图像领域处理效率低的问题。由于GPU能够进行并行计算，其架构本身较为适合深度学习算法。因此，通过对GPU的优化，能够进一步满足深度学习大量计算的需求。公司通过对GPU架构升级而推出的相关产品，其运算性能得到显著提升，广泛用于数据中心等计算密集领域。

表：公司典型数据中心GPU产品

产品型号	发布时间	制程	双精度浮点运算性能（TFLOPS）	单精度浮点运算性能（TFLOPS）	半精度浮点运算性能（TFLOPS）	整型定点运算性能（TOPS）	显存	显存带宽	最大功耗
H100 SXM	2022	4nm	67(Tensor Core)	989(TF32,Tensor Core)	1979(Tensor Core)	3958(Tensor Core)	80GB	3.35TB/s	700W
A100 SXM	2020	7nm	19.5(Tensor Core)	312(TF32,Tensor Core)	624(Tensor Core)	1248(Tensor Core)	80GB	2039GB/s	400W
V100S PCIe	2019	12nm	8.2	16.4	-	-	32GB	1134GB/s	250W

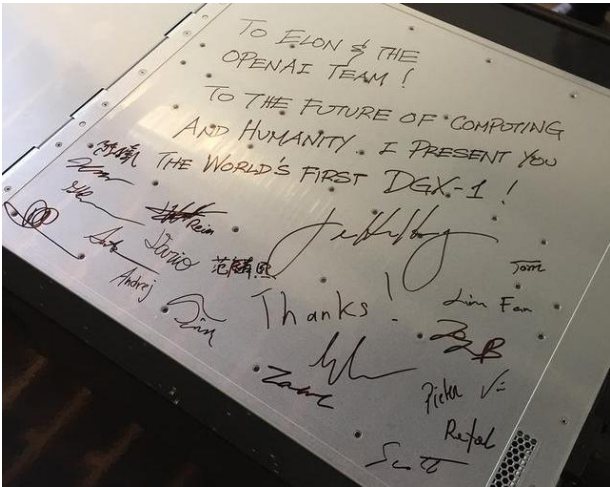
资料来源：公司官网，国信证券经济研究所整理

●2016年，英伟达CEO黄仁勋向OpenAI亲手交付了第一台NVIDIA DGX-1 AI超级计算机，这也是支持ChatGPT大型语言模型突破背后的引擎。

●2018年，OpenAI向微软提出“建立一个可以永远改变人机交互方式的人工智能系统”的想法。为了构建支持OpenAI项目的超级计算机，微软斥资数亿美元，在Azure云计算平台上将几万个NVIDIA A100芯片连接在一起，并改造了服务器机架。在这台超算上，OpenAI训练的模型不断强大，为后来诞生ChatGPT奠定基础。

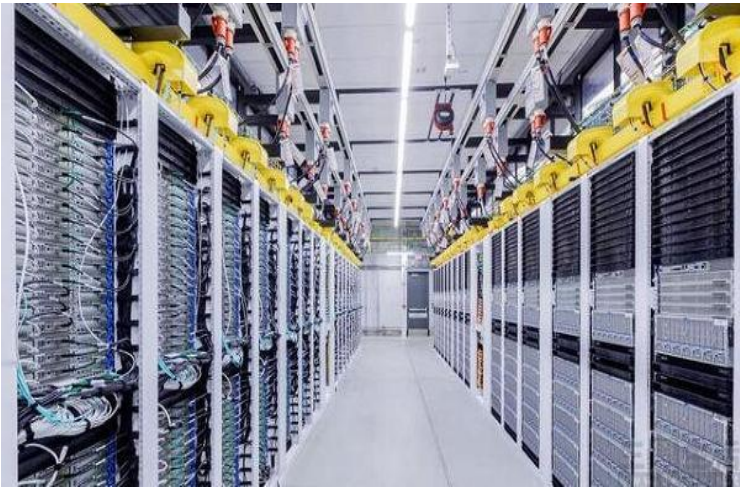
●2022年11月，微软宣布与英伟达联手构建“世界上最强大的AI超级计算机之一”，来处理训练和扩展AI所需的巨大计算负载。这台超级计算机基于微软的Azure云基础设施，使用了数以万计个Nvidia H100和A100 Tensor Core GPU，及其Quantum-2 InfiniBand网络平台，可用于研究和加速DALL-E和Stable Diffusion等生成式AI模型。2023年3月14日，微软宣布加强与英伟达的合作，将GPU从此前的A100升级到H100，推出专门针对人工智能开发的全新的ND H100 v5虚拟机。

图：英伟达赠送OpenAI第一台DGX-1



资料来源：CSDN，国信证券经济研究所整理

图：微软超级计算机



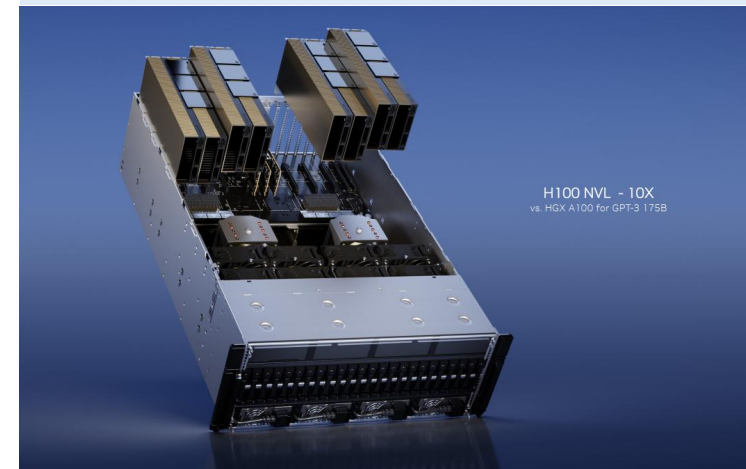
资料来源：微软官网，国信证券经济研究所整理

围绕不同的生成式AI工作负载提供推理平台及软硬件新品

●2023年3月，公司在GTC大会上推出4个针对各种生成式AI应用程序进行优化的推理平台，帮助开发者快速构建专业的人工智能应用程序。每个平台都包含针对特定生成式人工智能推理工作负载进行优化的GPU以及专用软件，包括AI视频、图像生成、大语言模型部署及推荐系统：

- NVIDIA L4 for AI Video可以提供比CPU高120倍的人工智能视频性能，同时能效提高99%；
- NVIDIA L40 for Image Generation针对图形和支持AI的2D、视频和3D图像生成进行了优化；
- NVIDIA H100 NVL for Large Language Model Deployment是大规模部署ChatGPT等大规模语言模型的理想选择；
- NVIDIA Grace Hopper for Recommendation Models是理想的图推荐模型、矢量数据库和图神经网络。

图：配备四对H100与双GPU NVLink的标准服务器速度较HGX A100最高可达10倍



资料来源：公司公告，国信证券经济研究所整理

图：推出4个针对各种生成式AI应用程序进行优化的推理平台



资料来源：公司公告，国信证券经济研究所整理

●公司在GTC大会上发布带有双GPU NVLink的H100 NVL加速计算卡，以支持ChatGPT类大型语言模型推理。H100 NVL拥有188GB HBM3内存（每张卡94GB），是目前发布的内存最大的加速计算卡。与适用于GPT-3处理的HGX A100相比，配备四对H100与双GPU NVLink的标准服务器的速度最高可达10倍。

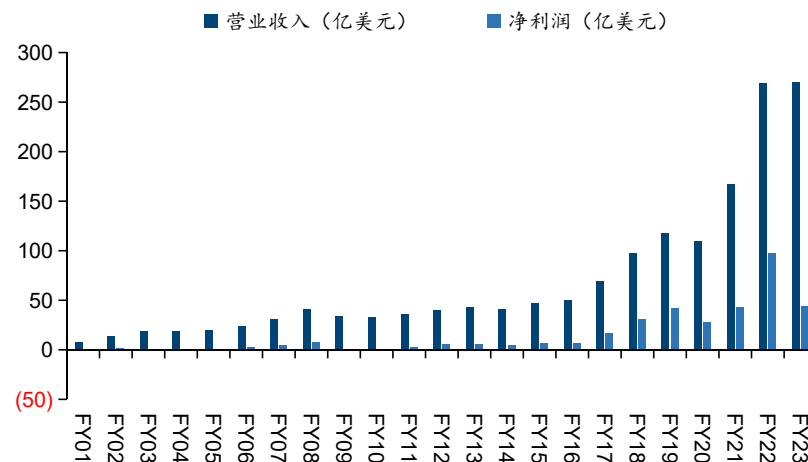
三、英伟达业绩实现与拆解

公司近年收入规模快速增长，毛利率水平持续提高

●FY01-FY16，公司收入由7.35亿美元稳步增长至50.10亿美元，对应CAGR约13.6%；净利润亦由0.98亿美元增长至6.14亿美元。自FY17后，受益于AI浪潮推动及游戏和数字中心业务需求旺盛，公司收入和利润水平快速增长，收入由FY17的69.10亿美元增长至FY22的269.14亿美元，对应CAGR达32.3%；净利润亦由FY17的16.66亿美元增长至FY22的97.52亿美元。

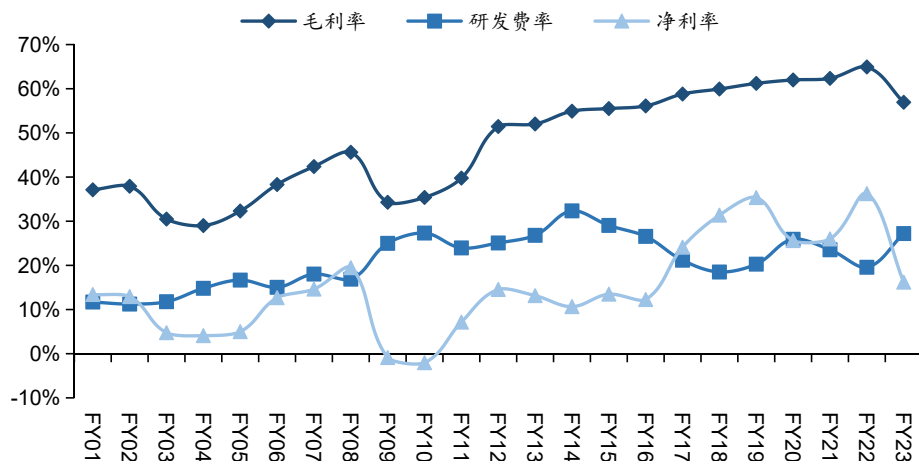
●FY23，公司收入为269.74亿美元，同比增长0.22%，主要原因为数据中心及汽车领域收入的高速增长抵消了游戏及专业可视化领域所带来的收入减少；净利润为43.68亿美元，同比减少55.21%，主要由于运营费用同比增长50%所致，包括与Arm交易相关的13.5亿美元收购终止费用。

图：公司FY01-FY23营业收入及净利润



资料来源：Bloomberg，国信证券经济研究所整理

图：公司FY01-FY23毛利率、净利率及研发费用率



资料来源：Bloomberg，国信证券经济研究所整理

●公司毛利率自FY12起保持50%以上并逐年持续增长，FY23降至56.9%；净利率自FY17起保持20%以上波动，FY23降至16.2%。FY23，公司毛利率同比下降8.0pct至56.9%，主要由于Ampere架构的游戏和数据中心产品供过于求带来的库存费用增加。净利率同比下降20.0pct至16.2%。

●公司研发费率FY01-FY14呈波动上升趋势，FY14达峰值32.3%之后波动下降；FY23为27.2%，同比上升7.6pct，主要由薪酬增加、员工增长、工程开发成本和数据中心基础设施等因素所致。

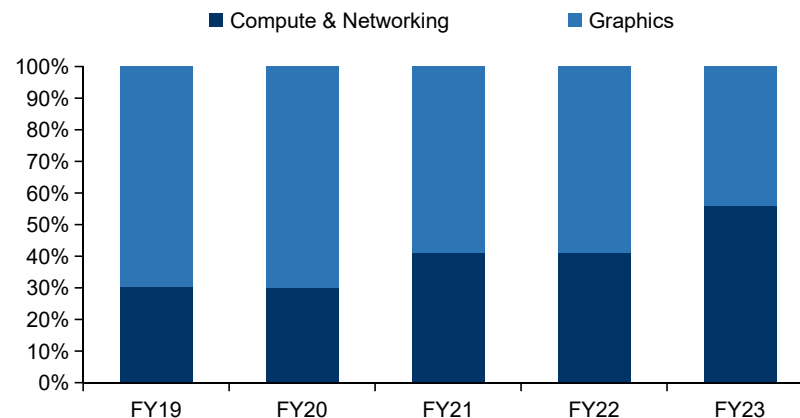
公司业务部门包括计算与网络事业部和图形事业部

●**计算与网络事业部 (Compute & Networking segment)** 涉及领域及产品包括数据中心加速计算平台、网络、汽车人工智能座舱、自动驾驶开发协议、自动驾驶汽车解决方案、电动汽车计算平台、用于机器人等的Jetson嵌入式平台、AI企业级等软件、加密货币挖掘处理器 (CMP) 等。

●**图形事业部 (Graphics segment)** 涉及领域及产品包括用于游戏和PC的GeForce GPU、GeForce NOW游戏流媒体服务及相关基础设施、游戏平台解决方案、用于企业工作站图形学的Quadro/NVIDIA RTX GPU、用于基于云可视化和虚拟计算的虚拟图形处理器软件 (vGPU)、用于信息娱乐系统的汽车平台、用于构建和操作元宇宙及3D互联网应用程序的Omniverse企业级平台。

●**公司计算与网络事业部逐渐成为收入占比最高的部门。** FY19-FY23, 公司计算与网络事业部收入占比呈上升趋势, 由30.36%上升至55.86%, 成为收入占比最高的部门。与此同时, 图形事业部收入占比由69.64%下降至44.14%。

图：公司FY19-FY23各业务部门收入占比

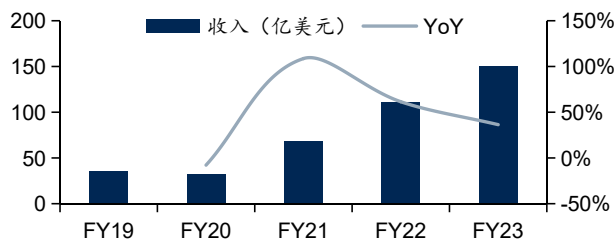


资料来源：Bloomberg，公司公告，国信证券经济研究所整理

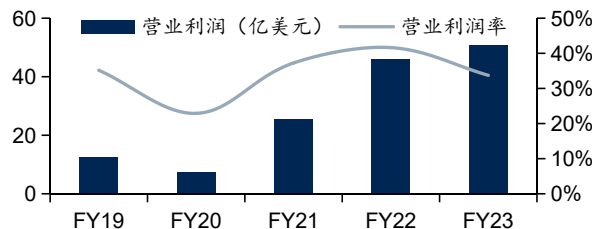
●**计算与网络事业部FY19-FY23收入快速增长，CAGR达43.46%。** 公司计算与网络事业部收入由FY19的35.57亿美元增长至FY23的150.68亿美元，对应CAGR达43.46%。其中，FY23收入同比增长36.4%，主要由于部分超大规模客户的增长，以及部分CSP（云服务提供商）合作方所带来的销量，此外收入的增长也受益于汽车行业的增长。

●**营业利润 (Operating Profit)** 为50.83亿美元，同比增长10.5%；营业利润率为33.7%，同比下降7.9pct。

图：公司FY19-FY23计算与网络事业部收入



图：公司FY19-FY23计算与网络事业部营业利润（率）

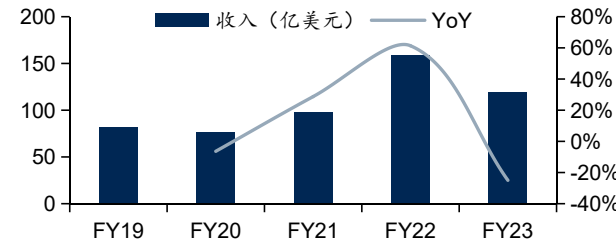


资料来源：Bloomberg，公司公告，国信证券经济研究所整理

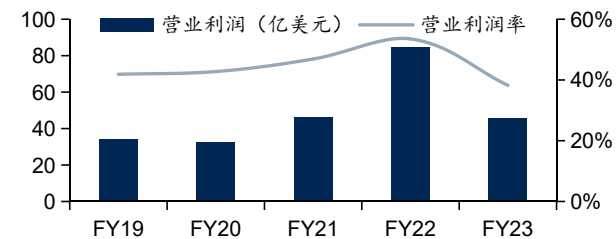
●**图形事业部FY19-FY22收入持续增长，FY23收入有所下降。** 公司图形事业部收入由FY19的81.59亿美元增长至FY22的158.68亿美元，呈现持续增长趋势；FY23收入为119.06亿美元，同比减少25.0%；减少主要由于全球宏观经济状况及疫情影响了游戏需求，渠道商为降低库存水平而减少了对其销售。

●**营业利润**为45.52亿美元，同比减少46.4%；营业利润率为38.2%，同比下降15.3pct。

图：公司FY19-FY23图形事业部收入



图：公司FY19-FY23图形事业部营业利润（率）



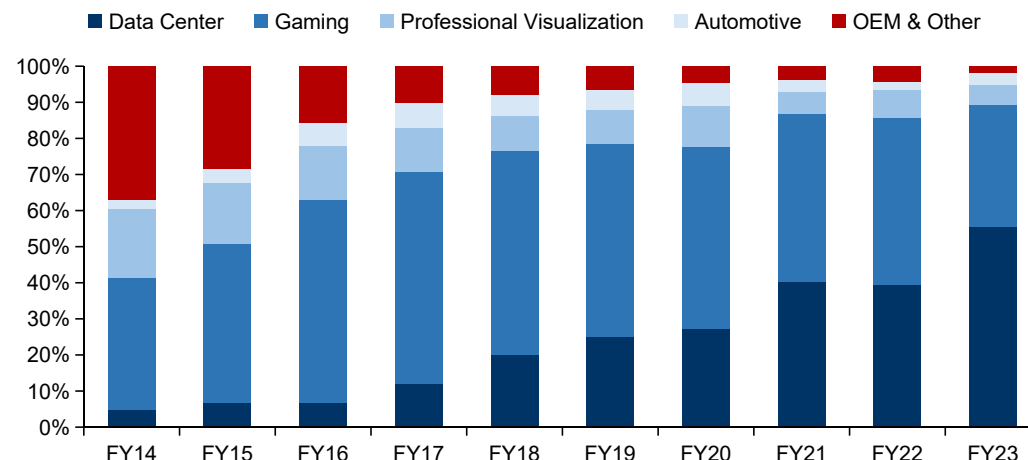
资料来源：Bloomberg，公司公告，国信证券经济研究所整理

FY23数据中心领域收入同比增长41.4%，游戏领域收入同比减少27.2%

●公司的平台及产品主要应用于数据中心、游戏、专业可视化、汽车等四大领域。

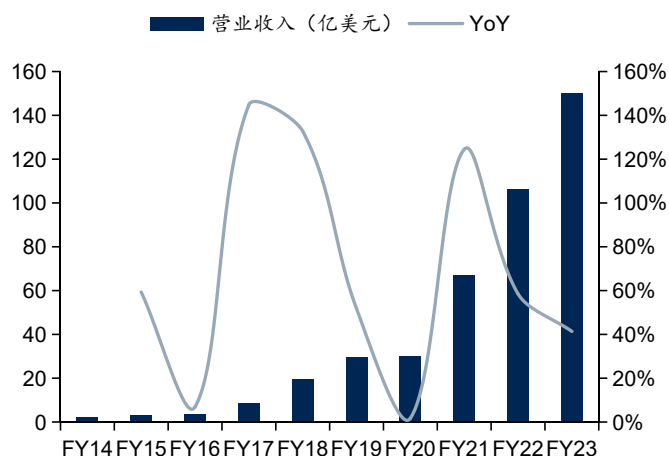
- 来自数据中心领域收入占比由FY14的4.82%持续提升至FY23的55.63%，已成为占比最高的领域。
- 来自游戏领域收入占比呈现先增后减的趋势，由FY14的36.59%先增长至FY17的极大值58.76%，而后不断降至FY23的33.61%。
- 来自专业可视化领域收入占比不断下降，由FY14的19.10%下降至FY23的5.72%。
- 来自汽车领域收入占比先增后降，FY23为3.35%。来自OEM及其他领域收入由FY14的37.09%降至FY23的1.69%。

图：公司FY14-FY23各市场领域收入占比



资料来源：Bloomberg，国信证券经济研究所整理

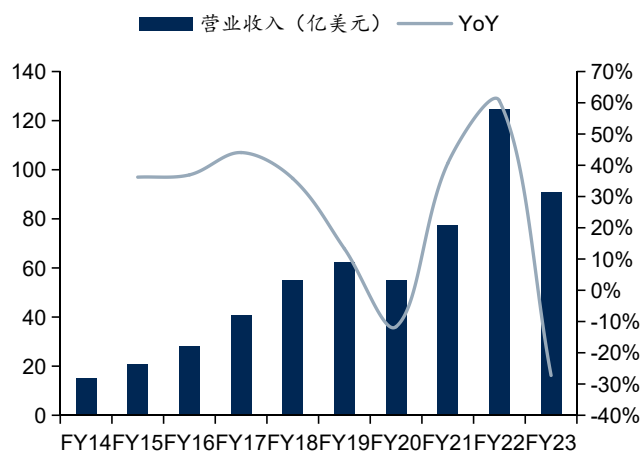
图：公司FY14F-Y23数据中心领域收入



资料来源：Bloomberg，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：公司FY14-FY23游戏领域收入



资料来源：Bloomberg，国信证券经济研究所整理

●数据中心（Data Center）领域中，公司计算平台专注于加速超大规模、云、企业级、公共部门和边缘数据中心的计算密集型工作负载，如人工智能、数据分析、图形和科学计算等。**FY23，数据中心领域收入为150.05亿美元，同比增长41.4%，**主要受益于加速计算平台和网络组合的超大规模及云端使用的影响。

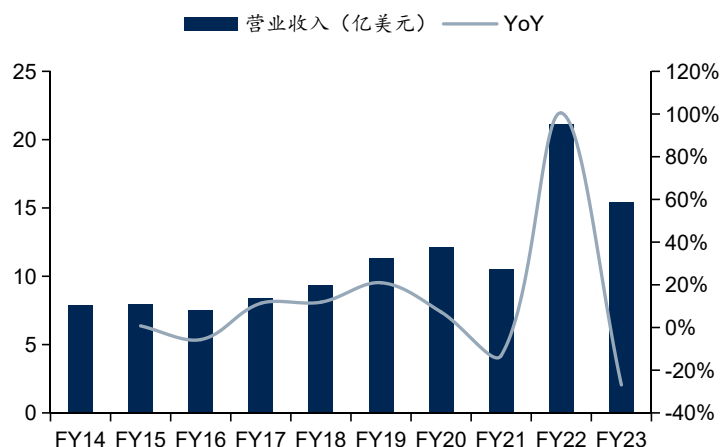
●游戏（Gaming）领域中，公司利用GPU产品和复杂软件，以更流畅、更高质量的图形来增强游戏体验，包括用于游戏台式机和笔记本电脑的GeForce RTX和GeForce GTX GPU等。**FY23，游戏领域收入为90.67亿美元，同比减少27.2%，**主要由于全球宏观经济疲软及疫情影响减少了游戏的需求以及渠道合作商较高的库存压力影响。

FY23专业可视化领域收入同比减少26.9%，汽车领域收入同比增长59.5%



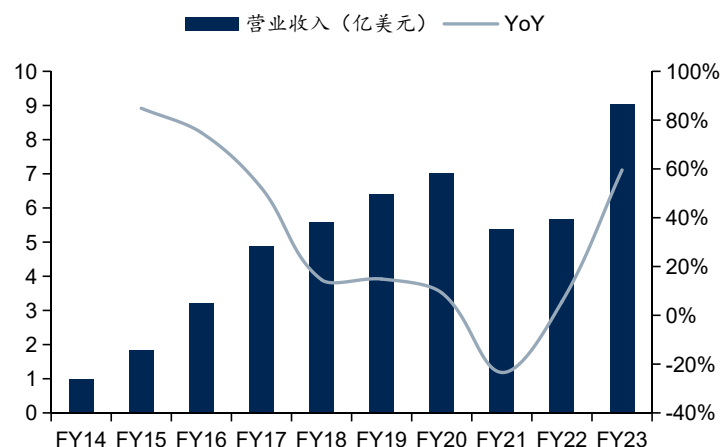
- 专业可视化（Professional Visualization）领域中，公司GPU计算平台为设计和制造以及数字内容创作等领域的关键工作流程引入新功能，包括计算机辅助设计、建筑设计、消费产品制造、医疗仪器和航空航天等。FY23，专业可视化领域收入为15.44亿美元，同比减少26.9%，主要由于受经济状况影响，合作商去库存压力较大。
- 汽车（Automotive）领域中，公司涉及自动驾驶、人工智能座舱、电动汽车计算平台和信息娱乐平台解决方案，旗下DRIVE Hyperion品牌为自动驾驶汽车市场提供完整的端到端解决方案。FY23，汽车领域收入为9.03亿美元，同比增长59.5%，主要由于针对自动驾驶解决方案、电动汽车制造商计算解决方案的销售增长，包括AI驾驶舱解决方案及汽车开发协议带来的增长；公司开始生产DRIVE Orin自动驾驶汽车SoC，并推出下一代DRIVE Thor。
- OEM及其他领域FY23收入为4.55亿美元，同比减少60.8%，主要受笔记本电脑OEM和CMP（加密货币挖掘处理器）减少的影响。

图：公司FY14-FY23专业可视化领域收入



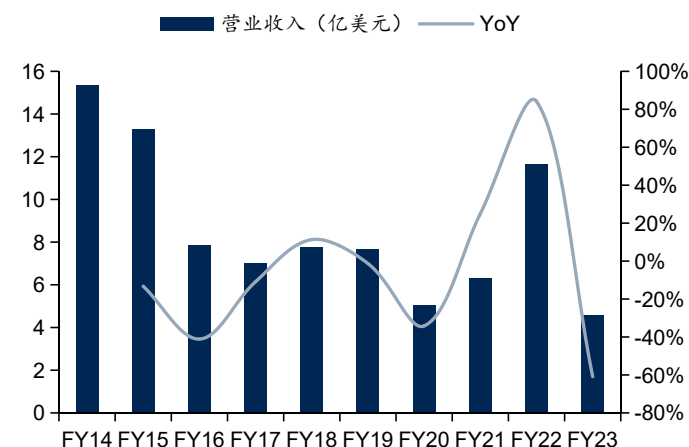
资料来源：Bloomberg，国信证券经济研究所整理

图：公司FY14-FY23汽车领域收入



资料来源：Bloomberg，国信证券经济研究所整理

图：公司FY14-FY23 OEM及其他领域收入

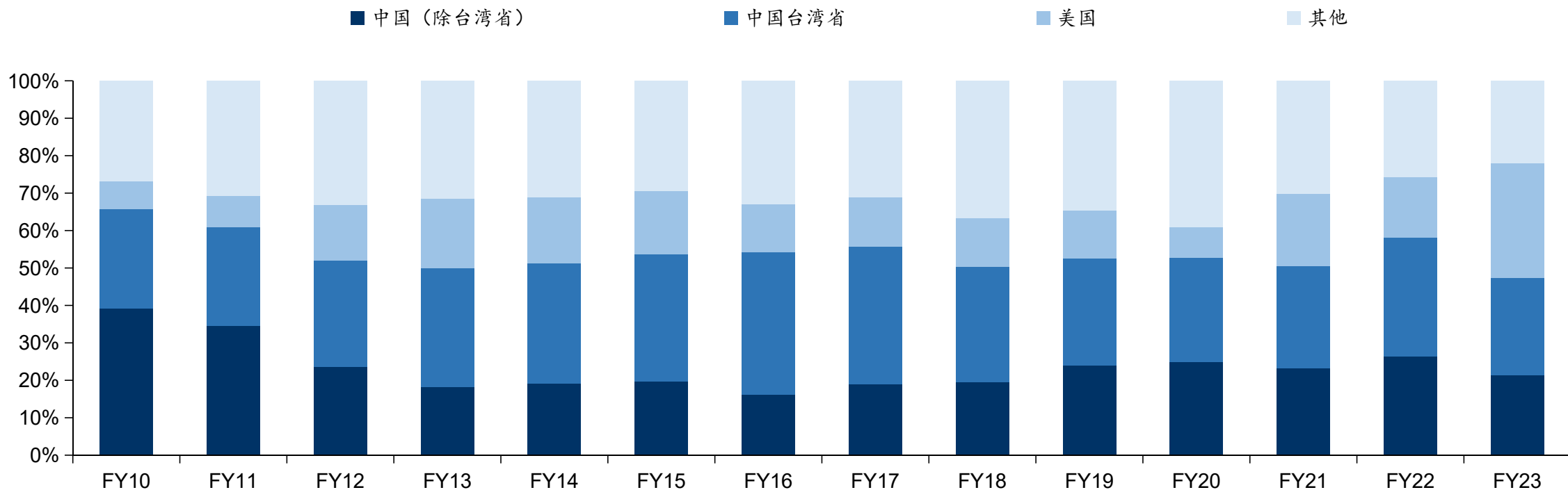


资料来源：Bloomberg，国信证券经济研究所整理

FY23来自美国地区的收入占比最高，达30.7%

●FY10-FY22，来自中国（除台湾省）及中国台湾省的收入占比通常达到50%以上，来自美国地区的收入占比通常不到20%，其他地区占比通常为20%-40%。FY23，由于美国以外地区特别是中国的数据中心和相关业务减少，使得美国地区收入占比提高，达到30.7%，成为收入来源的主要区域；中国（除台湾省）、中国台湾省、其他地区收入占比分别降至21.4%、25.9%、21.9%。

图：公司FY10-FY23分地区收入占比



公司FY2Q24收入指引为107.8–112.2亿美元



●公司FY1Q24实现收入71.92亿美元（YoY -13.2%，QoQ 18.9%）。在GAAP准则下，毛利率64.6%（YoY -0.9pct，QoQ 1.3pct），净利润20.43亿美元（YoY 26.3%，QoQ 44.5%）；在Non-GAAP准则下，毛利率66.8%（YoY -0.3pct，QoQ 0.7pct），净利润27.13亿美元（YoY -21.2%，QoQ 24.8%）。

- 分市场领域看，FY1Q24中，数据中心领域收入42.8亿美元（YoY 14.1%，QoQ 18.4%），占总收入59.5%；
- 游戏领域收入22.4亿美元（YoY -38.1%，QoQ 22.3%），占比31.1%；
- 专业可视化领域收入2.95亿美元（YoY -52.6%，QoQ 30.5%），占比4.1%；
- 汽车领域收入2.96亿美元（YoY 114.5%，QoQ 0.7%），占比4.1%；
- OEM及其他领域收入0.81亿美元（YoY -48.7%，QoQ -3.6%），占比1.1%。

图：公司FY2Q24业绩指引

Outlook

NVIDIA's outlook for the second quarter of fiscal 2024 is as follows:

- Revenue is expected to be \$11.00 billion, plus or minus 2%.
- GAAP and non-GAAP gross margins are expected to be 68.6% and 70.0%, respectively, plus or minus 50 basis points.
- GAAP and non-GAAP operating expenses are expected to be approximately \$2.71 billion and \$1.90 billion, respectively.
- GAAP and non-GAAP other income and expense are expected to be an income of approximately \$90 million, excluding gains and losses from non-affiliated investments.
- GAAP and non-GAAP tax rates are expected to be 14.0%, plus or minus 1%, excluding any discrete items.

资料来源：公司公告，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：公司FY1Q24业绩实现

Q1 Fiscal 2024 Summary

GAAP					
(\$ in millions, except earnings per share)	Q1 FY24	Q4 FY23	Q1 FY23	Q/Q	Y/Y
Revenue	\$ 7,192	\$ 6,051	\$ 8,288	Up 19%	Down 13%
Gross margin	64.6%	63.3%	65.5%	Up 1.3 pts	Down 0.9 pts
Operating expenses	\$ 2,508	\$ 2,576	\$ 3,563	Down 3%	Down 30%
Operating income	\$ 2,140	\$ 1,257	\$ 1,868	Up 70%	Up 15%
Net income	\$ 2,043	\$ 1,414	\$ 1,618	Up 44%	Up 26%
Diluted earnings per share	\$ 0.82	\$ 0.57	\$ 0.64	Up 44%	Up 28%

Non-GAAP					
(\$ in millions, except earnings per share)	Q1 FY24	Q4 FY23	Q1 FY23	Q/Q	Y/Y
Revenue	\$ 7,192	\$ 6,051	\$ 8,288	Up 19%	Down 13%
Gross margin	66.8%	66.1%	67.1%	Up 0.7 pts	Down 0.3 pts
Operating expenses	\$ 1,750	\$ 1,775	\$ 1,608	Down 1%	Up 9%
Operating income	\$ 3,052	\$ 2,224	\$ 3,955	Up 37%	Down 23%
Net income	\$ 2,713	\$ 2,174	\$ 3,443	Up 25%	Down 21%
Diluted earnings per share	\$ 1.09	\$ 0.88	\$ 1.36	Up 24%	Down 20%

资料来源：公司公告，国信证券经济研究所整理

●英伟达表示，数据中心收入激增主要是由于对使用基于Hopper和Ampere架构GPU的生成式AI和大语言模型的需求不断增长。由于众多云公司竞相部署AI芯片，锁定了数据中心芯片的大幅增长，英伟达称计划下半年大幅增加供应。

●公司FY2Q24收入指引为107.8–112.2亿美元（YoY 60.8%至67.4%，QoQ 49.9%至56.0%）。在GAAP准则下，预计毛利率68.1%–69.1%（YoY 24.6至25.6pct，QoQ 3.5至4.5pct）；在Non-GAAP准则下，预计毛利率69.5%–70.5%（YoY 23.6至24.6pct，QoQ 2.7至3.7pct）。

四、风险提示

- 1、宏观AI应用推广不及预期。**AI技术在应用推广的过程可能面临各种挑战，比如：（1）AI技术需要更多的时间来研发和调试，而且在应用过程中可能会受到数据质量、资源限制和技术能力等因素的制约；（2）AI技术的实施需要更多的资源和资金支持；（3）市场竞争可能也会影响企业在AI应用推广方面的表现。因此，投资者应审慎评估相关企业的技术实力、资金实力以及管理能力，相关企业的AI应用存在推广进度不及预期的风险。
- 2、AI投资规模低于预期。**尽管AI技术在过去几年中受到广泛关注，但AI相关领域的企业投资回报并不总是符合预期。部分企业在AI领域可能缺乏足够的经验和资源，难以把握市场机会。此外，市场竞争也可能会影响企业的投资力度。因此，存在AI领域投资规模低于预期，导致企业相关业务销售收入不及预期的风险。
- 3、AI服务器渗透率提升低于预期。**虽然AI服务器的应用已经较为广泛，但AI服务器渗透率提升的速度存在低于预期的风险，这与企业对AI技术的投资意愿有关，也可能与市场需求和技术进展的速度有关。
- 4、AI监管政策收紧。**由于AI技术的快速发展和广泛应用，监管机构可能会加强对AI技术的监管力度。监管机构可能会制定严格的AI技术使用规定，以保障人们的隐私和数据安全，这些监管政策可能会对企业的业务模式和发展战略造成影响。

国信证券投资评级		
类别	级别	定义
股票投资评级	买入	预计6个月内，股价表现优于市场指数20%以上
	增持	预计6个月内，股价表现优于市场指数10%-20%之间
	中性	预计6个月内，股价表现介于市场指数±10%之间
	卖出	预计6个月内，股价表现弱于市场指数10%以上
行业投资评级	超配	预计6个月内，行业指数表现优于市场指数10%以上
	中性	预计6个月内，行业指数表现介于市场指数±10%之间
	低配	预计6个月内，行业指数表现弱于市场指数10%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。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，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我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。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构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，并向客户发布的行为。



国信证券
GUOSEN SECURITIES

国信证券经济研究所

深圳

深圳市福田区福华一路125号国信金融大厦36层

邮编：518046 总机：0755-82130833

上海

上海浦东民生路1199弄证大五道口广场1号楼12楼

邮编：200135

北京

北京西城区金融大街兴盛街6号国信证券9层

邮编：100032