

电子

英伟达引领“AI 摩尔定律”，直击 AI 的 iPhone 时刻

事件点评

◆ 事件：

英伟达 3 月 22 日宣布推出 3 款全新推理 GPU，发布了 NVIDIA cuLitho 计算光刻库、AI 超级计算服务 DGX Cloud、加速企业创建大模型和生成式 AI 的云服务 NVIDIA AI Foundations 等。英伟达 CEO 黄仁勋称，“崭新的计算平台已经诞生，AI 的 iPhone 时刻已经到来”(We are at the iPhone moment of AI)。

◆ DGX H100 AI 超级计算机全面投入生产，简化 AI 训练

英伟达加速计算始于 DGX (AI 超级计算机)，这是大型语言模型实现突破背后的引擎。英伟达 H100 GPU 基于公司最新 Hopper 架构，在大型语言模型上比上一代 A100 提供了快 9 倍的 AI 训练和快 30 倍的 AI 推理。本次发布的 DGX H100 拥有 8 个 H100 GPU 模组，在 FP8 精度下可提供 32PetaFLOPS 的算力，并提供完整的英伟达 AI 软件堆栈，助力简化 AI 开发。黄仁勋宣布，NVIDIA DGX H100 AI 超级计算机已全面投入生产，很快将面向全球企业。微软宣布 Azure 将向其 H100 AI 超级计算机开放私人预览版。

GPU 高效互连技术为英伟达提供底层技术支撑。此次发布的计算平台是在 2022 年 Hopper 架构的基础上，进行升级和拓展。DGX H100 有 8 个 H100 协同工作，通过 NVLINK 和 NVSwitch 彼此相连，组成了一个巨型 GPU，实现全面无阻塞通信。借助 NVSwitch，NVLink 连接可在节点间扩展，以创建无缝、高带宽的多节点 GPU 集群，从而有效地形成数据中心大小的 GPU，快速完成大型 AI 作业。

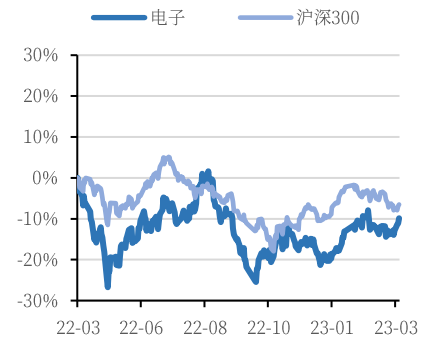
我们认为，在 ChatGPT 掀起的 AIGC 热潮下，大规模 AI 模型训练和推理量激增，催生对低延迟、高速率传输交换的光网络的迫切需求。当前英伟达

投资评级

增持

维持评级

行业走势图



作者

刘牧野

分析师

SAC 执业证书：S0640522040001

邮箱：liumy@avicsec.com

相关研究报告

科技专题研究：ChatGPT 狂飙破壁，现象级 AI 应用引发范式革命 —2023-02-16

电子行业周报：算力时代传统光模块功耗制约凸显，CPO 降本增效迎发展良机 —2023-03-12

股市有风险 入市需谨慎

中航证券研究所发布 证券研究报告

请务必阅读正文之后的免责声明部分

联系地址：北京市朝阳区望京街道望京东园四区2号楼中航产融大厦

中航证券有限公司

公司网址：www.avicsec.com

联系电话：010-59219558 传真：010-59562637

可以通过 NVLink-C2C 技术实现高速、低延迟、芯片到芯片的互连，NVSwitch 结构在本质上仍然是一种对 NUMA 直接放大的方法，仍然受到线缆布局的限制。光电共封（CPO）将进一步降低数据传输延迟和减少信号损失，是英特尔、英伟达等头部算力公司重点关注的技术方向。

◆ 针对垂直领域推出芯片及付费服务

针对不同的应用领域，英伟达发布了用于 AI 视频生成的通用 GPU L4；用于图像生成的 L40；针对 ChatGPT 等大型语言模型的大规模部署的 H100 NVL；为推荐系统和大型语言模型的 AI 数据库所设计的 Grace Hopper 超级芯片。

在计算服务端，英伟达推出了一项名为 DGX Cloud 的 AI 超级计算服务，与微软 Azure、谷歌 OCP、Oracle OCI 合作，通过一个 Web 浏览器就能访问，以便企业为生成式 AI 和其他开创性应用训练先进的模型。DGX Cloud 实例的起价为每个实例每月 36999 美元。此外，英伟达还推出了三种模型构建服务，分别为文本生成模型构建服务、视觉语言模型构建服务、生命科学模型构建服务。

我们在之前的报告指出，通用性递减，专用性增强，为 AI 芯片的主要发展方向，英伟达此次发布偏专用领域的推理平台，印证了我们的观点。2021 年中国以 GPU 为主实现数据中心计算加速，GPU 在算力芯片的市场占有率接近 90%。ASIC，FPGA，NPU 等非 GPU 芯片市场占有率超过 10%。国际科技网络巨头公司谷歌、脸书，亚马逊等等在 AI 芯片领域从云端训练到终端产品应用，在开源框架赋能产业行业上有一定的领先优势。受美国科技制裁的影响，部分高性能 GPU 对华禁售，如英伟达 A100 和 H100，AMD 的 MI250。国内人工智能产业的发展需要国内 GPU 的共同成长，国产替代势在必行，重点关注 GPU、ASIC、FPGA 芯片进展，以及有望突破制程限制的 Chiplet 技术。

◆ 工业元宇宙加速落地，构建 Omniverse 生态

Omniverse 是英伟达打造的工业数字化平台，旨在搭建数字化和物理世界之间的桥梁，让各个行业在投入实际生产之前先以数字化方式设计、建造、运营和优化物理产品和工厂。

英伟达宣布了专门为运行 Omniverse 而设计的硬件设备。第一个是由英伟达 Ada RTX GPU 和英特尔最新 CPU 驱动的新一代 NVIDIA RTX 工作站, 能使专业人士能随时随地访问 Omniverse 和工业元宇宙工作负载, 将从三月份开始 Boxx、戴尔、惠普和联想上市; 第二个是针对 Omniverse 优化的新型 Nvidia ovx 服务器, 将由戴尔、惠普、Quantn、技嘉、联想和 SuperMicro 提供, 企业可以利用 OVX 性能在可视化、虚拟工作站和数据中心处理工作流程等方面进行协作。

英伟达还推出了 Omniverse Cloud 的云服务, 支持即时、安全地访问托管的 Omniverse Cloud API。英伟达正在与微软合作, 将 Omniverse 云服务引入全球各个行业。同时, 英伟达正在将 Omniverse 云连接到 Microsoft 365 生产力套件, 包括 Teams、OneDrive、SharePoint 和 Azure IoT 数字孪生服务。

我们认为, 随着 AI 训练成本的降低, “AI 平民化” 并且赋能前行百业的时代即将到来。据 ARK 预测, 硬件和软件的融合可以推动人工智能训练成本在 2030 年前均以每年 70% 的速度下降。以 GPT-3 模型为例, 2020 年, 单次训练成本高达 460 万美金, 需要 1 万块 GPU 耗费 13 天。参考计算机从大型化、工程化应用, 到台式笔记本, 再到智能手机的发展路径, 随着人工智能在编程、办公等工业应用逐渐兴起, 以 GPT4 为代表的多模态 AIGC 模型将进一步挖掘消费端的应用市场。英伟达 Omniverse 服务降低了 AI 模型的成本, 使移动端 AI 应用成为可能, 将在更多的垂直领域引发产业革命。

◆ 推荐关注

随着算法迭代、算力增强和数据增长, 国内 AI 大模型将持续追赶国际头部公司, AI 产业链的估值将保持抬升的态势, 建议持续跟踪本轮科技浪潮。

相关产业链投资机会:

算法环节: 拓尔思、科大讯飞、大华股份、海康威视;

算力环节: 寒武纪、海光信息、复旦微电、紫光国微、拓维信息、中科曙光、商汤 (港股) 等;

CPO、光传输环节: 源杰科技、中际旭创、光迅科技、新易盛、光库科技、声光电科等;

Chiplet: 通富微电、长电科技、甬矽电子、芯原股份、兴森科技等;

AI 服务器电源: 杰华特、希荻微等。

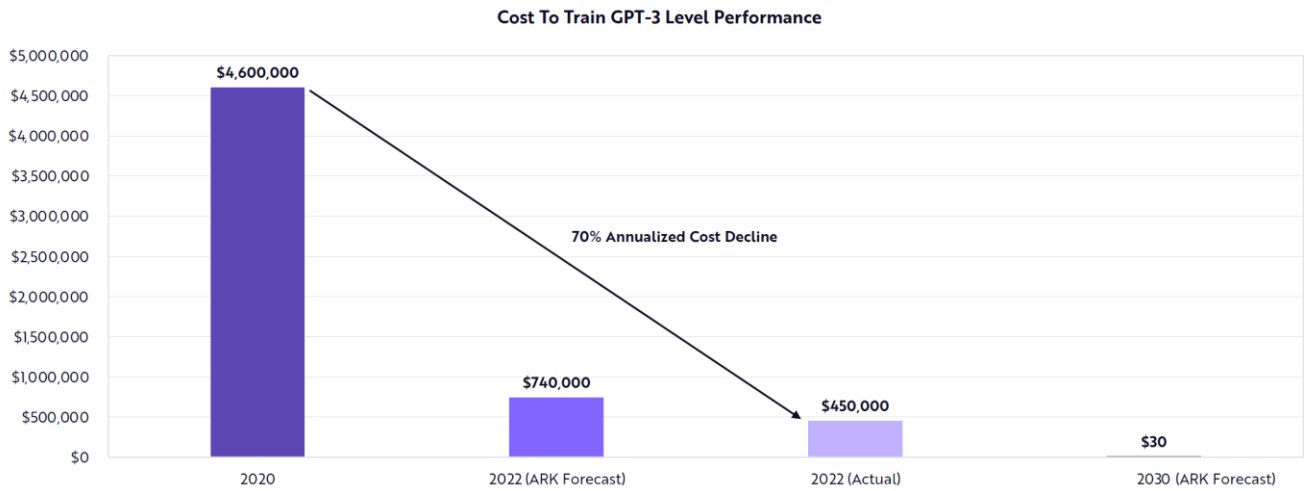
◆ 风险提示：AI 算法、模型存较高不确定性，AI 技术发展不及预期；ChatGPT 用户付费意愿弱，客户需求不及预期；针对 AI 的监管政策收紧

图1 AI 产业链投资地图



资源来源：中航证券研究所

图2 GPT3 大模型训练成本快速下降



资源来源：ARK，中航证券研究所

公司的投资评级如下:

买入: 未来六个月的投资收益相对沪深 300 指数涨幅 10%以上。

持有: 未来六个月的投资收益相对沪深 300 指数涨幅-10%~10%之间。

卖出: 未来六个月的投资收益相对沪深 300 指数跌幅 10%以上。

行业的投资评级如下:

增持: 未来六个月行业增长水平高于同期沪深 300 指数。

中性: 未来六个月行业增长水平与同期沪深 300 指数相若。

减持: 未来六个月行业增长水平低于同期沪深 300 指数。

研究团队介绍汇总:

首席: 赵晓琨 十六年消费电子及通讯行业工作经验, 曾在华为、阿里巴巴、摩托罗拉、富士康等多家国际级头部品牌终端企业, 负责过研发、工程、供应链采购等多岗位工作。曾任职华为终端半导体芯片采购总监, 阿里巴巴人工智能实验室供应链采购总监。长期专注于三大方向: 1、半导体及硬科技; 2、智慧汽车及机器人; 3、大势所趋的新能源。 分析师: 刘牧野 约翰霍普金斯大学机械系硕士, 2022 年 1 月加入中航证券。拥有高端制造、硬科技领域的投研经验, 从事科技、电子行业研究。

销售团队:

李裕淇, 18674857775, liyuq@avicsec.com, S0640119010012

李友琳, 18665808487, liyoul@avicsec.com, S0640521050001

曾佳辉, 13764019163, zengjh@avicsec.com, S0640119020011

分析师承诺:

负责本研究报告全部或部分内容的每一位证券分析师, 再次申明, 本报告清晰、准确地反映了分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告中的具体推荐或观点直接或间接相关。

风险提示: 投资者自主作出投资决策并自行承担投资风险, 任何形式的分享证券投资收益或者分担证券证券投资损失的书面或口头承诺均为无效。

免责声明:

本报告由中航证券有限公司(已具备中国证券监督管理委员会批准的证券投资咨询业务资格)制作。本报告并非针对意图送发或为任何就送发、发布、可得到或使用本报告而使中航证券有限公司及其关联公司违反当地的法律或法规或可致使中航证券受制于法律或法规的任何地区、国家或其它管辖区域的公民或居民。除非另有显示, 否则此报告中的材料的版权属于中航证券。未经中航证券事先书面授权, 不得更改或以任何方式发送、复印本报告的材料、内容或其复本给予任何其他人。未经授权的转载, 本公司不承担任何转载责任。

本报告所载的资料、工具及材料只提供给阁下作参考之用, 并非作为或被视为出售或购买或认购证券或其他金融票据的邀请或向他人作出邀请。中航证券未有采取行动以确保于本报告中所指的证券适合个别的投资者。本报告的内容并不构成对任何人的投资建议, 而中航证券不会因接受本报告而视他们为客户。

本报告所载资料的来源及观点的出处皆被中航证券认为可靠, 但中航证券并不能担保其准确性或完整性。中航证券不对因使用本报告的材料而引致的损失负任何责任, 除非该等损失因明确的法律或法规而引致。投资者不能仅依靠本报告以取代行使独立判断。在不同时期, 中航证券可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告仅反映报告撰写日分析师个人的不同设想、见解及分析方法。为免生疑, 本报告所载的观点并不代表中航证券及关联公司的立场。

中航证券在法律许可的情况下可参与或投资本报告所提及的发行人的金融交易, 向该等发行人提供服务或向他们要求给予生意, 及或持有其证券或进行证券交易。中航证券于法律容许下可于发送材料前使用此报告中所载资料或意见或他们所依据的研究或分析。

联系地址: 北京市朝阳区望京街道望京东园四区 2 号楼中航产融大厦中航证券有限公司

公司网址: www.avicsec.com

联系电话: 010-59219558

传 真: 010-59562637