

AI算力跟踪深度（四）：
怎么看算力的天花板和成长性？

证券分析师：张良卫

执业证书编号：S0600516070001

联系邮箱：zhanglw@dwzq.com.cn

联系电话：021-60199793

证券分析师：李博韦

执业证书编号：S0600525070006

联系邮箱：libw@dwzq.com.cn

2025年10月15日

请务必阅读正文之后的免责声明部分

我们认为AI产业的天花板、变现性、成长性、产业链友好度等是稀缺的，从宏观到微观进行对比分析，我们认为主要有以下几点因素：

1、行业天花板更高：AI的产出是可乘数放大的“智能”，由“算力—模型—应用—数据”飞轮驱动，OpenRouter等平台Token调用量呈指数级增长；对比通信流量的“加法式”增长与频谱/站址等物理边界，AI的可拓展空间显著更大，AI在各行业重构流程并催生新价值链与新市场，行业天花板由可被智能化的任务空间决定而非人口、设备数等底层数字。

2、见效更快：4G/5G“先覆盖后变现”的工程属性与ARPU增速受限导致回本周期达数年；AI模型、算力投入后可即插即用，并不断迭代满足新需求、新场景，进行滚动式兑现。

3、硬件迭代更快，优化价格、需求周期：AI算力“年更+软硬协同”的高频迭代，在12-18个月内刷新单位算力成本并催生新需求，从而在价格下行之前就以“更高代际的价值锚”重定价，打破了传统硬件的量价周期。当算力更便宜、更易得时，开发者会把更复杂的模型与系统作为新基准，提升参数量、上下文与并行度。模型架构的迭代有机会减小单次推理、训练所需算力，而AI产业发展过程中杰文斯悖论将演绎很多次，如Genie 3等生成视频的世界模型，或需要跨数量级提升的算力才可满足。

投资建议：我们认为AI算力更高的天花板与更好的竞争格局支撑起AI算力相对4G/5G的更高估值框架与更强Beta，相关标的——光互连：中际旭创，新易盛，天孚通信，光库科技，长芯博创，仕佳光子，源杰科技，长光华芯，太辰光；交换机：锐捷网络，盛科通信；铜互连：兆龙互连，中际旭创。

风险提示：AI算力需求不及预期；客户处份额不及预期；产品研发落地不及预期；行业竞争加剧。



■ AI的天花板与飞轮不一样

■ AI算力以更快的技术迭代，优化价格与需求周期

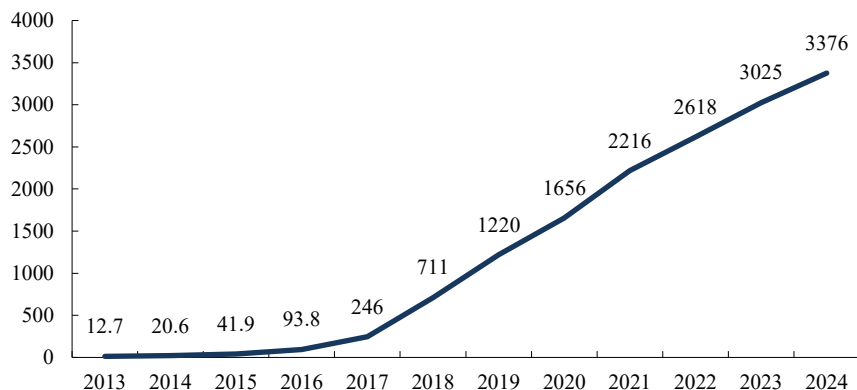
■ 投资建议及风险提示

一、 AI的天花板与飞轮不一样

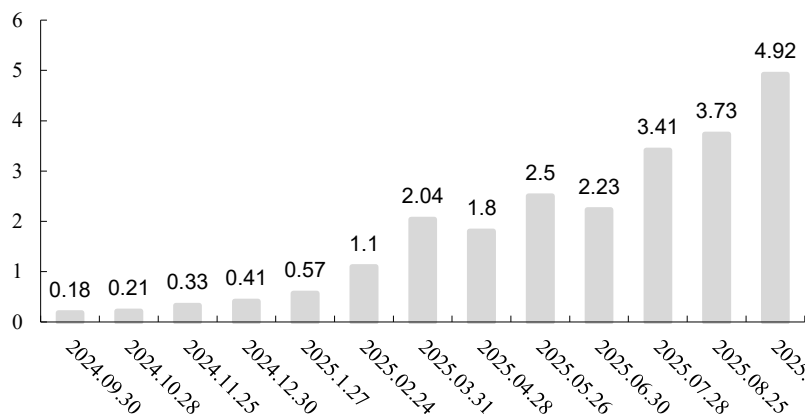
1.1 AI具有更高的天花板

- AI 算力呈“乘法式”扩张，行业天花板与技术迭代速度显著更高；4G/5G 呈“加法式”增长、受物理边界约束。
- 5G 的需求与可部署资源和连接总量高度相关，受频谱、站址、覆盖半径、功耗等硬约束影响，收入结构本质上由基建 CAPEX 与运营服务收入 相加决定，边界受人口与设备密度限制。
- AI 的核心产出是“智能”，作为生产要素与数据、劳动力、资本相乘提升全要素生产率，算力需求由模型规模、上下文长度、多模态维度与推理负载共同驱动，能力提升与单位算力成本下行形成“算力-模型-应用-数据”正反馈飞轮，带来指数级模型复杂度与场景渗透。
- 横向对比流量消耗量与token消耗量，从13年开始的4G时代到当前的5G时代，流量消耗量 CAGR（2013-2024年）约61.1%，而token消耗量则呈指数级增长，以openrouter平台为例，过去一年即增长约28倍。
- AI 在各行业重构流程并催生新价值链与新市场，行业天花板由可被智能化的任务空间决定而非人口与设备数，概念上更接近无界；5G 的连接上限虽高，但受频谱资源和物理部署所框定。

中国移动互联网流量消耗量（亿GB）



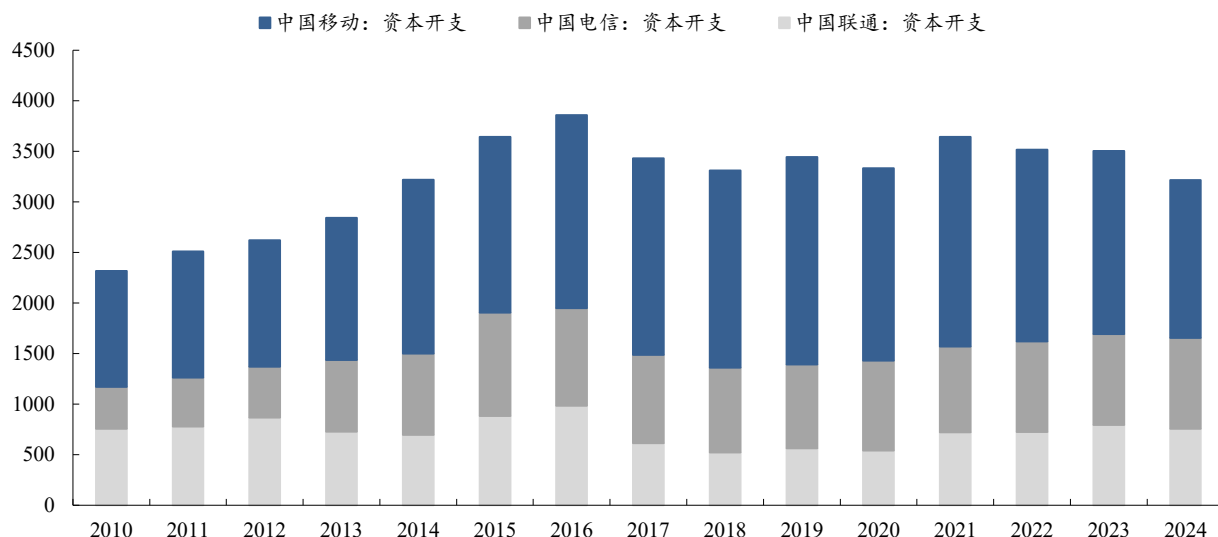
openrouter平台token消耗量（T）



1.1 AI具有更高的天花板

- 基础设施端，受“可计算的连接上限”约束，4G/5G呈现明显的周期特性。
- 从国内三大运营商资本开支轨迹可见，4G推进期（约2013–2016）CAPEX阶梯上行并在2016年前后见顶，随后2017–2019年进入消化回落；5G首轮规模建网（2020–2022）CAPEX再度抬升，伴随覆盖范围扩大、边际收益下行与折旧约束，2023–2024转为高位趋稳。建设周期取决于频谱/标准迭代速度与站址、光纤、供电等物理资源上限，以及人口与终端渗透的可测边界，使连接类基础设施的TAM收敛。

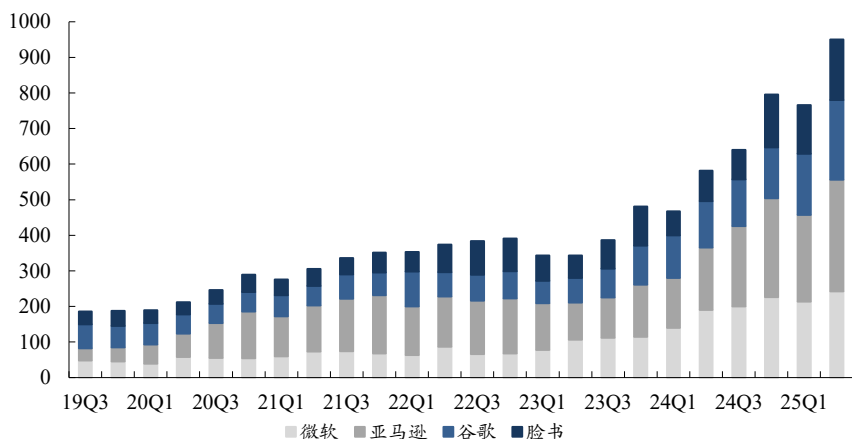
2010-2024年中国三大运营商资本开支（亿元）



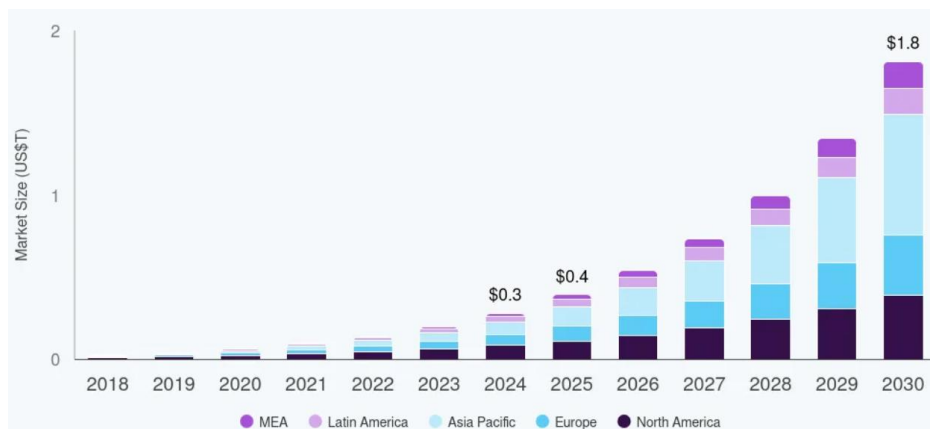
1.1 AI具有更高的天花板

- 我们认为AI算力的行业天花板远高于4G/5G并仍在上修，体现为数据中心Capex与上游加速芯片TAM的绝对规模与增速同时抬升。
- 北美四大云厂资本开支持续高增，主权AI等贡献新增量，英伟达指引Blackwell和Rubin平台有望共同推动全球AI基础设施规模在2030年左右达到3-4万亿美元。
- Grand View Research预测预计全球直接AI市场规模到2030年将达到18117.5亿美元，2025年至2030年的复合年增长率为35.9%。

海外四大云厂CapEx (亿美元)



全球直接AI市场规模预测 (十亿美元)



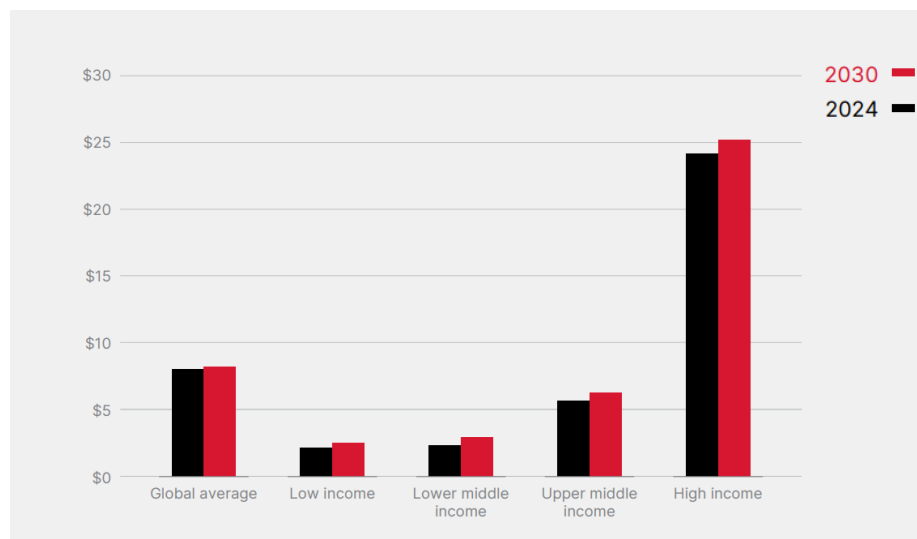
注：2025年起为预测数据

1.2 AI投资正滚动式兑现

4G/5G 的投资回收期显著偏长，源于前期重资本、后端慢变现的商业模式。根据 PatentPC，5G投资平均收支平衡时间预计为8-10年。

- “先覆盖、后变现”的工程属性决定 CAPEX 前置。
- 频谱拍卖与许可费用进一步抬升资本门槛，如德国2019年5G频谱竞拍总额65.5亿欧元，挤压早期现金回笼空间。
- 收益侧提升有限：ARPU提价空间受竞争与监管制约，GSMA预计至 2030 年高收入市场 ARPU年增仅约1%，行业收入增速温和，ROI 摊薄。
- 二次建设叠加：当一期覆盖接近“边际”，Capex/Revenue 比率见顶后趋稳，但 5G-A 与行业专网等二次建设仍需持续投入，使回收呈“建设—消化—再建设”的多周期拉长。

运营商移动服务ARPU预测（美元/月）



注：2030年为预测数据

1.2 AI投资正滚动式兑现

以云服务商为例，AI算力的回收期显著短于4G/5G，核心在于边部署边变现（按量计费）+ 高利用率。

- 变现即时性：算力以按秒/按小时计费的 On-Demand 形态，上线即产生现金流，避免“大网络先覆盖、后变现”的滞后。
- 单机收入高：主流 8×H100 机型对应每卡时价常见 \$2-\$4+ 区间（AWS p5.48xlarge 折算 \$3.933/GPU·h；专业 GPU 云最低 \$2.99/GPU·h），在高利用率场景下可快速摊薄前期投入。以一张H100价格为3万美元计算，大约需400天即可覆盖单卡成本。
- 高利用率：AI加速卡长期处于供不应求的状态，云服务厂商RPO激增，算力资源被提前锁定。

4G/5G与AI算力投资兑现对比

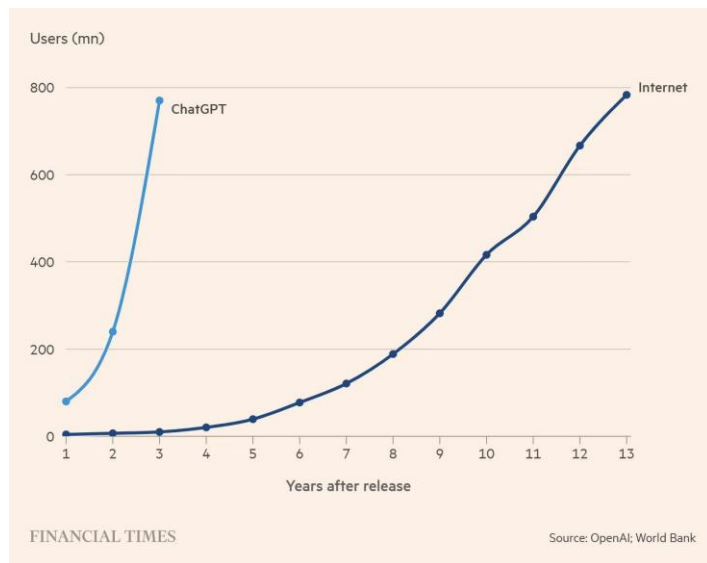
	4G/5G周期	AI 算力
主要投资者	全球数百家电信运营商	少数超大规模云服务商和专业GPU云商
主要资本支出资产	频谱、基站、光纤（长生命周期）	GPU、定制芯片、数据中心（短生命周期）
资产生命周期	10年以上	2-3年（GPU快速迭代）
预计回报周期	8-10年	2-5年
典型回报特征	稳定、低利润率、类似公用事业	高风险、高回报、指数级增长潜力

1.2 AI投资正滚动式兑现

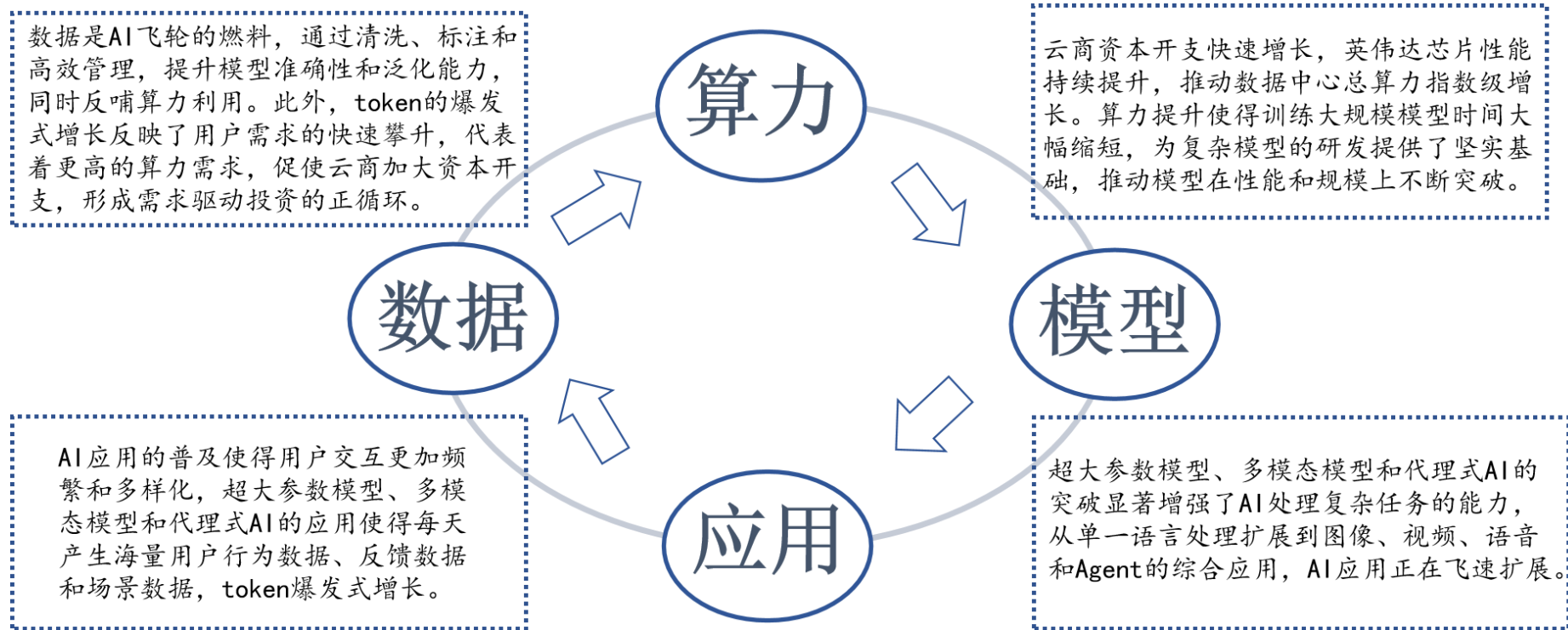
AI产品的用户渗透速度远超互联网，已拥有庞大的用户基础，潜在变现能力强，投资回收期相比4G/5G有望显著缩短。ChatGPT上线约3年用户已接近8亿，而互联网达到相近体量约用时12-13年，同等用户规模在更短时间达成，使单位获客成本更快被摊薄。

- ①用户金字塔完善：免费用户→高级订阅→企业版与API，用量计费与席位订阅并行，带来更高ARPU与更长生命周期价值；
- ②强网络效应与内容自传播降低获取成本，结合云化交付的近零边际成本，使规模扩张直接转化为毛利改善；
- ③产品梯度与功能包（更强模型、更高上下文、更快速率等）提供持续的付费升级与交叉销售空间。

AI用户增长速度远超互联网（百万用户）



算力-模型-应用-数据的飞轮加速转动，不断打开行业天花板



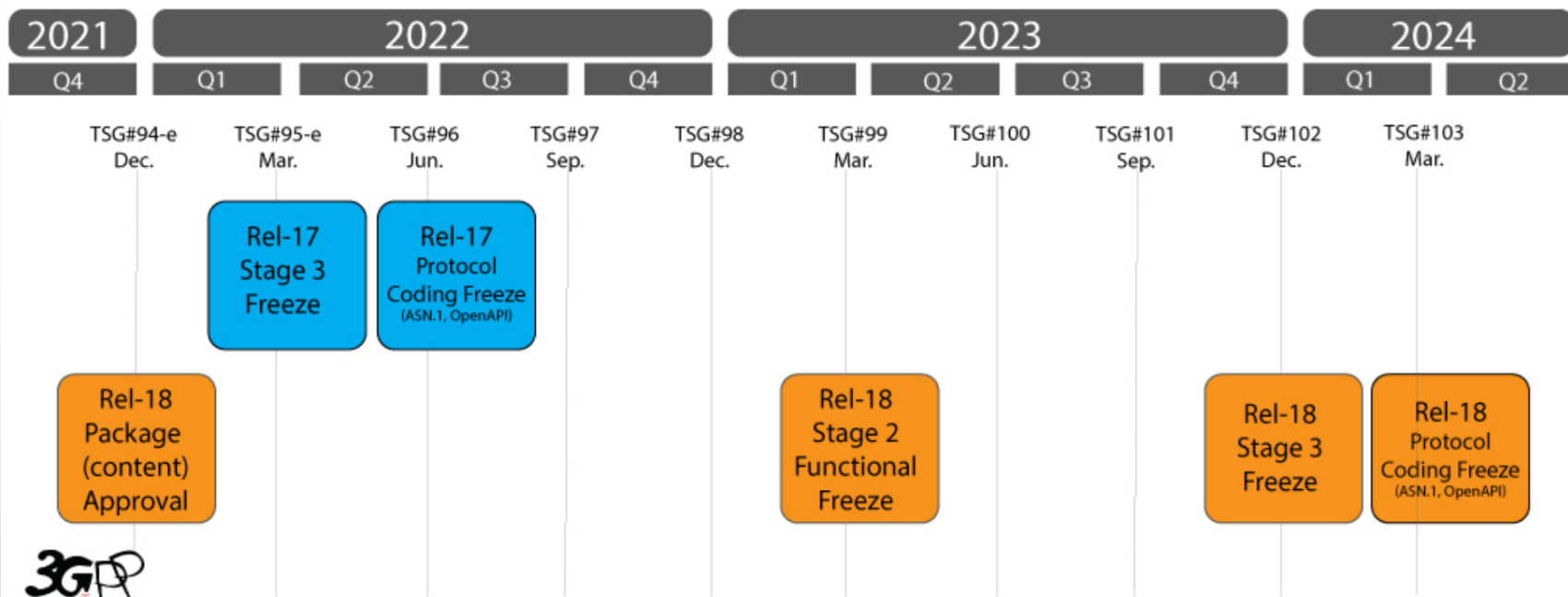
二、AI算力以更快的技术迭代，优化价格与需求周期

2.1 AI算力以更快的技术迭代，打破典型价格与需求周期

4G/5G的技术演进以 3GPP 标准为节拍，单次Release通常 18–24 个月、代际跨越 6–10 年。

- **标准节拍决定速度：**5G 第一阶段标准 Release-15 在2018年6月功能冻结，支撑 2019 年起的全球商用；面向 5G-Advanced 的 Release-18 在2024年6月冻结，Release-19/20 按既定规划推进。单次 Release通常保持18–24 个月的稳定节奏。
- **代际跨越长：**从 LTE (R8, 2008 年冻结) →LTE-A (R10, 2011 年冻结) →5G (R15, 2018 年冻结) →5G-A (R18, 2024 年冻结)，每次“代际/大阶段”跨度通常6–10 年，且需等待终端/频谱/站点改造的协同到位。
- **产业链约束：**蜂窝网络迭代不仅受芯片与软件影响，更受频谱规划、站点建设、回传与核心网演进等多要素制约，导致从标准冻结到规模商用需 1–2 年生态爬坡。

3GPP单次Release节奏

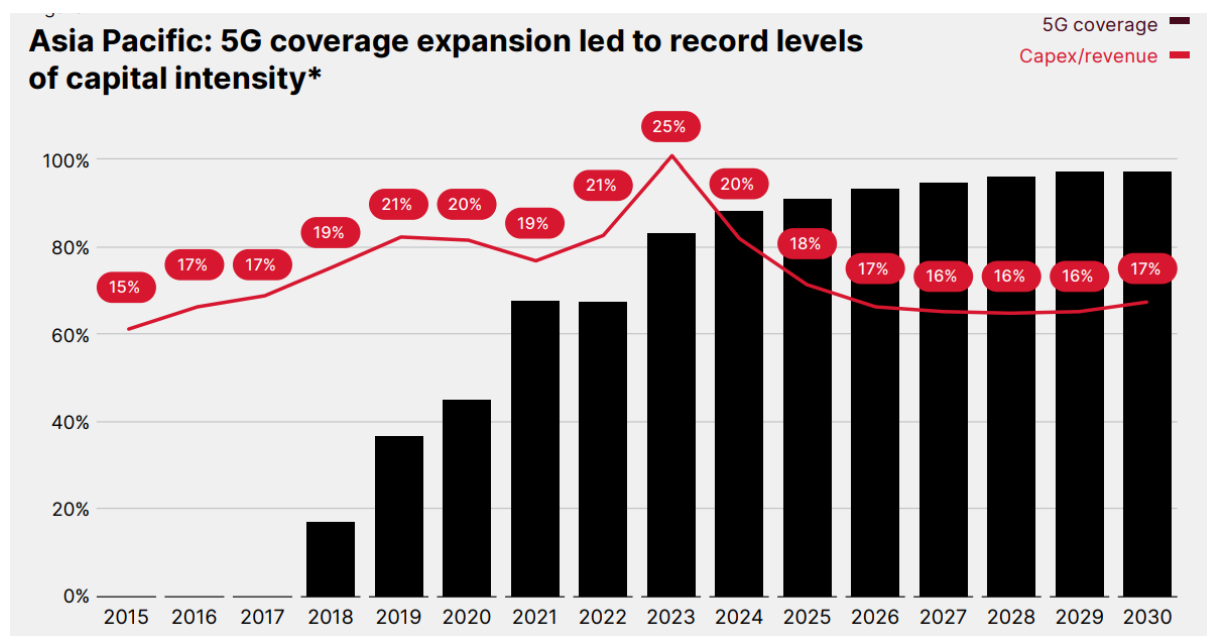


2.1 AI算力以更快的技术迭代，打破典型价格与需求周期

4G/5G在基础设施端呈现典型量价周期，在每轮建设周期中，出货量暴增、伴随价格年降是客观产业规律。

- 通信标准切换与牌照落地触发首轮覆盖建设（基站、载波、光传输、配电），运营商 CAPEX和设备商出货量快速上行；随后进入密集化与容量扩建期，量维持高位但集中招标+多厂竞价使 ASP 年度下修；当覆盖边际趋缓且折旧与能耗压力上升，转入优化与消化期，量显著放缓、设备 ASP 延续下探、价跌幅度大于量，链条利润率承压。

5G建设周期



注：2025年起为预测数据

2.1 AI算力以更快的技术迭代，打破典型价格与需求周期

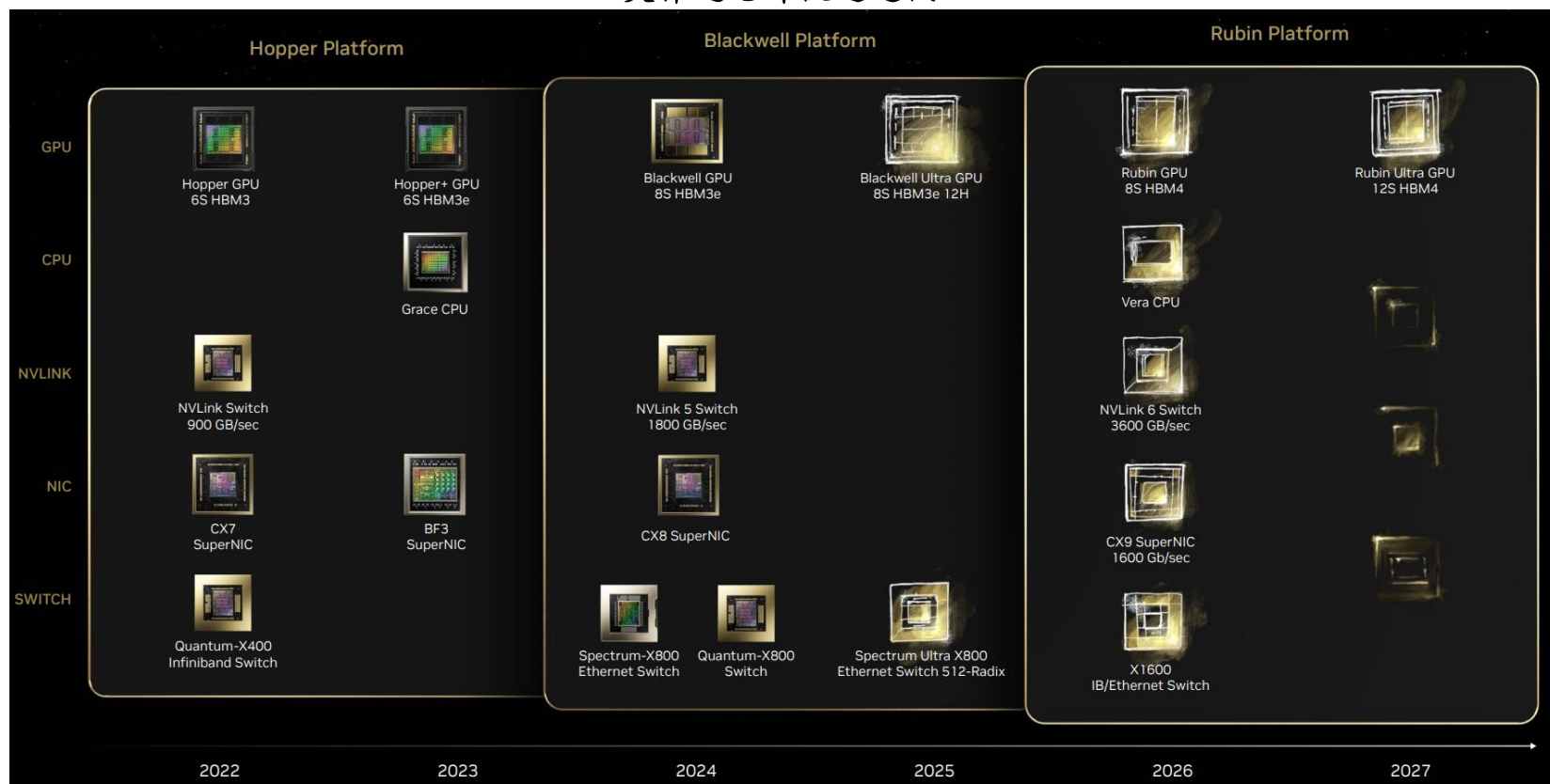
- AI算力“年更+软硬协同”的高频迭代，在12-18个月内刷新单位算力成本并催生新需求，从而在价格下行之前就以“更高代际的价值锚”重定价，打破了传统硬件的量价周期。
- GPU从2-3年一代逐渐切换为一年一代，且每代伴随互连（NVLink等）、高带宽内存、编译器与内核的系统性升级，将研发—生态—整机的爬坡窗口压缩，显著抬高进入门槛，竞争格局相对稳定。HBM、先进封装、算力配套（光模块、PCB）等产能需提前锁定、良率爬坡周期长，头部厂商通过长期协议占用大部分产能。价格的主驱动从“同代去库存”转为“新旧代际的价值差”，行业呈新代具提价能力、旧代有底价支撑的结构。
- **需求端推动更快迭代：**前沿模型训练所用计算量以每年数倍的速度增长，拉动硬件与存储迭代节奏向12-18个月收敛。

2.1 AI算力以更快的技术迭代，打破典型价格与需求周期

AI算力技术迭代速度显著快于4G/5G，系统性能呈12-18个月一次的阶跃式提升。

- 硬件迭代周期缩短为一到两年：英伟达在2024 COMPUTEX宣布新平台规划：Hopper→Blackwell→Rubin等。

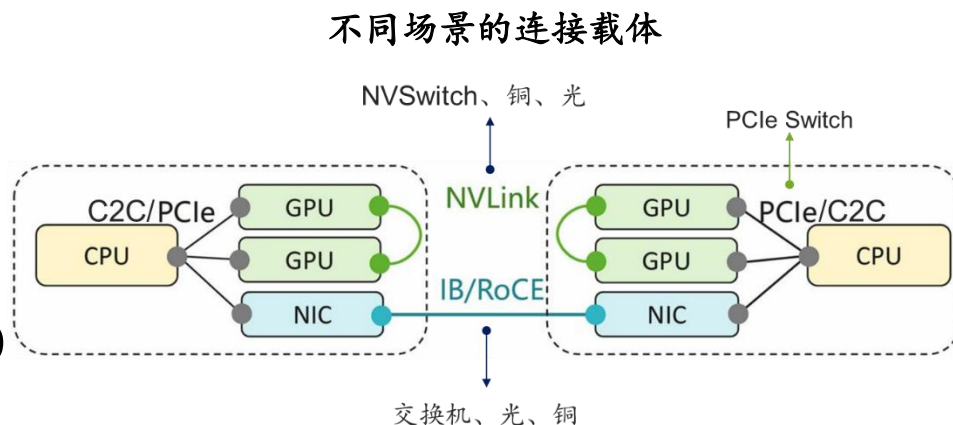
英伟达芯片快速迭代



注：2026年起为预测进展

2.1 AI算力以更快的技术迭代，打破典型价格与需求周期

- **PCB板上芯片间的连接、设备间的连接场景均需要通过不同载体实现性能升级。**
- **互连带宽迅速增长：**NVLink跟随每一代GPU架构进行升级，目前最新用于B系列GPU的NVLink 5.0可支持单卡7.2Tb的带宽，相较于用于H100的NVLink 4.0带宽翻倍。
- **存储代际压缩：**HBM3E已在2024Q2随H200上量，2025年内主流厂商已完成HBM4开发、内测并准备量产，存储带宽与容量几乎以1-2年一代推进，为大模型上下文与激活带来更快的可用带宽与更低能耗。

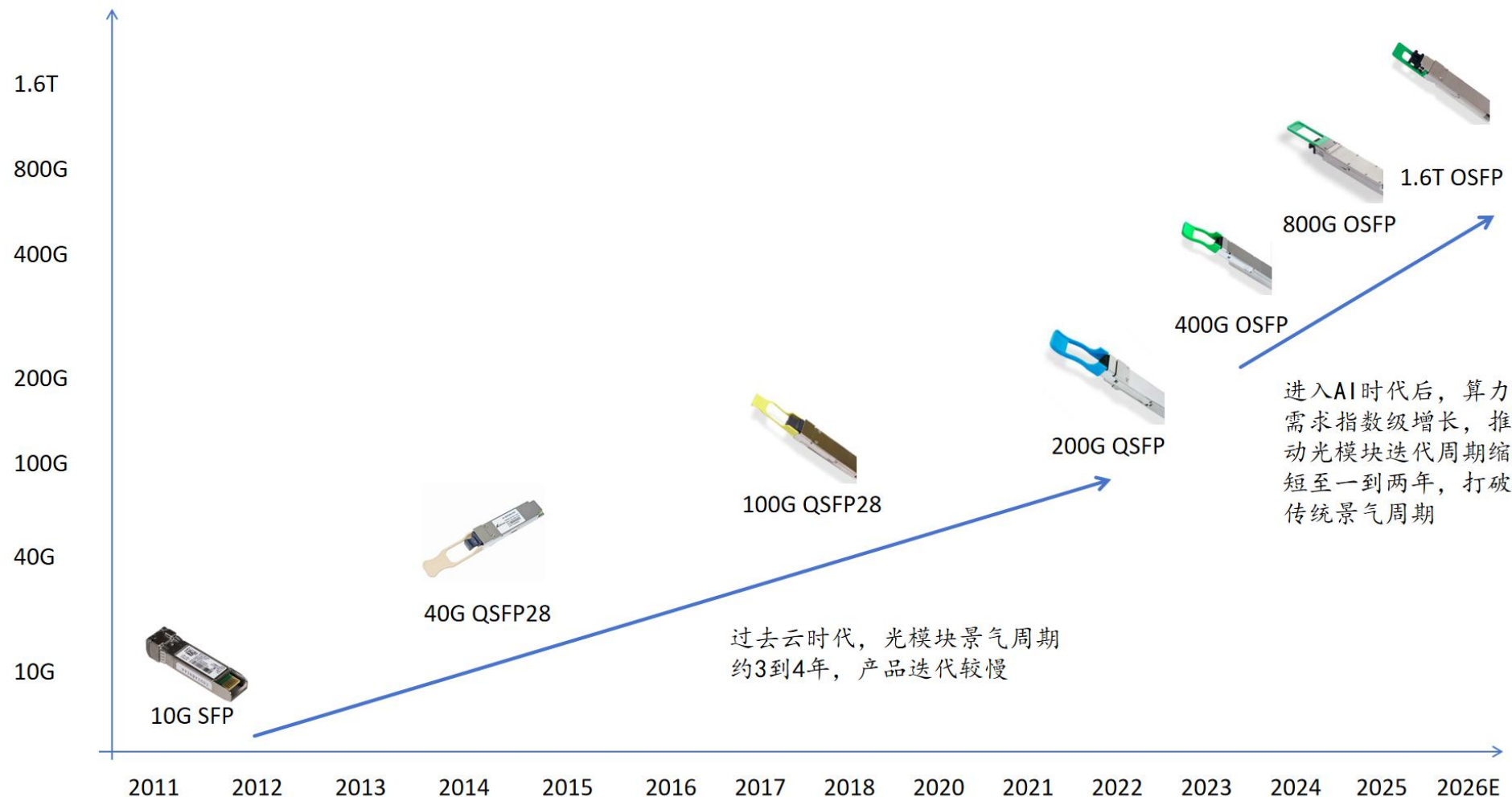


NVLink互联带宽迅速增长

版本	年份	GPU 架构	每链接带宽 (Gb/s)	链接数 (条)	总带宽 (Gb/s)
1.0	2016	Pascal	160	4	640
2.0	2017	Volta	200	6	1200
3.0	2020	Ampere	200	12	2400
4.0	2022	Hopper	400	18	3600
5.0	2024	Blackwell	800	18	7200

2.1 AI算力以更快的技术迭代，打破典型价格与需求周期

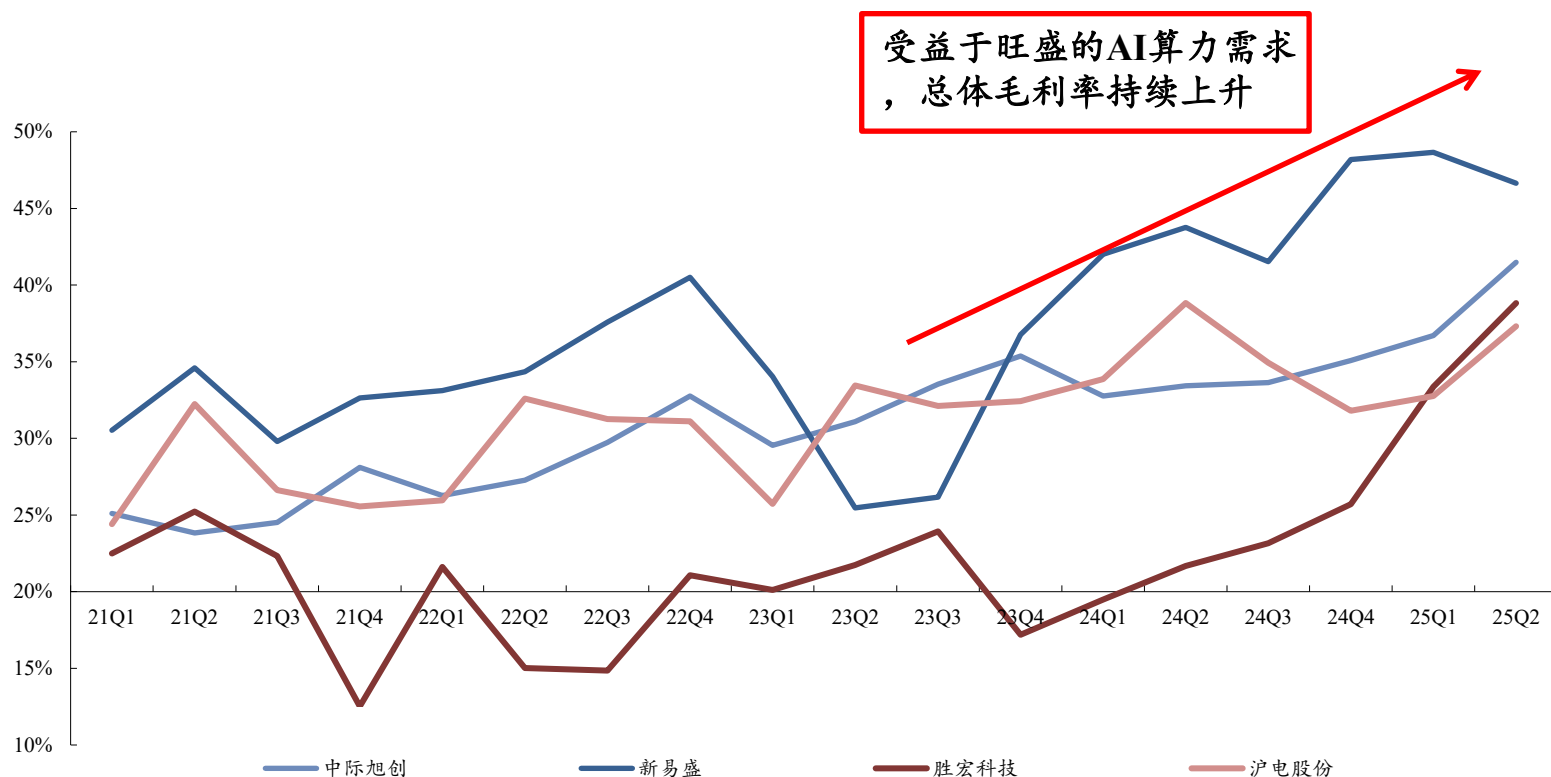
AI时代打破光模块传统景气周期



2.1 AI算力以更快的技术迭代，打破典型价格与需求周期

- 以A股海外算力链龙头中际旭创、新易盛、胜宏科技和沪电股份为例，受益于旺盛的AI算力需求与产品持续迭代，格局基本稳定，随着高端产品出货量增长与规模效应的产生，总体毛利率稳步提升。

海外算力链龙头毛利率情况

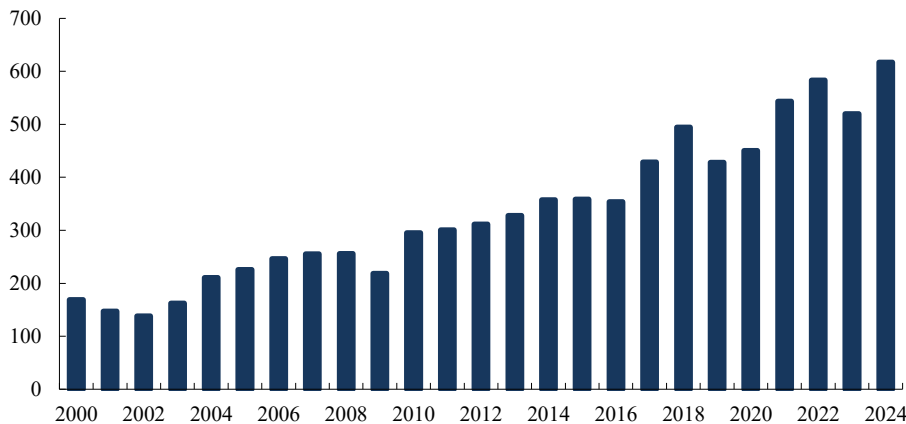
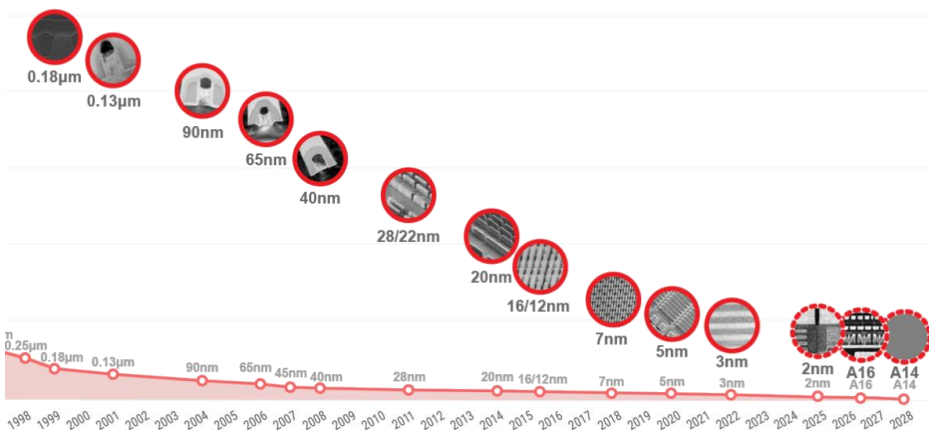


2.2 技术迭代有望满足越来越多尚未满足的新需求

- 参考半导体制程迭代，芯片在性能升级后可应用于更高规格应用场景，创造更多直线无法满足的需求，行业总空间可随摩尔定律的演绎持续增长。
- 当算力更便宜、更易得时，开发者会把更复杂的模型与系统作为新基准，提升参数量、上下文与并行度。
- 模型架构的迭代有机会减小单次推理、训练所需算力，而AI产业发展过程中杰文斯悖论将演绎很多次，**如Genie 3等生成视频的世界模型，或需要跨数量级提升的算力才可满足。**

先进制程迭代

全球半导体销售额高速增长（十亿美元）



2.2 技术迭代有望满足越来越多尚未满足的新需求

具体分析欢迎参考之前外发的深度报告《从英伟达的视角看算力互连板块成长性——Scale Up 网络的“Scaling Law”存在吗？》

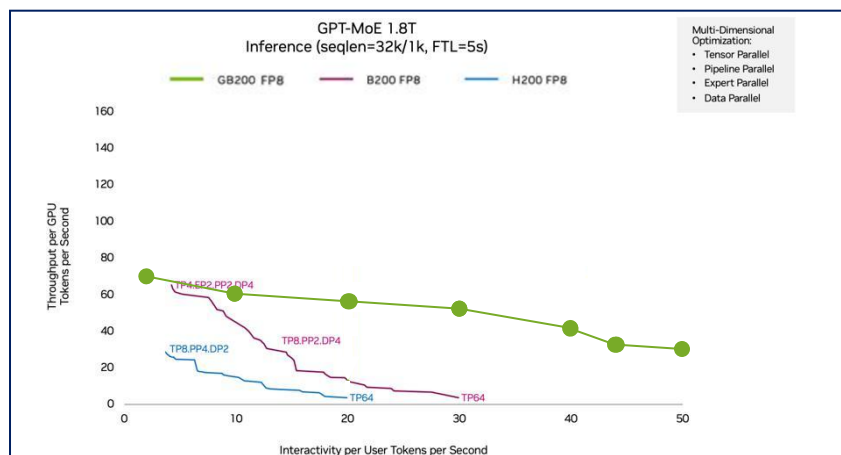
- 纵轴单卡实际性能越高，GB200在横轴上单用户TPS的绝对、相对优势均越小；
- 横轴单用户TPS越高，GB200的单卡实际性能越低；
- 随着用户体验提升、模型能力拓展，横轴将不断向右延伸，下降越慢的曲线在满足推理负载的同时能够实现更小的TCO，而搭建更大规模的Scale Up网络是缓解下降最有效的方法之一；
- 图中所需ScaleUp规模的增速高于单用户TPS及单卡性能增速，且横纵指标成反比，可以推测所需ScaleUp网络规模与单用户TPS、单卡实际性能及模型参数量间或存在Scaling Law，即：

$$S = aG^b U^c$$

其中S为Scale Up超节点规模，a为与模型参数、并行计算路数等有关的变量，G为单张GPU的实际性能（同款GPU对比），b为大于1的常数，U为单用户TPS，c为大于1的常数。

- 基于这一Scaling Law，我们认为随着推理时对单用户TPS、单GPU实际性能越来越高的要求，Scale Up规模会非线性增长。

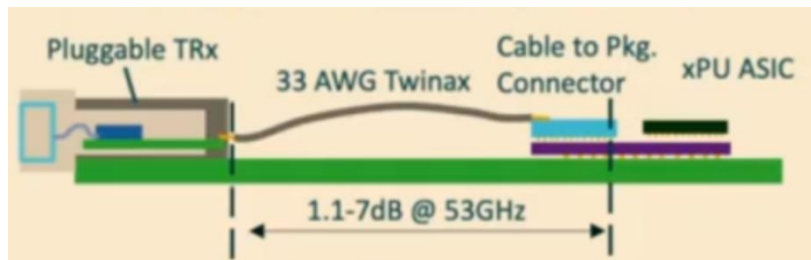
不同方案推理性能对比估算（GB200来自原数据估测存在一定误差）



2.2 技术迭代有望满足越来越多尚未满足的新需求

- Scale Up需要更高的信号密度，更低的功耗、延时、信号损耗；
- LPO相比DSP光模块功耗减半（~7.5w vs 15w），能效比介于DSP光模块与CPO之间（~20pJ/bit vs 10pJ/bit vs 6-7pJ/bit），CPC可代替Tray内PCB走线连接xPU与端口，LPO+降低功耗、延时、信号损耗；
- Scale Up单端口速率升级进展有望快于Scale Out，单端口3.2T需要400G/通道速率，异质集成TFLN-on-SiPho满足单通道400G等更高速率，实现更高信号密度。

CPC+LPO实现出Tray连接



薄膜铌酸锂与其它材料体系对比

Items	LiNbO ₃	InP	SiPh	TFLN	Performance compared to other materials
Optical Loss (dB)	☹☹☹	☹☹	☹	☺☺☺☺	Excellent
Max. Bandwidth (GHz)	☹☹	☹☹☹	☹☹	☺☺☺☺	Excellent
Half-Wave Voltage V _π (V)	☹☹☹	☹☹☹	☹☹	☺☺☺☺	Excellent
Extinction Ratio (dB)	☹☹☹☹	☹☹☹	☹☹	☺☺☺☺	Excellent
Chip Length (mm)	☹☹☹☹	☹☹	☹☹	☺☺☺	Median
Linearity	☹☹☹☹	☹☹	☹	☺☺☺☹	Excellent
Integration	☹	☹☹	☺☺☺☹	☺☹☹☹	Excellent
Cost	☹☹	☹☹☹☹	☹	☺☺☺ (Reduced after scaling)	Median (Reduced after scaling)

三、投资建议及风险提示

我们认为AI产业的天花板、变现性、成长性、产业链友好度等是稀缺的，可以支撑起AI算力行业的持续成长，相关标的：

- **光互连**：中际旭创，新易盛，天孚通信，光库科技，长芯博创，仕佳光子，源杰科技，长光华芯，太辰光
- **交换机**：锐捷网络，盛科通信
- **铜互连**：中际旭创，兆龙互连

- **算力需求不及预期：**若后续下游客户算力建设投入未达预期，或AI算力规模未达预期情况，各客户对于算力产品的需求也将不及预期，相关公司业绩表现将受到影响；
- **客户开拓与份额不及预期：**如果相关公司未如预期开拓潜在客户，或在客户处份额低于预期，公司业绩将受到影响；
- **产品研发落地不及预期：**如果相关公司在具有潜在应用前景的产品研发及量产应用上未达预期，将对公司业绩的表现造成影响；
- **行业竞争加剧：**如果行业竞争持续加剧，相关产品份额存在下降的可能。

免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。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。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中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投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

资评级基于分析师对报告发布日后6至12个月内行业或公司回报潜力相对基准表现的预期（A股市场基准为沪深300指数，香港市场基准为恒生指数，美国市场基准为标普500指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证50指数），具体如下：

公司投资评级：

买入：预期未来6个月个股涨跌幅相对基准在15%以上；

增持：预期未来6个月个股涨跌幅相对基准介于5%与15%之间；

中性：预期未来6个月个股涨跌幅相对基准介于-5%与5%之间；

减持：预期未来6个月个股涨跌幅相对基准介于-15%与-5%之间；

卖出：预期未来6个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

增持：预期未来6个月内，行业指数相对强于基准5%以上；

中性：预期未来6个月内，行业指数相对基准-5%与5%；

减持：预期未来6个月内，行业指数相对弱于基准5%以上。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所

苏州工业园区星阳街5号

邮政编码：215021

传真：（0512）62938527

公司网址：<http://www.dwzq.com.cn>

东吴证券 财富家园