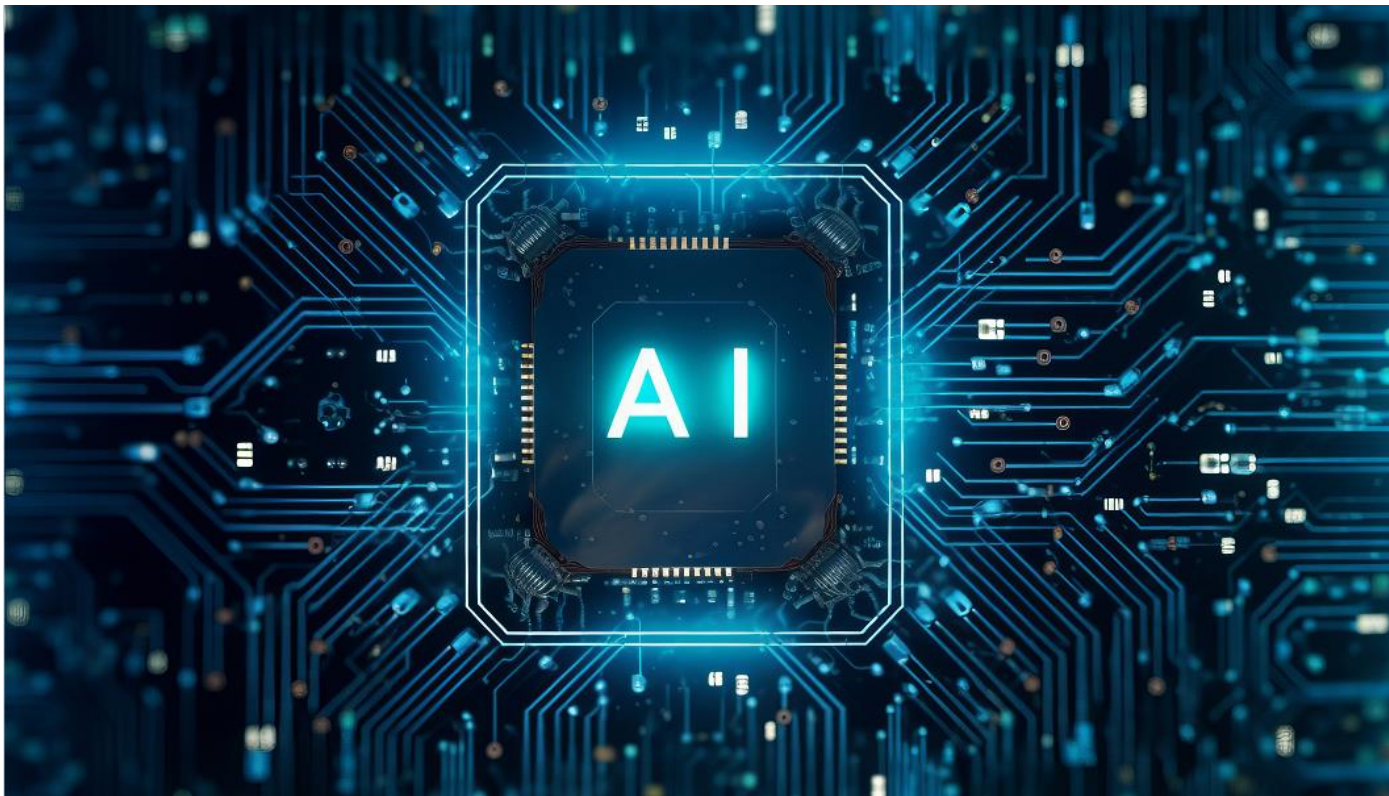


2024年7月2日



## 全球 AI 算力行业首次覆盖：从云到端，云端协同， AI 开启科技行业超级成长周期

**超配****首次覆盖**

随着 2022 年 11 月底 OpenAI 发布 ChatGPT 3.5，AI 相关行业在 2023 年进入爆发式发展。OpenAI 随后发布 ChatGPT 4.0，谷歌、Meta 等多家厂商也陆续推出 Gemini、Llama 等 AI 大模型。在中国，包括互联网头部企业和初创公司在内的玩家相继投入到大模型的研究和发布中。根据 IDC 数据，2020 年至 2023 年全球生成式 AI 市场空间上涨了约 6 倍，预计 2024 年至 2030 年期间复合增长率达到 40%，2030 年有望接近万亿美元。算力芯片作为 AI 大模型行业的最重要支撑，也正享受行业高增长红利。因此，我们首次覆盖全球 AI 算力行业，给予“超配”评级；考虑到英伟达受益于云侧 AI 高增长以及智能驾驶端侧渗透，首次覆盖英伟达 (NVDA.US)，给予“买入”评级；考虑到高通受益于智能手机和新能源车两大终端的端侧 AI 渗透，首次覆盖高通 (QCOM.US)，给予“买入”评级。

沈岱 (首席科技分析师)  
tony\_shen@spdbi.com  
(852) 2808 6435

马智焱 (科技分析师)  
ivy\_ma@spdbi.com  
(852) 2809 0300

黄佳琦 (科技分析师)  
sia\_huang@spdbi.com  
(852) 2809 0355



欢迎关注  
浦银国际研究

# 目录

全球 AI 算力行业投资要点 .....	4
全球 AI 算力行业概览：发展初期，高速增长 .....	11
全球 AI 大模型具备较大的成长空间 .....	11
AI 算力及 GPU 需求测算：需求空间仍旧巨大 .....	16
AI 算力芯片产能瓶颈逐步解决，供需趋于平衡 .....	19
中国的 AI 大模型快速跟进：AI 需求大幅上扬和大模型能力快速追赶 .....	21
AI 算力供应链解析：价值量集中度高，算力芯片端垄断竞争 .....	24
AI 服务器价值量集中度高，提升幅度大 .....	24
中国 AI 算力产业链 .....	25
AI 大模型：从云侧向端侧渗透 .....	27
AI 端侧大模型正在快速落地中 .....	27
AI 算力芯片竞争格局：云侧 vs 端侧 .....	30
AI 算力芯片竞争格局：云侧产品比较 .....	32
美股科技股价值投资回顾 .....	34
美进入降息周期有利于成长科技股估值上行 .....	34
美股纳斯达克以及 M7 具备长线的基本面推动成长的能力 .....	35
附录：AI 算力行业术语简介 .....	38
英伟达（NVDA.US）首次覆盖：AI 赛道坡长雪厚，增长可期 .....	40
英伟达公司速览 .....	42
估值 .....	44
SPDBI 乐观与悲观情景假设 .....	46
风险提示 .....	48
公司背景 .....	49
财务报表 .....	50
高通（QCOM.US）首次覆盖：端侧 AI 推动公司开启大成长周期 .....	53
高通公司速览 .....	55
估值 .....	57
SPDBI 乐观与悲观情景假设 .....	59
风险提示 .....	61
公司背景 .....	62
财务报表 .....	63



## 全球 AI 算力行业首次覆盖：从云到端， 云端协同，AI 开启科技行业超级成长周期

- 云侧 AI 算力需求持续高增长，端侧 AI 渗透率快速提升，启动新一轮科技增长周期：**自 2022 年底 Chat GPT 面向公众开放以来，生成式 AI 带动了新一轮的生产效率提升，也开启了新一轮 AI 投资机遇。根据 IDC 数据，2020 年至 2023 年全球生成式 AI 市场空间上涨了约 6 倍，预计 2024 年至 2030 年期间复合增长率达到 40%，2030 年有望接近万亿美元。作为 AI 大模型行业最重要支撑的算力芯片也正享受行业高增长红利。因此，我们首次覆盖全球 AI 算力行业，给予“超配”评级；考虑到英伟达受益于云侧 AI 高增长以及智能驾驶端侧渗透，高通受益于最大的两个终端——智能手机和新能源车的端侧 AI 渗透，首次覆盖英伟达 (NVDA.US) 和高通 (QCOM.US)，均给予“买入”评级。
- 全球 AI 算力行业仍然处于发展初期阶段，短中长期都具备扎实的成长动能：**AI 大模型行业正处于发展初期、包括巨头和初创企业在内的大量厂商百家争鸣的阶段。因此，AI 行业仍然处于由大模型厂商，即供应端，推动 AI 大模型快速迭代、性能加速提升、行业高速发展的时期。因此，大模型供应商，即 AI 算力芯片的直接需求方，正在大幅采购 AI 算力芯片。在当前逐步步入供需平衡的阶段，AI 算力芯片产能扩张会带来行业的有效增量，仍然是行业增长的“甜蜜期”。
- 端侧 AI 算力需求有望快速提升，并反哺云侧 AI 的算力需求：**AI 大模型从云侧向端侧渗透和延伸，会有效拓展 AI 大模型在 C 端用户的使用。AI PC、Gen-AI 智能手机、高阶智驾新能源车、ARVR，都是较强的触及用户的端侧 AI 落地载体。AI 大模型正在端侧加速渗透。端侧 AI 渗透有望加速 AI 大模型商业闭环以及厂商盈利能力提升。从今明年年的时间维度看，能够支持端侧 AI 大模型算力的芯片都将取得较高增速。这些端侧 AI 需求会在一定程度拉动整体 AI 算力芯片训练端的需求，拓展 AI 算力芯片的增长边界。
- 美股 AI 算力半导体估值仍有上行空间：**我们认为当前美股 AI 算力芯片依然还有小量的上行空间。这主要来自于两个方面的动能：1) 半导体周期基本上行的时间和空间都能给与支撑，2) AI 大模型初创企业传递到二级市场的估值溢价。目前，英伟达的未来 12 个月市盈率为 43.6x，高通的未来 12 个月市盈率为 19.1x。这两家公司的估值不便宜，但是估值距离过去 5 年周期顶部的峰值还有小量距离，更何况 AI 有望推动估值溢价超过历史峰值。最后，美国进入降息周期叠加 AI 带动的科技产业变革都有望推动美股 AI 算力半导体的估值中枢上移。
- 投资风险：**全球或美国经济面临增长压力，多个下游终端，包括服务器、智能手机、新能源车等，需求不及预期；AI 大模型需求爆发持续性弱于预期，大模型厂商商业闭环或盈利能力低于预期；全球半导体周期上行动能不足；行业竞争加剧，拖累利润表现；AI 算力芯片迭代不及预期。

沈岱

首席科技分析师

tony\_shen@spdbi.com

(852) 2808 6435

马智焱

科技分析师

ivy\_ma@spdbi.com

(852) 2809 0300

黄佳琦

科技分析师

sia\_huang@spdbi.com

(852) 2809 0355

2024 年 7 月 2 日

英伟达(NVDA.US)

买入

目标价(美元)

147.6

潜在升幅/降幅

+19%

目前股价(美元)

124.0

高通(QCOM.US)

买入

目标价(美元)

240.7

潜在升幅/降幅

+23%

目前股价(美元)

195.2

注：目前股价截至 2024 年 6 月 27 日

资料来源：Factset，浦银国际



# 全球 AI 算力行业投资要点

2022 年 11 月底，OpenAI 发布 ChatGPT 3.5。紧随其后，AI 相关的行业在 2023 年进入爆发式发展期。无论是 OpenAI 随后发布的 ChatGPT 4.0，还是谷歌、Meta 等多家厂商发布的 Gemini、Llama 等 AI 大模型，正是这一轮 AI 行业爆发式发展的体现。中国的玩家，包括互联网头部企业和初创公司，都相继投入到大模型的研究和发布中。根据调研，中国的 AI 大模型的能力有望在今年年底逼近 ChatGPT4.0。

我们清晰地看到，大模型的到来成为了 AI 行业发展的奇点，当前行业处于加速上行的高速发展阶段。这个阶段比较接近当年 iPhone 4 发布之后，大量安卓品牌相继跟进进入智能手机行业的阶段，也同样比较接近特斯拉进入中国量产后，多家中国新势力、传统车企的新能源车品牌相继涌入该赛道的阶段。因此，我们对 AI 行业未来的成长趋势持相当程度的乐观态度。

同时，AI 大模型正在快速向新能源车、智能手机等端侧快速渗透，从而促进 AI 进入 C 端领域，并进一步打开市场空间。

因此，我们在这篇报告中试图回答三个问题：1) 在 AI 大模型发展大爆发之后，目前行业发展处于怎样的阶段，是否还有上行空间？2) AI 大模型会给端侧带来多少的空间增量，可以拉动多大的存量空间？3) 如何看待当前全球头部 AI 玩家的预期、估值和股价空间？

**第一，在 AI 大模型发展大爆发之后，目前行业发展处于怎样的阶段？短中长期，AI 算力行业是否还有较大的上行空间？**

我们认为 AI 行业仍然处于发展较为早期的快速增长阶段。这个基本判断可以从两个方向来看，一是最能贴近 AI 行业的服务器出货量中 AI 的渗透率，目前只有个位数；二是全球相关的资金/资本正在快速涌入，包括巨头和初创公司。

这两个指标都将推动 AI 大模型在供应端为保持性能和技术领先进入快速迭代和发展，从而推动行业大幅上行。在这个过程中，AI 大模型背后所需要的 AI 算力服务器/芯片都会处于需求持续增长期。

根据 MIC，2023 年全球 AI 服务器出货量超过 125 万台，同比增长 47%，预计 2024 年 AI 服务器出货量将达到 194 万台，同比增长 55%，同比增速上行。根据中商产业研究院，中国 AI 服务器 2023 年规模达 134 亿元，同比增长约 101%，预计 2024 年中国 AI 服务器规模达 307 亿元，同比增长 128%，同比增速也在加速上扬。

从渗透率来看，2022 年、2023 年、2024 年，全球 AI 服务器出货量渗透率达到/预计达到 5%、6%、8%（图表 1）。与科技新兴产业发展趋势类似，目前 AI 服务器渗透率也处于加速上扬阶段。展望未来展望，AI 服务器渗透率有望持续上扬。

从大模型供应商的角度，即 AI 算力芯片的直接需求方，正在大量采购 AI 算力芯片。在近期新闻中，英伟达最新出货的 GB200 上调订单预测。大模型厂商为了取得行业初期成长红利，以及潜在进入成熟期之后更大的份额，需要保持自身产品性能领先和体验的差异化。

AI 算力背后最核心的是 GPU。GPU 是重要的计算芯片，也是重要的半导体逻辑芯片。AI 服务器渗透率的提升以及市场资本持续的涌入确实可以作为中长期行业增长动能的判断指标。但是，即使如此，分析 AI 算力芯片这种处于初期的高成长行业时，我们依然会在意半导体产业的周期性动能带来的影响，这也是短期去跟踪 AI 算力芯片行业的重要参考。**短期看，AI 算力芯片增长动能强劲，下行风险较小。**

首先，在经历了 2023 年 AI 大模型算力爆发增长带来的 AI 算力芯片供不应求之后，AI 算力芯片大体上逐步进入供需平衡的阶段，在局部（例如 GB 200）存在供应紧张。这个阶段仍然是行业增长的“甜蜜期”，尽管公司/行业基本面持续超出较高市场预期的空间缩小，但是仍然有较高动能推动成长，而且收入端的成长可以有效传递为利润端的增长。（图表 2）

根据市场信息，目前台积电的 CoWoS 的产能约 1.5 万片/月，并有望在今年年底提升至 3 万片/月。作为英伟达 AI 芯片的瓶颈之一的先进封装产能处于满载满负荷和加速扩产阶段。因此，我们相信 AI 算力芯片行业的短期动能依然强劲。（图表 3）

其次，库存作为半导体供需关系中一个重要指标，是衡量行业短期动能的参考之一。目前，英伟达的库存周转天数为 89 天，处于比较低的位置，说明产品的拉货动能比较强。因此，短期来看，AI 算力芯片维持在较为强劲的高位。（图表 5）

**从长期的时间维度，例如 3-5 年甚至更长，AI 算力行业的竞争格局会产生演变。**

当前 AI 大模型处于百家争鸣的阶段，行业初期红利叠加大量市场/巨头资金支持，推动行业加速发展。高增长、高容量的行业也会孕育较激烈的竞争环境。根据浦银国际互联网团队的调研结果，今年以来 AI 大模型厂商收费单价大幅度下滑，较降价前下滑 90%。当然，目前阶段并没有进入 AI 大模型厂商的“生死”时刻。

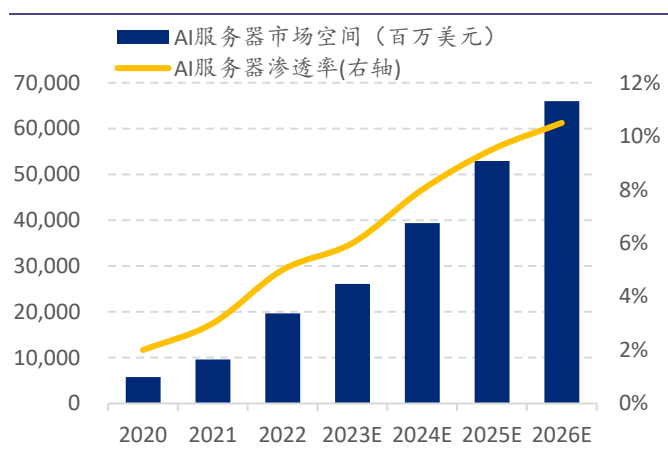
在行业商业模式落地和商业变现的过程中，行业对 AI 大模型厂商综合能力，包括技术、产品、营销、管理，提出更高的要求，也对厂商的利润能力提出要求。在这个阶段，部分或者大量的 AI 大模型厂商退出行业。

这会在一段时间内对 AI 算力芯片行业造成大量供给释放的冲击。即头部的 AI 大模型厂商有望以较低的价格获取已经退出市场的 AI 大模型玩家的算力，从而导致 AI 算力芯片短期的拉货动能下滑。我们认为在这个阶段，投资人应该保持更加谨慎的投资态度，等待行业消化剩余产能。

从AI算力芯片行业来看，英伟达接近垄断的市场格局也可能因为市场规模的大幅增长而发生改变。一方面，头部的AI大模型玩家会有意愿开发并自研AI算力芯片以取得客制化差异化性能和成本优势。特斯拉的Dojo超算、谷歌的TPU芯片、百度的昆仑AI芯片等，都存在取代/部分取代当前英伟达算力芯片的可能。另一方面，在AMD以及其他初创公司GPU等AI算力芯片性能存在部分优势的情况下，AI大模型厂商也会有意愿增加非英伟达芯片使用量以促进AI算力芯片的竞争，平衡供应商之间的成本。

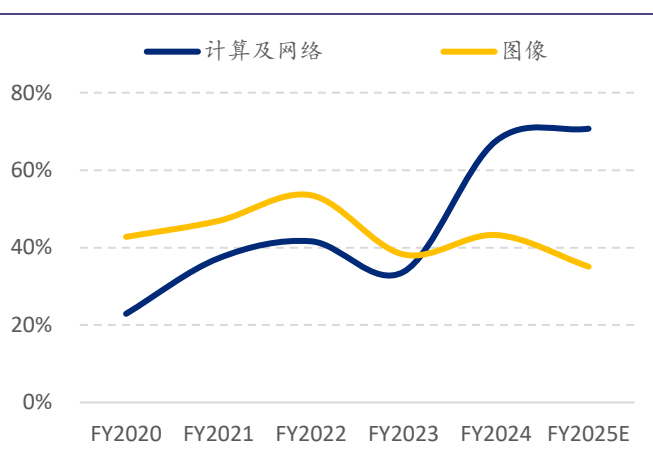
当然，从长期维度看，尽管英伟达的AI算力芯片份额下行风险大于维持近乎垄断地位的风险，但是基于头部优势，英伟达在维持大多数产品的性能和成本优势的同时保持相对健康的利润率是比较有保障的。

图表 1: AI服务器渗透率



注: E=IDC 预测  
资料来源: IDC、浦银国际

图表 2: 英伟达营业利润率



注: E=浦银国际预测, FY=Fiscal Year, 财务年  
资料来源: 公司公告、浦银国际

图表 3: 台积电 CoWoS 产能预测

台积电 CoWoS 产能预测	1H23	2H23E	2024E	2025E	2026E
万片/月	0.8	1.2	3	4.8	4.8
每片 CoWoS 晶圆芯片数量假设	29	29	28	26	25
总数 (万片/月)	23.2	34.8	84.0	124.8	120.0

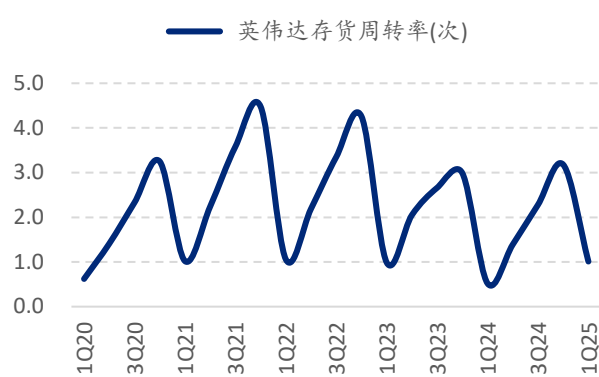
注: E=根据市场调研整理  
资料来源: DigiTimes, 台积电, 浦银国际

图表 4：英伟达存货周转天数



注：时间为英伟达财年（Fiscal Year），1Q25FY 截至 2024 年 4 月 28 日  
资料来源：iFind、公司公告、浦银国际

图表 5：英伟达存货周转率



注：时间为英伟达财年（Fiscal Year），1Q25FY 截至 2024 年 4 月 28 日  
资料来源：iFind、公司公告、浦银国际

## 第二，AI 大模型会给端侧带来多少的空间增量？端侧 AI 的应用可以拉动多大的存量空间进入增长？

端侧 AI 大模型会逐步拉动 AI 算力需求，并且逐步扩大对于云侧 AI 算力需求量。端侧对于 AI 大模型的需求主要有两个来源：一是对于 AI 大模型厂商，例如 OpenAI 的技术需求，二是端侧品牌自身构建更贴近用户需求的 AI 大模型。

由于 AI 大模型仍然处于发展的初期阶段，AI 大模型的商业闭环尚未完美形成。根据我们的估算，OpenAI 在 2023 年年化收入已经达到 10 亿至 20 亿美元，但是距离整体盈利仍有较大距离。因此，AI 大模型厂商仍然需要一定时间才能形成规模化稳定的盈亏平衡或者盈利。

AI 大模型从云侧向端侧渗透和延伸，会有效拓展 AI 大模型在 C 端用户的使用。端侧载体，包括电脑、智能手机、新能源车、ARVR 等，都能有效触达最广泛的人群。

以智能手机为例，全球智能手机用户超过 60 亿人，因此 AI 大模型在该终端的渗透和普及可以更有效地触达 C 端用户，促使 AI 大模型从供应驱动向需求驱动转型。同时，对于 C 端用户的有效触达，更有可能形成商业闭环。

此时，商业模式有两种。一种类似当前 ChatGPT 的付费使用模式，即消费者直接付费使用性能更优的 AI 大模型。另一类是将 AI 大模型的成本加入智能手机、新能源车等终端售价中。

微软、苹果借助与 OpenAI 的合作，快速地将 AI 大模型带入自身电脑和手机等电子终端，从而给消费者带来生产力的提升和用户体验的改善。

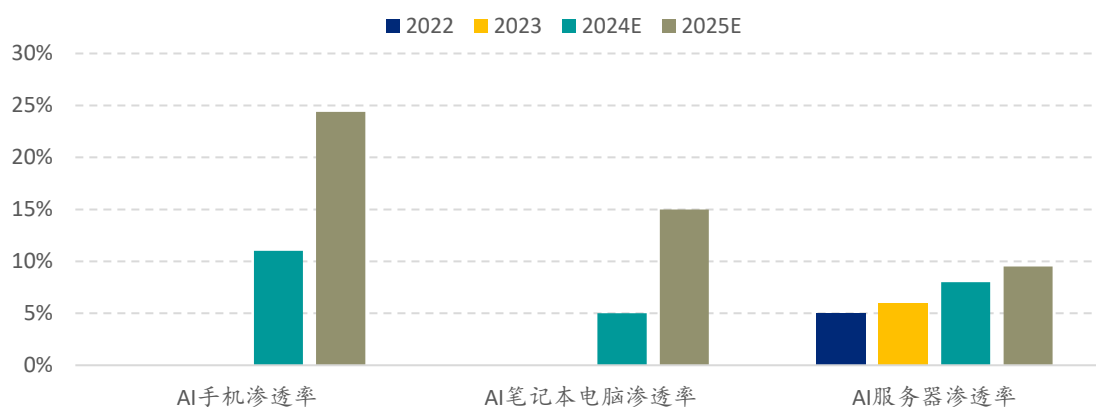
我们认为 AI PC、Gen-AI 智能手机、高阶智驾新能源车、ARVR，都是较强的触达用户的端侧 AI 落地载体。**AI 大模型正在端侧加速渗透。**

- 根据 Counterpoint 的预测，2024 年、2025 年具备高阶 AI 能力的笔记本电脑渗透率将达到 5%、15%。生成式 AI 最有望在笔记本电脑作为生产力工具。

- 根据 Counterpoint 的预测，2024 年、2025 年具备生成式 AI 能力的智能手机渗透率将达到 11%和 24%。作为触达 C 端用户最广的电子终端，AI 将有望带动智能手机换机需求，从而反哺对于云侧的 AI 算力的需求增长。
- 根据车百智库，2023 年中国高阶智驾的 NOA 的渗透率为中低个位数，而 2024 年该渗透率有望到达两位数。我们认为中国新能源车企对于智驾的推动将是带动 AI 大模型落地端侧最快最大的应用场景。并且，中国新势力已经搭建自身的 AI 算力平台，以促进 AI 大模型在智能座舱和智驾的应用。
- AR 作为热度很高且也正在不断发展的电子计算平台，与 AI 大模型的结合，相互促进、相互提高，对行业扩张有很大帮助。

所以，从今明两年的时间维度看，能够支持端侧 AI 大模型算力的芯片都将取得较高增速，包括智能手机上高通骁龙 Gen 系列、联发科天玑系列，智能驾驶中使用的英伟达 Orin 或 Thor 等。同时，这些端侧 AI 需求会一定程度拉动整体 AI 算力芯片训练端的需求，拓展 AI 算力芯片的增长边界。

图表 6：端侧 AI 渗透率



E= AI 服务器渗透率为 IDC 预测，AI 手机、AI PC 渗透率为 Counterpoint 预测；AI 笔记本电脑指具备高阶 AI 功能的笔记本电脑  
资料来源：IDC, Counterpoint, 浦银国际



### 第三，当前全球头部的 AI 玩家的股价已经反映了多少市场预期？如何看待当前全球头部 AI 玩家的估值和股价空间？

我们认为回答这个问题的核心在于海外 AI 主题相关的标的热度已经持续了很长一段时间，在高成长的预期中，估值也没有那么便宜，那么此时估值会不会面临较大的下行风险？

我们认为当前美股 AI 算力芯片的估值依然有小量的上行空间。AI 算力的估值可以拆解为 2 个部分：1) 半导体周期层面，2) AI 大模型初期爆发层面。

首先，从半导体周期层面来看，AI 算力芯片，不管是云侧还是端侧，都处于全球半导体上行周期中。而 AI 算力增量是本轮全球半导体上行的最重要的动能。我们观察到费城半导体指数在 2022 年 10 月触底上行以来，已经持续 20 个月，跟历史上约 2 年的上行周期比较，还有一些上行的空间和机会。但是，由于 AI 大模型对于多个科技子行业的带动，本轮周期估值上行的高度可能会较以往峰值更高。

另外，全球半导体基本面周期，即全球半导体月度销售额同比增速，在 2023 年 4 月触底，随后开启迅速上行。这轮基本面上行的斜率比较陡，可以支撑更高的估值峰值。同时，本轮半导体基本面上行才 12 个月，还有接近一年的上行空间。非 AI 相关半导体基本面上行将推动估值上行，给与 AI 相关的估值高溢价做好支撑。

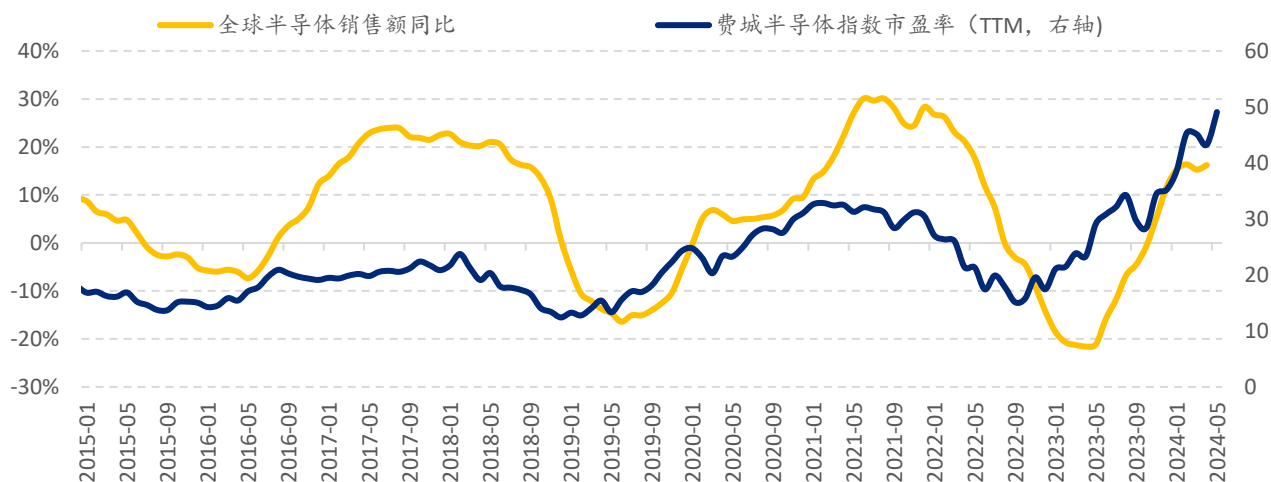
其次，本来 AI 行业再次爆发，不仅将带动各家巨头入局，也同样带动 OpenAI 等初创企业的估值水涨船高。根据新闻报道，2024 年 OpenAI 最新一轮的估值已经达到 800 亿美元。如果 OpenAI 或其他 AI 大模型初创企业在近两年上市，有可能会带动二级市场 AI 板块估值中枢上移，已经上市的 AI 标的估值则有进一步上行空间。

目前，英伟达的未来 12 个月市盈率为 43.6x，高通的未来 12 个月的市盈率为 19.1x。这两家公司的估值不便宜，但是估值距离过去 5 年周期顶部的峰值还有小量距离。进一步看，我们认为本轮 AI 带动科技产业革命有望带动行业估值突破历史新高。

另外，美国处于利率高点对于高成长的科技股的估值不利。未来美国进入降息通道后，美股的科技股有望享受估值加成上扬。

因此，综上所述，我们认为在 AI 算力芯片行业，云侧需求短期供需平衡基本面推动股价上行，端侧需求拉动新增应用领域有望持续贡献增量，估值不便宜但是依然具备上行空间。

图表 7：全球半导体三个月移动平均值销售额同比增速与费城半导体指数市盈率



资料来源：Wind、美国半导体协会、费城证券交易所、浦银国际

图表 8：海内外头部初创 AI 企业融资情况

时间	公司	募资	估值
2024 年 5 月	智谱 AI	智谱 AI 获 Prosperity7 Ventures 4 亿元美元 C 轮投资	200 亿人民币
2024 年 5 月	Scale AI	AI 数据标注公司 Scale AI 完成 10 亿美元的募资	138 亿美元
2024 年 5 月	xAI	马斯克为 xAI 投入 60 亿美元的融资	180 亿美元
2024 年 5 月	Kimi 月之暗面	截至 5 月，阿里巴巴 2024 财年合计投入 8 亿美元	25 亿美元
2024 年 4 月	Anthropic	亚马逊宣布完成了对 Anthropic (Claude) 总计 40 亿美元的融资。2023 年 9 月亚马逊提供了一笔 12.5 亿美元的投资，之后又追加了 27.5 亿美元	184 亿美元
2023 年 11 月	零一万物	完成 10 亿美元的新一轮融资，由阿里云领头	NA
2023 年 10 月	Anthropic	谷歌领投了 Anthropic 的 5 亿美元融资，此后又追加了 20 亿美元的投资	NA
2023 年 4 月	OpenAI	微软领投了约 100 亿美元，多家 PE 投资超 3 亿美元	800 亿美元

资料来源：公开资料整理，浦银国际

# 全球 AI 算力行业概览：发展初期，高速增长

## ● 全球 AI 大模型具备较大的成长空间

人工智能（AI, Artificial Intelligence）指通过计算程序或机器来模拟实现人类智能的技术和方法。伴随着 OpenAI 将人工智能再次推上风口浪尖，二级市场股价大幅上扬。

本轮 AI 大模型的技术，与前几年的 AI 技术路径有所不同。此前，AI 技术中提及比较多的是卷积神经网络（CNN, Convolutional Neural Network），在图像识别等领域已经较为成熟，更多还是规则制。而 OpenAI 则使用大语言模型（LLM, Large Language Model），针对自然语言处理（NLP, Natural Language Processing），实现文本生成、翻译、问答等能力。

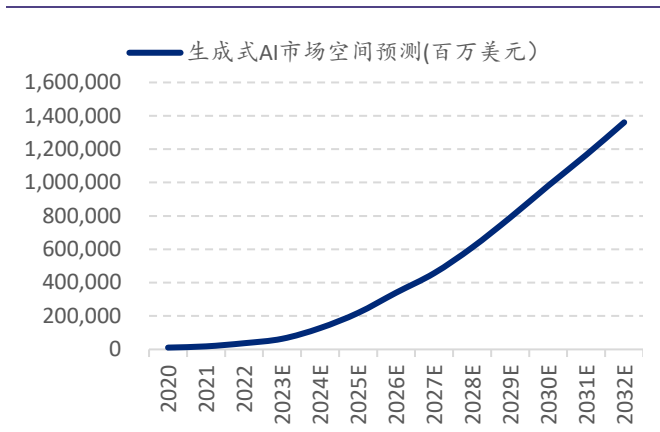
OpenAI 发布的 ChatGPT 的核心是 Transformer 架构，特别是自注意力机制，这使得模型能够并行处理输入数据，理解和生成具有上下文依赖性的文本。而这与前几年提及比较多的卷积神经网络技术有差异。此前，图片识别技术依赖于卷积神经网络等深度学习算法，通过特征提取和分类器来识别图像内容。

ChatGPT 通过预训练和微调的方式进行训练，预训练阶段使用大量文本数据，微调阶段则针对特定任务进行优化。因此，大模型数量和参数提升推动 AI 服务需求提升。

大模型推陈迭代，参数及语料提升带动算力需求提升。过去几年，大模型参数呈现出指数型增长。为了实现更高的精度，大模型参数从 2018 年 GPT-1 的 11.7 亿，提升到 2020 年 GPT-3 的 1,750 亿，再到 2023 年 GPT-4 的约 1.8 万亿。训练集大小也由 GPT-3 的 3,000 亿 tokens 显著提升至 GPT-4 的 12 万亿 tokens。我们预计未来模型参数将持续高速增长，从千亿级到万亿级提升。

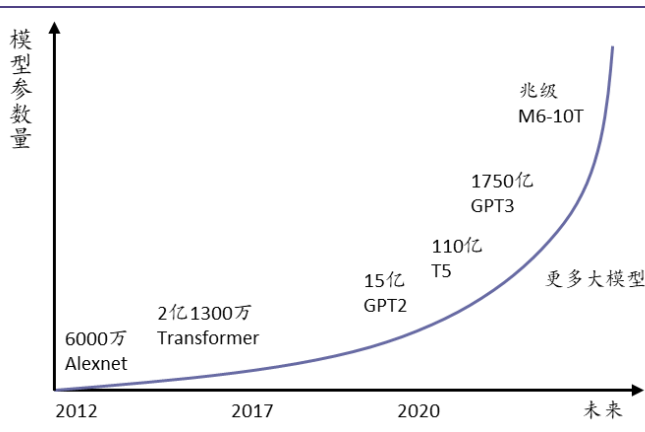
从长期展望来看，全球生成式 AI 有望迎来万亿美元市场空间。自 2022 年底 Chat GPT 面向公众开放以来，生成式 AI 带动了新一轮的生产效率提升，也开启了新一轮 AI 投资机遇。根据 IDC 数据，2020 年至 2023 年全球生成式 AI 市场空间上涨了约 6 倍，预计 2024 年至 2030 年期间复合增长率达到 40%，2030 年有望接近 10,000 亿美元。

图表 9：全球生成式 AI 市场空间及预测



注：E=IDC 预测  
资料来源：IDC、浦银国际

图表 10：模型数量与参数爆发性成长



资料来源：联发科发布会、浦银国际

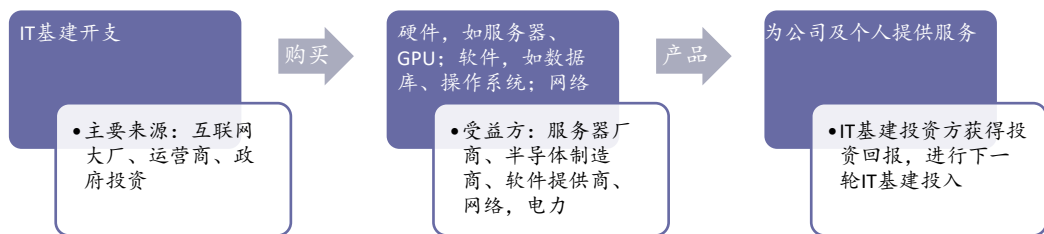
经过分析，我们认为 IT 基建开支、服务器收入、云服务收入，可以用来追踪 AI 产业链的成长趋势和动能。

从商业模式来看，AI 产业链由三个部分组成。最上游为 IT 基建开支。IT 基建开支主要来自于互联网企业、电信运营商和政府投入。IT 基础设施主要包括用于传输、存储、管理、分析数据等，为用户提供数据计算、网络连接、网络安全等功能。

中游环节涉及 IT 基建开支投入的具体方向，基本组成为硬件、软件、网络连接。硬件主要包括服务器、数据中心、个人终端设备；软件主要包括管理系统和应用程序；网络包括路由器、交换机等。中游环节的主要受益厂商包括服务器、半导体、光模块等。

下游应用环节，包括 Chat GPT、端侧 AI 大模型、互联网企业 CSP (Cloud Solution Provider) 云服务商。IT 基建的投资者在下游应用推广的过程中受益获得回报，并依据回报的多少和对未来的预期再进行下一轮 IT 基建投入。

图表 11：AI 产业链模式闭环



资料来源：浦银国际

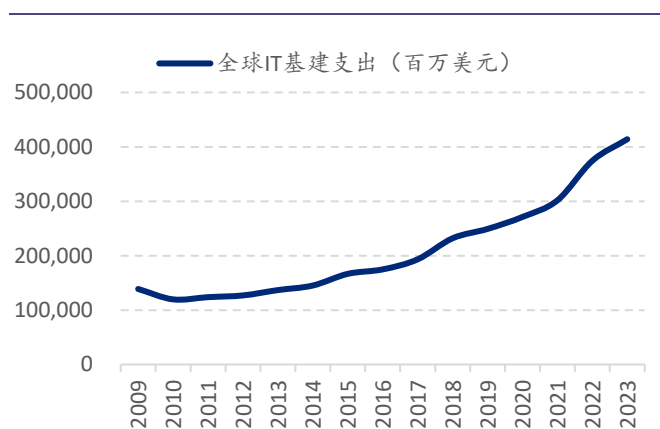


全球 IT 基建支出正在加速上行。2023 年美国四大互联网厂商（谷歌、微软、亚马逊、Meta）资本开支在全球 IT 基建占比超三分之一。全球 IT 基建支出由 2009 年的 1,388 亿美元增长至 2023 年的 4,140 亿美元（图表 12）。其中，2017 至 2023 年进入加速上行阶段，期间复合增长率达到 14%。美国四大互联网厂商在全球 IT 基建中占比最高达到 42%，2023 年占比 36%，是推动全球 IT 基建上行的重要力量。

美国四大互联网厂商资本支出的同比增速和全球 IT 基建支出同比增速大体保持一致。美国的头部四家资本开支存在一定的周期。从季度频率看，这四家公司资本支出同比增速在经历了 10 个季度的下行后，在 2023 年二季度见底触及-9%的增速，随后就开启上行，并在今年一季度实现+30%增长（图表 15）。本轮美国互联网大厂资本开支投向主要集中在 AI 算力的部署。

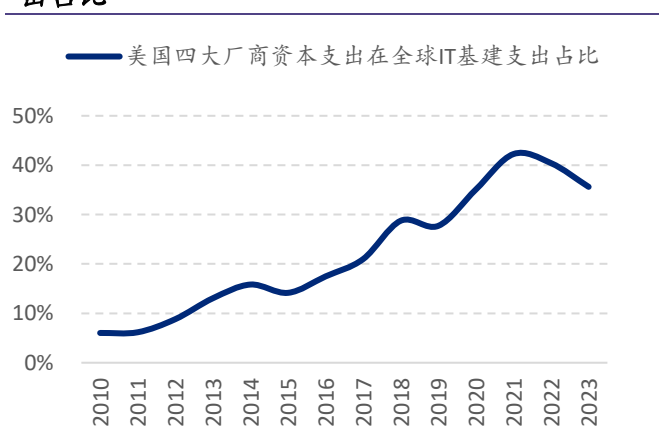
我们预期美国互联网厂商以及其他初创企业都在加速部署 AI 大模型算力，以避免在新技术商业落地过程中落后于同行和竞争对手。所以，我们预期这些企业的资本开支上行动能会推动 AI 算力需求大幅上扬。

图表 12：全球 IT 基建支出



资料来源：IDC、浦银国际

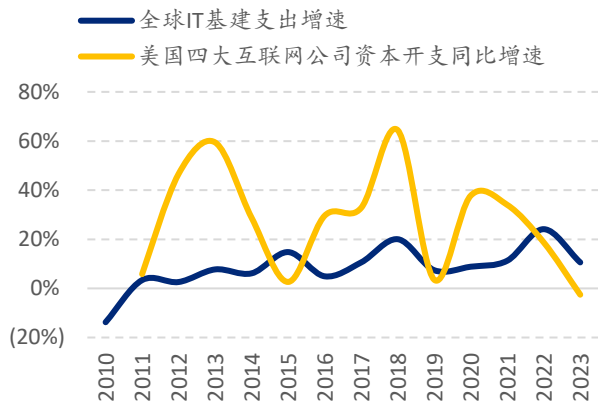
图表 13：美国四大厂商资本支出在全球 IT 基建支出占比



注：互联网厂商资本开支范围与 IT 基建较为一致

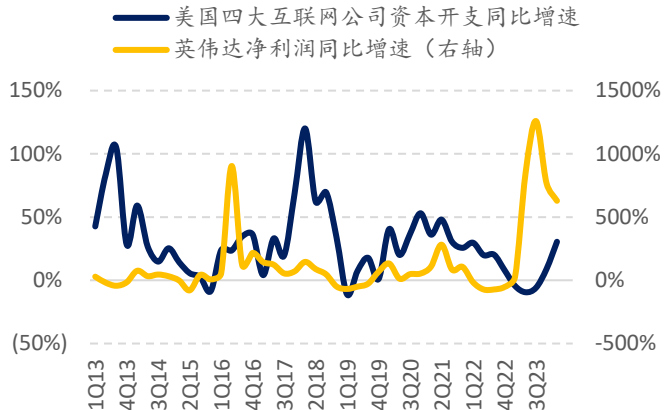
资料来源：IDC、浦银国际

**图表 14: 美国四大互联网厂商资本支出的同比增速和全球 IT 基建支出同比增速周期**



注：互联网厂商资本开支范围与IT基建较为一致  
资料来源：IDC、Factset、浦银国际

**图表 15: 英伟达净利润同比增长与美头部互联网公司资本开支相关**

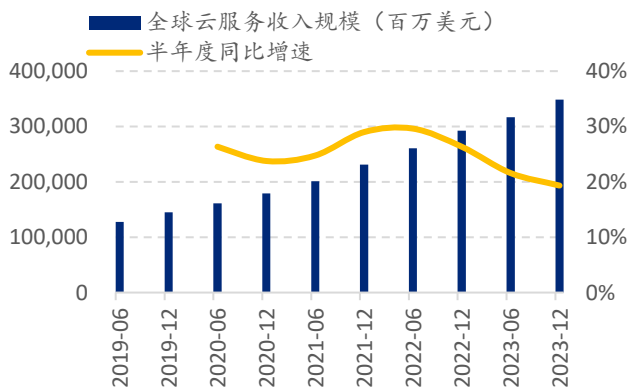


资料来源：IDC、Factset、浦银国际

全球云服务收入稳步增长，过去四年全球云服务收入年复合增长率达 25%。2019 年至 2023 年，全球云服务年收入规模由 2,724 亿美元增长至 6,655 亿美元（图表 16），期间年复合增速高达 25%，2023 年下半年全球云服务同比增速 19%。全球云服务收入最大的三家公司及份额分别为微软 17%、亚马逊 12%、谷歌 5%，CR3 合计份额为 34%（图表 17）。根据 IDC 预测，2024 至 2028 年全球云服务收入有望从 8,000 亿美元增长至 1.6 万亿美元，期间复合增长率 19%。

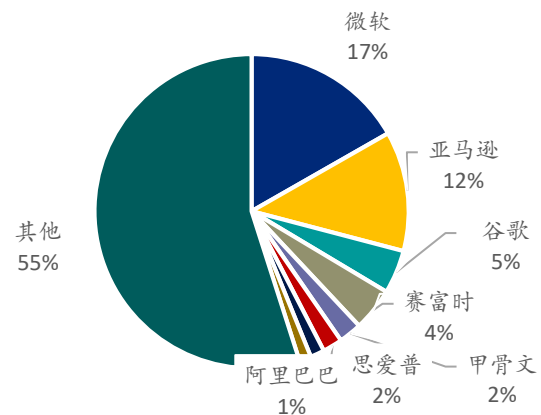
从经济角度出发，并不是所有的 AI 大模型厂商都适合自建 AI 算力。在考虑成本和规模效应时，规模较小的 AI 大模型厂商将可能使用租赁算力的方式来打磨自身大模型能力。因此，我们认为全球云服务行业规模和 AI 算力行业规模都将保持同步增长。

**图表 16: 全球云服务市场规模**



资料来源：IDC、浦银国际

**图表 17: 全球云服务主要玩家**

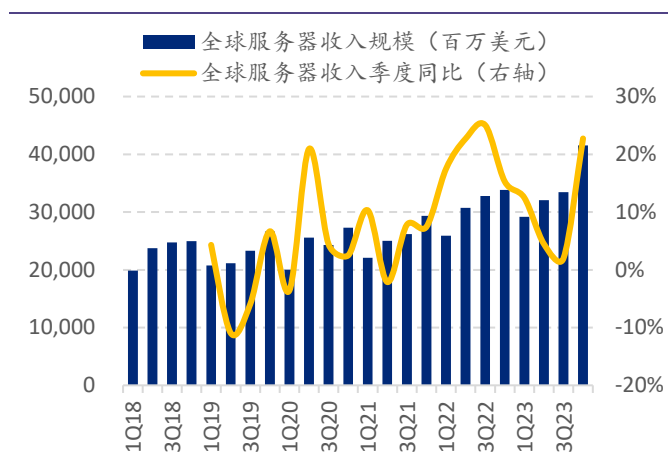


资料来源：IDC、浦银国际

**AI 应用助力服务器市场加速增长。**2018 年至 2023 年，全球服务器收入规模由 934 亿美元上升至 1,362 亿美元，期间复合增长率为 8%。其中，2022 年四季度 Chat GPT 推出后，全球服务器收入规模加速上行，4Q23 对比 4Q22 同比增长 23%。服务器市场中，除去占比最高的原厂直销的白牌服务器外，收入占比前三的玩家及份额分别为戴尔 12%，超微电脑 8%，惠普 7%。

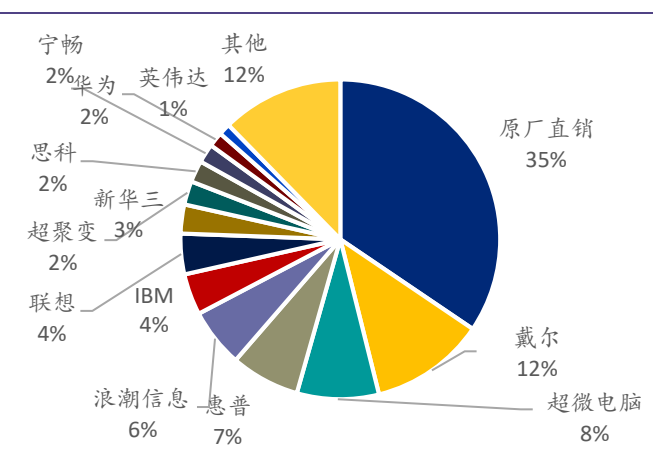
根据 IDC 预测，2024 年、2025 年全球 AI 服务器收入规模分别为 393.5 亿美元和 529.4 亿美元，同比增长 51%、35%，对应渗透率 8%、9.5%。

图表 18: 全球服务器市场规模



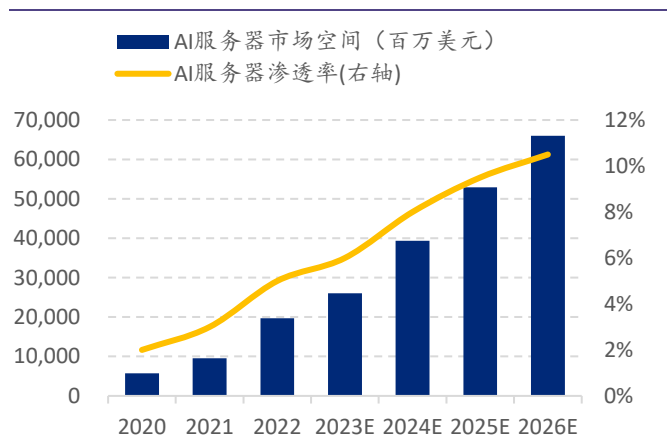
资料来源: IDC、浦银国际

图表 19: 全球服务器主要玩家



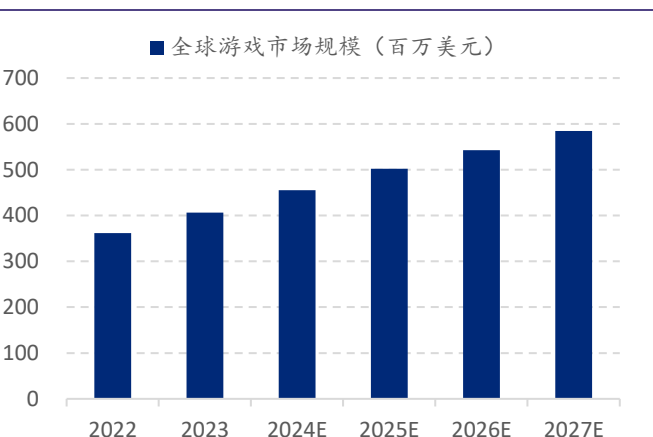
资料来源: IDC、浦银国际

图表 20: 全球 AI 服务器市场规模



注: E=IDC 预测  
资料来源: IDC、浦银国际

图表 21: 全球游戏市场规模



注: E=Statista 预测  
资料来源: Statista、浦银国际

## ● AI 算力及 GPU 需求测算：需求空间仍旧巨大

### 训练侧：

影响 AI 训练端算力需求的参数主要有 2 个，模型参数量和训练集大小。根据 OpenAI 的 Kaplan 发布的论文 *Scaling Laws for Neural Language Models*，AI 模型训练所需的浮点运算次数计算公式为： $6 \times \text{模型参数规模} \times \text{训练集大小}$ （PFLOPS）。根据 OpenAI，ChatGPT-3 包括 1,750 亿参数，在 3,000 亿 tokens 的训练集中完成训练。通常，1,000 个 tokens 约合 750 个英文单词，约合 500 个汉字。

我们假设 GPT-3 由英伟达 A100 训练完成。由于 GPU 参数标注为理想条件下的最高算力，我们需要效率乘数来还原真实的算力及服务器需求数量。GPU 的有效算力比率取决于 GPU 硬件型号及大语言模型架构等多方面因素。根据 OpenAI 论文，英伟达 A100 在 GPT-3 上的有效算力比率为 21.3%。根据英伟达官网，A100 服务器包含 8 个 A100 GPU，最高性能可达 5 petaFLOPS。由此，我们计算出 GPT-3 需要单个英伟达 A100 服务器计算 3,423 天。

根据我们的产业链调研和专家访谈，业内拥有 200-300B（B=bn）参数的大语言模型，在英伟达 A100 上通常需要 30 多天的训练。我们假设 GPT-3 目标训练用时为 30 天，如果同时训练一个模型，那么我们对 A100 服务器的需求为 114 台，折合 912 块 A100 GPU。

**ChatGpt3 向 ChatGpt4 的迭代，训练集大小提升拉动算力需求明显提升。**根据 SemiAnalysis 的测算，GPT-4 采用了 1.8 万亿参数，训练集大小约 12 万亿个 tokens。我们假设 GPT-4 采用 H100 服务器，有效算力比率为 30.2%，如果想要在 30 天内完成训练，需要的 H100 服务器数量上升至 5,174 个，H100 GPU 需求数量上升至 41,391 个。

2023 年 12 月，谷歌发布 Gemini 大模型。Gemini 的训练数据库为 Youtube 上 93.6 亿分钟的视频字幕，根据 SemiAnalysis 的测算，总数据集大小约为 GPT-4 的两倍。2024 年 4 月，Meta 旗下 Llama-3 发布，采用了 4,000 亿参数量和 15 万亿的训练集，数据量是 Llama-2 的七倍。模型训练集大小的快速提升，拉动训练算力需求及 AI 服务器需求成比例的快速提升。

### 推理侧：

**应用级爆款 AI 将会推动训练侧算力大幅增长。**根据 OpenAI 论文，AI 模型推理算力需求和模型参数规模及训练集大小正相关。公式为： $\text{AI 模型训练所需的浮点运算次数} = 2 \times \text{模型参数规模} \times \text{训练集大小}$ 。

根据 Similarweb，2024 年 5 月前三周 GPT-4 的日均访问量为 7,700 万次。我们假设单次对话 token 数量为 500（约 375 个单词、250 个汉字），每次访问对话次数为 10 轮。由于 GPT 返回答案的时长对用户体验的影响较大，我们假设目标用户等待时间为 1 秒，那么仅仅 GPT-4 就需要等同于 424,945 个 H100 GPU 的算力进行推理运算。我们预计，未来伴随着应用级爆款 AI 的 APP 的出现，访问人数、访问时长的激增还将推升推理侧 AI 训练集和推理侧算力的需求大幅增长。



图表 22：海外大模型训练侧算力需求

模型名称	GPT-3	GPT-4	Gemini	Llama 3
发布时间	2020年6月	2023年3月	2023年12月	2024年4月
参数量（亿）	1,750	18,000	18,000	4,000
假设训练集大小（亿 tokens）	3,000	120,000	240,000	150,000
<b>AI模型训练所需的浮点运算次数：6×模型参数规模×训练集大小（PFLOPS）</b>	<b>315,000,000</b>	<b>129,600,000,000</b>	<b>259,200,000,000</b>	<b>36,000,000,000</b>
使用硬件假设	NVIDIA A100	NVIDIA H100	NVIDIA H100	NVIDIA H100
有效算力比率假设	21.30%	30.20%	30.20%	30.20%
<b>实际算力需求（PFLOPS）</b>	<b>1,478,873,239</b>	<b>429,139,072,848</b>	<b>858,278,145,695</b>	<b>119,205,298,013</b>
服务器峰值算力（PFLOPS）	5	32	32	32
单位服务器包含 GPU 个数	8	8	8	8
训练时间需求（秒）	295,774,648	13,410,596,026	26,821,192,053	3,725,165,563
约合天数	3,423	155,215	310,430	43,115
<b>30天完成训练需要服务器数量</b>	<b>114</b>	<b>5,174</b>	<b>10,348</b>	<b>1,437</b>
<b>30天完成训练需要 GPU 数量</b>	<b>913</b>	<b>41,391</b>	<b>82,781</b>	<b>11,497</b>

资料来源：Scaling Laws for Neural Language Models, Semianalysis, 各公司官网, 浦银国际

图表 23：海外大模型推理侧算力需求

模型名称	GPT-3	GPT-4	Gemini	Llama 3
参数量（亿）	1,750	18,000	18,000	4,000
假设单次对话训练集大小（tokens）	500	500	500	500
假设日均访问数量（百万次）	53	77	77	77
假设每次访问对话轮数	10	10	10	10
<b>AI模型训练所需的浮点运算次数: 2× 模型参数规模×训练集大小 (PFLOPS)</b>	<b>92,750,000</b>	<b>1,386,000,000</b>	<b>1,386,000,000</b>	<b>308,000,000</b>
使用硬件假设	NVIDIA A100	NVIDIA H100	NVIDIA H100	NVIDIA H100
使用效率假设	21.30%	30.20%	30.20%	30.20%
<b>实际算力需求 (PFLOPS)</b>	<b>435,446,009</b>	<b>4,589,403,974</b>	<b>4,589,403,974</b>	<b>1,019,867,550</b>
服务器峰值算力 (PFLOPS)	5	32	32	32
单位服务器包含 GPU 个数	8	8	8	8
<b>训练时间需求 (秒)</b>	<b>87,089,202</b>	<b>143,418,874</b>	<b>143,418,874</b>	<b>31,870,861</b>
约合天数	1,008	1,660	1,660	369
假设用户的需求来源均匀分布在 24 小时，每秒的算力需求	5,040	53,118	53,118	11,804
如果我们需要用户的等待时间小于 1 秒，需要的服务器数量	5,040	53,118	53,118	11,804
<b>折合 GPU 个数</b>	<b>40,319</b>	<b>424,945</b>	<b>424,945</b>	<b>94,432</b>

资料来源：Scaling Laws for Neural Language Models, SemiAnalysis, Similarweb, 各公司官网, 浦银国际

## ● AI 算力芯片产能瓶颈逐步解决，供需趋于平衡

HBM（High Bandwidth Memory，高带宽内存）和先进封装 CoWoS（Chip on Wafer on Substrate）的产能是此前制约 AI 算力芯片 GPU 的瓶颈。

为实现逻辑芯片的最大效率，存储器到逻辑芯片的传输速度必须大于等于逻辑芯片的运算速度。同时，由于逻辑芯片的迭代速度更快，内存传输速度决定了一块芯片的算力速度上限。HBM 存储芯片是 GPU 想要实现高速计算的必需品。HBM 芯片通过纵向堆叠 DRAM 裸片的方式减少内存的表面积，由此来减少传输距离，实现更快的内存传输速度。当前 HBM 领域全球主要玩家为 SK 海力士、三星和美光三家。

HBM 为了将多层 DRAM 芯片相互连接、垂直堆叠，并安装在逻辑芯片旁，需要使用 TSV (Through Silicon Via) 和 3D 芯片堆叠等先进封装技术。由于制造难度极高，必须使用先进封装技术。

先进封装方面，主要产能来自于台积电，三星等其他晶圆厂也有部分产能。根据我们的调研，台积电的 CoWoS 产能在 2023 年年底达 1.2 万片/月，预计 2024 年台积电 CoWoS 产能将达到 3 万片/月，并在 2025 年继续扩大产能。

单片 CoWoS 产能能够封装的 GPU 数量取决于良率和裸片大小。例如，单片 CoWoS 产能可以用于封装约 30 片英伟达 H100 GPU。但由于英伟达 B100 是通过连接两个 GPU 裸片和 8 个 HBM 组成，其芯片面积约为 H100 的两倍，单片 CoWoS 可满足 15-18 片 B100 GPU。

我们认为伴随 AI 服务器需求爆发初期产能供不应求，HBM 和先进封装 CoWoS 产能成为制约 GPU 的瓶颈的困境正在得到改善。

图表 24：英伟达各环节供应商

功能	类型	主要厂商
代工	代工	台积电、三星
封测	封测	日月光、Amkor 安靠
存储	HBM	SK 海力士、三星、美光
	HDD	东芝、希捷
	SSD	西部数据、三星电子、SK 海力士、美光
通信	光模块	天孚通信、中际旭创、华工科技、Coherent、Fabrinet
	铜互联	安费诺
其他芯片	板内互联(以太网芯片)	博通, Marvell
	接口芯片	博通
其他	散热	Veritiv (维谛技术)、英业达
	电源	麦格米特
	PCB 板	沪电股份
ODM	ODM	纬创、工业富联、鸿海精密、伟创力、广达、英业达、华勤技术、和硕、华硕、技嘉、比亚迪电子 (Jabil)

资料来源：英伟达公开资料整理，浦银国际



## ● 中国的 AI 大模型快速跟进：AI 需求大幅上扬和大模型能力快速追赶

在 OpenAI 发布 ChatGPT 3.5 之后，中国互联网厂商、初创企业、手机品牌、新能源车企等，都开始加速进入布局 AI 大模型相关内容。这也大幅提升全球对于 AI 算力芯片的需求。

根据 IDC 预测，中国 AI 市场 IT 基建支出 2023-2027 年期间复合增长率约 20%（图表 25）。中国 IT 基建开支由 2013 年的 146.5 亿美元，增长至 2023 年的 900.2 亿美元，期间复合增长率约 20%。根据 IDC 的预测，中国 AI 市场 IT 基建支出将有望从 2023 年约 190 亿美元增长至 2027 年约 400 亿美元，期间复合增长率约 20%。

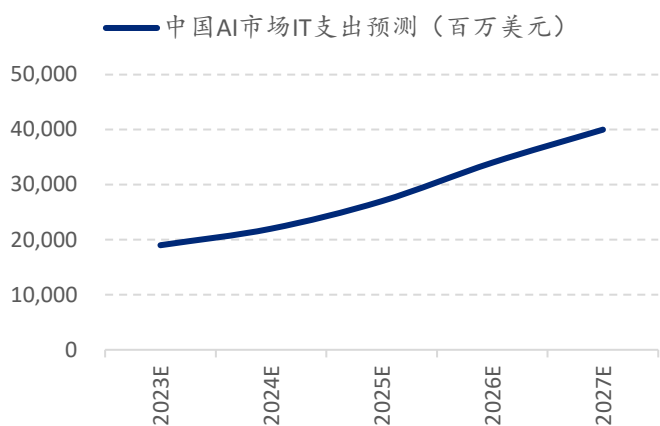
作为中国比较大的云服务厂商，中国运营商都在加大对于 AI 算力的基建投入，奠定新质生产力基础。中国的四家互联网公司（腾讯、阿里、百度、京东）资本支出在中国 IT 基建中占比于 2018 年达到峰值 43%，而后下行至 2023 年占比 18%。

根据公开资料，今年前四个月，全国规划、开工、落成或启用的智算中心项目约有 40 个。其中，由中国电信、中国联通、中国移动投资建设的项目 14 个。中国电信计划 2024 年总投资 960 亿元，在云和算力相关投资 180 亿元，智算算力提升 10EFLOPS，达到 21EFLOPS 以上。中国联通规划 2024 年总投资 650 亿元，投资重点由稳基础的联网通信业务转向高增长的算网数智业务。

再以中国移动为例，中国移动正在加快算力多元供给，不断完善新型基础设施。截至 2023 年，中国移动已经布局的算力包括通算 F32 8.0 EFLOPS，增长 14%，和智算 F16 10.1 EFLOPS，增长 206%。公司将重点布局智算与超算算力的建设，预计在 2024 年，实现增加通算算力 1 EFLOPS 和智算算力 7 EFLOPS 以上（增长 70%）。中国移动 2024 年算力规划投资达 475 亿元，累计智算算力超过 17 EFLOPS。

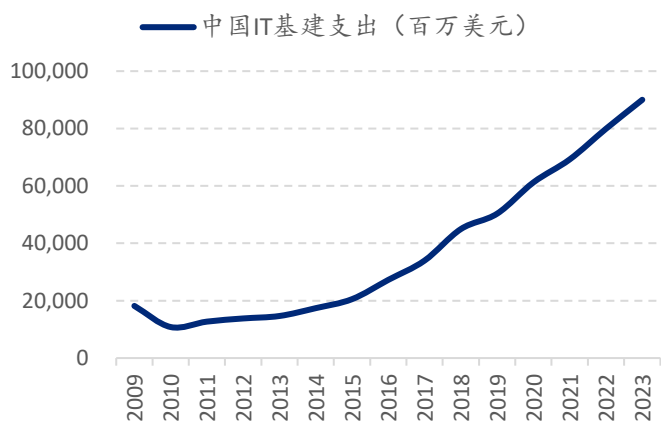
我们认为中国三大运营商在这个方向的投入也有望带动相关 GPU 算力芯片的需求的增长，包括潜在的英伟达针对中国市场的“特供版 H20 芯片”，以及部分中国 GPU 算力国产化需求。

图表 25: 中国 AI 市场 IT 支出预测



注: E=IDC 预测, 口径为针对 AI 市场的 IT 支出  
资料来源: IDC、浦银国际

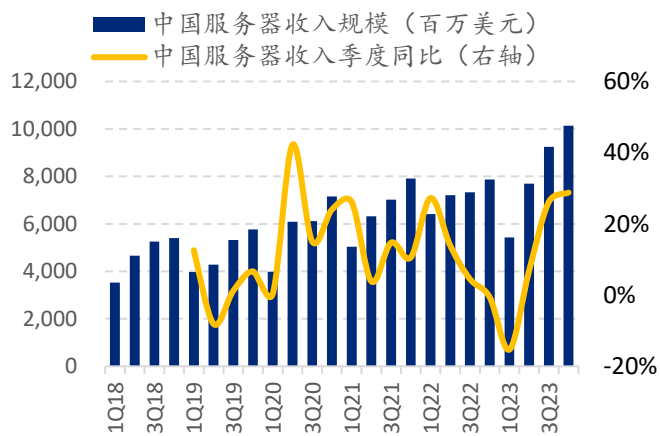
图表 26: 中国 IT 基建支出



资料来源: IDC、浦银国际

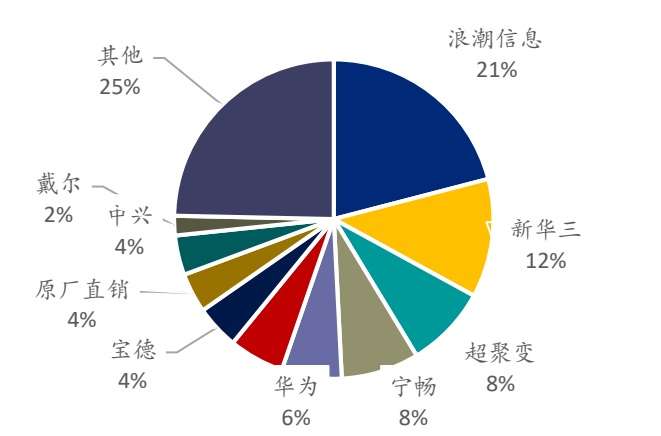
2023 年下半年中国服务器市场收入显著增长。中国服务器市场规模由 2018 年的 188.6 亿美元增长至 2023 年的 325.2 亿美元, 期间复合增长率 11.5%。2023 年下半年, 中国服务器收入同比增速明显提升, 4Q23 收入同比增长 28.8%。我们认为 2023 年下半年中国服务器市场收入增长主要由 AI 相关需求拉动。中国服务器市场前三大玩家及份额分别为浪潮信息 21%、新华三 12%、超聚变 8%, 均为中国本土玩家。

图表 27: 中国服务器市场规模



资料来源: IDC、浦银国际

图表 28: 中国服务器主要玩家



资料来源: IDC、浦银国际

国产大模型的性能及算力需求在向 GPT-4 靠拢。我们假设国产大模型在性能向 GPT-4 靠拢的同时, 其参数和算力也会向 GPT-4 靠拢。以当前国内可以购买到的英伟达 H20 为例, 约合 H100 算力的 20%。如果国产大模型同样希望在 30 天完成单个模型的测算, 即需要  $44,840 \times 5 = 224,200$  块英伟达 H20。

当前国内约有 200 个大模型。在 Llama 开源后, 很多公司以 Llama 模型为基础进行精调, 但发展也受限制于 Llama 模型的迭代。我们认为相当一部分的大模型厂商会选择自训练。

2024年6月25日，OpenAI向部分地区开发者发送警告信，表示从7月9日开始不再支持访问，收信者主要来自中国地区。我们认为该事件机遇大于挑战。国内外大模型在中文内容生成能力的差异较小，这将会促使更多的开发者转向国产大模型。智谱AI已经表示，将向OpenAI API用户提供“特别搬家计划”，并且将为高用量客户提供与OpenAI使用规模对等的token赠送计划。

同理，类似我们在海外大模型推理侧算力需求的测算，我们认为伴随更多的AI应用和爆款AI APP的出现，国内大模型的推理侧算力需求将显著提升。

图表 29：国产大模型训练侧算力及 GPU 需求

模型名称	文心一言	豆包	混元	通义千问
发布时间	2023年3月	2023年8月	2023年9月	2023年10月
参数量（亿）	2,600	2,000	1,000	720
假设训练集大小（亿 tokens）	3,000	3,000	3,000	3,000
<b>AI模型训练所需的浮点运算次数：<math>6 \times</math>模型参数规模<math>\times</math>训练集大小（PFLOPS）</b>	<b>468,000,000</b>	<b>360,000,000</b>	<b>180,000,000</b>	<b>129,600,000</b>
使用硬件假设	NVIDIA A100	NVIDIA A100	NVIDIA A100	NVIDIA A100
有效算力比率假设	21.30%	21.30%	21.30%	21.30%
<b>实际算力需求（PFLOPS）</b>	<b>2,197,183,099</b>	<b>1,690,140,845</b>	<b>845,070,423</b>	<b>608,450,704</b>
服务器峰值算力（PFLOPS）	5	5	5	5
单位服务器包含 GPU 个数	8	8	8	8
训练时间需求（秒）	439,436,620	338,028,169	169,014,085	121,690,141
约合天数	5,086	3,912	1,956	1,408
<b>30天完成训练需服务器数量</b>	<b>170</b>	<b>130</b>	<b>65</b>	<b>47</b>
<b>30天完成训练所需 GPU 数量</b>	<b>1,356</b>	<b>1,043</b>	<b>522</b>	<b>376</b>

资料来源：公开资料整理、浦银国际

# AI 算力供应链解析：价值量集中度高，算力芯片端垄断竞争

## ● AI 服务器价值量集中度高，提升幅度大

英伟达 AI 服务器产业链环节涉及 GPU 代工、封装、存储（HBM）、光模块、散热、电源管理、服务器组装等等。其中，GPU 代工和先进封装主要交给台积电完成，部分订单交由三星完成。2024 年以前，英伟达 HBM 均由 SK 海力士供应，但今年三星和美光都将加入 HBM 供应商队列。

AI 服务器的价值量大头集中在芯片成本本身。AI 服务器中，GPU 芯片成本占比高达 72.6%，是新增价值量最大的模块。因此，在这轮 AI 需求爆发过程中，AI 算力芯片供应商英伟达是当前的最大受益者。

而且，与传统服务器相比，GPU 芯片新增价值量最高。根据 SemiAnalysis 的拆分，我们对比了典型的 AI 服务器和通用服务器的 BOM (Bill of Materials)。典型的 AI 服务器英伟达 DGX H100 总售卖成本约 26.8 万美元，是通用服务器的约 25 倍。AI 服务器中，GPU 芯片成本占比高达 72.6%，是新增价值量最大的模块。智能网卡 NIC、存储的价值量占比也较为靠前，分别为 4.1% 和 4.2%。其他单机价值量增幅较高，值得关注的环节包括电源+300%，组装测试+200%，CPU +181%。

图表 30：AI 服务器价值量拆分

零件类型	通用服务器（以 2x Intel Sapphire Rapids Server 为例）		AI 服务器（以 Nvidia DGX H100 为例）		价值增幅
	价格（美元）	占比	价格（美元）	占比	单机增幅
CPU	1,850	17.7%	5,200	1.9%	181%
8GPU 4 NVSwitch Baseboard		0.0%	195,000	72.6%	
内存 DRAM	3,930	37.7%	7,860	2.9%	100%
硬盘 NAND	1,536	14.7%	3,456	1.3%	125%
网卡 SmartNIC	654	6.3%	10,908	4.1%	1568%
机箱（外壳、背板、电缆）	395	3.8%	563	0.2%	43%
主板	300	2.9%	360	0.1%	20%
散热（散热器+风扇）	275	2.6%	463	0.2%	68%
电源	300	2.9%	1,200	0.4%	300%
组装测试	495	4.7%	1,485	0.6%	200%
Markup	689	6.6%	42,000	15.6%	5996%
<b>总成本</b>	<b>10,424</b>	<b>100.0%</b>	<b>268,495</b>	<b>100.0%</b>	<b>2476%</b>

资料来源：SemiAnalysis，浦银国际

## ● 中国 AI 算力产业链

受到美对中半导体限制，英伟达 A100、H100 等芯片性能超过美国商务部性能密度阈值要求，芯片无法进入中国大陆。因而尽管中国大陆具备强大的电子制造和组装能力，但是在英伟达 AI 服务器产业链中的占比较少。AI 服务器的组装中中国台湾厂商是供应主力。

美对中半导体限制也为中国 AI 芯片国产替代带来机遇。华为是国产 AI 加速卡主要玩家。根据 IDC 的数据，2023 年，中国加速芯片的市场规模达到近 140 万张，其中 GPU 占 85% 的市场份额，约 119 万张。中国本土品牌 AI 芯片出货量超过了 20 万张，占整体 AI 芯片市场的 14%。根据 IDC 数据，2022 年华为在中国 AI 加速卡市场的市占率约 10%，是国产 AI 加速卡的最大玩家；百度、寒武纪、燧原科技也有少量份额。

除计算芯片外，我们同时看好在 DDR5 和 HBM 市场有布局的相关中国企业。根据 TrendForce 的调研显示，2022 年三大原厂 HBM 市占率分别为 SK 海力士 50%、三星约 40%、美光约 10%。国产 HBM 芯片处于起步阶段。

图表 31：美国对中国半导体行业的制裁持续加码

时间	政策概况
2018-04	禁止中兴通讯在未来 7 年内向美国企业购买敏感商品
2018-07	美国与中兴通讯和解，但需要支付 14 亿美元罚金
2019-05	美国将华为及其 70 个分支机构纳入“实体清单”，美国产品及美国技术含量 > 25% 的外国产品受到限制
2019-06	华为被列为美国和其盟邦国家的安全威胁
2019-08	白宫宣布禁止美国政府部门购买华为的设备和服
2019-10	28 家中国实体纳入出口管制清单
2020-05	美国商务部 BIS 发布公告对华为管制升级，限制其采购美国产品及采用美国技术 (>0%) 的外国商品、代工服务
2020-08	增加 38 家华为附属公司进入实体清单 美国商务部发布公告，将中芯国际及附属公司加入实体清单，采购含美技术设备需美国批准，14nm 及以下原则上
2020-12	不批准
2021-11	将多家中国半导体企业，如飞腾信息、申威、国科微、中科微、云芯微、新华三半导体等列入实体清单 美国半导体设备企业 (AMAT、LRCX 等) 收到美商务部函件，要求向中国大陆禁售用于 14nm 及以下先进制程的
2022-07	设备
2022-08	美国芯片法案颁布，对在美国建厂给予补贴，获得补贴的企业禁止到中国大陆建设先进制程产线
2022-10	美国商务部 BIS 公布了《对向中国出口的先进计算和半导体制造物项实施新的出口管制》，美国对中国半导体产业制裁的再次升级。
2023-03	日本、荷兰同意美国针对中国的半导体制裁，并修改相关法规
2023-10	美国商务部 BIS 更新了芯片出口禁令新规，进一步限制中国购买高端计算芯片及先进半导体设备，并将 13 家中国 GPU 企业列入实体清单
2024-4	美国 4 月 4 日开始实施对华芯片出口管制的新规定，装载人工智能芯片的电脑出口也将受到限制
2024-5	英特尔和高通被停止了向华为供货的许可，这可能意味着华为将无法使用高通跟英特尔的芯片

资料来源：公开资料收集、浦银国际



图表 32: 华为昇腾供应商

功能	类型	主要厂商
计算	GPU	华为海思 (昇腾 910/310)
	CPU	华为海思 (鲲鹏 920)
存储	HBM	SK 海力士
	DRAM	长鑫存储
	HDD/SSD	百维存储、朗科、江波龙、希捷
通信	网卡	星云智联、台湾研华
其他芯片	PMIC	圣邦股份、矽力杰、晶丰明源、杰华特、台达电
	BMC	卓易信息
	板内互联/接口芯片	华为海思、澜起科技
配套模块	电源模块	中国长城、台达电
	散热	英维克、高澜、川润
	PCB 板	沪电股份、胜宏科技、深南电路、生益科技
	被动元器件/线缆	风华高科、三环集团、微容科技
代工生产	ODM	宝德、华鲲振宇、工业富联、闻泰科技

资料来源: 公开资料资料, 浦银国际

图表 33: 国产 AI 芯片竞争格局

公司	产品	应用类型	算力 (TFLOPS)	制程 (nm)	带宽 (GB/s)	功耗 (w)
寒武纪	MLU370-X8	训练+推理	256(INT8)	7	614	250
	MLU370-X4	训练+推理	256(INT8)	7	307	150
景嘉微	JM9	图形	1.5(FP32)	14	128	30
华为海思	昇腾 310	推理	16(INT8)	12	/	8
	昇腾 910	训练	640(INT8)	N7+	/	310
燧原科技	T20	训练	256(INT8)	/	1638	300
	T21	训练	256(INT8)	/	1638	300
摩尔线程	MTTS3000	图形	15.2(FP32)	12	448	250
海光信息	深算一号	训练	/	7	1024	350
壁仞科技	BR100	训练	240(FP32)	/	128	550
百度	昆仑芯 2 代	训练+推理	256(INT8)	7	512	120

资料来源: 公开资料资料, 浦银国际

# AI 大模型：从云侧向端侧渗透

## ● AI 端侧大模型正在快速落地中

端侧 AI 距离用户更近，是 AI 大模型商业化闭环重要的应用之一。

端侧 AI 打造个性化助手，提升用户体验。端侧大模型实现 AI 个人助手功能。端侧 AI 模型可以实现多种大众期待的 AI 功能，打造了解个人习惯，能够有针对性地调整回答的个人助理，功能包括健康监测、实时语音翻译、摄影时的自动修图等。对比云侧 AI，端侧 AI 具备更低的回答延时、离线 AI 功能、个人信息保密等优势。

大模型算力及内存需求庞大，通过两方面实现落地。一方面，通过模型量化处理来减少精度，从而达到减少模型大小、减少内存消耗、加快模型推理的目标。比如，将 32 位的浮点数据，通过量化处理将数字映射到 8 位浮点数据，在最大限度保留信息的同时，模型大小将会降低到原有大小的约 1/4。在模型大小减少后，存储空间和内存耗用都会自然减少，需要计算的浮点数量降为原来的 1/4，推理速度和设备功耗也会自然减少。根据我们的调研，一个端侧模型压缩后可能需要 7-9G 的内存。

另一方面，可以通过采用混合专家模型 MoE (Mixture of Experts) 架构，在推理中调用大模型中的部分小型专家模型来实现。MoE 架构由多个小型专家模型组合而成，在推理过程中，只需要调用大模型中的部分小型专家模型，而非整个大模型，这样可以大幅减少推理过程中的参数量和算力成本。例如，在其他条件不变的情况下，如果一个 1,000 亿参数的大模型通过 MoE 架构在推理时将参数减少到了 50 亿，那么根据推理算力计算公式，推理算力需求也减少到了 1/20。

端侧 AI 模型落地有两种可能的方式，一是手机搭载手机厂自研大模型，允许各 APP 进行调用。二是 APP 自带大模型。我们认为，由于压缩后单个端侧模型对内存的占用仍然为 7-9G，而当前手机内存通常为 8-24G，所以 APP 自带本地化端侧 AI 模型较为不现实。我们认为未来的发展路径可能主要为 APP 调用手机厂本地大模型，或者调用云端模型。

手机厂、SoC 芯片厂共同合作助力体验升级，苹果自研 SoC 或将占据优势。我们认为苹果等具备自研 SoC 能力的手机品牌，将会具备更加顺滑的跨应用平台调用能力。苹果在 WWDC2024 中发布 Apple Intelligence。除具备其他 AI 终端具备的理解和创建语言、图像的功能，Apple Intelligence 还可以实现 ChatGPT 跨苹果平台的调用和集成，将大幅提升苹果终端设备的智能化程度和用户体验。

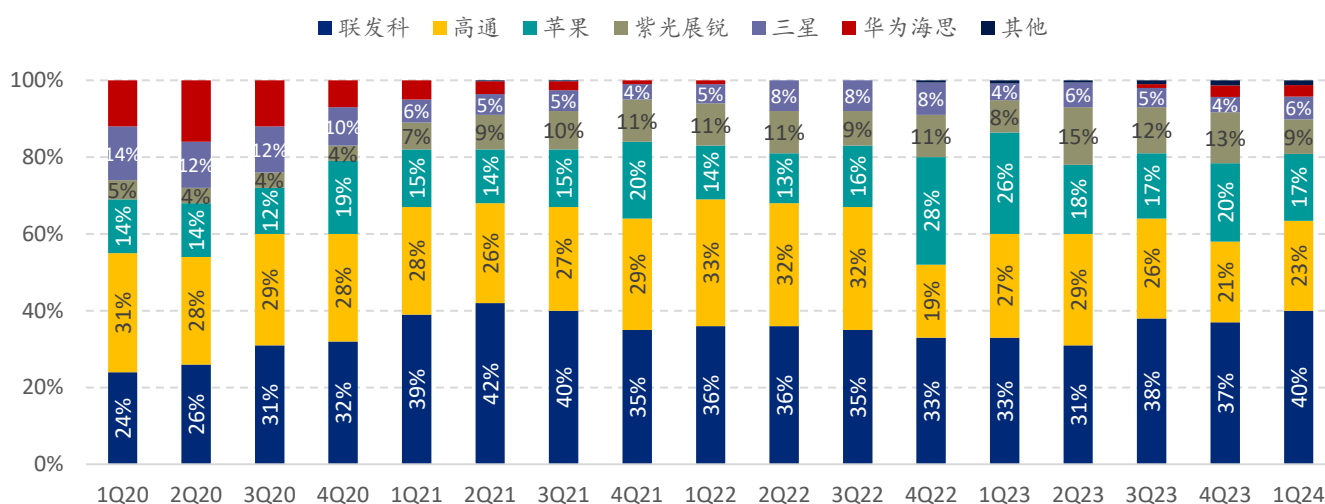
具备 AI 能力的手机芯片价格或将大幅提升。近期，高通宣布骁龙 8 Gen 3 芯片售价从 160 美元上涨至 200 美元，2024 年下半年即将发布的 8 Gen 4 定价达到 250 美元，人民币价格约 1,700 元，涨幅达到 25%。而 2016 年发布的旗舰 SoC 高通骁龙发售价仅为约 40 美元，约人民币 300 元。2016 年至今高通旗舰 SoC 单价提升 5.6 倍，期间年复合增长率超过 21%。联发科也有类似的价格上涨趋势，联发科 2020 年 5G SoC 均价在 40-45 美元左右，而 4G SoC 在 15-20 美元左右，价格提升幅度超 100%。

我们认为高端芯片价格的持续上涨，主要由于芯片性能不断大幅提升，以及制造工艺升级带来的成本提升。AI SoC 芯片的设计及工艺难度较普通高端手机 SoC 更大，我们认为具备 AI 能力的手机芯片价格或将大幅提升。

今年上半年，在 CES、MWC 等全球电子展会上我们看到越来越多具备 Gen-AI 功能的智能手机亮相。例如，三星的 Galaxy S24 可实现通话实时翻译功能，小米 14 Ultra 可实现 AI 计算摄影功能。三星、小米等高端机型在今年上半年也有较好的出货量表现。

高通和联发科分别发布的高通骁龙 8 Gen3 芯片和天玑 9300+ 芯片都具备 AI 功能。高通的 8 Gen3 芯片支持终端设备运行 100 亿参数模型。联发科的天玑 9300+ 芯片具备 22 TOPS 性能算力。这些 SoC 厂商都在为 Gen-AI 端侧手机提供算力基础，这些产品将在今年下半年和明年逐步上市，推动行业成长。

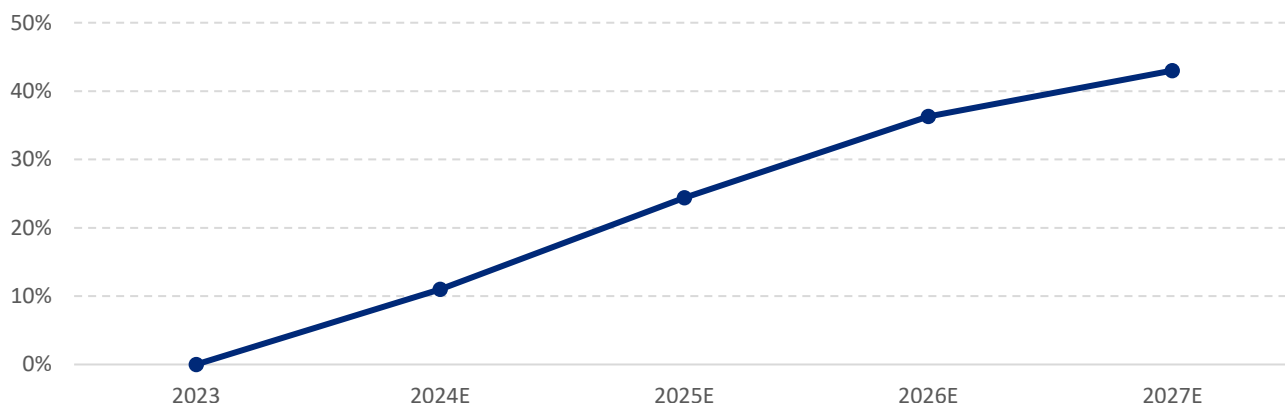
图表 34：智能手机 SoC 出货量占比按品牌划分



资料来源：Counterpoint，浦银国际

图表 35: AI 手机渗透率

AI手机渗透率



注: E= Counterpoint 预测

资料来源: Counterpoint, 浦银国际

图表 36: 苹果 M4 芯片 AI 算力高达 38 TOPS

	M4	M1	M2	M3
发布日期	2024 年 5 月 7 日	2020 年 11 月 11 日	2022 年 6 月 7 日	2023 年 10 月 31 日
晶体管数量	280 亿	160 亿	200 亿	250 亿
制程工艺	台积电 N3B	台积电 N5	台积电 N5P	台积电 N3B
CPU 性能	-	3.2GHz	3.5GHz	4.1GHz
	10 核	8 核	8 核	8 核
GPU 性能	10 核	7/8 核 (2.6 TFLOPS)	8/10 核 (3.6 TFLOPS)	8/10 核
NPU 性能	16 核 (38 TOPS)	16 核 (11 TOPS)	16 核 (15.8 TOPS)	16 核 (18 TOPS)
内存带宽	120GB/s	68.25GB/s	100GB/s	100GB/s

资料来源: 公司官网、浦银国际

图表 37: AI 手机芯片 AI 生成速度可达每秒 22 个 tokens

发布日期	AI 手机芯片	品牌	AI 生成速度 (token/秒)
2023 年 10 月 25 日	骁龙 8 Gen3	高通	20
2023 年 11 月 6 日	天玑 9300	联发科	20
2024 年 3 月 18 日	骁龙 8S Gen3	高通	-
2024 年 5 月 7 日	天玑 9300+	联发科	22

注: 生成速度基于 70 亿参数大模型

资料来源: 公司官网、浦银国际

图表 38: AI PC 芯片算力高达 45 TOPS

公司	型号	NPU 算力 (TOPS)
英特尔	Meteor Lake	11
	Lunar Lake	45
AMD	Hawk point	16
高通	高通骁龙 8cx Gen 3	29
	高通骁龙 X Elite	45
苹果	M3	18
	M4	38

资料来源: 公司官网、浦银国际

## ● AI 算力芯片竞争格局：云侧 vs 端侧

在云侧 AI 算力芯片市场中，英伟达处于并且未来一段时间都将保持垄断地位。英伟达潜在的竞争对手为 AMD 以及 AI 大模型厂商。但是，这些竞争对手的体量合起来仍比较小。虽然中国有华为这种形成一定体量和规模的 AI 算力芯片玩家，但是中美关系紧张使得英伟达和华为几乎无法处于直接的竞争中。

在云侧向端侧渗透的过程中，原来的手机 SoC 设计厂商，例如高通、联发科、苹果等，仍旧保持寡头竞争。在这个领域中，英伟达想要渗透的难度比较高，因此端侧更多还是原智能手机 SoC 厂商之间的竞争。

我们认为云侧和端侧玩家有望同台竞技的领域是新能源车的智能驾驶和智能座舱领域。由于英伟达牢牢把控云侧训练用的算力芯片，因此中国新能源车企有较强意愿采用其新能源车端侧算力芯片，例如 Orin 系列芯片，来实现智能驾驶。而智能座舱更多涉及接近智能手机用户体验，以及手机端 APP 移植到车机的体验，因而高通的 8295 等芯片更受到车企青睐。

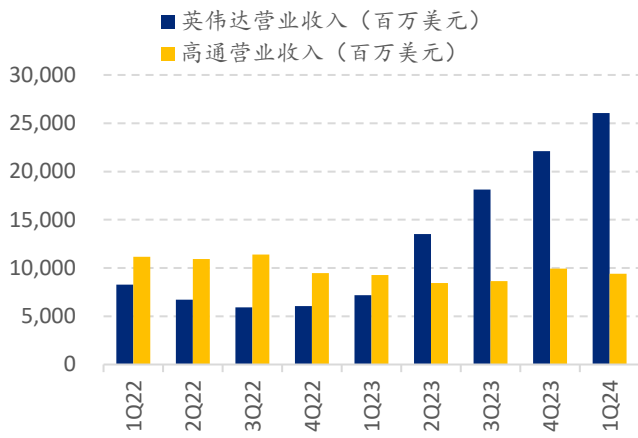
从新能源车长期发展来看，舱驾一体是未来趋势。因而，使用单一 SoC 实现智能驾驶和智能座舱功能将使得英伟达和高通进入更加直接的竞争中。英伟达下一代的 Thor 系列芯片和高通的 8755 芯片可能形成直接的对抗竞争。在这场竞争中谁能取得相对优胜，就看哪一个可以更快地补足相对劣势，提供整体解决方案。而在形成明显优势之前，两种方案都将在市场取得一席之地。

在 2024 年 6 月 18 日，英伟达收盘市值达 3.3 万亿美元，登顶全球市值第一。在 AI 大模型渗透端侧的过程中，各个品牌仍然需要英伟达的云侧算力来训练模型。这也隐含了英伟达未来一段时间收入利润增长幅度可能更大。

英伟达这两年收入端增长大爆发，收入绝对值也快速超过手机芯片领域最重要的玩家高通。这在一定程度意味着科技产业最前沿的技术正在从手机向 AI 大模型转移。

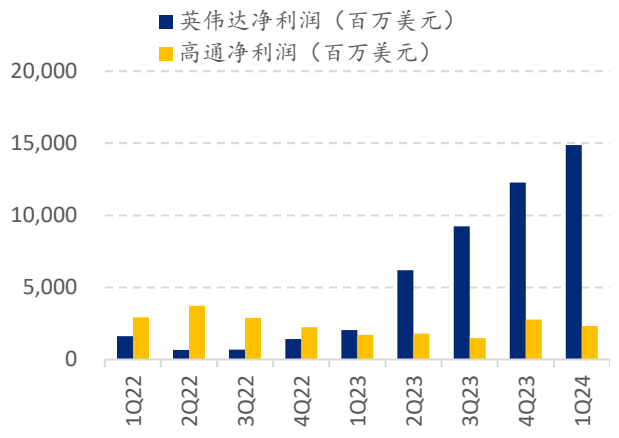


图表 39: 季度收入: 英伟达 vs 高通



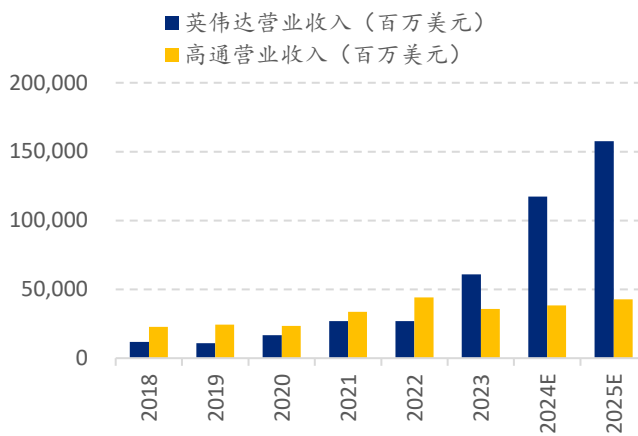
注: 图中时间为日历年 (Calendar Year); 英伟达财务年 (FY, Fiscal Year) 为上一年 2 月至当年 1 月, 高通财务年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月。  
资料来源: Factset、公司公告、浦银国际

图表 40: 季度利润: 英伟达 vs 高通



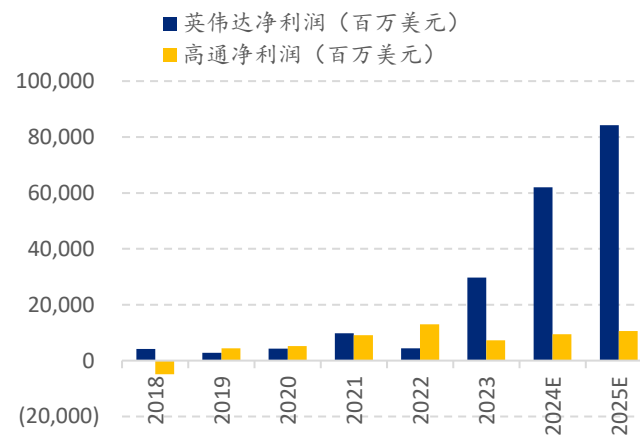
注: 图中时间为日历年 (Calendar Year); 英伟达财务年 (FY, Fiscal Year) 为上一年 2 月至当年 1 月, 高通财务年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月。  
资料来源: Factset、公司公告、浦银国际

图表 41: 年度收入: 英伟达 vs 高通 2018-2025E



注: 图中时间为日历年 (Calendar Year); 英伟达财务年 (FY, Fiscal Year) 为上一年 2 月至当年 1 月, 高通财务年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月。E=浦银国际预测  
资料来源: Factset、公司公告、浦银国际

图表 42: 年度净利润: 英伟达 vs 高通 2018-2025E



注: 图中时间为日历年 (Calendar Year); 英伟达财务年 (FY, Fiscal Year) 为上一年 2 月至当年 1 月, 高通财务年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月。E=浦银国际预测  
资料来源: Factset、公司公告、浦银国际

## ● AI 算力芯片竞争格局：云侧产品比较

在今年 3 月的 GTC，英伟达发布了新一代 Blackwell 平台的 GPU（Graphics Processing Unit，图形处理器），包括 B200 和 GB200 等产品系列。B200 由两个超大型裸片封装组合而成，内含超过 2080 个晶体管。这颗芯片还封装了 192GB 的高速 HBM3e 的显存。与前一代 H100 相比，B200 的每秒输出 token 数量提升 15 倍，Supercharged AI 训练表现提升 3 倍。与当前市场上量产的 AI 算力芯片比较，英伟达的产品仍然具备明显的性能优势。

从未来两年来看，英伟达计划在 2025 年推出 Blackwell 平台 Ultra 芯片，并在 2026 年推出 Rubin 平台的产品。我们预期英伟达有能力在 AI 算力芯片领域保持较强的产品迭代和升级能力，保持性能和成本的优势。

从更长期的竞争格局看，我们预期英伟达在 AI 算力近乎垄断的份额会有所减少，但是，英伟达在维持大多数产品的性能和成本优势的同时保持相对健康的利润率是比较有保障的。一方面，头部的 AI 大模型玩家会有意愿开发并自研 AI 算力芯片以取得定制化差异化性能和成本优势。特斯拉的 Dojo 超算、谷歌的 TPU 芯片、百度的昆仑 AI 芯片等，都存在取代/部分取代当前英伟达算力芯片的可能。另一方面，在 AMD 以及其他初创公司 GPU 等 AI 算力芯片性能存在部分优势的情况下，AI 大模型厂商也会有意愿增加非英伟达芯片使用量以促进 AI 算力芯片的竞争，平衡供应商之间的成本。

图表 43：英伟达 GPU 产品参数

品牌	英伟达				
产品	A100 SXM	H100 SXM	H200 SXM	H20	B200
发布时间	2020 年	2022 年	2024 年	2024 中国特供	2024 年
架构	Ampere	Hopper	Hopper	Hopper	Blackwell
GPU 显存	80GB	80GB	141GB	96GB	192GB
GPU 禁带宽度	2.0TB/s	3.3TB/s	4.8TB/s	4.0TB/s	8.0TB/s
最大 TDP（热设计功耗）	400W	700W	700W	400W	1000w
FP64 (TFLOPS)	9.7	34	34	1	-
FP64 Tensor Core (TFLOPS)	19.5	67	67	0	40
FP32 (TFLOPS)	19.5	67	67	44	-
TF32 Tensor Core (TFLOPS)	156	495	989	74	1100
FP16 Tensor Core (TFLOPS)	312	990	1979	148	2250
FP8 Tensor Core (TFLOPS)	624	1979	3958	296	4500
INT8 Tensor Core (TFLOPS)	624	1979	3958	296	4500

注：“-”指暂无相关信息

资料来源：公司官网、浦银国际

图表 44：英伟达及 AMD 核心 GPU 产品对比

品牌	AMD	英伟达
产品	MI300X	H200 SXM
发布时间	2023 年	2024 年
GPU 显存	192GB	141GB
GPU 禁带宽度	5.3TB/s	4.8TB/s
最大 TDP (热设计功耗)	750W	700W
TF32 Tensor Core	10.5PFLOPs	989TFLOPS
FP16 Tensor Core	20.9PFLOPs	1,979TFLOPS
FP8 Tensor Core	41.8POPs	3,958TFLOPS
INT8 Tensor Core	41.8PFLOPs	3,958TFLOPS

资料来源：公司官网、浦银国际

# 美股科技股价值投资回顾

## ● 美进入降息周期有利于成长科技股估值上行

浦银国际的策略组今年发布两篇关于降息周期与科技革命交汇下的资产配置（[上](#)、[下](#)）报告，重点回顾了1995-1996年的降息周期（图表45），当时的宏观背景与当下类似，而且遇上了互联网的第一波投资热潮。

与策略组的判断类似，我们认为1995-1996年美联储实施的降息更多是预防性降息，美国经济也有望实现“软着陆”。当前美股走势有望接近1995-1996年降息结束后，取得较好正回报（图表46）。与1995年互联网浪潮的起点类似，当前我们正在经历由AI引发的全球科技革命的新阶段。在未来两至三年内，AI技术有望在应用和普及上迎来爆发，或将带来难得一遇的投资机遇。在上一轮科技革命，互联网技术迭代与资本市场的走势息息相关。目前的投资者情绪和风险溢价水平与当时的水平也较为接近，这或能为我们布局这一轮AIGC引发的全球科技革命带来了一些启示（图表47）。

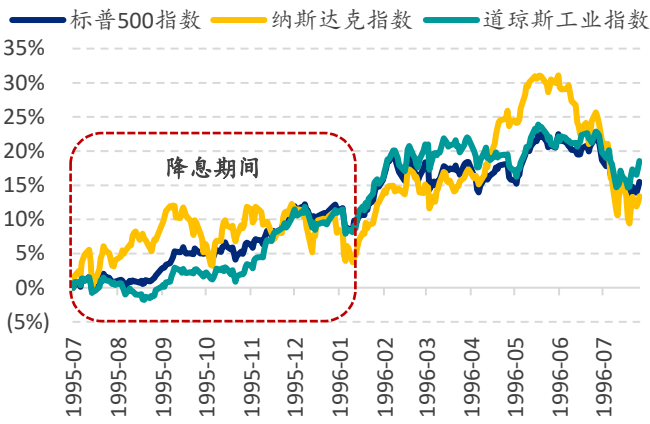
去年至今美股估值持续扩张，可能已经提前反映了美联储即将降息和通胀持续下滑的利好，降息周期开始之后实际利率下降给估值释放带来的空间可能已经有限。但是，盈利有望接棒成为支持美股向上的核心驱动力。在AIGC的驱动下，科技龙头公司的盈利增长预计仍较强劲，有望驱动股价向上（图表48）。

图表 45：1990 年至今，美联储共开启八轮降息周期，“降息预期”往往提前反映在资产价格上



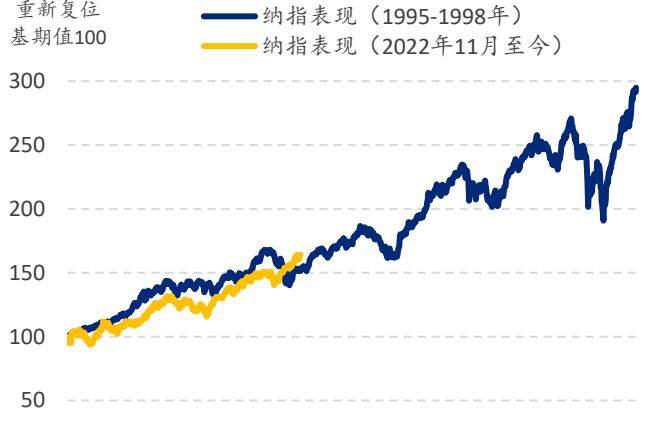
资料来源：CEIC、Bloomberg、Wind、浦银国际

**图表 46: 1995-1996 年降息期间及结束后短期美股三大指数均表现强劲**



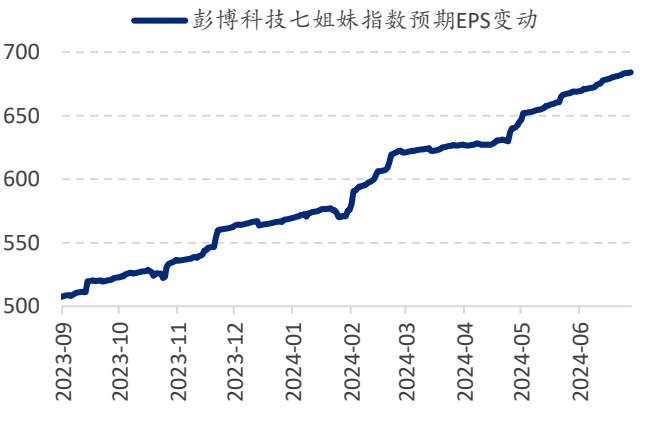
资料来源: Bloomberg、Wind、浦银国际

**图表 47: 1995-1998 年 vs 2022 年 11 月至今纳斯达克指数走势**



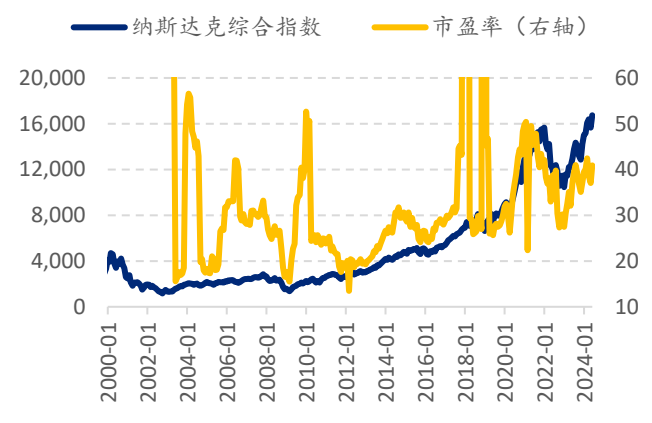
资料来源: Bloomberg、Wind、浦银国际

**图表 48: 科技龙头预期盈利增速不断得到上调**



资料来源: Bloomberg、浦银国际

**图表 49: 纳斯达克指数与估值**



资料来源: Bloomberg、浦银国际

## ● 美股纳斯达克以及 M7 具备长线的基本面推动成长的能力

Magnificent 7 (M7) 指美国 7 大科技巨头, 即英伟达、苹果、谷歌、亚马逊、Meta、微软和特斯拉, 这 7 只股票具备庞大的市场规模、技术能力和财务实力, 在美国科技市场占主导地位。

我们复盘了 M7 过去 30 年的股价表现, 我们发现 M7 的股价主要由每股盈利 (EPS) 驱动 (图表 50)。2012 年 5 月, M7 全部完成上市。2012 年 5 月至今, M7 股价上涨了 16 倍, 其中 EPS 上涨了接近 10 倍, 市盈率提升了 1.6 倍。

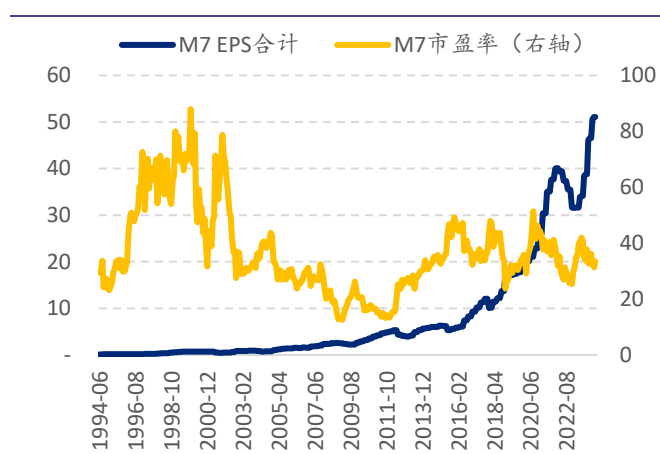
纳斯达克股价表现也和 M7 类似, 2003 年 1 月至今, 纳斯达克股价上涨了 12 倍, 其中 EPS 上涨了 26 倍, 市盈率为当时的 0.5 倍。



2022 年底开启的 AI 行情也主要由 EPS 驱动。当前 M7 股价较 2022 年 12 月低点上涨了 114%，其中 EPS 增长 62%，股价剩下部分由市盈率驱动。

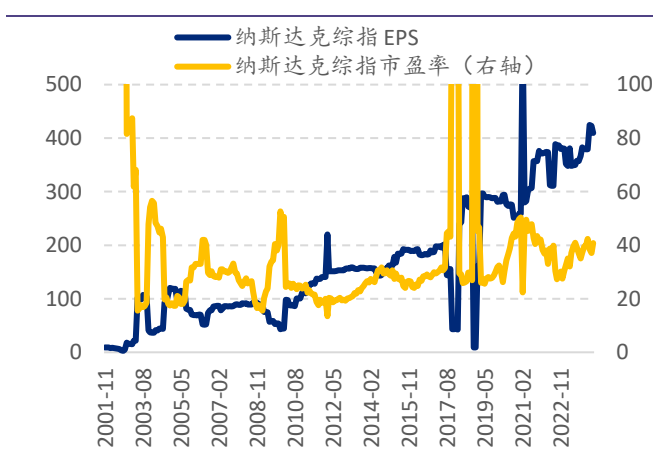
我们认为纳斯达克包括 M7 持续科技创新与技术进步带来的 EPS 提升，是美国纳斯达克指数及 M7 股价长牛最重要的因素。现在是新一轮 AI 革命的起点，人工智能对于社会生产力的提升已经初具苗头，预计美股纳斯达克以及 M7 股价将会长期受基本面推动成长。因此，英伟达和高通都有望享受这波 AI 带动的科技成长动能。

图表 50: 过去三十年 M7 股价主要由 EPS 驱动



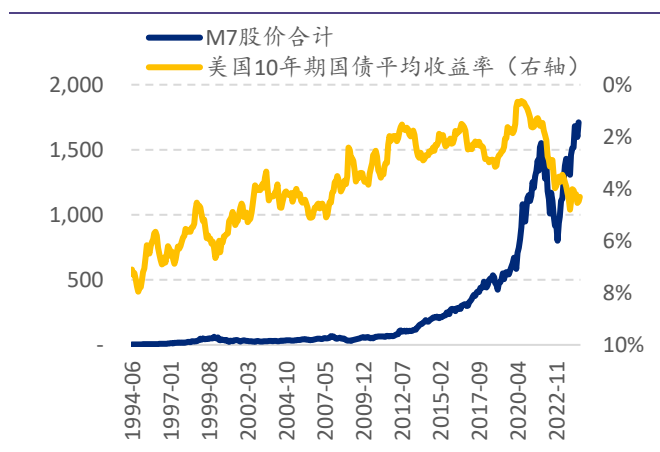
资料来源: Factset、浦银国际

图表 51: 2001 年至今纳斯达克综合指数股价主要由 EPS 驱动



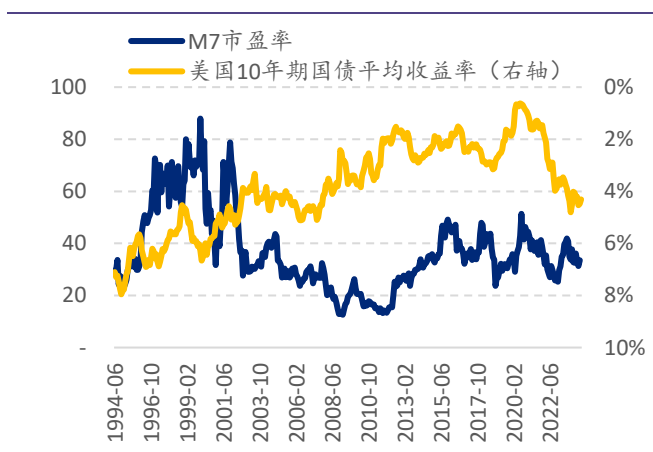
资料来源: Factset、浦银国际

图表 52: M7 市盈率与十年期美债收益率大体呈现负相关关系



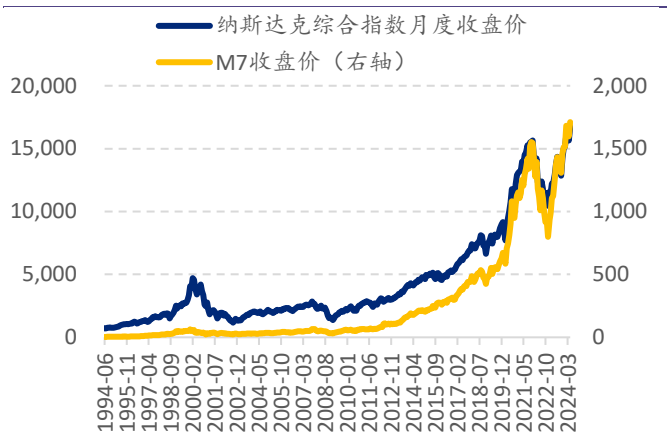
资料来源: Factset、浦银国际

图表 53: 2022 年 10 月以来涨幅主要来自 EPS 提升



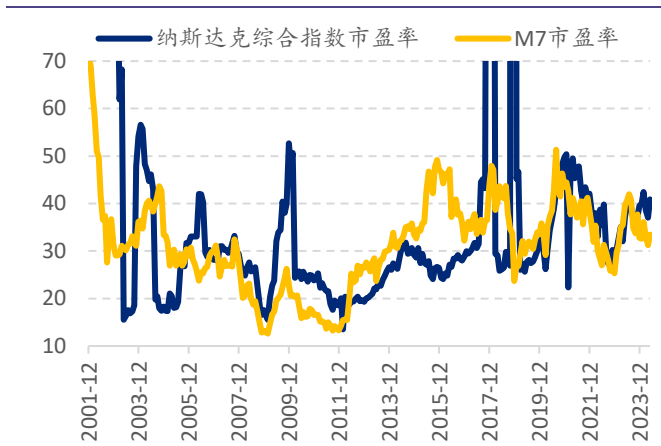
资料来源: Factset、浦银国际

图表 54: 纳斯达克综合指数 vs M7



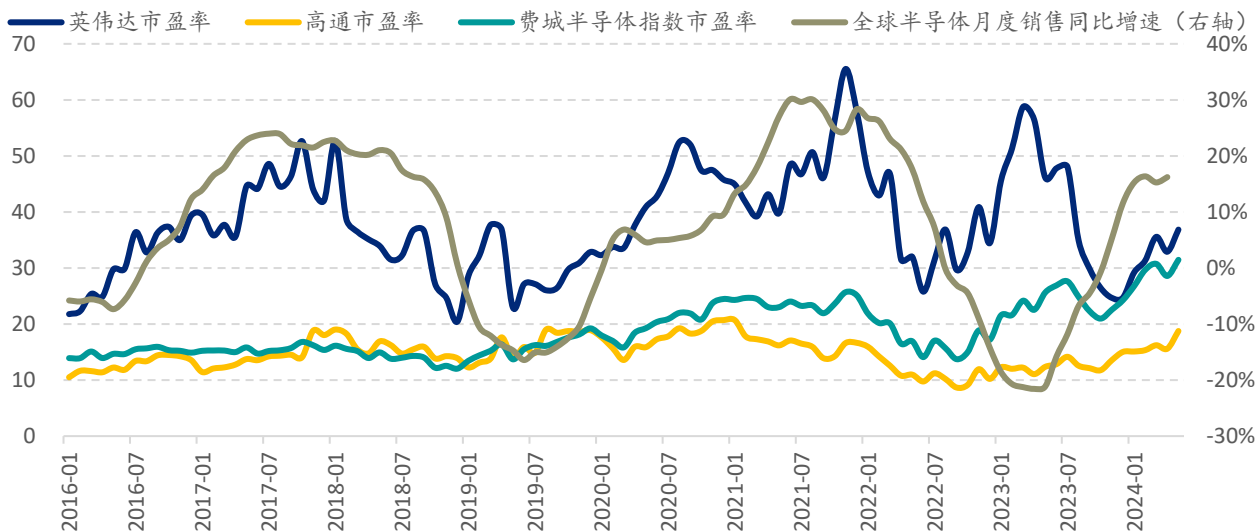
资料来源: Factset、浦银国际

图表 55: 市盈率: 纳斯达克综合指数 vs M7



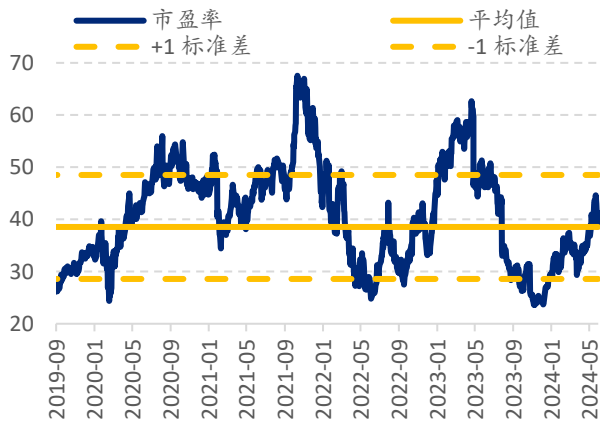
资料来源: Factset、浦银国际

图表 56: 英伟达市盈率 vs 高通市盈率 vs 全球半导体月度销售同比增速 vs 费城半导体市盈率



资料来源: Factset、浦银国际

图表 57: 英伟达市盈率



资料来源: Factset、浦银国际

图表 58: 高通市盈率



资料来源: Factset、浦银国际

## 附录：AI 算力行业术语简介

**AI 服务器**是指专门为人工智能应用而设计和优化的服务器。普通服务器主要设计用于提供网络服务、数据存储和数据处理等功能，而 AI 服务器则是为了满足人工智能应用的高性能计算、存储和数据处理需求。

**万卡集群**，是指规模数量由一万张及以上的加速卡（包括 GPU、TPU 及其他专用 AI 加速芯片）组成的高性能计算系统，通过集成巨量规模的单卡算力芯片，用于加速人工智能模型的训练和推理过程。

**GPU**（graphics processing unit）即图形处理器，主要用于图像和图形相关运算工作的微处理器。在浮点运算、并行计算等部分计算方面，GPU 可以提供数十倍乃至上百倍于 CPU 的性能。

**FLOPS**（floating-point operations per second）指计算机每秒钟能够执行的浮点运算次数，是一种评估计算速度的单位。FP32 被定义为单精度（32 位），FP64 为双精度（64 位）。从数据精度来看，FP64>FP32>TF32>FP16>INT8。数据精度越高，占据的带宽越大。

**带宽**：在数字芯片中指每秒可以传输的数据量。

**HBM（高带宽存储器）**采用了先进的封装技术，如通过硅穿孔（TSV）和微凸点（micro-bump）互连，相比传统的 DRAM（动态随机存取存储器）实现了更高的数据传输速率和更低的功耗。HBM 的设计允许多个存储器堆叠在一起，形成一个紧凑的模块，这不仅提高了存储密度，还减少了与处理器之间的物理距离，从而降低了延迟。

**光模块**是进行光电和电光转换的光电子器件。光模块的发送端把电信号转换为光信号，接收端把光信号转换为电信号。光模块是光纤通信系统的核心器件之一，可以提供更快速的数据传输和更高效的计算能力，一颗 GPU 往往需要 6-8 颗光模块。400G 光模块是指速率为 400G 的光模块，G 是光信号传输速率（Gbps）的单位。

**英伟达 GPU 分为 PCIe 和 SXM 两种不同接口：**

**PCIe**（Peripheral component interconnect express）是英特尔公司在 2001 年提出的一种高速串行计算机扩展总线标准。优点是兼容性比较好，数据传输速率高、潜力大。

**SXM** 是由英伟达设计，目的是为高性能计算和数据中心提供更强的计算能力和传输速度。SXM 接口的 GPU 通常是存在于 DGX 系统板上，该 DGX 系统板支持 4 张或 8 张 SXM GPU，每个 GPU 之间通过 NVLink 进行通信。

**SXM 方案对比 PCIe 的优点:** SXM 接口采用了 NVLink 的技术,可以提供更高的带宽和更低的延迟,可以加快 GPU 之间的数据传输。SXM 能够实现 4/8 个 GPU 之间的通信。PCIe 在使用桥接器后虽然能够达到和 SXM 相同的带宽,但只能实现 2 块 GPU 卡之间的通信。

**NVLink:** 是英伟达推出的一种高速互联技术,使多个英伟达 GPU 可以相互连接。多卡连接后可以实现并行计算,提高数据处理性能。

**Chiplet** 也称芯粒,是将不同功能芯片裸片的拼搭,用封装技术整合在一起,借此在摩尔定律趋缓下的情况下实现类似先进制程的迭代,在提升性能的同时实现低成本和高良率。采用 Chiplet 时大概率需使用先进封装,比如 2.5D 封装 CoWoS。

**先进封装:** 传统封装的封装效率(裸芯面积/基板面积)较低,浪费面积较多。在后摩尔定律时代,芯片制程迭代受限的情况下,先进封装便是另一条实现更高密度的集成的出路。先进封装技术包括 2.5D 封装 CoWoS,以及 3D 封装技术硅通孔(TSV)等。

**SoC (System on Chip)** 是指在一块芯片上集成一整个信息处理系统。智能手机 SoC 芯片通常为 CPU、GPU、NPU 计算芯片,通信基带芯片、影音处理、显示、内存芯片集合,包含一套完整处理系统。**基带芯片 (Baseband Chip)** 内含调制调节器 (Modem)、信道编码器、数字信号编码器等。

**调制调节器 (Modem)**, 俗称“猫”,可以将数字信号转换为模拟信号以通过电话传播。

**微控制器 (MCU, Micro Control Unit)**, 将 CPU、RAM、ROM 集成在一片芯片上。MCU 对比 SoC 的区别在于,MCU 是芯片级的芯片,SoC 是系统级的芯片。SoC 一个完整的单芯片计算机系统,可能包含许多 MCU。

**IP 核 (Intellectual Property Core)**, 指已验证、可重复利用、具有某种确定功能的芯片设计模块。



# 英伟达 (NVDA.US) 首次覆盖：AI 赛道坡长雪厚，增长可期

我们首次覆盖英伟达(NVDA.US)，给予“买入”评级，目标价为 147.6 美元，潜在升幅 19%。

- AI 算力最大受益者，享受科技行业初期成长动能：**英伟达是 AI 大模型需求爆发后最大最直接的受益者。目前，英伟达在云侧 AI 训练和推理芯片几乎占据垄断地位。中短期来看，AI 大模型算法厂商出于竞争目的依然有较高意愿购买算力，以支撑其模型的迭代和性能领先。长期来看，AI 大模型正在渗透千行百业，推动全球经济增长。虽然英伟达目前 43.6x 的未来 12 个月的市盈率不是最低价，但是美国降息叠加 AI 科技革命有望给与估值溢价支撑。因此，我们对英伟达保持较为乐观态度。
- AI 算力芯片需求高速增长，供需平衡全力推动释放增量：**借助代工厂产能加速扩建，英伟达有望跟上 AI 大模型带来的 AI 算力的高速增长。同时，借助 GPU 持续迭代，英伟达有能力在市场中保持较高的竞争力，以维护自身利润空间。我们预期来自于数据中心的收入在 2025 财年和 2026 财年将分别同比增长 118% 和 39%，是英伟达最重要的收入和利润推动器。
- 半导体周期上行，推动游戏、自动驾驶等重回增长：**全球半导体行业基本面持续上行，有望借助游戏、新能源车等下游需求复苏，而维持增长。因此，我们看到游戏、新能源车行业本身的增长将会带动英伟达业务板块保持较高增长动能。进一步看，新能源车可能是 AI 大模型落地端侧最显性的应用，因此我们看好公司智驾业务的未来三年都将保持 30% 以上的增速。
- 估值：**我们采用 DCF (Discounted Cash Flow, 现金流量贴现法) 估值方法。我们假设英伟达 FY2030-FY2034 的营收成长率为 30%-35%，永久增长率为 3%，WACC (Weighted Average Cost of Capital, 加权平均资金成本) 是 16.3%，得到英伟达目标价为 147.6 美元，潜在升幅 19%，对应 FY2025 市盈率为 59.3x，首予“买入”评级。
- 投资风险：**全球或美国经济下行，多个下游需求动能不足；AI 需求爆发持续性弱于预期，大模型厂商盈利低于预期；半导体周期上行动能不足；行业竞争加剧，拖累利润表现；研发等费用率增长较快；芯片等迭代速度和性能低于预期。

图表 59：盈利预测和财务指标 (FY2023-FY2027E)

美元百万	FY2023	FY2024	FY2025E	FY2026E	FY2027E
营业收入	26,974	60,922	117,311	157,568	188,605
营收同比增速	0%	126%	93%	34%	20%
净利润	4,368	29,760	61,972	84,243	100,935
净利润增速	-55%	581%	108%	36%	20%
目标 PE (x)	84.1	12.3	59.3	43.6	36.4

注：英伟达财年 (FY, Fiscal Year) 为上一年 2 月至当年 1 月；E=浦银国际预测；资料来源：公司公告、浦银国际

## 英伟达(NVDA.US)

**买入**

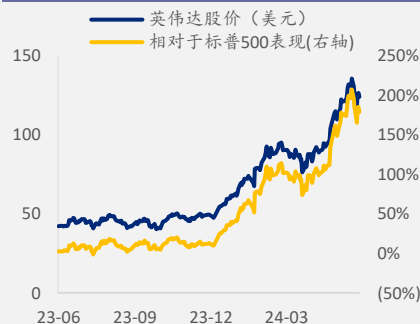
目标价 (美元)	147.6
潜在升幅/降幅	+19%
目前股价 (美元)	124.0
52 周内股价区间 (美元)	39.2-140.8
总市值 (百万美元)	3,050,154
近 90 日日均成交额 (百万美元)	435

注：截至 2024 年 6 月 27 日收盘价

## 市场预期区间



## 股价相对表现



截至 2024 年 6 月 27 日收盘价

资料来源：Factset、浦银国际



## 财务报表分析与预测

### 利润表

美元百万	FY2023	FY2024	FY2025E	FY2026E	FY2027E
<b>营业收入</b>	<b>26,974</b>	<b>60,922</b>	<b>117,311</b>	<b>157,568</b>	<b>188,605</b>
营业成本	(11,618)	(16,621)	(29,040)	(39,421)	(47,073)
<b>毛利</b>	<b>15,356</b>	<b>44,301</b>	<b>88,271</b>	<b>118,147</b>	<b>141,531</b>
<b>营业支出</b>	<b>(11,132)</b>	<b>(11,329)</b>	<b>(15,713)</b>	<b>(21,103)</b>	<b>(25,324)</b>
销售及行政及一般费用	(2,440)	(2,654)	(4,022)	(5,399)	(6,458)
研发费用	(7,339)	(8,675)	(11,692)	(15,704)	(18,866)
重组及其他费用/收入	(1,353)	-	-	-	-
<b>营业利润</b>	<b>4,224</b>	<b>32,972</b>	<b>72,558</b>	<b>97,044</b>	<b>116,207</b>
<b>营业外收支</b>	<b>(43)</b>	<b>846</b>	<b>1,214</b>	<b>1,167</b>	<b>1,159</b>
利息收入/(支出)净额	5	609	964	937	935
其他收益/(支出)净额	(48)	237	250	230	225
<b>税前利润</b>	<b>4,181</b>	<b>33,818</b>	<b>73,772</b>	<b>98,211</b>	<b>117,367</b>
所得税	187	(4,058)	(11,800)	(13,968)	(16,431)
<b>净利润(不含少数股东权益)</b>	<b>4,368</b>	<b>29,760</b>	<b>61,972</b>	<b>84,243</b>	<b>100,935</b>
少数股东和可赎回少数股东损益	-	-	-	-	-
<b>归属普通股股东净利润</b>	<b>4,368</b>	<b>29,760</b>	<b>61,972</b>	<b>84,243</b>	<b>100,935</b>
基本股数(百万)	2,490	2,490	24,900	24,900	24,900
摊销股数(百万)	2,490	2,490	24,900	24,900	24,900
<b>基本每股收益(美元)</b>	<b>1.75</b>	<b>11.95</b>	<b>2.49</b>	<b>3.38</b>	<b>4.05</b>
<b>稀释每股收益(美元)</b>	<b>1.75</b>	<b>11.95</b>	<b>2.49</b>	<b>3.38</b>	<b>4.05</b>

### 资产负债表

美元百万	FY2023	FY2024	FY2025E	FY2026E	FY2027E
<b>货币资金</b>	<b>3,389</b>	<b>7,280</b>	<b>17,331</b>	<b>52,314</b>	<b>99,930</b>
现金及其等价物	9,907	18,704	36,016	48,376	57,905
应收账款	3,827	9,999	19,254	25,861	30,955
存货	5,159	5,282	9,229	12,528	14,959
其他流动资产	791	3,080	5,931	4,505	5,218
<b>流动资产合计</b>	<b>23,073</b>	<b>44,345</b>	<b>87,760</b>	<b>143,584</b>	<b>208,968</b>
固定资产-物业, 厂房及设备	3,807	3,914	6,534	10,203	14,452
无形资产	1,676	1,112	1,112	1,112	1,112
商誉	4,372	4,430	4,430	4,430	4,430
租赁权	1,038	1,346	1,346	1,346	1,346
其他非流动资产	7,216	10,581	10,581	10,581	10,581
<b>资产总计</b>	<b>41,182</b>	<b>65,728</b>	<b>111,763</b>	<b>171,256</b>	<b>240,889</b>
短期借款	5,370	7,932	9,518	11,422	13,706
应付账款	1,193	2,699	4,716	6,401	7,644
其他流动负债	-	-	-	-	-
<b>流动负债合计</b>	<b>6,563</b>	<b>10,631</b>	<b>14,234</b>	<b>17,823</b>	<b>21,350</b>
长期借款	9,703	8,459	8,459	8,459	8,459
其他非流动负债	2,815	3,660	4,392	5,270	6,324
<b>负债合计</b>	<b>19,081</b>	<b>22,750</b>	<b>27,085</b>	<b>31,553</b>	<b>36,134</b>
实收资本	11,973	13,134	13,134	13,134	13,134
资本公积	10,171	29,817	71,517	126,542	191,594
留存收益	(43)	27	27	27	27
其他综合收益	-	-	-	-	-
少数股东权益	-	-	-	-	-
<b>股东权益合计</b>	<b>22,101</b>	<b>42,978</b>	<b>84,678</b>	<b>139,703</b>	<b>204,755</b>
<b>负债及股东权益合计</b>	<b>41,182</b>	<b>65,728</b>	<b>111,763</b>	<b>171,256</b>	<b>240,889</b>

E=浦银国际预测

注: 英伟达财年(FY, Fiscal Year)为上一年2月至当年1月

### 现金流量表

美元百万	FY2023	FY2024	FY2025E	FY2026E	FY2027E
<b>经营活动产生的现金流量净额</b>	<b>5,641</b>	<b>28,090</b>	<b>49,431</b>	<b>79,306</b>	<b>96,305</b>
净利润	4,368	29,760	61,972	84,243	100,935
折旧及摊销	1,544	1,508	1,495	1,858	2,366
其他营业活动现金流	1,943	822	(2,851)	1,425	(713)
<b>营运资金变动</b>	<b>(2,214)</b>	<b>(4,000)</b>	<b>(11,185)</b>	<b>(8,221)</b>	<b>(6,283)</b>
应收账款(增加)/减少	822	(6,172)	(9,255)	(6,607)	(5,094)
存货(增加)/减少	(2,554)	(98)	(3,947)	(3,299)	(2,432)
应付账款增加/(减少)	1,042	4,070	2,017	1,686	1,243
其他经营资金变动	(1,524)	(1,800)	-	-	-
利息收入(支出)	-	-	-	-	-
<b>投资活动产生的现金流量净额</b>	<b>7,375</b>	<b>(10,566)</b>	<b>(21,427)</b>	<b>(17,887)</b>	<b>(16,144)</b>
资本支出	(1,959)	(2,137)	(4,115)	(5,527)	(6,616)
短期投资	9,334	(8,429)	(17,312)	(12,360)	(9,529)
其他	-	-	-	-	-
<b>融资活动产生的现金流量净额</b>	<b>(11,617)</b>	<b>(13,633)</b>	<b>(17,953)</b>	<b>(26,436)</b>	<b>(32,545)</b>
借款	-	(1,250)	1,586	1,904	2,284
发行股份	(9,684)	(9,130)	(17,581)	(23,614)	(28,265)
发行债券	-	-	-	-	-
发放现金股利	(398)	(395)	(2,691)	(5,604)	(7,618)
其他	(1,535)	(2,858)	732	878	1,054
外汇损益	-	-	-	-	-
<b>现金及现金等价物净流量</b>	<b>1,399</b>	<b>3,891</b>	<b>10,051</b>	<b>34,984</b>	<b>47,616</b>
期初现金及现金等价物	1,990	3,389	7,280	17,331	52,314
<b>期末现金及现金等价物</b>	<b>3,389</b>	<b>7,280</b>	<b>17,331</b>	<b>52,314</b>	<b>99,930</b>

### 主要财务比率

	FY2023	FY2024	FY2025E	FY2026E	FY2027E
<b>营运指标增速</b>					
营业收入增速	0%	126%	93%	34%	20%
毛利润增速	(12%)	188%	99%	34%	20%
营业利润增速	(58%)	681%	120%	34%	20%
净利润增速	(55%)	581%	108%	36%	20%
<b>盈利能力</b>					
净资产收益率	19.8%	69.2%	73.2%	60.3%	49.3%
总资产报酬率	10.6%	45.3%	55.4%	49.2%	41.9%
投入资本回报率	11.9%	48.9%	59.4%	52.2%	44.0%
<b>利润率</b>					
毛利率	56.9%	72.7%	75.2%	75.0%	75.0%
营业利润率	15.7%	54.1%	61.9%	61.6%	61.6%
净利润率	16.2%	48.8%	52.8%	53.5%	53.5%
<b>营运能力</b>					
<b>现金循环周期</b>					
应收账款周转天数	57	41	46	52	55
存货周转天数	122	115	91	101	107
应付账款周转天数	47	43	47	51	54
净债务(净现金)	11,684	9,111	647	(32,433)	(77,765)
自由现金流	1,739	25,131	48,167	72,354	90,402

## ● 英伟达公司速览

2020年，英伟达推出 A100。作为首款采用 Ampere 架构的 GPU，A100 较前一代 V100 在 HPC 上有 1.5x-2x 的效率提升。A100 允许将单个 GPU 分割成多个小的独立 GPU，大大提升了云和数据中心的资源分配效率。A100 能在 FP32 到 NT4 的精度范围内进行加速，提升深度学习推理效率。

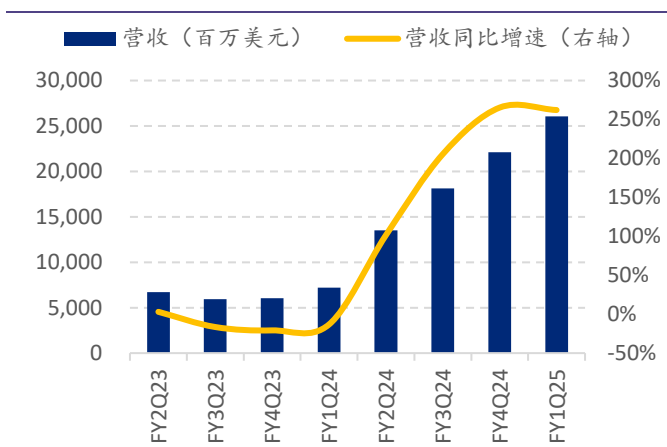
2022年，英伟达推出 H100。H100 升级了 Tensor 核心，支持双精度 (FP64)、单精度 (FP32)、半精度 (FP16) 和整数 (INT8) 计算负载，显著提高了 AI 训练和推理的速度。以 GPT-3 为例，H100 在 1750 亿参数的大模型上性能较 A100 快 4x。如果采用 NVLink 方案，H100 在 3950 亿参数的大模型上速度为 A100 的 6x。

2024年，英伟达推出 H200。H200 是首款采用 HBM3e 的 GPU。在 Llama 大模型上，其性能较 H100 提升了 1.4x-1.9x。由于 H200 基于 Hopper 架构，与 H100 相互兼容，因此人工智能公司无需更改其服务器系统或软件即可使用新版本。

2024年 GTC 大会上，英伟达推出旗舰款 GPU B100，搭载 Blackwell 架构，采用先进的 3nm 制程技术。将两个计算芯片连接到 8 个 8-Hi HBM3e 显存堆栈，总容量高达 192GB。这种 2.5D 封装技术不仅提高了 B100 处理能力，同时也节省了空间并降低了功耗，B100、B200 和 GB200，其主要区别在于功率和性能。据英伟达介绍，这些芯片的工作功率范围可在 700W 至 1200W 之间，视具体型号和冷却方式而定。

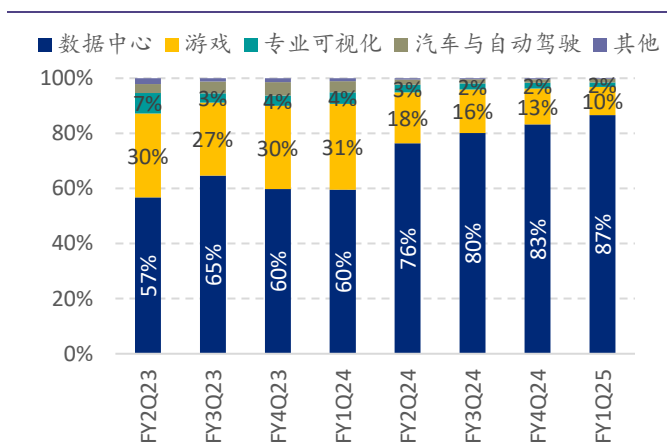
最早，NVIDIA 针对美国禁售的 A100、H100，为中国市场设计了特供版 A800、H20。虽然这两款特供版芯片在传输带宽上有所“阉割”，但是其算力并未减少。理论上有可能还原到 A100、H100 原有的性能。

图表 60: 英伟达季度收入及同比增速



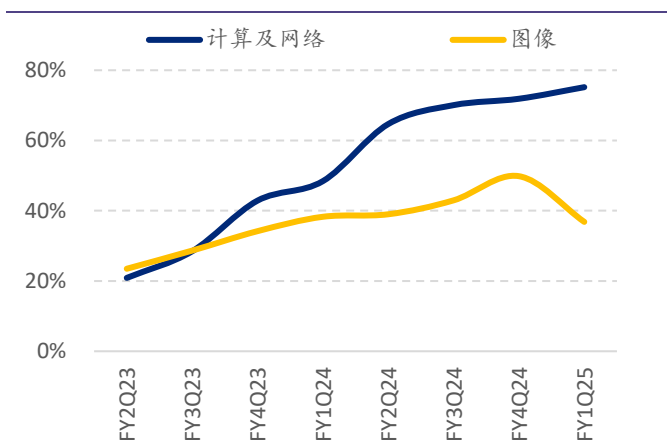
注: FY = Fiscal Year, 财历年  
资料来源: Wind、浦银国际

图表 61: 英伟达季度收入拆分



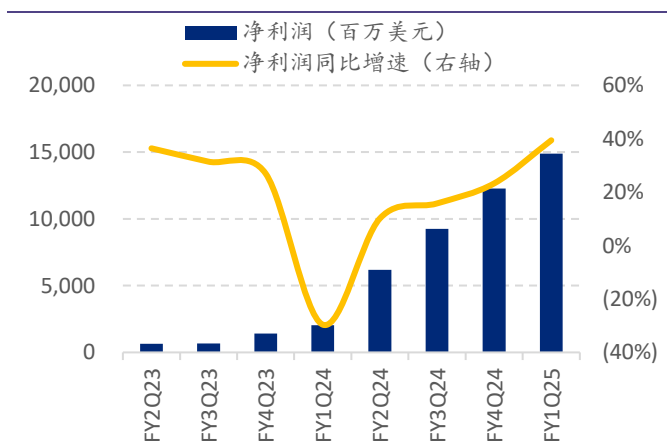
注: FY = Fiscal Year, 财历年  
资料来源: Wind、浦银国际

图表 62: 英伟达季度营业利润率拆分



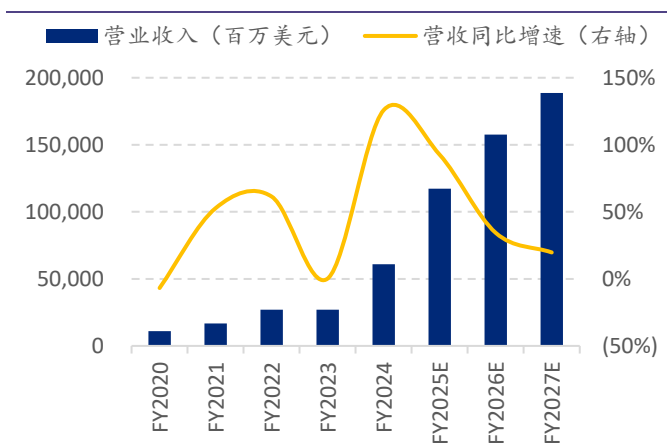
注: FY = Fiscal Year, 财历年  
资料来源: Wind、浦银国际

图表 63: 英伟达季度利润及同比增速



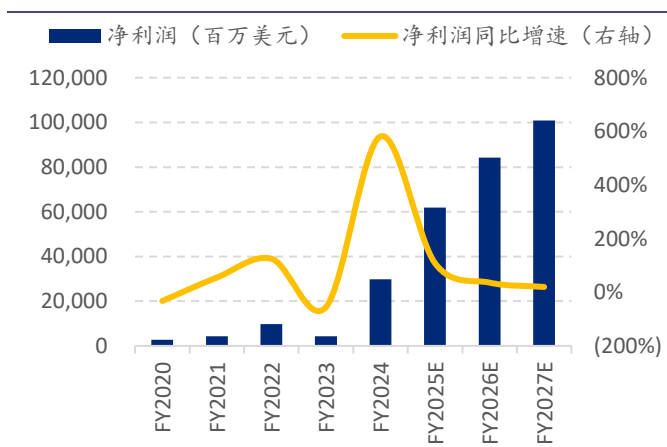
注: FY = Fiscal Year, 财历年  
资料来源: Wind、浦银国际

图表 64: 英伟达年度收入及增速预测



注: E=浦银国际预测; FY = Fiscal Year, 财历年  
资料来源: Wind、浦银国际

图表 65: 英伟达年度利润及增速预测



注: E=浦银国际预测; FY = Fiscal Year, 财历年  
资料来源: Wind、浦银国际

## ● 估值

我们采用 DCF（Discounted Cash Flow，现金流量贴现法）估值方法。我们假设英伟达 FY2030-FY2034 的营收成长率为 30%-35%，永久增长率为 3%。另外，我们假设 WACC（Weighted Average Cost of Capital，加权平均资金成本）是 16.3%，得到英伟达目标价为 147.6 美元，潜在升幅 19%，对应 FY2025 市盈率为 59.3x，首予“买入”评级。

图表 66：英伟达 WACC 假设

WACC 计算			
Beta	1.6	债务成本	12.3%
无风险利率	4.6%	债务股本比	17.5%
股权风险溢价	8.1%	所得税率	14.0%
股本成本	<b>17.5%</b>	<b>WACC</b>	<b>16.3%</b>

注：WACC，WeightedAverageCostofCapital，加权平均资金成本

资料来源：浦银国际预测

图表 67：英伟达自由现金流预测

美元百万	FY2025E	FY2026E	FY2027E	FY2028E	FY2029E	FY2030E	FY2031E	FY2032E	FY2033E	FY2034E	FY2035 往后
营业收入	117,311	157,568	188,605	263,557	363,117	490,208	661,781	860,316	1,118,411	1,453,934	
营收增速	93%	34%	20%	40%	38%	35%	35%	30%	30%	30%	
经营利润	72,558	97,044	116,207	165,771	232,142	314,372	425,726	555,164	723,950	944,043	
经营利润率	61.9%	61.6%	61.6%	62.9%	63.9%	64.1%	64.3%	64.5%	64.7%	64.9%	
加：折旧及摊销	1,495	1,858	2,366	2,954	3,825	5,164	6,971	9,063	11,782	15,316	
EBITDA	74,053	98,903	118,573	168,726	235,967	319,536	432,697	564,227	735,732	959,359	
EBITDA 率	63.1%	62.8%	62.9%	64.0%	65.0%	65.2%	65.4%	65.6%	65.8%	66.0%	
所得税率	16.0%	14.2%	14.0%	14.0%	14.0%	14.0%	14.0%	14.0%	14.0%	14.0%	
资本支出	(4,115)	(5,527)	(6,616)	(9,245)	(12,737)	(14,648)	(16,845)	(18,530)	(20,383)	(22,421)	
资本支出占营收比	3.5%	3.5%	3.5%	3.5%	3.5%	3.0%	2.5%	2.2%	1.8%	1.5%	
净营运资本变动	(11,185)	(8,221)	(6,283)	(8,734)	(17,531)	(16,567)	(15,656)	(14,247)	(9,260)	(6,019)	
自由现金流	70,360	98,957	121,944	173,954	238,198	332,333	459,798	609,173	807,442	1,063,085	9,584,809
永续增长率											3.0%

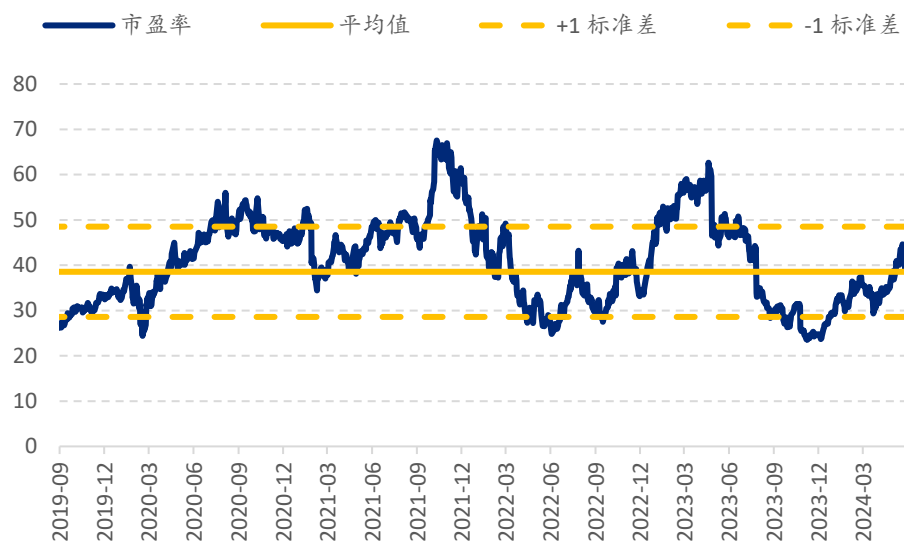
资料来源：浦银国际预测

图表 68：英伟达 DCF 估值预测（FY2025）

WACC	自由现金流现值 (美元百万)	净现金 (美元百万)	权益价值 (美元百万)	股数 (百万)	每股价值 (美元)
16.3%	3,675,803	(647)	3,675,156	24,900	147.6

资料来源：浦银国际预测

图表 69：英伟达历史 PE：2019 年以来均值 38.5x



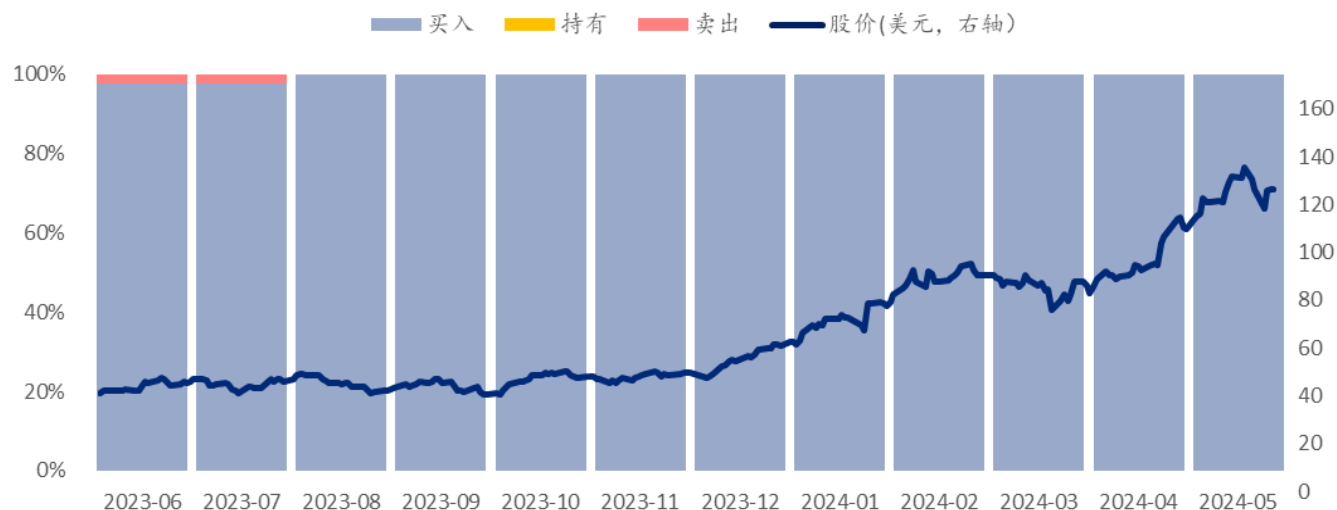
注：截至 2024 年 6 月 27 日收盘价

资料来源：Factset、浦银国际



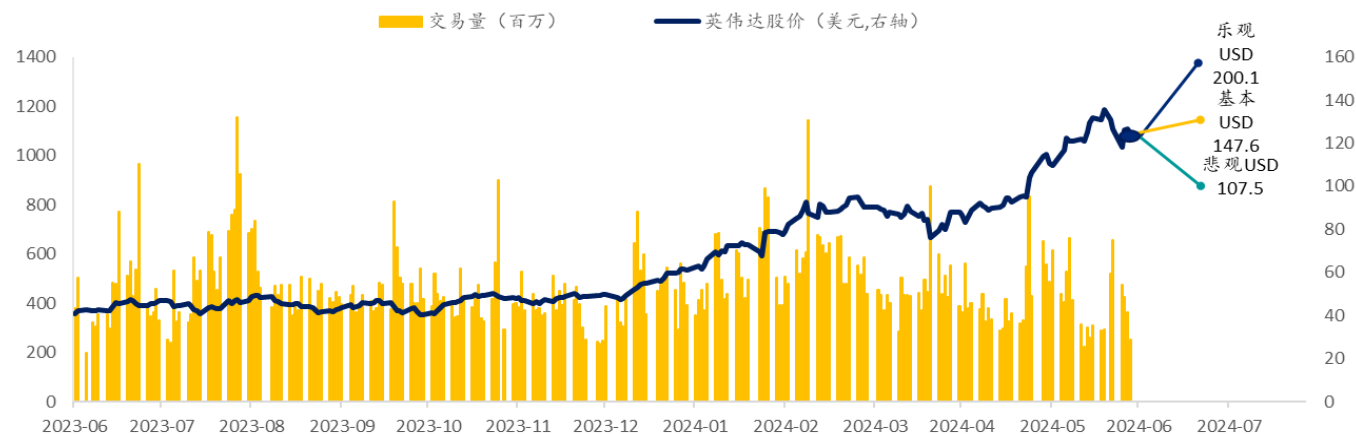
## ● SPDBI 乐观与悲观情景假设

图表 70: 市场普遍预期: 英伟达 (NVDA.US)



资料来源: Factset、浦银国际

图表 71: SPDBI 情景假设: 英伟达 (NVDA.US)



### 乐观情景: 公司收入增长好于预期

目标价: 200.1 美元 (概率: 20%)

- AI 算力需求增长好于预期, 需求持续性动能充裕
- 维持供需紧平衡, 价格有上行动能
- 新能源车智驾渗透率大幅提升
- 公司产品持续快速迭代, 推动 GPU 保持市场领先

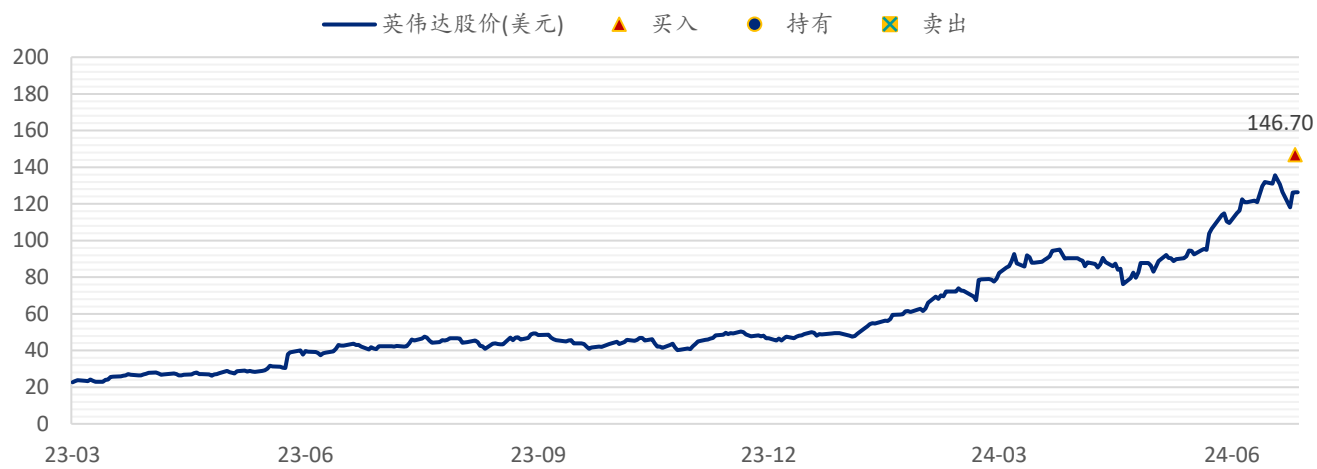
### 悲观情景: 公司收入增长不及预期

目标价: 107.5 美元 (概率: 10%)

- AI 基建投入不及预期, GPU 及 AI 服务器需求提升速度不及预期
- 竞争对手或者 AI 大模型厂商的产品发力, 公司在 AI 训练或推理市场份额下行速度较快
- 费用等投入较高, 拖累利润表现

资料来源: 浦银国际预测

图表 72: SPDBI 目标价: 英伟达 (NVDA.US)



资料来源: Factset、浦银国际

## ● 风险提示

### 下行风险

- 美国或全球经济下行，多个下游需求动能不足；
- AI 需求爆发持续性弱于预期，大模型厂商盈利低于预期；
- 半导体周期上行动能不足；
- 行业竞争加剧，拖累利润表现；
- 研发等费用率增长较快；
- GPU 芯片等迭代速度和性能低于预期。

## ● 公司背景

图表 73: 英伟达发展里程碑

年份	里程碑
1993 年	英伟达成立于美国加州
1999 年	英伟达进行了一项开创性的发明：图形处理单元（GPU）。同年挂牌上市
2006 年	英伟达推出 CUDA 架构，将 GPU 的并行处理能力开放给科学研究和计算
2012 年	英伟达开始布局人工智能
2015 年	英伟达发布支持高级驾驶辅助系统的 NVIDIA DRIVE
2016 年	英伟达推出 PASCAL 架构
2016 年	英伟达推出 DGX-1，为全球首款深度学习一体化超级计算机
2018 年	推出 RTX GPU，是第一个支持实时光线追踪的产品
2020 年	收购 Mellanox
2020 年	推出 Ampere 架构与首款采用 Ampere 架构的 GPU A100
2022 年	推出 H100，升级了 Tensor 核心，显著提高了 AI 训练和推理的速度
2023 年	推出 GH200 芯片 DGX 超级计算系统
2024 年	推出 H200 芯片，是首款采用 HBM3e 的 GPU。推出旗舰款 GPU B100，搭载 Blackwell 架构，采用先进的 3nm 制程技术

资料来源：公开资料整理、浦银国际

## ● 财务报表

图表 74：英伟达：损益表

美元百万	FY2023	FY2024	FY2025E	FY2026E	FY2027E
<b>营业收入</b>	<b>26,974</b>	<b>60,922</b>	<b>117,311</b>	<b>157,568</b>	<b>188,605</b>
营业成本	(11,618)	(16,621)	(29,040)	(39,421)	(47,073)
<b>毛利</b>	<b>15,356</b>	<b>44,301</b>	<b>88,271</b>	<b>118,147</b>	<b>141,531</b>
<b>营业支出</b>	<b>(11,132)</b>	<b>(11,329)</b>	<b>(15,713)</b>	<b>(21,103)</b>	<b>(25,324)</b>
销售、行政及一般费用	(2,440)	(2,654)	(4,022)	(5,399)	(6,458)
研发费用	(7,339)	(8,675)	(11,692)	(15,704)	(18,866)
重组及其他费用/收入	(1,353)	-	-	-	-
<b>营业利润</b>	<b>4,224</b>	<b>32,972</b>	<b>72,558</b>	<b>97,044</b>	<b>116,207</b>
<b>营业外收支</b>	<b>(43)</b>	<b>846</b>	<b>1,214</b>	<b>1,167</b>	<b>1,159</b>
利息收入/(支出)净额	5	609	964	937	935
其他收益/(支出)净额	(48)	237	250	230	225
<b>税前利润</b>	<b>4,181</b>	<b>33,818</b>	<b>73,772</b>	<b>98,211</b>	<b>117,367</b>
所得税	187	(4,058)	(11,800)	(13,968)	(16,431)
<b>净利润</b>	<b>4,368</b>	<b>29,760</b>	<b>61,972</b>	<b>84,243</b>	<b>100,935</b>
少数股东和可赎回少数股东损益	-	-	-	-	-
<b>归属普通股股东净利润</b>	<b>4,368</b>	<b>29,760</b>	<b>61,972</b>	<b>84,243</b>	<b>100,935</b>
基本股数(百万)	2,490	2,490	24,900	24,900	24,900
摊万股数(百万)	2,490	2,490	24,900	24,900	24,900
<b>基本每股收益(美元)</b>	<b>1.75</b>	<b>11.95</b>	<b>2.49</b>	<b>3.38</b>	<b>4.05</b>
<b>稀释每股收益(美元)</b>	<b>1.75</b>	<b>11.95</b>	<b>2.49</b>	<b>3.38</b>	<b>4.05</b>

注：E=浦银国际预测；英伟达财务年（FY, Fiscal Year）为上一年2月至当年1月

资料来源：公司公告、浦银国际



图表 75：英伟达：资产负债表

美元百万	FY2023	FY2024	FY2025E	FY2026E	FY2027E
货币资金	3,389	7,280	17,331	52,314	99,930
受限制资金	9,907	18,704	36,016	48,376	57,905
应收账款	3,827	9,999	19,254	25,861	30,955
存货	5,159	5,282	9,229	12,528	14,959
其他流动资产	791	3,080	5,931	4,505	5,218
<b>流动资产合计</b>	<b>23,073</b>	<b>44,345</b>	<b>87,760</b>	<b>143,584</b>	<b>208,968</b>
固定资产-物业, 厂房及设备	3,807	3,914	6,534	10,203	14,452
无形资产	1,676	1,112	1,112	1,112	1,112
商誉	4,372	4,430	4,430	4,430	4,430
租赁权	1,038	1,346	1,346	1,346	1,346
其他非流动资产	7,216	10,581	10,581	10,581	10,581
<b>资产总计</b>	<b>41,182</b>	<b>65,728</b>	<b>111,763</b>	<b>171,256</b>	<b>240,889</b>
短期借款	5,370	7,932	9,518	11,422	13,706
应付账款	1,193	2,699	4,716	6,401	7,644
其他流动负债	-	-	-	-	-
<b>流动负债合计</b>	<b>6,563</b>	<b>10,631</b>	<b>14,234</b>	<b>17,823</b>	<b>21,350</b>
长期借款	9,703	8,459	8,459	8,459	8,459
其他非流动负债	2,815	3,660	4,392	5,270	6,324
<b>负债合计</b>	<b>19,081</b>	<b>22,750</b>	<b>27,085</b>	<b>31,553</b>	<b>36,134</b>
实收资本	11,973	13,134	13,134	13,134	13,134
资本公积	10,171	29,817	71,517	126,542	191,594
留存收益	(43)	27	27	27	27
其他综合收益	-	-	-	-	-
少数股东权益	-	-	-	-	-
<b>股东权益合计</b>	<b>22,101</b>	<b>42,978</b>	<b>84,678</b>	<b>139,703</b>	<b>204,755</b>
<b>负债及股东权益和计</b>	<b>41,182</b>	<b>65,728</b>	<b>111,763</b>	<b>171,256</b>	<b>240,889</b>

注：E=浦银国际预测；英伟达财务年（FY, Fiscal Year）为上一年2月至当年1月

资料来源：公司公告、浦银国际

图表 76：英伟达：现金流量表

美元百万	FY2023	FY2024	FY2025E	FY2026E	FY2027E
<b>经营活动产生的现金流量净额</b>	<b>5,641</b>	<b>28,090</b>	<b>49,431</b>	<b>79,306</b>	<b>96,305</b>
净利润	4,368	29,760	61,972	84,243	100,935
折旧及摊销	1,544	1,508	1,495	1,858	2,366
其他营业活动现金流	1,943	822	(2,851)	1,425	(713)
<b>营运资金变动</b>	<b>(2,214)</b>	<b>(4,000)</b>	<b>(11,185)</b>	<b>(8,221)</b>	<b>(6,283)</b>
应收账款(增加)/减少	822	(6,172)	(9,255)	(6,607)	(5,094)
存货(增加)/减少	(2,554)	(98)	(3,947)	(3,299)	(2,432)
应付账款增加/(减少)	1,042	4,070	2,017	1,686	1,243
其他经营资金变动	(1,524)	(1,800)	-	-	-
利息收入(支出)	-	-	-	-	-
<b>投资活动产生的现金流量净额</b>	<b>7,375</b>	<b>(10,566)</b>	<b>(21,427)</b>	<b>(17,887)</b>	<b>(16,144)</b>
资本支出	(1,959)	(2,137)	(4,115)	(5,527)	(6,616)
短期投资	9,334	(8,429)	(17,312)	(12,360)	(9,529)
其他	-	-	-	-	-
<b>融资活动产生的现金流量净额</b>	<b>(11,617)</b>	<b>(13,633)</b>	<b>(17,953)</b>	<b>(26,436)</b>	<b>(32,545)</b>
借款	-	(1,250)	1,586	1,904	2,284
发行股份	(9,684)	(9,130)	(17,581)	(23,614)	(28,265)
发行债券	-	-	-	-	-
发放现金股利	(398)	(395)	(2,691)	(5,604)	(7,618)
其他	(1,535)	(2,858)	732	878	1,054
外汇损益	-	-	-	-	-
<b>现金及现金等价物净流量</b>	<b>1,399</b>	<b>3,891</b>	<b>10,051</b>	<b>34,984</b>	<b>47,616</b>
期初现金及现金等价物	1,990	3,389	7,280	17,331	52,314
<b>期末现金及现金等价物</b>	<b>3,389</b>	<b>7,280</b>	<b>17,331</b>	<b>52,314</b>	<b>99,930</b>

注：E=浦银国际预测；英伟达财年（FY，Fiscal Year）为上一年2月至当年1月

资料来源：公司公告、浦银国际

# 高通 (QCOM.US) 首次覆盖：端侧 AI 推动公司开启大成长周期

我们首次覆盖高通 (QCOM.US)，给予“买入”评级，目标价为人民币 240.7 美元，潜在升幅 23%。

- 端侧 AI 的最大受益标的，开启大成长周期：**高通正在进入大成长周期，主要驱动力来源于：1) AI 智能手机渗透率快速上扬，带动出货量的价格上升，2) 新能源车业务维持高份额高增长的增长动能，3) 全球半导体基本面周期依然在上行，仍然有较大的上行时间和空间机会。我们预期高通 2024 财年和 2025 财年净利润将分别同比增长 30% 和 13%。当前市盈率 19.1x，在美降息预期以及端侧 AI 带来的估值溢价支撑下，我们维持对于公司比较乐观的态度。
- 智能手机需求复苏，AI 智能手机拉动平均单价成长：**作为 AI 落地端侧最大的终端，AI 智能手机渗透率有望在 2024 年和 2025 年达到 11% 和 24%，并且带动智能手机换机需求，推动出货量进入温和增长。进一步看，我们预计端侧 AI 带动 SoC 价值上浮约 25%。因此，高通已经步入手机芯片量价齐升的大周期，是端侧 AI 的受益标的。
- 新能源车业务有望提供长期增量：**新能源车的智能驾驶和智能座舱也是端侧 AI 落地的重要领域。高通的智能座舱芯片已经取得较高市场份额，并可以享受行业持续增长动能。进一步看，未来新能源车舱驾一体需求有望提升高通在新能源车中的价值量，推动公司该业务板块的长期增长。
- 估值：**我们采用 DCF (Discounted Cash Flow, 现金流量贴现法) 估值方法。我们假设高通 FY2029-FY2033 的成长率为 12%，永久增长率为 3%，WACC (Weighted Average Cost of Capital, 加权平均资金成本) 是 11.2%，得到高通目标价为 240.7 美元，潜在升幅 23%，对应 FY2024 PE 为 28.8x，首予“买入”评级。
- 投资风险：**美国或全球经济下行，智能手机、新能源车等下游需求动能不足；AI 智能手机渗透率上扬速度较慢，渗透中低端价格段手机较慢；半导体周期上行动能不足；行业竞争加剧，拖累利润表现；研发等费用率增长较快；芯片等迭代速度和性能低于预期。

图表 77：盈利预测和财务指标 (FY2022-FY2026E)

美元百万	FY2022	FY2023	FY2024E	FY2025E	FY2026E
营业收入	44,200	35,820	38,415	42,763	45,713
营收同比增速	32%	-19%	7%	11%	7%
净利润	12,936	7,232	9,389	10,633	11,447
净利润增速	43%	-44%	30%	13%	8%
目标 PE (x)	20.8	37.2	28.8	25.2	23.4

E=浦银国际预测 资料来源：公司公告、浦银国际

## 高通 (QCOM.US)

买入

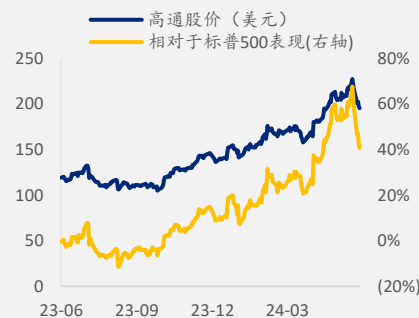
目标价 (美元)	240.7
潜在升幅/降幅	+23%
目前股价 (美元)	195.2
52 周内股价区间 (美元)	102.7-230.6
总市值 (百万美元)	217,787
近 90 日日均成交额 (百万美元)	9.51

注：截至 2024 年 6 月 27 日收盘价

## 市场预期区间



## 股价相对表现



注：截至 2024 年 6 月 27 日收盘价

资料来源：Factset、浦银国际

浦银国际

首次覆盖

高通 (QCOM.US) 首次覆盖

## 财务报表分析与预测

### 利润表

美元百万	FY2022	FY2023	FY2024E	FY2025E	FY2026E
<b>营业收入</b>	<b>44,200</b>	<b>35,820</b>	<b>38,415</b>	<b>42,763</b>	<b>45,713</b>
营业成本	(18,635)	(15,869)	(17,014)	(19,190)	(20,984)
<b>毛利</b>	<b>25,565</b>	<b>19,951</b>	<b>21,401</b>	<b>23,573</b>	<b>24,729</b>
<b>营业支出</b>	<b>(9,705)</b>	<b>(12,163)</b>	<b>(11,275)</b>	<b>(11,633)</b>	<b>(11,864)</b>
销售、行政及一般费用	(2,570)	(2,483)	(2,591)	(2,885)	(3,084)
研发费用	(8,194)	(8,818)	(8,712)	(8,748)	(8,780)
重组及其他费用/收入	1,059	(862)	28	-	-
<b>营业利润</b>	<b>15,860</b>	<b>7,788</b>	<b>10,126</b>	<b>11,940</b>	<b>12,864</b>
<b>营业外收支</b>	<b>(862)</b>	<b>(345)</b>	<b>51</b>	<b>(263)</b>	<b>(258)</b>
利息收入/(支出)净额	(490)	(694)	(699)	(697)	(697)
其他收益/(支出)净额	(372)	349	749	434	439
<b>税前利润</b>	<b>14,998</b>	<b>7,443</b>	<b>10,177</b>	<b>11,677</b>	<b>12,607</b>
所得税	(2,012)	(104)	(795)	(1,044)	(1,160)
<b>净利润(不含少数股东权益)</b>	<b>12,986</b>	<b>7,339</b>	<b>9,382</b>	<b>10,633</b>	<b>11,447</b>
少数股东和可赎回少数股东损益	-	-	-	-	-
<b>归属普通股股东净利润</b>	<b>12,936</b>	<b>7,232</b>	<b>9,389</b>	<b>10,633</b>	<b>11,447</b>
基本股数(百万)	1,119	1,118	1,123	1,112	1,112
摊销股数(百万)	1,119	1,118	1,123	1,112	1,112
<b>基本每股收益(美元)</b>	<b>11.56</b>	<b>6.47</b>	<b>8.36</b>	<b>9.56</b>	<b>10.30</b>
<b>稀释每股收益(美元)</b>	<b>11.56</b>	<b>6.47</b>	<b>8.36</b>	<b>9.56</b>	<b>10.30</b>

### 资产负债表

美元百万	FY2022	FY2023	FY2024E	FY2025E	FY2026E
<b>货币资金</b>	<b>2,773</b>	<b>8,450</b>	<b>8,020</b>	<b>10,067</b>	<b>12,118</b>
现金及其等价物	3,609	2,874	2,874	2,874	2,874
应收账款	5,643	3,183	3,414	3,800	4,062
存货	6,341	6,422	6,885	6,213	5,435
其他流动资产	2,358	1,535	1,947	1,741	1,844
<b>流动资产合计</b>	<b>20,724</b>	<b>22,464</b>	<b>23,139</b>	<b>24,694</b>	<b>26,333</b>
固定资产-物业、厂房及设备	5,168	5,042	4,869	5,031	5,176
无形资产	1,882	1,408	1,408	1,408	1,408
商誉	10,508	10,642	10,642	10,642	10,642
租赁权	-	-	-	-	-
其他非流动资产	10,732	11,484	13,781	16,537	19,844
<b>资产总计</b>	<b>49,014</b>	<b>51,040</b>	<b>53,839</b>	<b>58,312</b>	<b>63,403</b>
短期借款	1,945	914	715	814	765
应付账款	3,796	1,912	2,050	2,312	2,528
其他流动负债	6,125	6,802	6,802	6,802	6,802
<b>流动负债合计</b>	<b>11,866</b>	<b>9,628</b>	<b>9,567</b>	<b>9,929</b>	<b>10,095</b>
长期借款	13,537	14,484	14,484	14,484	14,484
其他非流动负债	5,598	5,347	5,347	5,347	5,347
<b>负债合计</b>	<b>31,001</b>	<b>29,459</b>	<b>29,398</b>	<b>29,760</b>	<b>29,926</b>
实收资本	195	490	490	490	490
资本公积	17,840	20,733	23,593	27,704	32,629
留存收益	(22)	358	358	358	358
其他综合收益	-	-	-	-	-
少数股东权益	-	-	-	-	-
<b>股东权益合计</b>	<b>18,013</b>	<b>21,581</b>	<b>24,441</b>	<b>28,552</b>	<b>33,477</b>
<b>负债及股东权益合计</b>	<b>49,014</b>	<b>51,040</b>	<b>53,839</b>	<b>58,312</b>	<b>63,403</b>

©浦银国际预测

注: 高透财年(FY, Fiscal Year)为上一年8月至当年8月

### 现金流量表

美元百万	FY2022	FY2023	FY2024E	FY2025E	FY2026E
<b>经营活动产生的现金流量净额</b>	<b>9,096</b>	<b>11,299</b>	<b>10,176</b>	<b>13,130</b>	<b>13,836</b>
净利润	12,986	7,339	9,382	10,633	11,447
折旧及摊销	1,762	1,809	1,761	1,743	1,760
其他营业活动现金流	2,148	1,003	(412)	206	(103)
<b>营运资金变动</b>	<b>(7,800)</b>	<b>1,148</b>	<b>(556)</b>	<b>548</b>	<b>732</b>
应收账款(增加)/减少	(2,066)	2,472	(231)	(386)	(262)
存货(增加)/减少	(3,137)	8	(463)	672	778
应付账款增加/(减少)	1,036	(1,880)	138	262	216
其他经营资金变动	(3,633)	548	-	-	-
利息收入(支出)	-	-	-	-	-
<b>投资活动产生的现金流量净额</b>	<b>(5,804)</b>	<b>762</b>	<b>(3,884)</b>	<b>(4,661)</b>	<b>(5,213)</b>
资本支出	(2,262)	(1,323)	(1,588)	(1,905)	(1,905)
长期投资	1,208	898	-	-	-
长期投资	(4,780)	(215)	-	-	-
其他	30	1,402	(2,297)	(2,756)	(3,307)
<b>融资活动产生的现金流量净额</b>	<b>(7,196)</b>	<b>(6,663)</b>	<b>(6,721)</b>	<b>(6,422)</b>	<b>(6,572)</b>
借款	(352)	(498)	(199)	100	(50)
发行股份	(3,539)	(3,060)	(3,060)	(3,060)	(3,060)
发行债券	(63)	434	-	-	-
发放现金股利	(3,212)	(3,462)	(3,462)	(3,462)	(3,462)
其他	(30)	(77)	-	-	-
外汇损益	-	-	-	-	-
<b>现金及现金等价物净流量</b>	<b>(3,904)</b>	<b>5,398</b>	<b>(430)</b>	<b>2,046</b>	<b>2,052</b>
期初现金及现金等价物	7,116	3,099	8,527	8,097	10,144
<b>期末现金及现金等价物</b>	<b>3,099</b>	<b>8,527</b>	<b>8,097</b>	<b>10,144</b>	<b>12,195</b>

### 主要财务比率

	FY2022	FY2023	FY2024E	FY2025E	FY2026E
<b>营运指标增速</b>					
营业收入增速	32%	(19%)	7%	11%	7%
毛利润增速	32%	(22%)	7%	10%	5%
营业利润增速	62%	(51%)	30%	18%	8%
净利润增速	43%	(44%)	30%	13%	8%
<b>盈利能力</b>					
净资产收益率	71.8%	33.5%	38.4%	37.2%	34.2%
总资产报酬率	26.4%	14.2%	17.4%	18.2%	18.1%
投入资本回报率	41.0%	20.8%	23.6%	24.8%	24.0%
<b>利润率</b>					
毛利率	57.8%	55.7%	55.7%	55.1%	54.1%
营业利润率	35.9%	21.7%	26.4%	27.9%	28.1%
净利润率	20.2%	24.4%	24.9%	25.0%	27.0%
<b>营运能力</b>					
<b>现金循环周期</b>					
应收账款周转天数	38	45	31	31	31
存货周转天数	94	147	143	125	101
应付账款周转天数	64	66	42	41	42
净债务(净现金)	12,709	6,948	7,179	5,232	3,130
自由现金流	4,686	8,973	9,000	11,019	12,034

## ● 高通公司速览

SoC (System on Chip) 是指在一块芯片上集成一整个信息处理系统。智能手机 SoC 芯片通常为 CPU、GPU、NPU 计算芯片，通信基带芯片、影音处理、显示、内存芯片集合，包含一套完整处理系统。基带芯片 (Baseband Chip) 内含调制调节器 (Modem)、信道编码器、数字信号编码器等。

基带芯片的追赶难度极大。除了芯片设计外，由于 5G 基带芯片同时要兼容 2G、3G、4G 的信号，因此目前全球通信协议极为复杂。高通等头部玩家积累了大量的技术优势和专利，后来者很容易遇到专利诉讼。当前一代技术研发成功后，高通、联发科等已经开始下一代产片、研发投入。

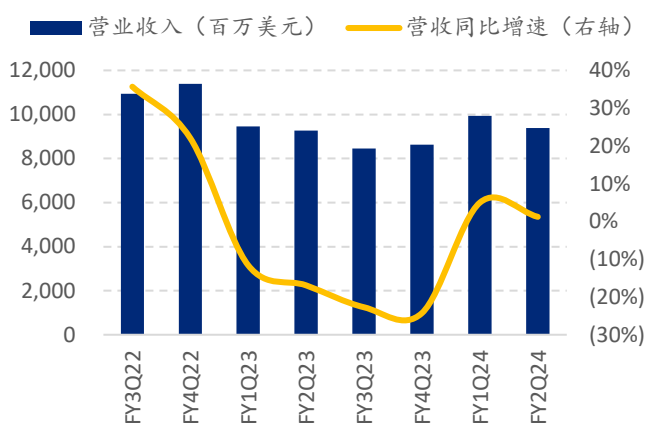
苹果 iPhone 15 系列采用了高通 X70 的调制调节器，iPhone 14 系列使用了高通的 X65。根据 SpeedSmart 网站测试，iPhone 15 系列相比 iPhone 14 系列 5G 速度提升幅度达到 24%。根据高通的业绩会，苹果将 QTL 授权向后延续了 2 年至 2027 年 3 月。公司预期 AI 将会带来其芯片平均售价提升 10% 左右。

2016 年 iPhone 7 系列，苹果开始引入英特尔基带，打破了之前和高通签订的独供协议。2016-2018 年，高通与苹果产生了多起诉讼与反诉讼摩擦。2018 年苹果 Xs 系列由英特尔独供基带芯片。由于英特尔基带芯片性能不稳定，同时苹果在 5G 开发上较为落后，因此在 2019 年 4 月，苹果与高通达成和解，支付高通 45 亿美元和解费用，并重新签订了一份授权协议。2020 年 iPhone 12 开始，苹果基带部分继续由高通独供。

2023 年，高通在智能手机 SoC 出货量位列第二，联发科排在第一。但是，高通在高端市场稳居第一。2023 年，联发科平均单价约为高通的三分之一。5G 基带芯片玩家有高通、联发科、紫光展锐、三星、华为海思等。手机 SoC 芯片的设计难点主要在于通信生态和专利。根据 TechInsights 统计，2022 年高通在全球基带芯片市场营收占比达到 60.9%，仍然维持霸主地位。

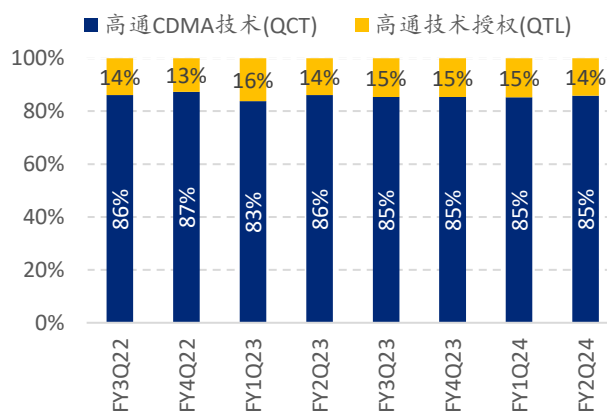


图表 78: 高通季度收入及同比增速



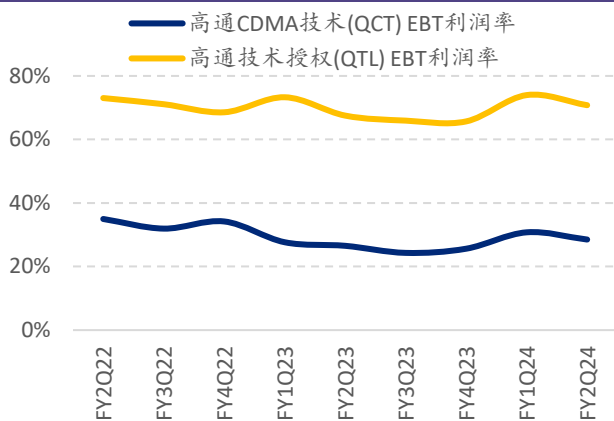
注: 高通财年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月  
资料来源: Wind、浦银国际

图表 79: 高通季度收入拆分



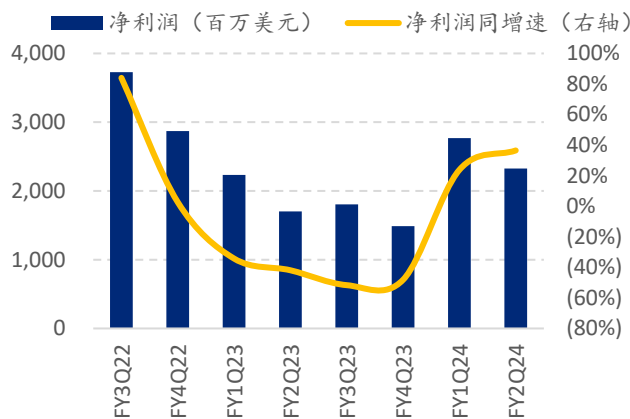
注: 高通财年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月  
资料来源: Wind、浦银国际

图表 80: 高通季度营业利润率拆分



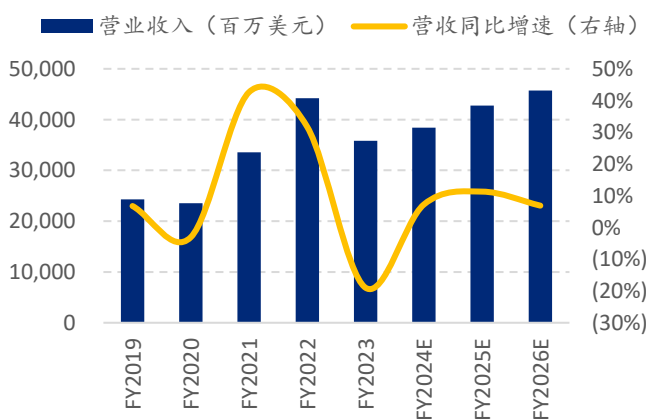
注: 高通财年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月  
资料来源: Wind、浦银国际

图表 81: 高通季度净利润及同比增速



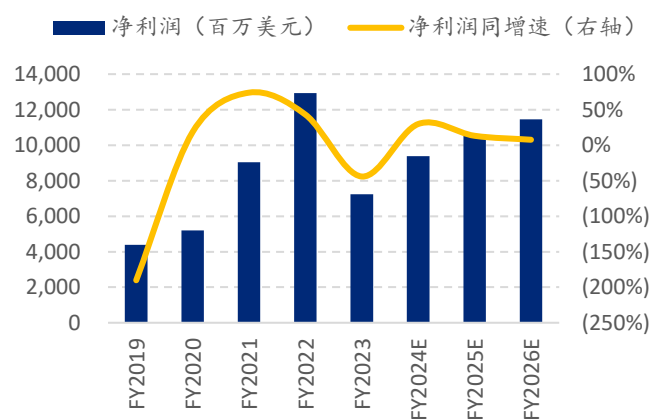
注: 高通财年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月  
资料来源: Wind、浦银国际

图表 82: 高通年度收入及增速预测



注: E=浦银国际预测; 高通财年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月  
资料来源: Wind、浦银国际

图表 83: 高通年度净利润及增速预测



注: E=浦银国际预测; 高通财年 (FY, Fiscal Year) 为上一年 8 月至当年 9 月  
资料来源: Wind、浦银国际

## ● 估值

我们采用 DCF（Discounted Cash Flow，现金流量贴现法）估值方法。我们假设高通 FY2029 年-FY2033 年的营收成长率为 12%，永久增长率为 3%。另外，我们假设 WACC（Weighted Average Cost of Capital，加权平均资金成本）是 11.2%。其他基本假设可以参考下方两个表格，得到目标价为 240.7 美元，潜在升幅 23%，对应 FY2024 年市盈率为 28.8x，首予“买入”评级。

图表 84：高通 WACC 假设

WACC 计算			
Beta	1.3	债务成本	9.1%
无风险利率	4.3%	债务股本比	38.3%
股权风险溢价	6.7%	所得税率	8.7%
股本成本	<b>13.0%</b>	<b>WACC</b>	<b>11.2%</b>

注：WACC，Weighted Average Cost of Capital，加权平均资金成本

资料来源：浦银国际预测

图表 85：高通自由现金流预测

美元百万	FY2024E	FY2025E	FY2026E	FY2027E	FY2028E	FY2029E	FY2030E	FY2031E	FY2032E	FY2033E	FY2034 往后
营业收入	38,415	42,763	45,713	48,769	53,967	60,443	67,696	75,819	84,918	95,108	
营收增速	7%	11%	7%	7%	11%	12%	12%	12%	12%	12%	
经营利润	10,126	11,940	12,864	14,739	16,560	18,608	20,909	23,494	26,398	29,661	
经营利润率	26.4%	27.9%	28.1%	30.2%	30.7%	30.8%	30.9%	31.0%	31.1%	31.2%	
加：折旧及 摊销	1,761	1,743	1,760	1,775	1,788	2,003	2,243	2,512	2,814	3,151	
EBITDA	11,887	13,683	14,624	16,513	18,349	20,611	23,152	26,006	29,212	32,812	
EBITDA 率	30.9%	32.0%	32.0%	33.9%	34.0%	34.1%	34.2%	34.3%	34.4%	34.5%	
所得税率	7.8%	8.9%	9.2%	9.2%	9.2%	9.1%	9.0%	8.9%	8.8%	8.7%	
资本支出	(1,588)	(1,905)	(1,905)	(1,905)	(1,905)	(2,038)	(2,181)	(2,334)	(2,497)	(2,672)	
资本支出占 营收比	4.1%	4.5%	4.2%	3.9%	3.5%	3.4%	3.2%	3.1%	2.9%	2.8%	
净营运资本 变动	(556)	548	732	(325)	(743)	(582)	(456)	358	281	314	
自由现金流	10,535	13,394	14,634	15,639	17,224	19,683	22,396	26,120	29,317	33,034	461,225
永续增长率											3.0%

资料来源：浦银国际预测

图表 86：高通 DCF 估值预测（FY2024 年）

WACC	自由现金流现值 (美元百万)	净现金 (美元百万)	权益价值 (美元百万)	股数 (百万)	每股价值 (美元)
11.2%	277,612	(7,179)	270,433	1,123	240.7

资料来源：浦银国际预测

图表 87：高通历史市盈率：过去五年均值 15.1x

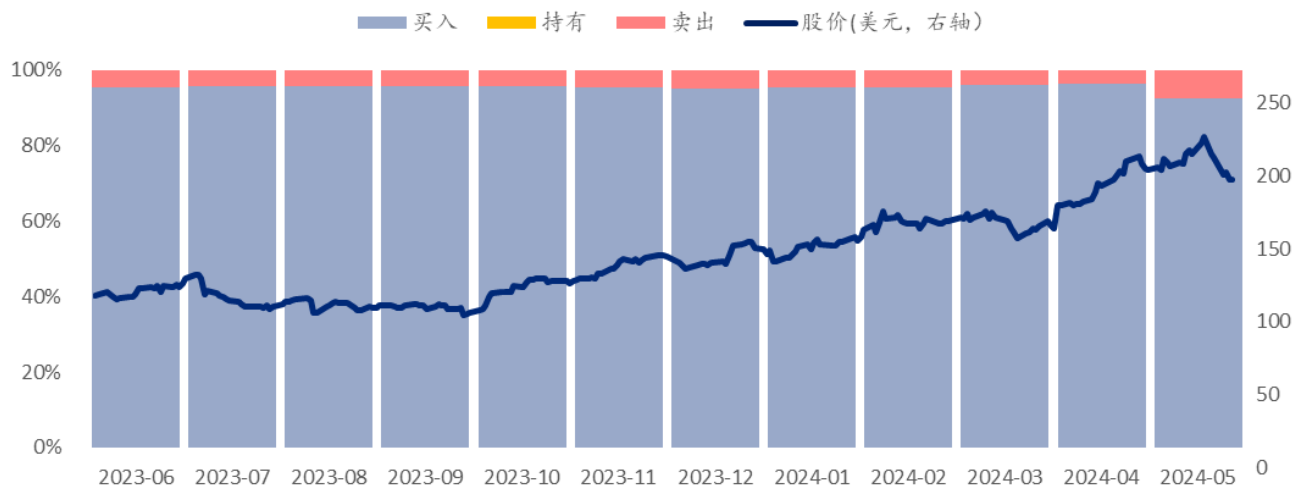


注：截至 2024 年 6 月 27 日收盘价；

资料来源：Factset、浦银国际

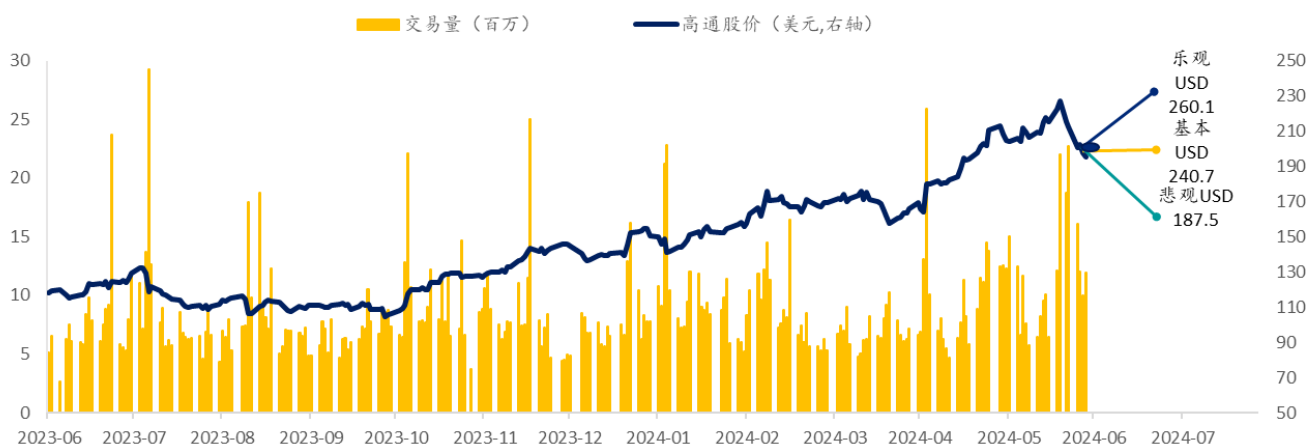
## ● SPDBI 乐观与悲观情景假设

图表 88：市场普遍预期：高通 (QCOM.US)



资料来源：Factset、浦银国际

图表 89：SPDBI 情景假设：高通 (QCOM.US)



### 乐观情景：公司收入增长好于预期

目标价：260.1 美元（概率：25%）

- AI 手机拉动换机周期，出货量增长强劲
- 端侧 AI 用户体验改善显著，用户愿意买单，提升平均售价
- 智能座舱和智能驾驶渗透率提升速度较快，拉动公司相关业务增长超预期

### 悲观情景：公司收入增长不及预期

目标价：187.5 美元（概率：10%）

- 智能手机需求走弱，复苏不及预期
- 端侧 AI 渗透中低端速度较慢
- 行业竞争加剧，拖累公司利润表现
- 舱驾一体方案订单落地延迟，增量业务贡献低于预期

资料来源：浦银国际预测

图表 90: SPDBI 目标价: 高通 (QCOM.US)



资料来源: Factset、浦银国际



## ● 风险提示

### 下行风险

- 美国或全球经济下行，智能手机、新能源车等下游需求动能不足；
- AI 智能手机渗透率上扬速度较慢，渗透中低端价格段手机较慢；
- 半导体周期上行动能不足；
- 行业竞争加剧，拖累利润表现；
- 研发等费用率增长较快；
- SoC 等芯片等迭代速度和性能低于预期。

## ● 公司背景

图表 91: 高通发展里程碑

年份	里程碑
1985	公司成立于加州
1991	高通上市
1995	高通公司成立 CDMA ASIC 产品单元
2007	推出骁龙 (Snapdragon) 处理器, 结合无线连接、多媒体播放、超快数据处理等任务
2009	参与 LTE 标准的研发和制定
2017	与苹果产生了多起诉讼与反诉讼摩擦
2018	高通发布首款支持 5G 的芯片骁龙 855
2019	高通与苹果和解, 获得 45 亿美元赔偿, 并重新签订授权协议
2021	高通在骁龙技术峰会发布全新一代骁龙 8 移动平台 8 Gen 1
2022	Meta 与高通宣布, 将联手开发用于虚拟现实产品的定制芯片组
2023	高通发布全球首个 5.5G 基带芯片骁龙 X75, 同时整合了微信连接功能
2023	发布具备 AI 功能的旗舰智能手机 SoC 骁龙 8 Gen3, 算力达 20 TOPS/秒
2024	推出骁龙 X Plus 平台; 发布旗舰级 5G 基带芯片 Snapdragon 骁龙 X80

资料来源: 公司官网、公开资料整理、浦银国际

## ● 财务报表

图表 92：高通：损益表

美元百万	FY2022	FY2023	FY2024E	FY2025E	FY2026E
<b>营业收入</b>	<b>44,200</b>	<b>35,820</b>	<b>38,415</b>	<b>42,763</b>	<b>45,713</b>
营业成本	(18,635)	(15,869)	(17,014)	(19,190)	(20,984)
<b>毛利润</b>	<b>25,565</b>	<b>19,951</b>	<b>21,401</b>	<b>23,573</b>	<b>24,729</b>
<b>营业支出</b>	<b>(9,705)</b>	<b>(12,163)</b>	<b>(11,275)</b>	<b>(11,633)</b>	<b>(11,864)</b>
销售、行政及一般费用	(2,570)	(2,483)	(2,591)	(2,885)	(3,084)
研发费用	(8,194)	(8,818)	(8,712)	(8,748)	(8,780)
重组及其他费用/收入	1,059	(862)	28	-	-
<b>营业利润</b>	<b>15,860</b>	<b>7,788</b>	<b>10,126</b>	<b>11,940</b>	<b>12,864</b>
<b>营业外收支</b>	<b>(862)</b>	<b>(345)</b>	<b>51</b>	<b>(263)</b>	<b>(258)</b>
利息收入/(支出)净额	(490)	(694)	(699)	(697)	(697)
其他收益/(支出)净额	(372)	349	749	434	439
<b>税前利润</b>	<b>14,998</b>	<b>7,443</b>	<b>10,177</b>	<b>11,677</b>	<b>12,607</b>
所得税	(2,012)	(104)	(795)	(1,044)	(1,160)
<b>净利润</b>	<b>12,986</b>	<b>7,339</b>	<b>9,382</b>	<b>10,633</b>	<b>11,447</b>
少数股东和可赎回少数股东损益	-	-	-	-	-
<b>归属普通股股东净利润</b>	<b>12,936</b>	<b>7,232</b>	<b>9,389</b>	<b>10,633</b>	<b>11,447</b>
基本股数(百万)	1,119	1,118	1,123	1,112	1,112
摊销股数(百万)	1,119	1,118	1,123	1,112	1,112
<b>基本每股收益(美元)</b>	<b>11.56</b>	<b>6.47</b>	<b>8.36</b>	<b>9.56</b>	<b>10.30</b>
<b>稀释每股收益(美元)</b>	<b>11.56</b>	<b>6.47</b>	<b>8.36</b>	<b>9.56</b>	<b>10.30</b>

注：E=浦银国际预测；高通财历年（FY, Fiscal Year）为上一年8月至当年9月  
资料来源：公司公告、浦银国际

图表 93：高通：资产负债表

美元百万	FY2022	FY2023	FY2024E	FY2025E	FY2026E
货币资金	2,773	8,450	8,020	10,067	12,118
受限制资金	3,609	2,874	2,874	2,874	2,874
应收账款	5,643	3,183	3,414	3,800	4,062
存货	6,341	6,422	6,885	6,213	5,435
其他流动资产	2,358	1,535	1,947	1,741	1,844
<b>流动资产合计</b>	<b>20,724</b>	<b>22,464</b>	<b>23,139</b>	<b>24,694</b>	<b>26,333</b>
固定资产-物业, 厂房及设备	5,168	5,042	4,869	5,031	5,176
无形资产	1,882	1,408	1,408	1,408	1,408
商誉	10,508	10,642	10,642	10,642	10,642
租赁权	-	-	-	-	-
其他非流动资产	10,732	11,484	13,781	16,537	19,844
<b>资产总计</b>	<b>49,014</b>	<b>51,040</b>	<b>53,839</b>	<b>58,312</b>	<b>63,403</b>
短期借款	1,945	914	715	814	765
应付账款	3,796	1,912	2,050	2,312	2,528
其他流动负债	6,125	6,802	6,802	6,802	6,802
<b>流动负债合计</b>	<b>11,866</b>	<b>9,628</b>	<b>9,567</b>	<b>9,929</b>	<b>10,095</b>
长期借款	13,537	14,484	14,484	14,484	14,484
其他非流动负债	5,598	5,347	5,347	5,347	5,347
<b>负债合计</b>	<b>31,001</b>	<b>29,459</b>	<b>29,398</b>	<b>29,760</b>	<b>29,926</b>
实收资本	195	490	490	490	490
资本公积	17,840	20,733	23,593	27,704	32,629
留存收益	(22)	358	358	358	358
其他综合收益	-	-	-	-	-
少数股东权益	-	-	-	-	-
<b>股东权益合计</b>	<b>18,013</b>	<b>21,581</b>	<b>24,441</b>	<b>28,552</b>	<b>33,477</b>
<b>负债及股东权益和计</b>	<b>49,014</b>	<b>51,040</b>	<b>53,839</b>	<b>58,312</b>	<b>63,403</b>

注：E=浦银国际预测；高通财务年（FY, Fiscal Year）为上一年8月至当年9月

资料来源：公司公告、浦银国际

图表 94：高通：现金流量表

美元百万	FY2022	FY2023	FY2024E	FY2025E	FY2026E
<b>经营活动产生的现金流量净额</b>	<b>9,096</b>	<b>11,299</b>	<b>10,176</b>	<b>13,130</b>	<b>13,836</b>
净利润	12,986	7,339	9,382	10,633	11,447
折旧及摊销	1,762	1,809	1,761	1,743	1,760
其他营业活动现金流	2,148	1,003	(412)	206	(103)
<b>营运资金变动</b>	<b>(7,800)</b>	<b>1,148</b>	<b>(556)</b>	<b>548</b>	<b>732</b>
应收账款(增加)/减少	(2,066)	2,472	(231)	(386)	(262)
存货(增加)/减少	(3,137)	8	(463)	672	778
应付账款增加/(减少)	1,036	(1,880)	138	262	216
其他经营资金变动	(3,633)	548	-	-	-
利息收入(支出)	-	-	-	-	-
<b>投资活动产生的现金流量净额</b>	<b>(5,804)</b>	<b>762</b>	<b>(3,884)</b>	<b>(4,661)</b>	<b>(5,213)</b>
资本支出	(2,262)	(1,323)	(1,588)	(1,905)	(1,905)
短期投资	1,208	898	-	-	-
长期投资	(4,780)	(215)	-	-	-
其他	30	1,402	(2,297)	(2,756)	(3,307)
<b>融资活动产生的现金流量净额</b>	<b>(7,196)</b>	<b>(6,663)</b>	<b>(6,721)</b>	<b>(6,422)</b>	<b>(6,572)</b>
借款	(352)	(498)	(199)	100	(50)
发行股份	(3,539)	(3,060)	(3,060)	(3,060)	(3,060)
发行债券	(63)	434	-	-	-
发放现金股利	(3,212)	(3,462)	(3,462)	(3,462)	(3,462)
其他	(30)	(77)	-	-	-
外汇损益	-	-	-	-	-
现金及现金等价物净流量	(3,904)	5,398	(430)	2,046	2,052
<b>期初现金及现金等价物</b>	<b>7,116</b>	<b>3,099</b>	<b>8,527</b>	<b>8,097</b>	<b>10,144</b>
<b>期末现金及现金等价物</b>	<b>3,099</b>	<b>8,527</b>	<b>8,097</b>	<b>10,144</b>	<b>12,195</b>

注：E=浦银国际预测；高通财年（FY，Fiscal Year）为上一年8月至当年9月

资料来源：公司公告、浦银国际

图表 95: SPDBI 科技行业覆盖公司

股票代码	公司	现价 (LC)	评级	目标价 (LC)	评级及目标价发布日期	行业
1810 HK Equity	小米集团-W	16.5	买入	23.6	13/5/2024	手机品牌
688036 CH Equity	传音控股	76.5	买入	179.4	26/2/2024	手机品牌
285 HK Equity	比亚迪电子	39.0	买入	36.7	1/3/2024	结构件、组装
600745 CH Equity	闻泰科技	28.3	买入	37.1	14/5/2024	ODM、功率半导体
002475 CH Equity	立讯精密	39.3	买入	34.5	24/10/2023	结构件、组装
300433 CH Equity	蓝思科技	18.3	买入	18.6	7/6/2024	结构件、组装
2018 HK Equity	瑞声科技	30.7	买入	28.8	25/3/2024	声学、光学器件
2382 HK Equity	舜宇光学科技	48.3	买入	80.8	10/11/2023	手机光学、车载光学
1478 HK Equity	丘钛科技	4.1	买入	5.1	10/11/2023	手机光学
603501 CH Equity	韦尔股份	99.4	买入	127.9	10/11/2023	手机 CIS、车载 CIS
NIO US Equity	蔚来	4.4	买入	5.9	7/6/2024	新能源汽车
9866 HK Equity	蔚来-SW	34.1	买入	48.5	7/6/2024	新能源汽车
XPEV US Equity	小鹏汽车	7.7	买入	9.6	7/6/2024	新能源汽车
9868 HK Equity	小鹏汽车-W	29.6	买入	37.6	7/6/2024	新能源汽车
LI US Equity	理想汽车	18.3	买入	24.5	7/6/2024	新能源汽车
2015 HK Equity	理想汽车-W	70.3	买入	95.6	7/6/2024	新能源汽车
9863 HK Equity	零跑汽车	26.8	买入	35.3	7/6/2024	新能源汽车
TSLA US Equity	特斯拉(TESLA)	197.4	持有	202.9	26/1/2024	新能源汽车
1211 HK Equity	比亚迪股份	232.0	买入	232.2	28/3/2024	新能源汽车
002594 CH Equity	比亚迪	250.3	买入	248.4	28/3/2024	新能源汽车
981 HK Equity	中芯国际	17.1	买入	17.9	7/2/2024	晶圆代工
688981 CH Equity	中芯国际	46.1	买入	52.6	7/2/2024	晶圆代工
1347 HK Equity	华虹半导体	22.1	买入	20.6	10/5/2024	晶圆代工
688396 CH Equity	华润微	35.9	买入	43.6	10/5/2024	功率半导体
600460 CH Equity	士兰微	37.4	买入	64.9	20/9/2023	功率半导体
300373 CH Equity	扬杰科技	17.5	买入	30.5	20/9/2023	功率半导体
688187 CH Equity	时代电气	38.9	买入	43.4	25/10/2023	功率半导体
3898 HK Equity	时代电气	49.4	买入	53.4	27/10/2023	功率半导体
603290 CH Equity	斯达半导	30.8	买入	38.7	27/10/2023	功率半导体
605111 CH Equity	新洁能	86.1	买入	230.6	20/9/2023	功率半导体
688711 CH Equity	宏微科技	30.6	买入	44.4	20/9/2023	功率半导体
NVDA US Equity	英伟达	124.0	买入	147.6	2/7/2024	AI 芯片
QCOM US Equity	高通	195.2	买入	240.7	2/7/2024	AI 芯片

注: A股、港股截至2024年6月28日收盘价;美股截至2024年6月27日收盘价

资料来源: Bloomberg、浦银国际预测



## 免责声明

本报告之收取者透过接受本报告(包括任何有关的附件),表示及保证其根据下述的条件下有权获得本报告,且同意受此中包含的限制条件所约束。任何没有遵循这些限制的情况可能构成法律之违反。

本报告是由从事证券及期货条例(香港法例第 571 章)中第一类(证券交易)及第四类(就证券提供意见)受规管活动之持牌法国-浦银国际证券有限公司(统称“浦银国际证券”)利用集团信息及其他公开信息编制而成。所有资料均搜集自被认为是可靠的来源,但并不保证数据之准确性、可信性及完整性,亦不会因资料引致的任何损失承担任何责任。报告中的资料来源除非另有说明,否则信息均来自本集团。本报告的内容涉及到保密数据,所以仅供阁下为其自身利益而使用。除了阁下以及受聘向阁下提供咨询意见的人士(其同意将本材料保密并受本免责声明中所述限制约束)之外,本报告分发给任何人均属未授权的行为。

任何人不得将本报告内任何信息用于其他目的。本报告仅是为提供信息而准备的,不得被解释为是一项关于购买或者出售任何证券或相关金融工具的要约邀请或者要约。阁下不应将本报告内容解释为法律、税务、会计或投资事项的专业意见或为任何推荐,阁下应当就本报告所述的任何交易涉及的法律及相关事项咨询其自己的法律顾问和财务顾问的意见。本报告内的信息及意见乃于文件注明日期作出,日后可作修改而不另通知,亦不一定会更新以反映文件日期之后发生的进展。本报告并未包含公司可能要求的所有信息,阁下不应仅仅依据本报告中的信息而作出投资、撤资或其他财务方面的任何决策或行动。除关于历史数据的陈述外,本报告可能包含前瞻性的陈述,牵涉多种风险和不确定性,该等前瞻性陈述可基于一些假设,受限于重大风险和不确定性。

本报告之观点、推荐、建议和意见均不一定反映浦银国际证券的立场。浦银国际控股有限公司及其附属公司、关联公司(统称“浦银国际”)及/或其董事及/或雇员,可能持有在本报告内所述或有关公司之证券、并可能不时进行买卖。浦银国际或其任何董事及/或雇员对投资者因使用本报告或依赖其所载信息而引起的一切可能损失,概不承担任何法律责任。

浦银国际证券建议投资者应独立地评估本报告内的资料,考虑其本身的特定投资目标、财务状况及需要,在参与有关报告中所述公司之证券的交易前,委任其认为必须的法律、商业、财务、税务或其它方面的专业顾问。惟报告内所述的公司之证券未必能在所有司法管辖区或国家或供所有类别的投资者买卖。对部分的司法管辖区或国家而言,分发、发行或使用本报告会抵触当地法律、法则、规定、或其它注册或发牌的规例。本报告不是旨在向该等司法管辖区或国家的任何人或实体分发或由其使用。

### 美国

浦银国际不是美国注册经纪商和美国金融业监管局(FINRA)的注册会员。浦银国际证券的分析师不具有美国金融监管局(FINRA)分析师的注册资格。因此,浦银国际证券不受美国就有研究报告准备和分析师独立性规则的约束。

本报告仅提供给美国 1934 年证券交易法规则 15a-6 定义的“主要机构投资者”,不得提供给其他任何个人。接收本报告之行为即表明同意接受协议不得将本报告分发或提供给任何其他人士。接收本报告的美国收件人如想根据本报告中提供的信息进行任何买卖证券交易,都应仅通过美国注册的经纪交易商来进行交易。

### 英国

本报告并非由英国 2000 年金融服务与市场法(经修订)(「FSMA」)第 21 条所界定之认可人士发布,而本报告亦未经其批准。因此,本报告不会向英国公众人士派发,亦不得向公众人士传递。本报告仅提供给合格投资者(按照金融服务及市场法的涵义),即(i)按照 2000 年金融服务及市场法 2005 年(金融推广)命令(「命令」)第 19(5)条定义在投资方面拥有专业经验之投资专业人士或(ii)属于命令第 49(2)(a)至(d)条范围之高净值实体或(iii)其他可能合法与之沟通的人士(所有该等人士统称为「有关人士」)。不属于有关人士的任何机构和个人不得遵照或倚赖本报告或其任何内容行事。

本报告的版权仅为浦银国际证券所有,未经书面许可任何机构和人士不得以任何形式转发、翻版、复制、刊登、发表或引用,浦银国际证券对任何第三方的该等行为保留追述权利,并且对第三方未经授权行为不承担任何责任。

### 权益披露

- 1) 浦银国际并没有持有本报告所述公司逾 1%的财务权益。
- 2) 浦银国际跟本报告所述公司在过去 12 个月内并没有任何投资银行业务的关系。
- 3) 浦银国际并没有跟本报告所述公司为其证券进行庄家活动。

## 评级定义

### 证券评级定义:

“买入”: 未来 12 个月, 预期个股表现超过同期其所属的行业指数

“持有”: 未来 12 个月, 预期个股表现与同期所属的行业指数持平

“卖出”: 未来 12 个月, 预期个股表现逊于同期其所属的行业指数

### 行业评级定义 (相对于 MSCI 中国指数):

“超配”: 未来 12 个月优于 MSCI 中国 10%或以上

“标配”: 未来 12 个月优于/劣于 MSCI 中国少于 10%

“低配”: 未来 12 个月劣于 MSCI 中国超过 10%

## 分析师证明

本报告作者谨此声明:(i) 本报告发表的所有观点均正确地反映作者有关任何及所有提及的证券或发行人的个人观点, 并以独立方式撰写;(ii) 其报酬没有任何部分曾经, 是或将会直接或间接与本报告发表的特定建议或观点有关;(iii) 该等作者没有获得与所提及的证券或发行人相关且可能影响该等建议的内幕信息/非公开的价格敏感数据。

本报告作者进一步确定 (i) 他们或其各自的关联人士 (定义见证券及期货事务监察委员会持牌人或注册人操守准则) 没有在本报告发行日期之前的 30 个历日内曾买卖或交易过本报告所提述的股票, 或在本报告发布后 3 个工作日 (定义见《证券及期货条例》(香港法例第 571 章)) 内将买卖或交易本文所提述的股票;(ii) 他们或其各自的关联人士并非本报告提述的任何公司的雇员; 及 (iii) 他们或其各自的关联人士没有拥有本报告提述的证券的任何金融利益。

### 浦银国际证券机构销售团队

#### 杨增希

essie\_yang@spdbi.com

852-2808 6469

### 浦银国际证券财富管理团队

#### 王玥

emily\_wang@spdbi.com

852-2808 6468

### 浦银国际证券有限公司

SPDB International Securities Limited

网站: [www.spdbi.com](http://www.spdbi.com)

地址: 香港轩尼诗道 1 号浦发银行大厦 33 楼

