

被低估的CPU与爆发的Agent时代： 全球服务器CPU市场规模测算和趋势分析

行业研究 · 海外市场专题

互联网 · 互联网 II

投资评级：优于大市（维持）

证券分析师：张伦可

0755-81982651

zhanglunke@guosen.com.cn

S0980521120004

证券分析师：刘子谭

liuzitan@guosen.com.cn

S0980525060001

- **需求端：AI Agent驱动CPU范式转移**，从配套配角到核心主力，配比、需求、价格、产业格局全面重构。根据GPU出货、柜内柜外划分、各场景CPU厂商份额和ASP假设（下同），**测算2030年全球服务器CPU市场将突破2000亿美元**。英伟达、谷歌、Intel等巨头均确认配比趋势，英伟达VeraCPU\ RubinGPU采用1:2配比，谷歌TPUV8与CPU为1:2，Intel CEO预判未来将走向1:1均衡，甚至可达4:1。
- **收入端：AI服务器CPU市场呈现量价齐升的高景气特征**。**测算2030年全球服务器CPU总出货量达6700万颗，较2025年增长3倍；市场规模从2025年312亿美元增至2070亿美元，年复合增长率46%**。AI服务器CPU需求由柜内配套+柜外扩容双轮拉动，**测算总CPU:GPU配比从2026年1:3升至2030年0.93:1，接近1:1均衡**。
- **格局端：全球服务器CPU市场从Intel、AMD X86双雄争霸，转向X86阵营守成、ARM阵营快速崛起的三足鼎立格局**，架构差异决定场景适配边界。**测算2030年Intel份额39%、AMD 21%、ARM 从2025年的11%提升至27%**。
 - ① X86：采用封闭授权模式，生态兼容性强、企业客户基础深厚，适配柜外扩容与传统服务器场景；
 - ② ARM：采用开放IP授权模式，支持深度定制，低功耗、高能效，成为英伟达、谷歌、亚马逊等云厂商柜内AI算力的首选。
 - 两生态场景分化明确，柜内AI训练/推理以ARM为主流，柜外扩容场景依托X86生态兼容性，由Intel、AMD占据主导。行业并非ARM线性替代X86，巨头间生态合作重塑格局，形成双架构共存的稳定生态。
- **供给端：四大核心CPU厂商依托不同商业模式，释放差异化业绩弹性**。
 - ① **Intel**：服务器CPU稳居行业龙头，IDM（自研自产）模式下产能优势与经营杠杆明显。当前市场份额触底企稳，产品逐步高端化转型。英特尔并未效仿AMD“CPU+GPU”协同出货的模式，也因此与各大自研芯片厂商形成错位竞争、合作自研CPU芯片。代工业务（IFS）成为第二增长曲线。
 - ② **AMD**：四家同业厂商中AI业务（服务器CPU+AI GPU）暴露度最高，业绩弹性明显，测算其数据中心GPU未来五年收入CAGR可达52%。CPU+GPU软硬件深度协同，产品生态布局完善。可能存在产能限制。
 - ③ **ARM**：受益与ARM生态规模快速扩张，轻资产版税与授权双轮驱动增长。自研CPU开辟新增量，产能成为短期约束。依赖消费电子周期，目前AI业务权重偏低。
 - ④ **高通**：作为消费级CPU赛道龙头，是Arm生态“架构级定制者+跨终端复用者”，低功耗与定制化能力强。把消费级CPU锁定的台积电先进产能、存储产能转向高价值CPU/加速器。可提供芯片、软件、场景优化一体化全栈方案，已获头部AI科技厂商意向订单。

• **风险提示：**CPU技术进度风险、AI需求增长不及预期风险、地缘政治扰动风险

请务必阅读正文之后的免责声明及其项下所有

更多免费行业报告

并购家

www.ipoipo.cn

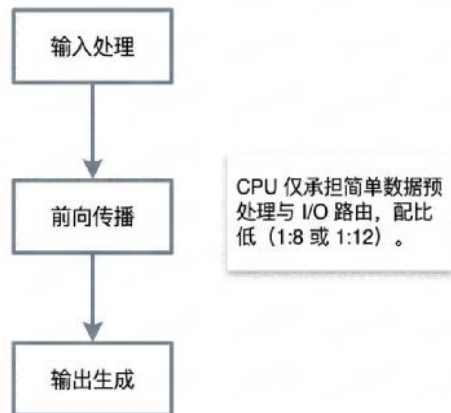
- [01] Agent背景下CPU配比变化
- [02] 服务器CPU市场规模测算
- [03] CPU市场格局测算
- [04] Intel、AMD、ARM、高通业务与弹性比较
- [05] 风险提示

范式转移：从“模型推理”到“ workflow 执行”

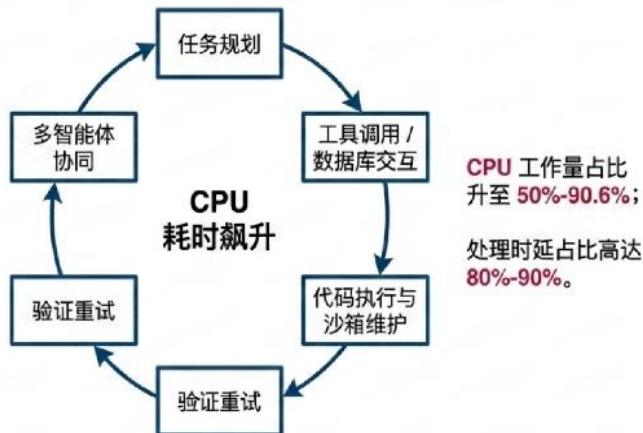
- AI 工作负载从“生成”走向“执行”之后，系统瓶颈由 GPU 单点算力转向 CPU 主导的编排、调度、工具调用、验证与 I/O。Agentic AI 将任务拆解为规划、调用工具、数据库/API 交互、代码执行、验证、重试、多智能体协同等多步骤工作流，这些环节大量依赖 CPU 处理。
- 因此 Agentic AI 阶段后，对应的是 CPU 侧耗时占比抬升至 50%-90%（引用下述论文结论）。Agent 需要加载到 CPU 内核的沙箱中运行，任务完成后才会释放资源。同时，CPU 环节价值量和性能提升，要求核心为高核数/多线程、大三级缓存，单核主频 3.5G-4.0G 即可满足需求。

图：CPU 范式转移

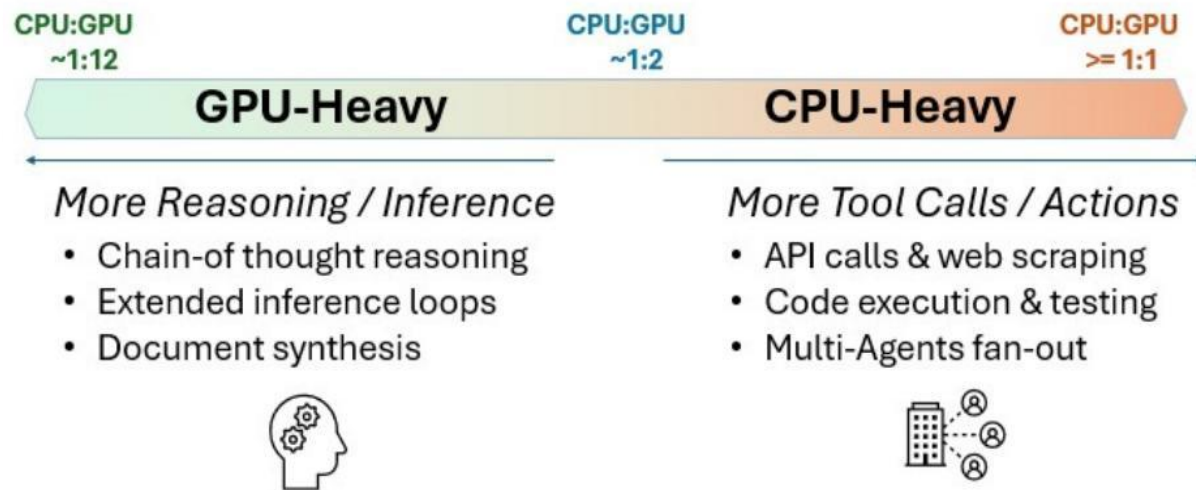
传统生成式 AI - GPU 主导



Agentic AI - CPU 主导编排



图：AI 任务从推理走向执行后 CPU:GPU 配比抬升



资料来源：阿里云、各公司官网，Ritik Raj, 《Towards Understanding, Analyzing, and Optimizing Agentic AI Execution: A CPU-Centric Perspective》, arXiv:2511.00739. 国信证券经济研究所整理

资料来源：Greater China Semiconductors, 国信证券经济研究所整理测算

Agent进程最少需要一个核，越复杂度Agent任务、CPU瓶颈约明显



- 一个Agent必须配一个沙箱作为“容器”来保证进程和内存独立、安全，一个沙箱最少需要1个核(高并发、多工具调用需更多)，该核在用户打开Agent时持续被占用。
- 不同复杂度任务的CPU资源消耗显著差异。100万并发任务需要320万核CPU，测算依据为CPU利用率40%，预留25%-30%的冗余buffer。CPU需求提升不是所有AI负载等比例扩张，而是与 workflow 复杂度、工具调用密度、代码验证频率和多Agent协作深度高度相关。
- 越复杂度Agent任务、CPU瓶颈约明显，如数据库检索、代码执行、搜索摘要等CPU任务处理时延占总时延80%-90%。

表：任务复杂度与CPU资源消耗诊断矩阵			
任务类型	核心与时长消耗	CPU角色与系统瓶颈	硬件规格偏好
低复杂度 (单轮问答、查询美食)	1核/约5秒	预处理 / 路由 (无明显瓶颈)	基础核数即可
中复杂度 (RAG/API调用/预定早餐)	4核 / 约30秒	调度、缓存、数据库交互	多核、高I/O
高复杂度 (代码执行、多工具调用)	6核 / 约300秒	多工具调用、代码执行、沙箱维护 (CPU成为主瓶颈)	高核数、高内存容量
超高复杂度 (复杂任务并行、生成APP)	8核/约20-25分钟	大规模调度、长时间运行状态维护	机架级互联、超大内存CPU集群

资料来源： Gartner、IDC，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有

更多免费行业报告

并购家

www.ipoipo.cn

权威声音：CPU：GPU配比将收紧至1:1-1:2，甚至4:1

- 核心趋势：从“GPU独大”向“CPU+GPU协同”转变，智能体时代配比从1:4-1:8收紧至1:1-1:2，CPU需求激增4倍。

① CEO黄仁勋在2026年GTC大会（26年3月16日）：“我们正迈入推理时代，CPU必须与GPU协同优化，以适配推理与智能体工作负载的复杂调度需求；VeraCPU与Rubin GPU采用1:2配比，核心目的是实现系统级算力均衡。”

② 谷歌技术团队在2026年Google Cloud Next大会（26年4月22日）：展示拓扑图TPU V8与CPU采用1:2配比设计，“Google Cloud 整体模型 API 的调用量已达到每分钟处理超过 160亿Tokens。”

③ Intel CEO在26Q1财报会（26年4月23日）：“CPU与GPU的配比过去是1:8，现在是1:4，我认为未来会走向1:1均衡，甚至更有利于CPU。”

④ 大摩科技年会（26年5月20日）：英特尔公司CEO陈立武表示，Intel-18A良率每月提升约7%，超内部预期。随着AI从训练转向推理，CPU与GPU的配置比例从1:8向1:1靠拢，甚至可达4:1。

图：CPU：GPU架构配比

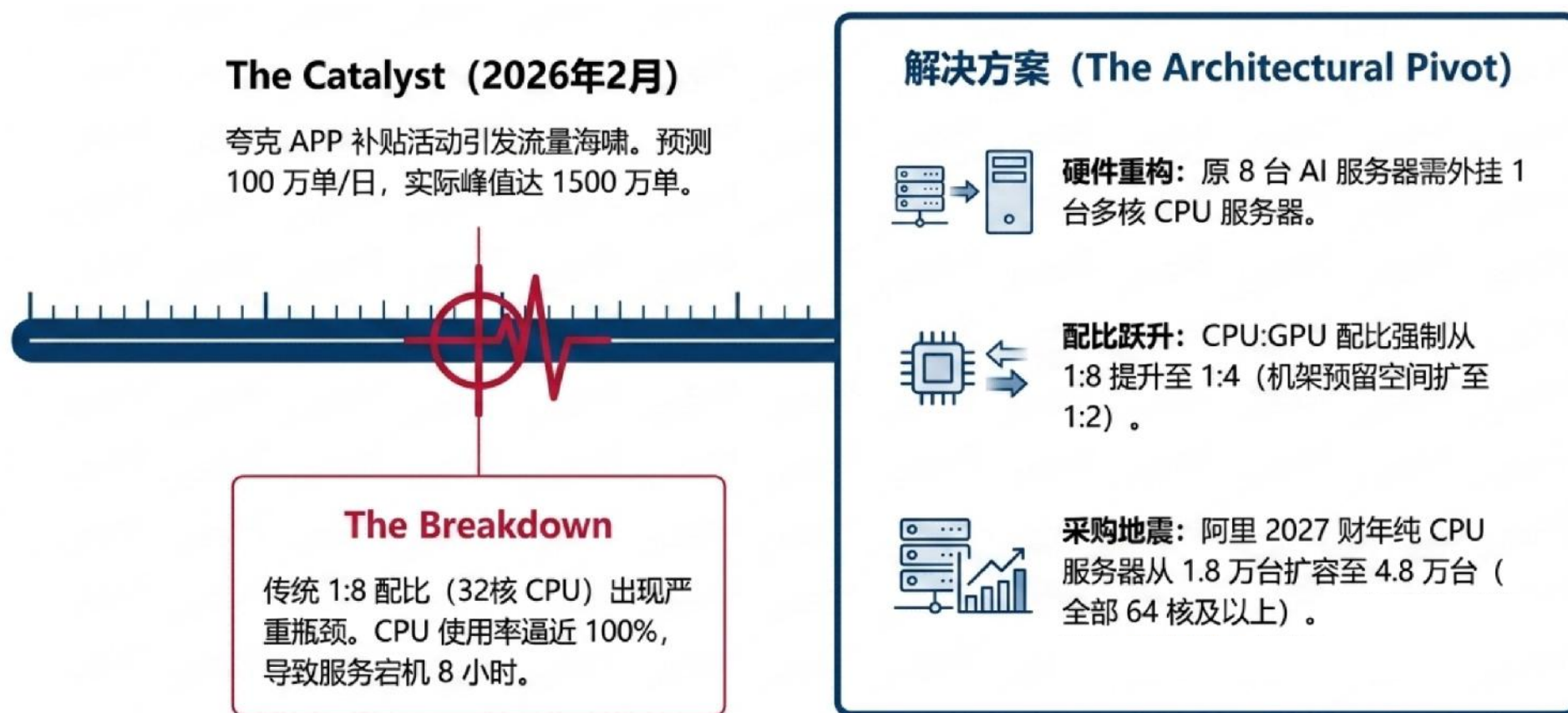


资料来源：各公司财报及公开会议纪要，国信证券经济研究所整理

案例：Agent背景下，阿里AI服务器配比来到1:4

- 据阿里云和钛媒体报道，2025年阿里AI服务器采用32核CPU与GPU 1:8的配比，可满足chatbot类推理业务需求，存量CPU以8核、16核、32核的低核产品为主。新架构下8台AI服务器需外挂多核CPU服务器，预计CPU与GPU配比提升至1:4；机架预留扩展槽，后续甚至可根据需求进一步扩容至1:2。

图：阿里Agent架构



资料来源：阿里云，钛媒体，国信证券经济研究所整理

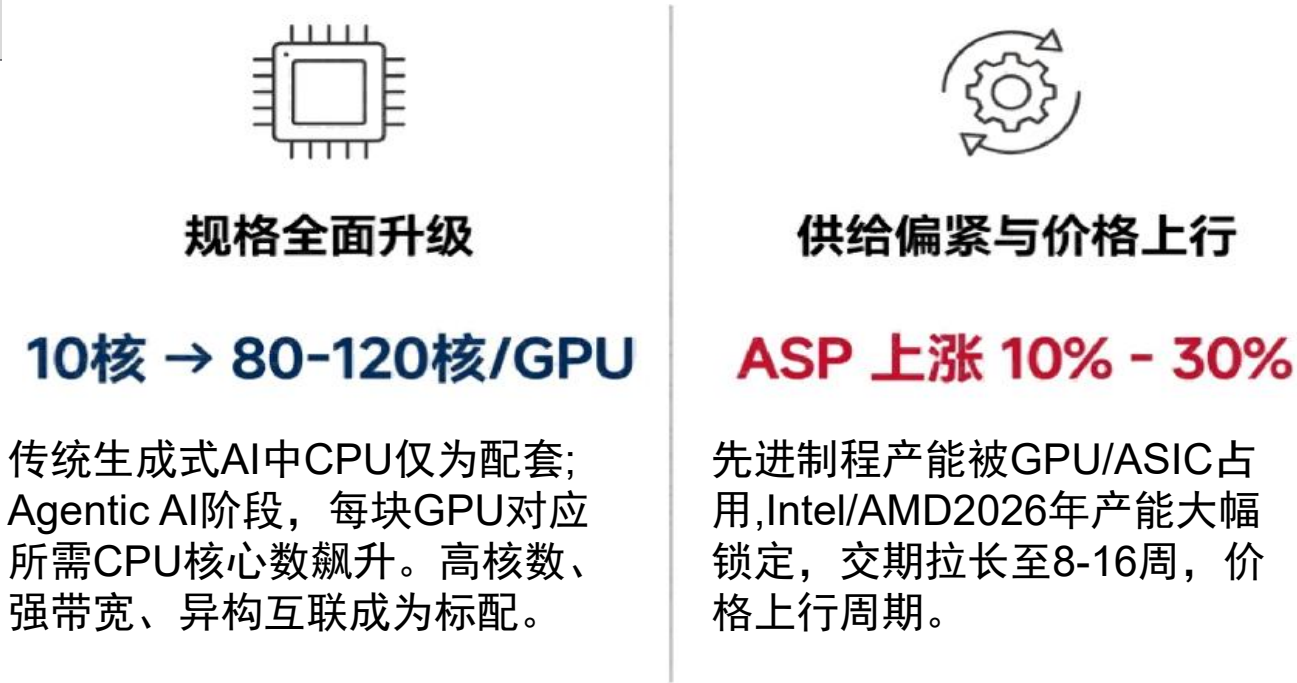
CPU价值重估传递到终端，CPU量价齐升

- 英特尔26Q1业绩会与高管CRN独家专访官方披露CPU价格较2025年底累计上涨25%-30%，当前涨价计划未发生变化，TrendForce预估今年还会启动新一轮涨价，2026年全年涨幅目标为30%，2027年计划在此基础上再上涨10%以上。
- 根据TrendForce，英特尔2026年Q1商用服务器CPU出货量同比2025年Q1增长20%；Q2出货量同比增长30%，当前Q3订单已全部接满，Q2到2026年末出货量预计维持30%左右的同比增速。

表：2026年CPU涨价情况

产品类型	涨价幅度	提价时间	说明
服务器CPU	已涨10%-20%， 高端AI型号 25%。	2026年3月 起	Intel 2月首轮涨幅10%-15%，3月16日在此基础上再涨15%，累计较1月涨价20%。计划5月第三次涨价，全年累计涨价约30%。
消费级CPU	5%-10%	2026年3月 起	受产线被大幅挤压影响，缺货导致代理商交期大幅拉长（部分交期长达6个月）。

图：CPU价值重估



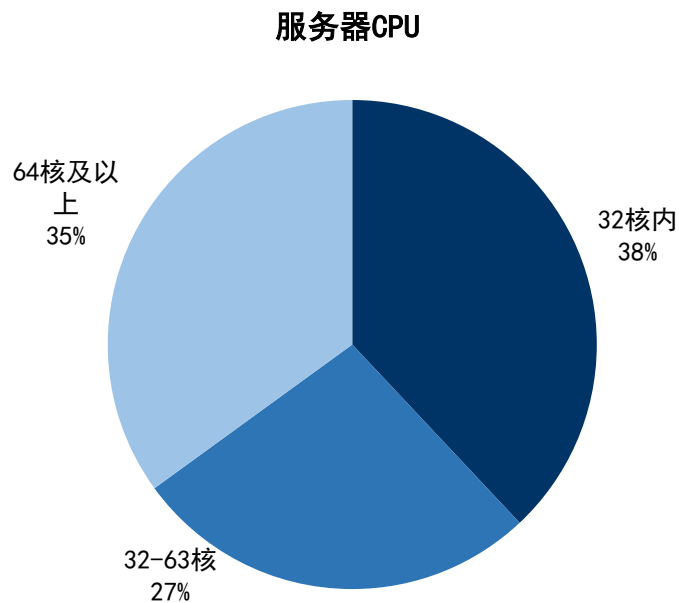
资料来源：华尔街新闻，TrendForce，国信证券经济研究所整理

资料来源：Intel、Ritik Raj, 《Towards Understanding, Analyzing, and Optimizing Agentic AI Execution: A CPU-Centric Perspective》, X: 2511-00730，国信证券经济研究所整理测算

算力重构的底层逻辑：真正的扩容变量“核”

- Agentic AI带来的需求上行不仅在于CPU数量增加，更在于单CPU核数持续提升。
 - ① 传统AI基础设施中，每GPU仅需8-32个CPU核心；
 - ② Agentic AI中，每GPU对应CPU核心需求升至80-120个，CPU:GPU比重甚至超过1:1。
 - ③ 每吉瓦数据中心所需CPU核心数从3000万提升至1.2亿，相当于4倍提升。

图：2025年服务器CPU不同核数占比



资料来源：IDC、SemiAnalysis、JPR，国信证券经济研究所整理测算

图：算力总需求推演公式与GW测算

$$\left[\text{CPU 核心总需求} \right] = \left[\text{全球 GPU 部署总量} \right] \times \left[\text{每块 GPU 所需 CPU 核心数} \right]$$

传统AI：每GPU约需8-12个CPU核心；
Agentic AI：每GPU对应CPU核心需求升至**80-120个**；

同等功率规模下，每吉瓦（GW）数据中心所需的**CPU核心数从3000万跃升至1.2亿**（测算每GW需要GPU约60-110W）。

即使GPU数量保持不变，单节点核心需求10倍跃升与1:4部署折算，直接带来整体CPU算力需求约4倍的结构性爆发。

- [01] Agent背景下CPU配比变化
- [02] 服务器CPU市场规模测算
- [03] CPU市场格局测算
- [04] Intel、AMD、ARM、高通业务与弹性比较
- [05] 风险提示

高资本开支下，GPU/加速器芯片出货量2030年有望达4000万颗



- 2026年全球科技资本开支持续增长，预计其中北美四大云与科技巨头（微软/谷歌/亚马逊/Meta）贡献超7000亿美元，全球资本开支约万亿美金。其他主权国家、企业资本开支持续加码，26CYQ1英伟达数据中心收入中Hyperscale（公有云+全球最大消费互联网公司）与ACIE（AI云、工业、企业）各占一半。黄仁勋在26CYQ1业绩会表示**超大规模企业资本支出2030年底年投入将达3-4万亿美元**。
- 根据各个企业对于其自研芯片产品路线与产能测算得：
- 2026全球算力芯片出货量约1700多万增长到2030年约4000万颗**，其中GPU（NVDA、AMD为主）约占比55%，ASIC芯片占比45%。谷歌TPU贡献了核心增量，测算2030年出货占比总出货量40%，其次为英伟达NVDA。

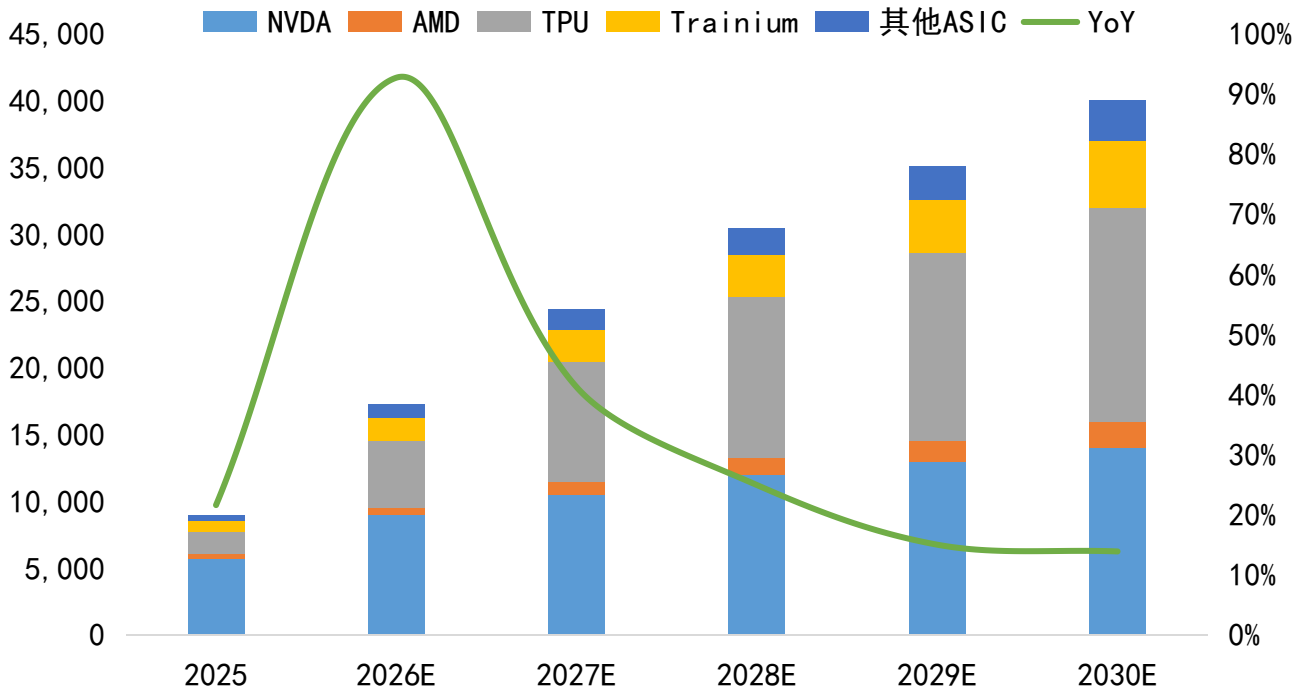
表：全球GPU/ASIC 出货量与YoY

Accelerator加速器/千颗*	2024	2025	2026E	2027E	2028E	2029E	2030E	
NVDA	4,850	5,700	9,000	10,500	12,000	13,000	14,000	25年测算英伟达GB约出货200W颗GPU，剩余为单独的H和B卡服务器出货，对应自研CPU约100W。2026年预计GB出货500W颗GPU，对应250颗CPU。
AMD	384	410	570	1,000	1,300	1,600	2,000	
TPU	1,280	1,700	5,000	9,000	12,000	14,000	16,000	预计2027年TPU V8采用Axion 自研CPU（ARM架构），CPU：GPU=1:2
Trainium	750	750	1,700	2,400	3,200	4,000	5,000	2025年Trainium3采用自研Graviton4（ARM 架构），CPU：GPU=1:4
其他ASIC	100	400	1,000	1,500	2,000	2,500	3,000	
Total	7,364	8,960	17,270	24,400	30,500	35,100	40,000	
YoY		22%	93%	41%	25%	15%	14%	

高资本开支下，GPU/加速器芯片出货量2030年有望达4000万颗

- 根据各个企业对于其自研芯片产品路线与产能测算得：
- **2026全球算力芯片出货量约1700多万增长到2030年约4000万颗**，其中GPU（NVDA、AMD为主）约占比55%，ASIC芯片占比45%。谷歌TPU贡献了核心增量，测算2030年出货占比总出货量40%，其次为英伟达NVDA。
- 同时，CPU与GPU的配比随着Agent和推理需求的增长快速提升，因此CPU行业增速显著快于GPU/加速器板块。

图：全球GPU/ASIC 出货量与YoY



资料来源：各公司业绩会，国信证券经济研究所整理测算 *根据各家产品迭代路线与代工产能测算

表：AI服务器CPU/GPU配比及型号

应用场景	CPU/GPU配比	推荐 CPU 型号	核心数	适用场景说明
大规模训练	1:8	Intel Xeon Platinum 8490H/AMD EPYC 9654	64-96 核	数据预处理、模型并行调度
常规训练	1:4	Intel Xeon Gold 6442Y/AMD EPYC 9554	48-64 核	中大型模型训练，参数更新频繁
批量推理	1:4	Intel Xeon Gold 6430/AMD EPYC 9454	32-64 核	图像识别、语音合成等固定模型
在线推理	1:2	AMD EPYC 9354/Intel Xeon 5318Y	24-32 核	低延迟要求，实时响应
Agentic AI	1.2:1	NVIDIA Grace（双芯片）/AMD EPYC Bergamo	96-144 核	智能体任务，多线程并行处理

资料来源：Intel、AMD及NVIDIA芯片产品白皮书，国信证券经济研究所整理

柜内、柜外双重拉动，不同产品/场景第三方CPU份额差距明显



- 将AI服务器CPU出货拆分为柜内（按照当前机柜方案CPU/GPU配比假设）、柜外测算（根据需求场景CPU需求量假设）：**2030年全球AI服务器CPU柜内出货量预计达2500多万颗、五年CAGR为55%；总AI服务器CPU 2030出货量预计达3700多万颗、五年CAGR为68%；**
- CPU与GPU的配比：**柜内从2026年出货的约1:3提升至0.63:1；加上柜外的CPU数量，**总的CPU：GPU配比从26年的1:3到30年的0.93:1。**
- 注：**蓝色底色为GPU出货量采用第三方CPU供货、而非自研CPU的情况。根据TrendForce，柜外CPU主要负责 Agentic AI工作流的非张量计算任务，其对应CPU:GPU约为1.2:1，假设Agent任务占比AI训练/推理占比得到柜外比例。

表：各家芯片方案对应CPU数量变化测算（千颗/%）							
芯片方案对应CPU数量（千）	2024	2025	2026E	2027E	2028E	2029E	2030E
NVDA自研CPU方案	116	1,425	2,250	2,625	3,360	2,925	3,696
NVDA外供CPU方案	1,154	713	1,485	2,100	3,600	5,915	7,840
谷歌自研CPU方案				3,150	4,680	5,250	6,400
谷歌外供CPU方案	160	425	1,250	1,080	2,400	4,550	6,000
AMD	95	100	143	250	429	880	1,400
亚马逊	94	94	280	720	1,600	2,600	4,000
其他ASIC	13	50	150	375	800	1,500	2,250
柜内CPU总数量（千颗）	1,632	2,807	5,558	7,150	12,189	18,370	25,186
YoY		72%	98%	29%	70%	51%	37%
机柜级 CPU:GPU比例	22%	31%	32%	29%	40%	52%	63%
柜外CPU总数量（千颗）	—	—	864	2,440	5,490	8,775	12,000
机柜外CPU:GPU比例	0%	0%	5%	10%	18%	25%	30%
总AI服务器CPU需求量（千颗）	1,632	2,807	6,421	9,590	17,679	27,145	37,186
YoY		72%	129%	49%	84%	54%	37%
总CPU:GPU比例	22%	31%	37%	39%	58%	77%	93%

资料来源：官方发布会、TrendForce，国信证券经济研究所整理。*NV外供方案：NV CPU出货中不采用其自研CPU的量，包括英伟达采用第三方CPU或客户只购买英伟达GPU、CPU采购第三方

加速器巨头逐步切换自研CPU，带动配套CPU增长

- 加速器巨头逐步切换自研CPU：英伟达2024年出货主力B系列和后续产品机柜方案切换为自研CPU（此前为Intel 至强芯片），对于Intel在AI服务器市场份额带来影响。谷歌2027年主力出货的TPU V8版本也将切换为自研Axion CPU。
- AMD与亚马逊加速器配套方案均采用自研CPU。

表：代表产品CPU与加速器的配比情况

厂商	发布时间	产品名称	CPU: 加速器配比	核心说明
英伟达	2022年	DGX H100	1:4 (2×Xeon CPU+8×H100 GPU)	CPU采用Intel 至强芯片（X86架构），2024年出货主力，配比1:4，生成式AI训练经典机型，CPU仅做基础调度
英伟达	2024年3月	GB200 NVL72	1:2 (36×Grace CPU+72×Blackwell GPU)	CPU采用自研Grace CPU（ARM架构），2025年出货，B系列为25年主力产品，配比1:2，面向万亿参数模型推理
亚马逊	2025年12月	Trainium3 Ultra	1:4 (36×Graviton4 CPU+144×Trainium3)	CPU采用自研Graviton4（ARM 架构），配比 1:4，面向万亿参数模型训练与 AI 智能体，CPU 为编排核心
英伟达	2026年3月（GTC）	VeraRubin NVL72	预计1:2 (36×Vera CPU+72×Rubin GPU)	CPU采用自研Vera CPU（ARM架构），配比1:2，预计2027年主力出货，AI智能体专用，NVLink-C2C协同，CPU为编排核心
谷歌	2026年4月（Cloud Next）	TPU PodV8 (8t/8i)	预计1:2 (自研Axion CPU+TPU V8)	CPU采用自研CPU（ARM架构），配比1:2，预计2027年主力出货。智能体推理/训练拆分，CPU负责集群编排
行业通用	2026年（智能体集群）	智能体专属集群	≥1:1（CPU数≥GPU数）	多步骤工具调用/多智能体场景，CPU成瓶颈

资料来源：各公司官方发布会、各公司官网、Google Cloud Next 26、SemiAnalysis、JPR、国信证券经济研究所整理

测算AI服务器CPU市场规模五年复合增速超70%

- AI服务器CPU是核心引擎：根据前页GPU出货，柜外柜内场景划分可得CPU年出货量变化。假设柜内、柜外不同公司的CPU份额和产品的单价（详见第三章、第四章），**测算其市场规模将从2025年的114亿美元，增长至2030年的1773亿美元，年复合增速超70%**，AI服务器在CPU市场占比将超过84%。
- 出货量与ASP双升：出货端，AI服务器CPU占比将从2025年的约12%，跃升至30年的55%；价格端，伴随AI需求爆发增长、价格稳健提升，且带动传统服务器产品ASP上行。

表：服务器CPU市场出货量与收入拆分预测

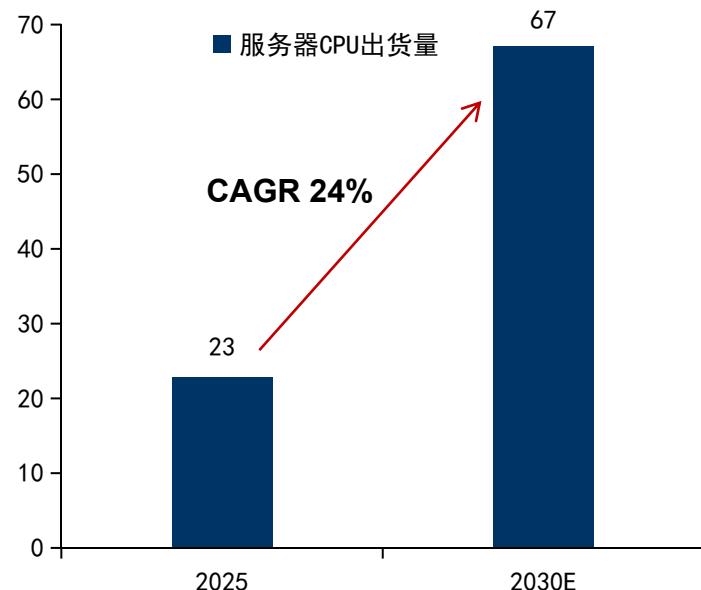
	2024	2025	2026E	2027E	2028E	2029E	2030E
柜内CPU总数量（百万颗）	1.6	2.8	5.6	7.2	12	18	25
ASP（美元）	3515	4076	4222	5683	5462	5117	4956
柜外CPU总数量（百万颗）			0.9	2.4	5.5	8.8	12
ASP（美元）			4185	4427	4482	4430	4373
AI服务器CPU总需求（百万颗）	1.6	2.8	6.4	9.6	17.7	27.1	37.2
AI服务器平均ASP（美元）	3515	4076	4217	5364	5158	4895	4993
AI服务器CPU市场规模（亿美元）	57	114	271	514	912	1329	1773
YoY		100%	137%	90%	77%	46%	33%
传统服务器CPU需求（百万颗）	19	20	22	24	26	28	30
ASP（美元）	868	990	987	993	1014	1035	1057
传统服务器CPU市场规模（亿美元）	165	198	217	240	265	292	316
YoY		20%	10%	11%	10%	10%	8%
服务器CPU总出货量（百万颗）	21	23	28	34	44	55	67
服务器CPU总市场（亿美元）	222	312	488	755	1177	1621	2089
YoY		41%	56%	55%	56%	38%	29%

资料来源：IDC、Gartner、TrendForce，国信证券经济研究所整理测算

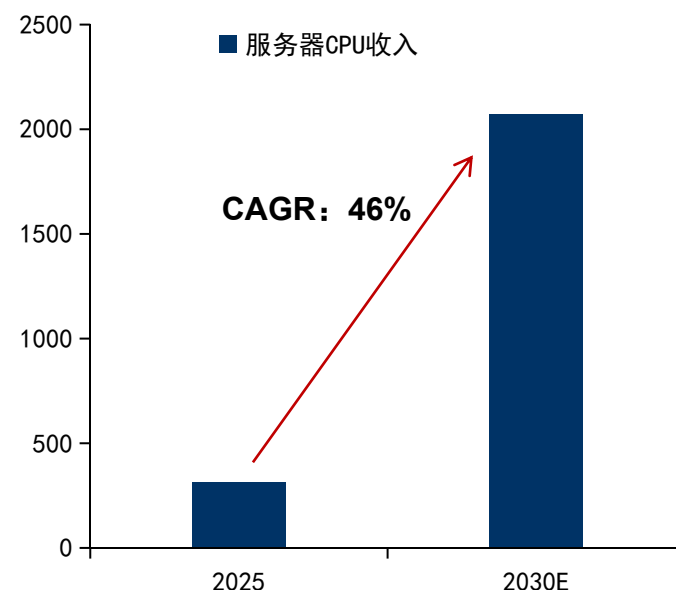
测算2030年全球服务器CPU市场规模超2000亿美金

- **市场出货量：**根据前页GPU出货，柜外柜内场景划分测算，**2030年全球服务器CPU市场总出货量预计达6700万颗，是2025年约2300万出货的三倍，CAGR 24%**。其中传统服务器CPU出货量占比约44%，AI服务器出货量从2025年的280万颗提升至2030年的3720万颗，占比总出货量56%。
- **市场规模：**假设柜内、柜外不同公司的CPU份额和产品的单价（具体见第三章、第四章），测算2030年全球**服务器CPU市场规模有望提升至2070亿美金，CAGR约46%**。其中AI服务器从2025年的约114亿美金收入提升至2030年的1700多亿美元，CAGR 73%。2030年AI服务器收入占比总收入85%。

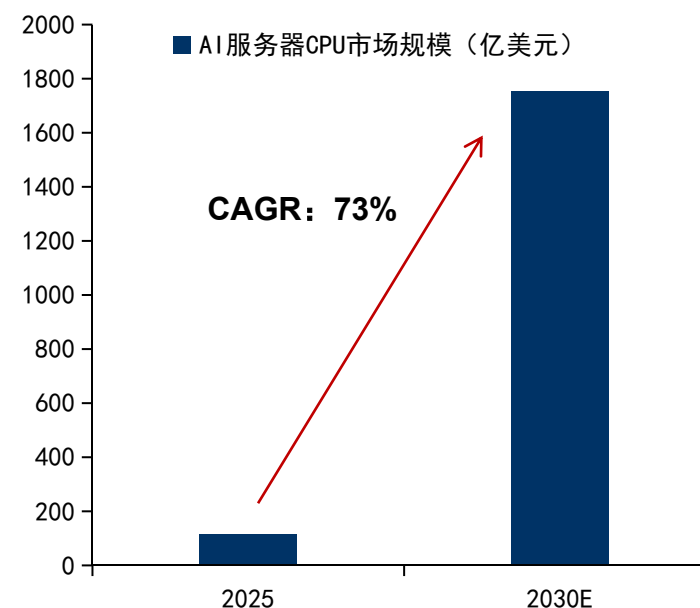
图：全球服务器CPU出货量（百万台）



图：2025 VS 2030 全球服务器CPU收入（亿美元）



图：2025 VS 2030 全球AI服务器CPU收入（亿美元）



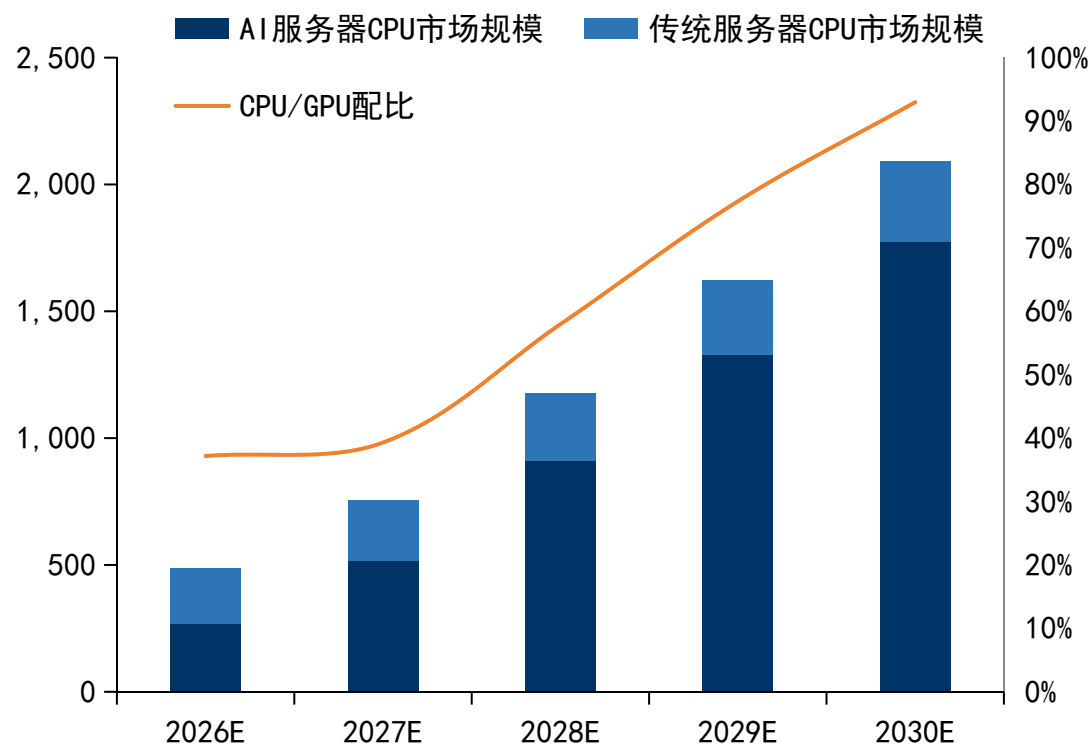
资料来源：IDC，国信证券经济研究所整理测算

国信证券经济研究所整理测算

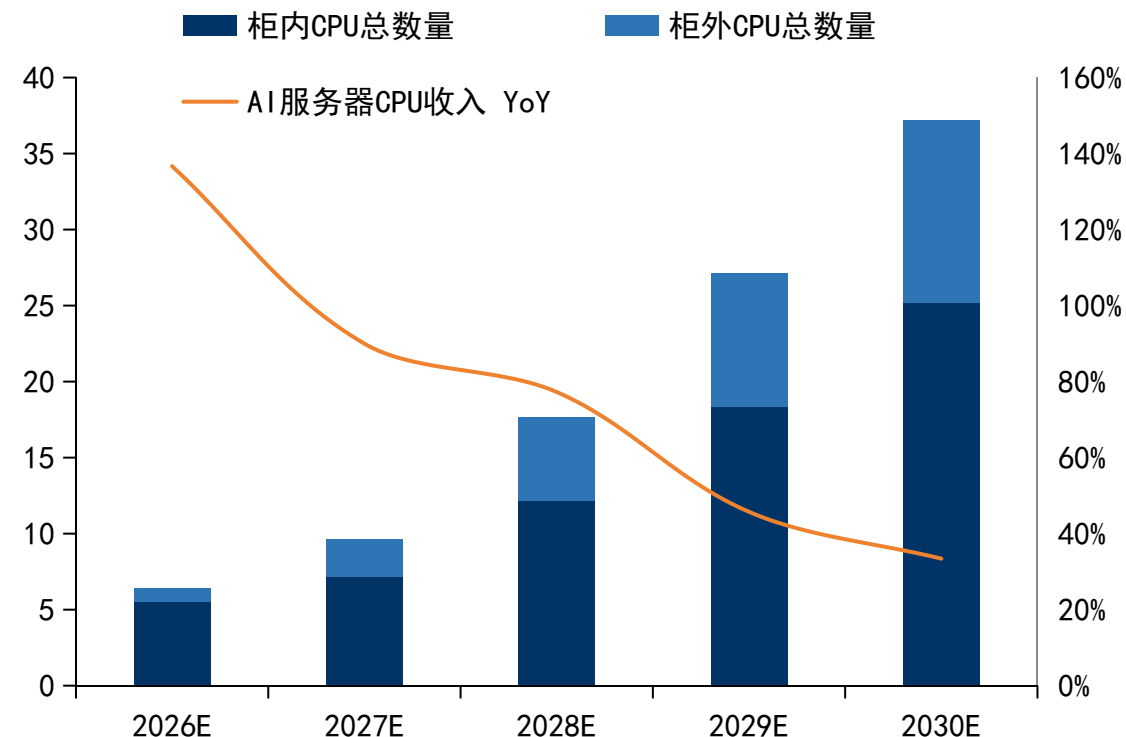
测算AI服务器CPU市场规模五年复合增速超70%

- AI服务器CPU是核心引擎：假设传统服务器市场维持当前增速平稳扩张，根据前页GPU出货、柜外柜内场景划分，**测算AI服务器在CPU市场占比从2025年的36%增长至2030年约85%**，彻底改写服务器CPU市场的格局。其中核心的驱动是2026年开始的CPU与GPU配比数的提升，从2025年的1:3提升至2030年的0.93:1。
- 柜内、柜外结构变化：当前AI服务器CPU仍以柜内为主，但外置独立CPU机柜/外置插槽随着Agent渗透快速发展，**预计柜外的CPU数量占比将从2026年的约14%跃升至30年的32%**，柜外的CPU将成为独立CPU供应商的主要市场（与柜内GPU协同性弱）。

图：AI服务器及传统服务器CPU市场规模与配比（亿美元）



图：AI服务器CPU收入与同比和柜内柜外拆分（亿美元，百万颗）



资料来源：IDC、TrendForce，国信证券经济研究所整理测算

资料来源：IDC、TrendForce，国信证券经济研究所整理测算

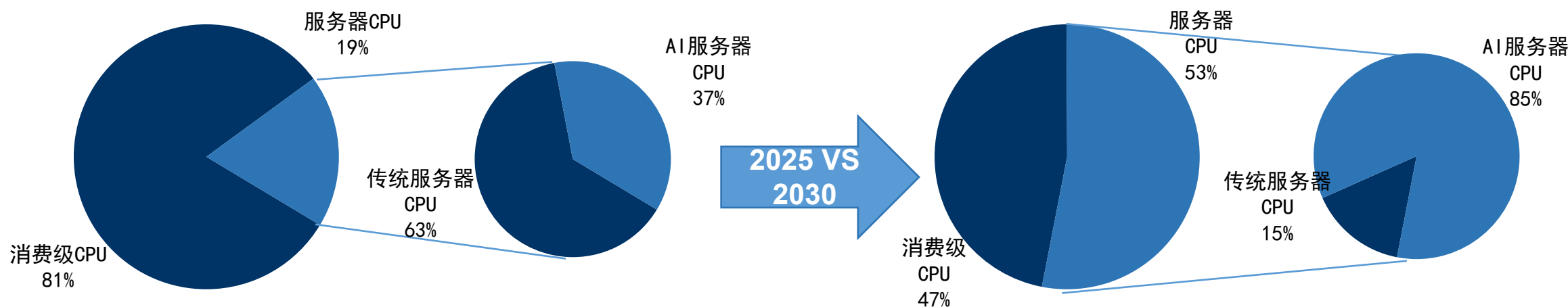
- [01] Agent背景下CPU配比变化
- [02] 服务器CPU市场规模测算
- [03] CPU市场格局测算
- [04] Intel、AMD、ARM、高通业务与弹性比较
- [05] 风险提示

整体CPU市场收入份额变化：服务器CPU占比显著提升

- **服务器CPU在整个CPU市场占比快速提升**：假设消费级服务器市场维持当前增速平稳扩张，根据GPU出货、柜外柜内场景、各场景CPU厂商份额和ASP假设（详见第二章、第四章）。**测算2025年全球CPU市场规模中服务器约占比19%、2030年提升至53%**，其余为消费级CPU，包括桌面端（PC、移动/笔记本）和移动端（手机、汽车、IoT等）。
- **当前CPU整体市场x86仍主导PC/服务器，ARM垄断移动**：根据Mercury Research，2025年底X86在全球整体CPU市场份额约75%，ARM约占比24%。其中x86由于其通用高性能与向下兼容性在桌面和笔记本市场占比超90%，ARM凭借其低功耗与高集成度在移动手机市场占比超99%。

图：整体CPU市场2025年 VS 2030年收入份额变化（%）

- ① 服务器CPU市场以X86生态（Intel/AMD）垄断；
- ② 消费级CPU以ARM生态主导，尤其垄断手机、IoT等移动端，X86仅在桌面端有一定份额。



失落的十年：X86生态Intel与AMD服务器CPU市场竞争复盘

- ① 2015-2018年：Intel垄断、AMD重返服务器。Intel仍凭Broadwell/Skylake维持统治，AMD直到Naples才重新打开局面。
- ② 2019-2022年：AMD凭借制程和技术优势抢占份额。Intel被110nm推迟所拖累，而AMD率先切入7nm、后续切入5nm，产品节奏连续兑现。
- ③ 2023-2026年：Intel加快产品迭代但仍未彻底扭转路线劣势，AMD把架构与平台优势进一步放大。

表：Intel与AMD服务器CPU迭代与竞争路线

年份	竞争判断	Intel主打产品	Intel关键规格	AMD主打产品	AMD关键规格
2015	Intel绝对垄断，AMD在核心数、内存通道、PCIe通道上已形成纸面优势，但生态与验证仍弱。	Broadwell EP/EX	14nm，最多24核，DDR4，PCIe 3.0	当期无EPYC主力	服务器端尚未形成竞争
2017		Skylake-SP	14nm+，最多28核56线程，6通道DDR4-2666，PCIe 3.0x48	Naples (EPYC 7001)	14nm，最多32核64线程，8通道DDR4-2666，PCIe 3.0x128
2019	关键拐点：AMD Chiplet方案，首次在制程、核心数、I/O、TCO全面领先，率先切入7nm、5nm，产品节奏连续兑现。Intel自研自产 IDM 战略，受到10nm推迟拖累。	Cascade Lake-SP	14nm++，最多28核56线程，6通道DDR4-2933，PCIe 3.0x48	Rome (EPYC 7002)	7nm+14nm Chiplet，最多64核128线程，8通道DDR4-3200，PCIe 4.0x128
2020		Cooper Lake-SP	14nm++，最多28核，6通道DDR4-3200	Rome延续	64核128线程持续放量
2021	Intel 10nm落地、但7nm翻车，提出IDM 2.0把高端CPU外包给台积电；AMD靠Zen3继续扩大性能与份额。	Ice Lake-SP	10nm，最多40核80线程，8通道DDR4-3200，PCIe 4.0x64	Milan (EPYC 7003)	7nm，最多64核128线程，8通道DDR4-3200，PCIe 4.0x128
2022		Sapphire Rapids准备中	多次延期，DDR5/PCIe5/CXL是卖点，但发布时间拖后	Genoa (EPYC 9004)	5nm，最多96核192线程，率先完成DDR5+PCIe5+高核心数升级，12通道DDR5-4800，PCIe 5.0x128
2023	Intel加快产品迭代但仍未扭转路线劣势，AMD开始Classic/ Dense/边缘化分层布局，仍在线程密度、能效、TCO上占优。	Sapphire Rapids / Emerald Rapids	Intel 7，最多60核112线程/64核128线程，8通道DDR5-4800/5600	Bergamo / Siena	5nm，最多128核256线程（Bergamo），6通道SP6（Siena）
2024		Granite Rapids / Sierra Forest	Intel 3，最多128核256线程P核 / 144核E核，12通道/8通道DDR5-6400	Turin / Turin Dense	3/4nm，最多128核256线程 / 192核384线程，12通道DDR5-6000
2026-	AMD新产品性能仍领先，Intel对标的 Diamond Rapids产品快速追赶，预计明年出货。	Clearwater Forest / Diamond Rapids（预期）	18A，Clearwater最高288 E核；Diamond预期192-256 P核、16通道DDR5、PCIe 6.0，但存Venice在推迟至2027风险		TSMC N2/2nm，最高256核512线程，16通道DDR5，约1.6TB/s，PCIe 6.0，相对Turin约1.7倍性能/能效

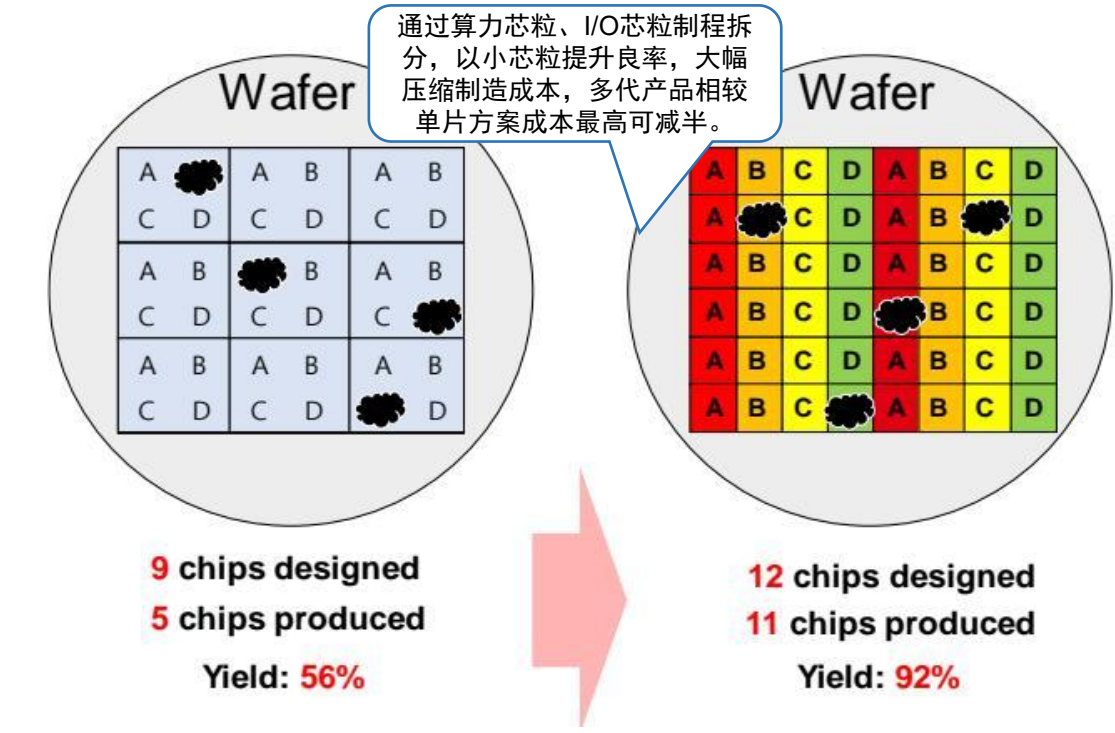
资料来源：各公司官网、公开技术白皮书，国信证券经济研究所整理

- Intel依赖IDM（自研自产），落后核心为10nm制程长期延期、路线反复:造成2015–2021年长达六年的制程迭代停滞。同期AMD完成从14nm到7nm的技术跃迁，实现弯道超车。延期带来三重核心冲击：Intel长期压榨14nm工艺，产品核心、I/O性能迭代乏力；AMD凭借7nm+Chiplet架构、PCIe4.0率先落地，冲击Intel高端服务器溢价体系；产品延期透支了数据中心客户最看重的交付信誉。
- Intel 2021年推出的IDM2.0战略，贴合AI时代制造与供应链的长期战略需求，但转型投入与体系重构，对短期产品迭代形成进一步拖累。
- AMD核心依托“Chiplet架构、台积电先进制程、产品分层迭代”三位一体的体系。AMD 2019年后逐步把“制程领先”转化为“核心数领先、能效领先、TCO领先、份额领先”。技术前瞻，AMD是首个将Chiplet技术大规模商业化并取得市场成功的CPU厂商，叠加清晰分层的SKU迭代策略，精准覆盖通用、高密度、边缘等全场景，稳步夯实长期竞争优势。

表：Intel与AMD各维度对比

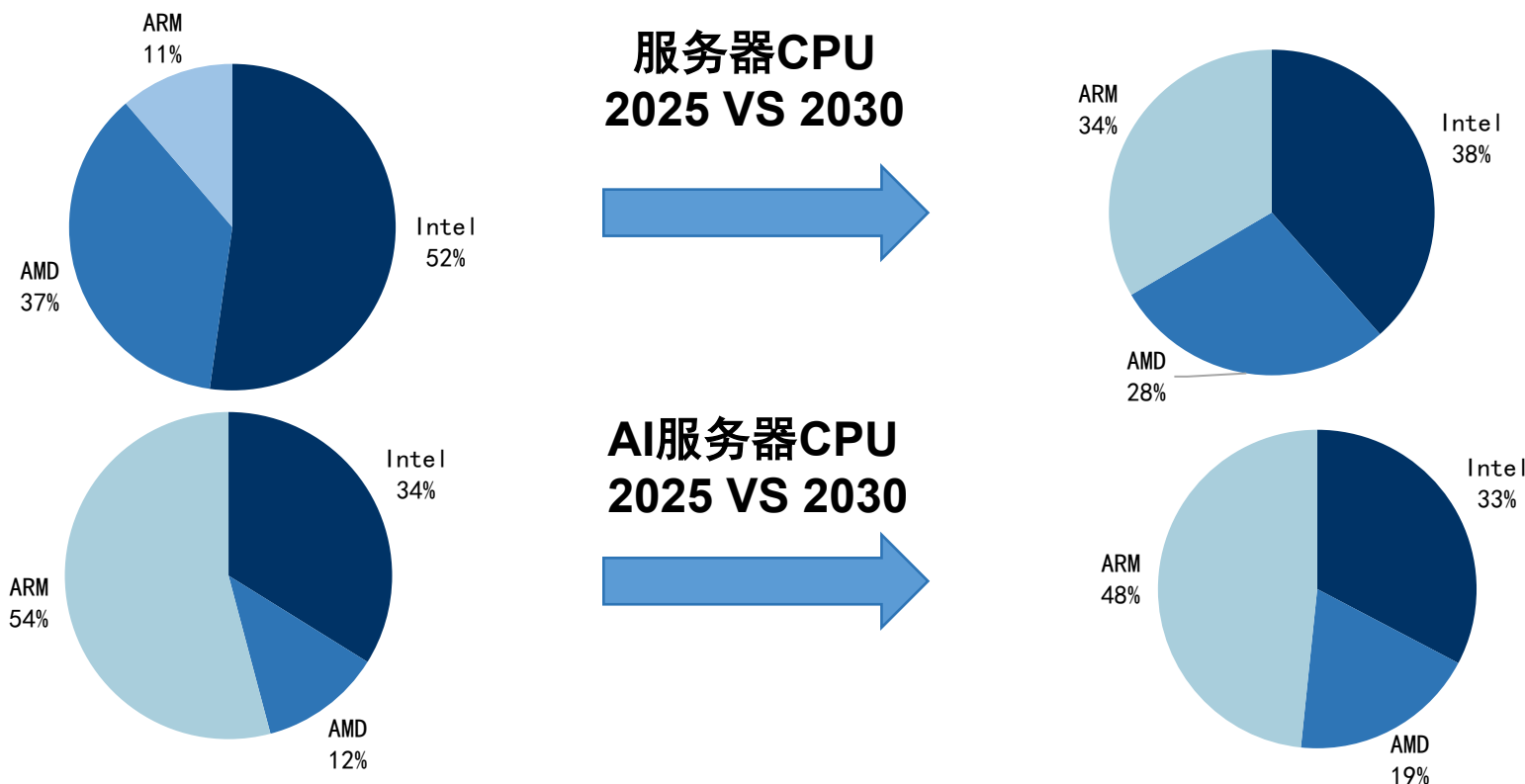
维度	Intel	AMD
组织模式	IDM/IDM 2.0，设计制造一体，追求全栈控制	Fabless+TSMC，专注架构/平台/产品分层
产品路线	P核/E核双线并行，平台区分明显，强调兼容性和企业客户验证	Classic/Dense/Edge分层，强调核心密度、TCO与云场景适配
主要优势	生态兼容性强、企业客户基础深、平台验证成熟、软件适配广	核心数高、功耗低、Chiplet成本优、制程领先、节奏稳定
主要短板	制程与封装执行风险高，路线图可信度受损，Foundry拖累利润	跨芯粒延迟与NUMA复杂性仍是天然约束，新平台切换需要客户重新验证
典型客户强项	企业、运营商、兼容性敏感场景	云、超大规模、高密度服务器、AI head node

图：AMD Chiplet架构介绍



- 未来五年市场格局将从“双雄争霸”转向“三足鼎立”：ARM架构凭借低功耗与高算力优势，在整体市场与AI场景中快速崛起，根据GPU出货、柜外柜内场景、各场景CPU厂商份额假设（详见第二章、第四章），**测算ARM架构份额将从2025年的11%跃升至2030年的28%**，对传统x86阵营（Intel+AMD）形成显著冲击。
- AI服务器CPU市场ARM维持50%以上的主导地位**：2025年ARM出货量占比较高的原因是加速器市场英伟达GPU一家独大，而其机柜方案CPU采用自研Grace CPU（ARM架构）。后续，伴随着ASIC厂商陆续出货，其部分方案将采用第三方CPU（X86架构）。同时，Agent背景下外置CPU机柜多采用X86架构产品，便于与传统服务器X86 CPU生态协同。

图：服务器CPU市场出货量份额变化（%）



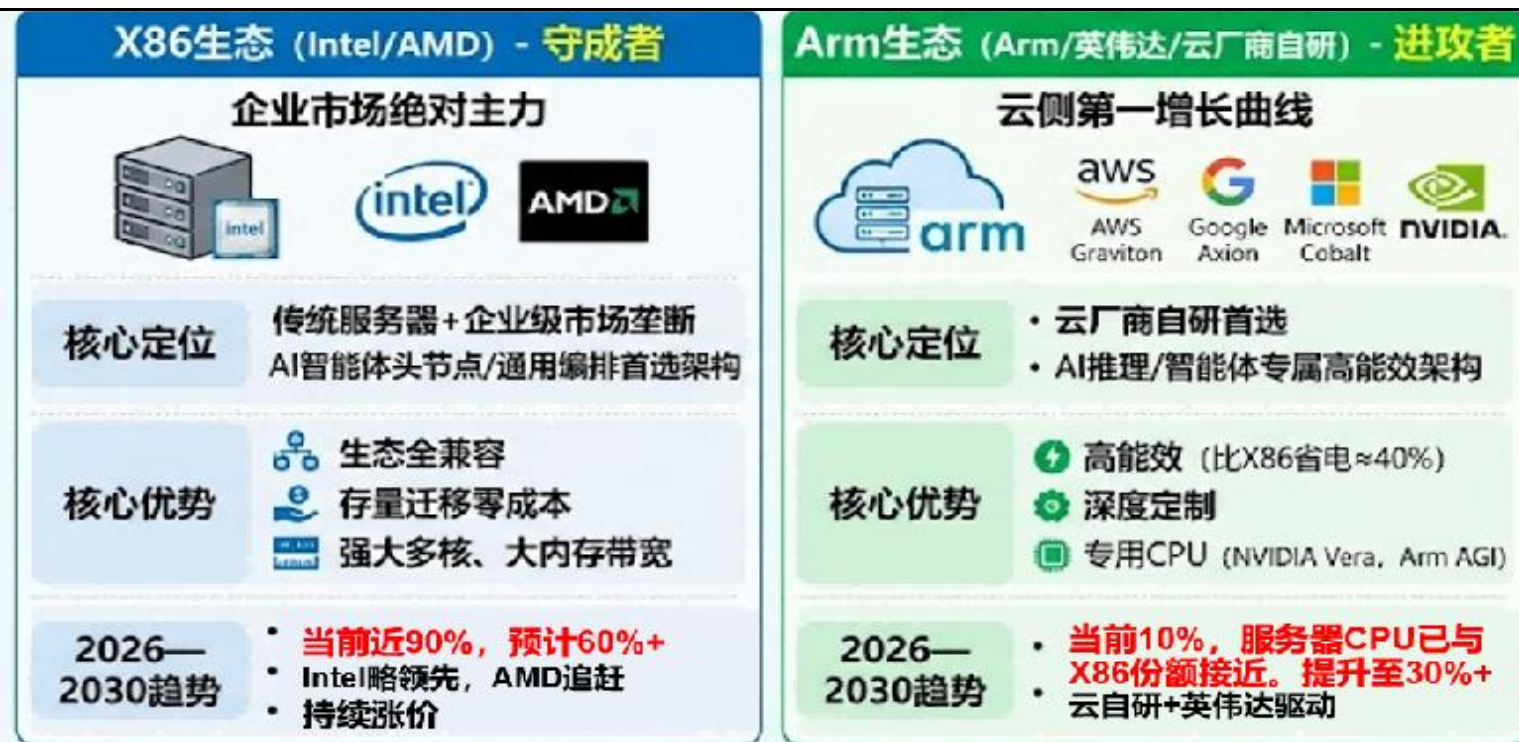
资料来源：Mercury Research, IDC, 国信证券经济研究所整理。根据GPU出货、柜外柜内场景、各场景CPU厂商份额假设（详见第二章、第四章）

商业模式与特性差异决定生态扩张差异，ARM因英伟达和云厂商自研芯片爆发：

① **x86**：采用封闭授权模式，仅英特尔和AMD两家拥有完整设计生产能力，客户只能采购标准化芯片，无法深度定制。同时，通用高性能与向下兼容性，单指令可完成复杂操作，硬件复杂度高但峰值性能强。

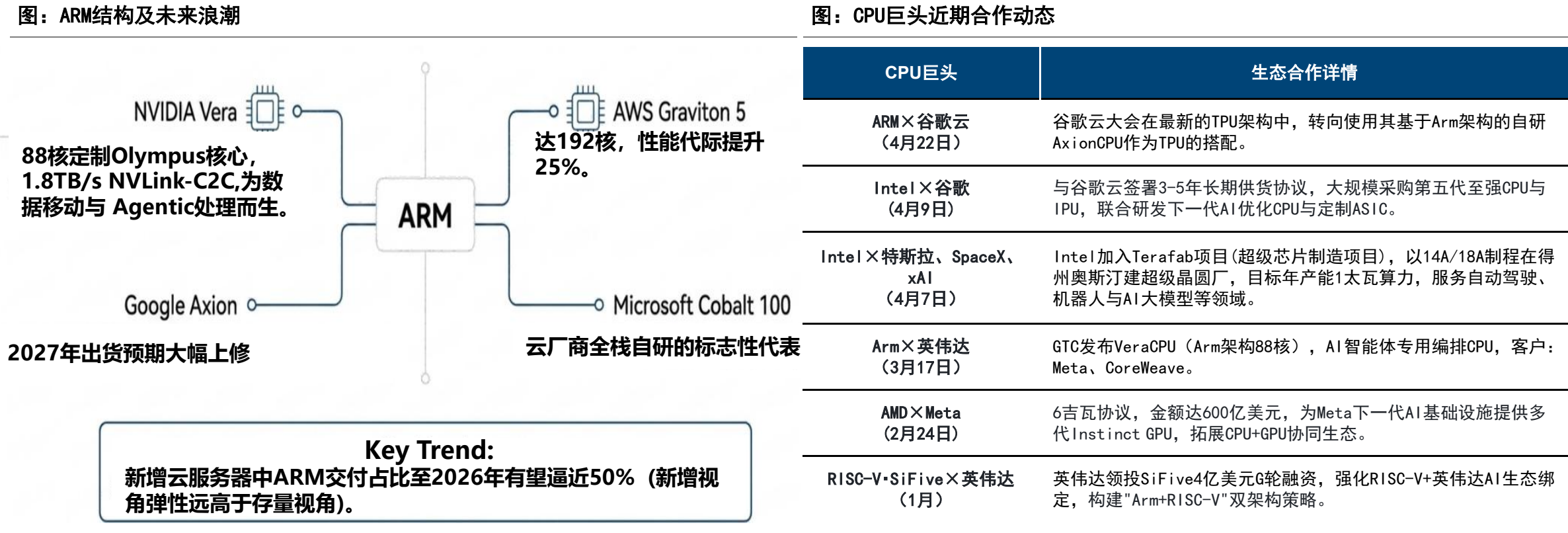
② **ARM**：采用开放IP授权模式，不卖芯片只卖设计蓝图，允许苹果、谷歌、亚马逊等客户根据自身需求定制核心数、缓存、AI加速单元等。起源于嵌入式系统，核心基因是低功耗与高集成度，指令规整、解码效率高。在数据中心电力成本成为核心约束时，ARM的能效比优势（每瓦性能是x86的2倍）凸显。

图：服务器CPU市场格局与收入份额变化



资料来源：公司公开报告、行业调研、公开数据、假设

- ARM不再仅仅是架构授权方，而是借助科技巨头直接演进为完整的AI基础设施平台方案。依靠能效、可定制化、云厂商自研和NVIDIA/Google/Microsoft/AWS等推动，在新增部署中快速提升占比。
- 除了巨头基于ARM自研外，同时看到巨头与CPU厂商之间的丰富合作，如“Intel与谷歌云签署3-5年长期供货协议”、“Intel与英伟达的联合研发CPU合作”、“AMD与Meta合作的协同生态”的**再平衡格局，因此格局可能不是“ARM替代x86”的线性叙事。**



资料来源：各公司官网、公开技术白皮书，国信证券经济研究所整理测算

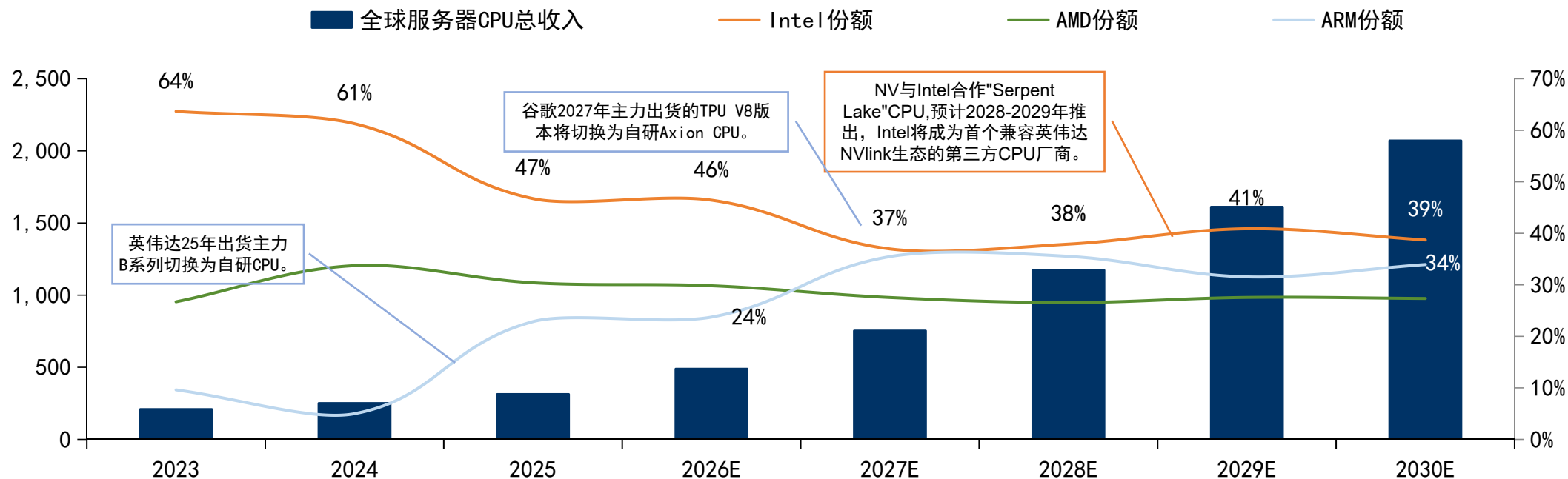
资料来源：各公司发布会，国信证券经济研究所整理

测算服务器CPU市场收入变化：X86生态主导，ARM快速崛起

•服务器CPU市场呈现X86双雄主导+ARM快速崛起的格局。根据GPU出货、柜内柜外划分、各场景CPU厂商份额假设，测算2026年全球服务器CPU市场x86份额约76%、ARM占比约24%（包含英伟达自研），到2030年x86份额约65%、ARM占比约35%。

- ① Intel：因产品迭代和制程问题，收入份额从21年的87%持续下滑，预计27年开始公司凭借大客户合作CPU和产能优势、份额相对稳定。
- ② AMD：2021年到25年凭借高端产品优势持续抢占服务器CPU市场份额，后续份额基本维持稳定。
- ③ ARM：伴随AI服务器兴起份额从个位数快速攀升。

图：全球服务器CPU总收入与份额（亿美元，%）

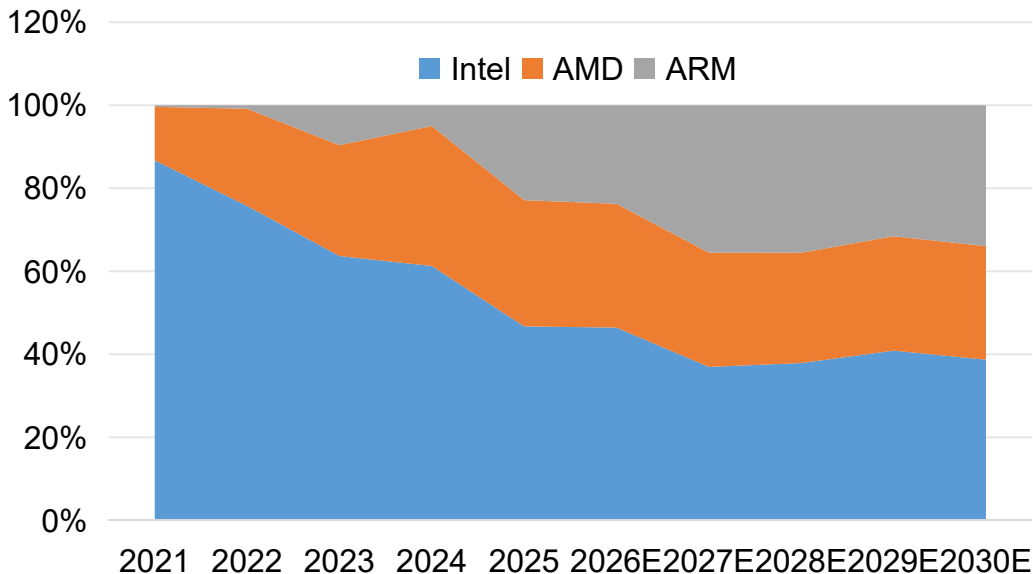


服务器市场格局演绎判断：中长期英特尔、AMD、ARM三分天下

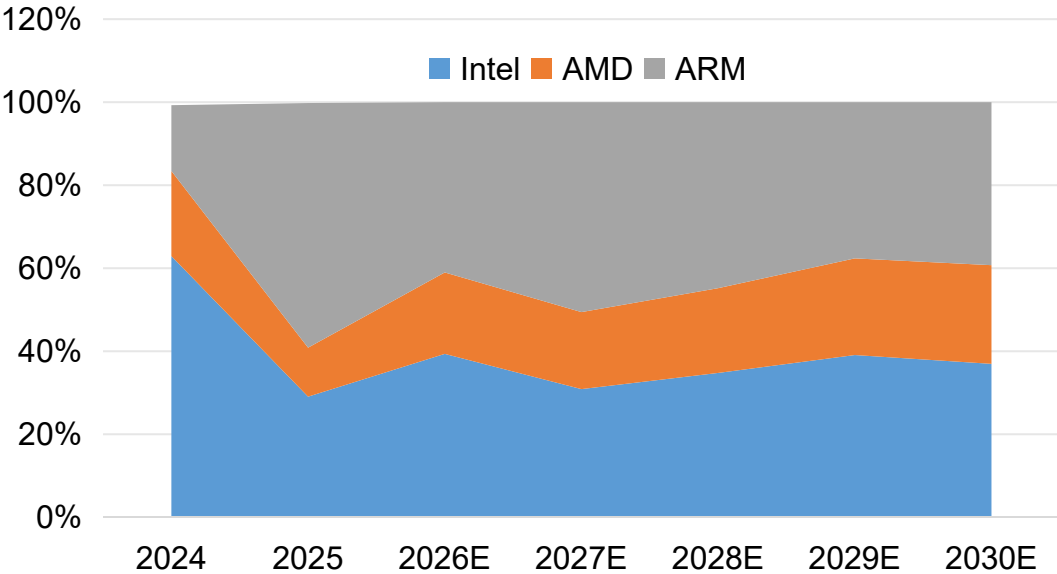


- 2025年ARM AI服务器CPU份额飙升至60%因英伟达GraceHopper架构芯片大规模出货，其CPU部分基于ARM架构自研。
- 后续行业格局迎来结构性修复：英伟达GPU市场份额逐步稀释，Grace架构带来的增量红利持续弱化。同时，英特尔、AMD的X86架构AI服务器CPU产品力持续迭代升级，如英特尔第六代至强芯片产能充分释放，叠加其与英伟达、谷歌联合研发的定制化CPU芯片陆续落地，X86阵营有望逐步收复流失份额。
- 中长期看，AI服务器CPU市场将形成英特尔、AMD、ARM三分天下的格局（根据GPU出货、柜内柜外划分、各场景CPU厂商份额和ASP假设测算，份额和ASP假设基于现有水平和后续生态合作、产品迭代计划给定），两类场景差异化优势明确：
 - ① 柜内AI算力场景：AWS、谷歌、微软等头部云厂商自研ARM架构CPU集中落地，ARM架构将成为柜内AI训练、推理场景的主流选择。
 - ② 柜外扩容场景：随着AI Agent应用快速普及，算力需求持续扩容，CPU与GPU配比显著提升，行业需依托外置机柜承载多进程算力需求。该场景下，产品性价比、与存量X86服务器生态的兼容性成为核心考量，英特尔、AMD为代表的X86阵营具备显著竞争优势。

图：整个CPU服务器收入份额变化（%）



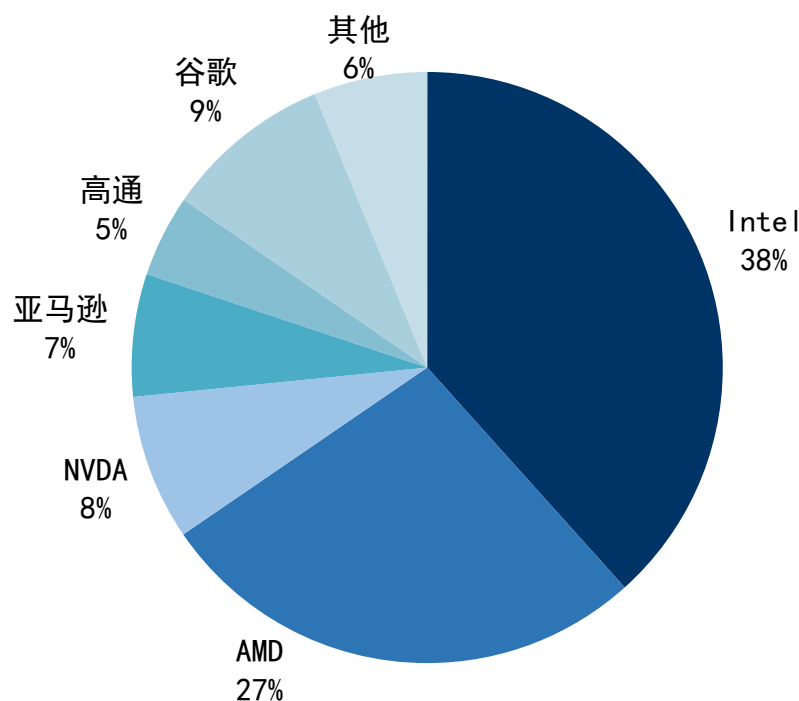
图：AI CPU服务器收入份额（%）



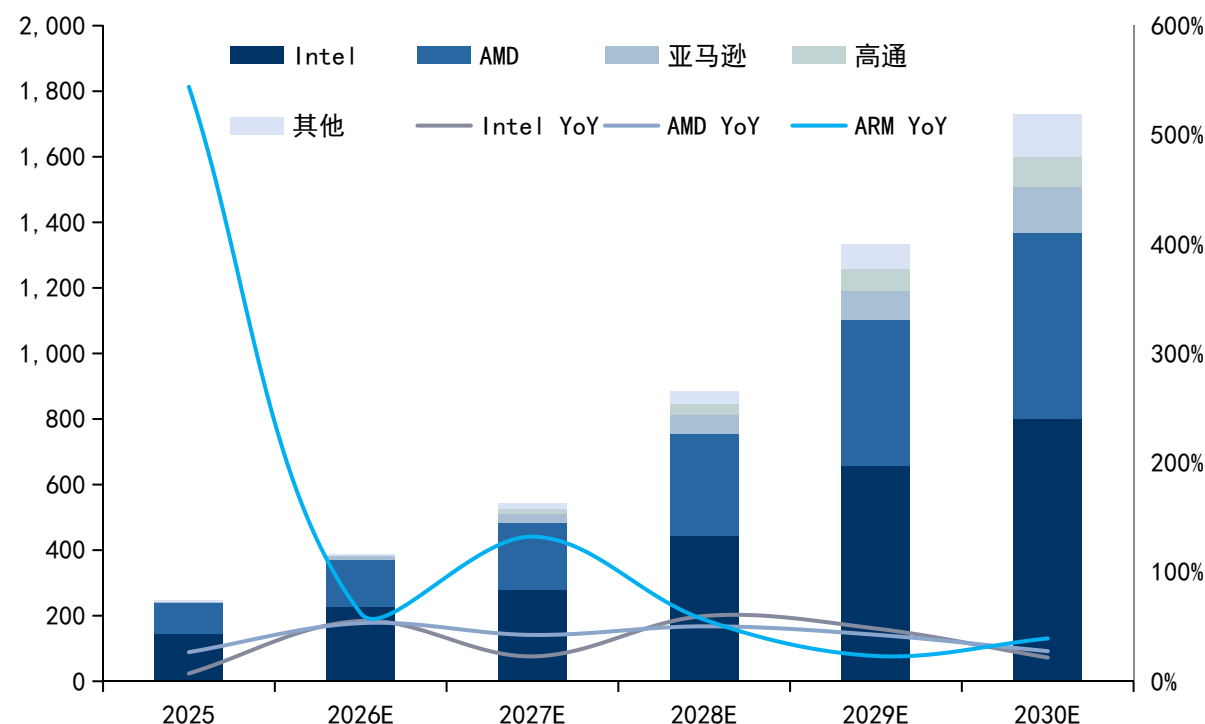
服务器CPU各公司收入份额变化

- **市场份额格局：**根据GPU出货、柜内柜外划分、各场景CPU厂商份额和ASP假设，测算2030年服务器CPU中，Intel仍以38%的收入、38%出货量位居第一，传统x86合计占比65%；Arm阵营从收入端看，**核心为谷歌9%、英伟达8%、亚马逊7%为核心构成，反映了CPU在AI时代与GPU生态紧密绑定的协同关系。**ARM阵营合计占比将达35%，成为不可忽视的力量。
- **收入规模与增速：**全球服务器CPU市场收入预计快速增长至2030年近2089亿美元，其中，Arm架构前期低基数同比增速显著，后续虽有回落但仍维持较高水平，体现出AI算力需求驱动下，**Arm阵营凭借云厂商自研与专用芯片优势快速渗透；**Intel与AMD的增速也处在较快水平，更多受益于整体市场扩容与高端化趋势（量价齐升）。

图：2030年CPU服务器收入份额（%）



图：CPU收入与同比增速（亿美元，%）



资料来源：IDC、Gartner预测，国信证券经济研究所整理测算

各公司主力服务器产品介绍

表：服务器产品介绍

厂商	细分市场	产品名称	发布时间	核数	价格	性能	代工厂/制程
Intel	AI服务器CPU核心产品	XEON 6 (Granite Rapids, 性能核)	2024年9月	72-128核	\$8,000-17,800+	支持AMX与MCR DIMM以提供极高单核吞吐量，专为AI推理设计。	Intel 3
	AI服务器CPU核心产品	XEON 5 (Emerald Rapids)	2023年末	最高64核	\$5,000-11,600+	具有较大L3缓存的存量市场过渡产品。	Intel 7
	传统服务器CPU核心	XEON 6+ E-Core(Clearwater Forest)	预计2026年H1	最高288核	\$5000+	采用背面供电技术，提供极致能效比的云原生横向扩展处理器。	Intel 18A
AMD	AI服务器CPU核心产品	EPYC 9005系列(Turin, Zen 5)	2024年10月	最高192核	\$10,000-14,000+	具备全512位AVX-512数据路径且IPC显著提升，为AI算力调度首选。	台积电 (3nm/4nm)
	AI服务器CPU核心产品	EPYC 9004X系列(Genoa-X, Zen 4)	2023年6月	最高96核	\$13,000-14,700	搭载3D V-Cache提供超大三级缓存，专为EDA与大型数据库优化。	台积电 (5nm)
	传统服务器CPU核心	EPYC 9005c系列(Turin Dense, Zen 5c)	2024年10月	最高192核	\$10,000-14,000+	通过较小L3缓存换取更高核心密度的云原生优化处理器。	台积电 (3nm/4nm)
英伟达	AI服务器CPU	Grace CPU	2023年	144核	捆绑GH200/GB200系统，无单卖价	提供极高内存带宽并可通过NVLink与GPU内存统一寻址的低功耗核心。	台积电 (4nm)
	AI服务器CPU	Vera CPU	2026年3月	88核	捆绑Vera Rubin系统，无单卖价	提供超高芯片间互连带宽，专为智能体AI和强化学习系统优化。	台积电 (3nm)
亚马逊	AI服务器CPU	Graviton5	2025年底	192核	AWS云服务按量/包年计费，未披露售价	核心延迟大幅降低且L3缓存倍增，主打重载AI推理和大型数据库。	台积电 (3nm)

- [01] Agent背景下CPU配比变化
- [02] 服务器CPU市场规模测算
- [03] CPU市场格局测算
- [04] Intel、AMD、ARM、高通业务与弹性比较
- [05] 风险提示

Intel、AMD、ARM、高通核心业务与商业模式对比

表：各家公司产品核心业务与商业模式介绍

维度	英特尔（Intel）	超威半导体（AMD）	安谋控股（ARM）	高通（Qualcomm）
核心业务模式	垂直整合制造（IDM），自产自销，以“设计+制造+交付”全流程为核心	无晶圆厂（Fabless），专注CPU/GPU芯片设计	轻资产IP授权模式，收取前期授权费+出货提成，边际成本极低	专注芯片定制设计与全栈优化，消费级CPU龙头，切入AI CPU赛道，主打低功耗、高性价比定制方案。
业务架构	① 数据中心AI（DCAI）：服务器CPU与GPU业务。核心产品是至强（Xeon）系列服务器CPU，以及Gaudi系列AI加速器芯片 ② 客户端计算（CCG）：消费级CPU业务，包括桌面级、移动级场景的酷睿系列处理器。 ③ 代工服务(IFS):目前97%收入为自身生产。	① 数据中心：服务器CPU与GPU业务，核心产品是EPYC（霄龙）系列CPU，以及Instinct系列AI GPU。 ② 客户端游戏：消费级CPU业务，现金流较好。核心产品包括面向桌面级、移动级的锐龙（Ryzen）系列。 ③ 嵌入式：工业级、端侧SoC产品。	① IP授权/Licensing：客户在获得架构使用权时，一次性支付的初始费用。核心产品包括自研的Neoverse CPU架构、Cortex-A系列处理器架构、配套的算子系统（CSS）； ② Royalty/版税：客户在基于ARM架构的芯片量产出货后，按照实际出货量支付提成。 ③ 自研服务器CPU：2026年新增业务。	① 半导体QCT：包括消费级与服务器CPU收入。核心为手机、汽车、IoT骁龙系列SoC芯片。服务器CPU业务25年底发布，预计26年底产生收入 ② 专利授权QTL：自研基础移动通讯技术、软硬件技专利授权。
26年服务器CPU市场收入份额测算	约 46%	约 30%	ARM生态收入份额为24%	无（2026年底出货）
技术架构基底	x86 架构	x86 架构	ARM 架构 IP授权	ARM 架构
核心服务器CPU产品系列	Xeon（至强）系列，部分配套 Gaudi AI 加速器	EPYC（霄龙）系列，部分配套 Instinct AI 加速器	Neoverse 架构 IP、AGI CPU 自研芯片	定制化 SD1 系列、AI200/AI250 推理芯片
主要制造合作方	自身晶圆厂，以 Intel7、Intel4、18A 工艺为主，部分外包给台积电	台积电，以 3nm、2nm 等先进工艺为主	台积电（仅自研芯片部分）	为台积电大客户（消费级CPU出货量），以 3nm、2nm 等先进工艺为主
核心客户群体	全球头部云服务商、企业级数据中心运营商	头部云服务商、头部互联网企业，GPU与CPU协同方案需求商	授权给自研芯片的头部云服务商、头部互联网企业	对功耗敏感的云服务商、边缘计算场景运营商

资料来源：公司官网，各公司业绩会，国信证券研究所整理。假设基于现有水平和后续生态合作、产品迭代计划

请务必阅读正文之后的免责声明及其项下所有

英特尔：作为龙头份额企稳，制造与生态仍有底牌

• 英特尔将深度受益于本轮AI浪潮，核心逻辑如下：

1) 服务器CPU稳居行业龙头，赛道高景气带来强业绩弹性

• 英特尔是全球服务器CPU领域绝对龙头，**2025年在全球服务器CPU市场份额维持半数以上**。x86生态兼容性强、企业存量客户深。从内部营收结构来看，**2025年服务器CPU业务占公司总营收约1/3**；受晶圆代工等业务持续亏损影响，该**板块利润贡献占比突破100%**。结合测算，预计到2030年服务器CPU业务营收占比将攀升至 56%。公司核心业务将充分享受板块增长红利，业绩弹性突出。

2) 市场份额触底企稳，产品高端化转型、构建差异化竞争壁垒

• 过去四年，受先进制程推进偏慢、产品迭代节奏滞后等因素影响，英特尔服务器 CPU市场份额持续下滑，这一态势在2025年终止、份额逐步企稳。**竞争策略层面，英特尔并未效仿AMD “CPU+GPU” 协同出货的模式，也因此与各大自研芯片厂商形成错位竞争；目前公司已深度绑定英伟达、谷歌等全球头部AI厂商，联合研发定制化 X86架构AI服务器芯片**（CRN、华尔街新闻报道），进一步巩固行业格局。

• 公司传统产品以中低端通用服务器CPU 、为主，通过优化产品结构，新一代至强6代芯片深度适配 AI推理等场景，下游需求增长强劲。高附加值高端芯片出货占比稳步提升，带动整体盈利质量改善。

3) IDM代工模式下产能优势与经营杠杆明显

• 依托IDM垂直整合模式，英特尔具备显著的产能壁垒。当前AI算力需求全面上行，公司自有产能有序释放、产能利用率维持高位，叠加18A先进制程良率快速爬坡，**内部生产端盈利能力持续优化**。旗下**晶圆代工业务（IFS）依托成熟产能与技术能力稳步拓展外部客户**，成为公司继服务器 CPU 之后的第二增长曲线。

• IDM属于典型重资产模式，前期高额资本开支与持续研发投入会阶段性压制利润兑现；但在行业景气上行周期，**营收规模扩张将充分释放固定成本下的经营杠杆效应**，进一步放大公司整体业绩弹性。

表：Intel业务拆分预测（亿美元/%）

Intel/亿美元	2025	2026E	2030E	备注
1) 消费CPU	322	312	340	五年CAGR 1%，下同
YoY	6%	-3%	3%	假设弱周期收入持平
2) DCAI	169	242	821	CAGR 37%
YoY	32%	43%	21%	假设主要为服务器CPU，按照量价拆分
AI服务器CPU收入	33	107	648	CAGR 81%
YoY	-8%	221%	26%	
出货量（百万颗）	1.0	2.6	14	CAGR 72%
ASP（美元）	3500	4130	4550	
传统服务器CPU收入	110	120	153	CAGR 7%
YoY	10%	9%	7%	
出货量（百万颗）	11	12	14	CAGR 5%
ASP（美元）	1000	1000	1093	
3) IFS	178	247	1087	CAGR 44%
YoY	2%	39%	36%	随着18A良率提升，假设对外收入占比从26年的96%提升至30年的60%
其他业务	36	26	30	
部门间抵消	-177	-237	-652	内部代工结算，假设与CPU收入正相关
YoY	55%	34%	17%	
总收入	529	589	1626	CAGR 25%
YoY	0%	11%	27%	

毛利率	35%	43%	63%	18-19年强周期为60%，叠加此轮代工业务优化成本
经调营业利润率	6%	18%	38%	18-19年强周期为35%
经调净利润	19	104	496	CAGR 92%
经调净利率	4%	16%	35%	18-19年收入增速10%时为30%

服务器CPU收入占比	27%	35%	57%	
服务器CPU经调营业利润率	27%	30%	42%	假设营业利润高于代工业务
服务器CPU 经调净利润	39	62	297	CAGR 50%，假设税率14%

资料来源：公司财报、国信证券经济研究所整理测算 *根据GPU出货、柜内柜外划分、各场景CPU厂商份额

英特尔IFS产线与客户合作情况

- 截至2026年5月，Intel在全球拥有15 座在产品圆厂，分布于10个地点，同时拥有7座封测厂。其中，8座300mm 晶圆厂构成了先进制程代工的核心产能，其余为 200mm 及以下成熟制程产线。
- Intel采用 “内部优先 + 选择性开放” 的客户策略，目前 70%-75%的芯片产量用于内部产品。
- 26Q1外部代工营收占比IFS总收入约3%，其中85%收入为先进封装业务。因为先进制程客户订单仍在爬坡阶段，当前收入主要来自成熟制程和先进封装业务。26Q1业绩会与谷歌达成多年合作，联合开发定制ASIC，同时加入SpaceX、xAI、特斯拉联合发起的Terafab项目。微软Maia2和亚马逊Trainium 3预计在26Q4进入大规模量产，届时 IFS 外部收入将显著增长。

表：英特尔产线情况			
工艺节点	工厂名称	分布区域	投产时间
18A(1.8nm)	Fab 52、D1X	美国	2025 年 10 月 (全面量产)
14A(1.4nm)	Fab 62、Fab 27 (俄亥俄一号)、D1X	美国	建设中，预计 2027 年 Q4 量产
			2026 年 4 月试产，预计 2028 年量产
			2026 年 Q2 (研发试产)
Intel 3(3nm)	Fab 34、Fab 42	美国、爱尔兰	2024 年
Intel 4(4nm)	Fab 34、Fab 42	爱尔兰、美国	2023 年
Intel 7 (10nm 增强)	Fab 28、Fab 12/22/32、Fab 24	以色列、爱尔兰、美国	1996-2008 年 (陆续投产，后升级)
Intel 16(16nm)	Fab 12/22	美国	2019 年 (工艺导入)
成熟制程 (22nm 及以下)	Fab 18、Fab 16、Fab 1/2、Fab 11X	以色列、美国	1980 年-2010年
先进封装	Fab 9、Fab 11X	美国、波兰	2024 、2025年改造完成

资料来源：公司官网，彭博新闻，国信证券经济研究所整理

表：英特尔代工业务合作客户情况				
客户名称	合作阶段	采用工艺 / 技术	产品类型	订单 / 预付款金额
微软	小批量生产	18A	下一代 Maia 2 AI 芯片	终身价值超 150 亿美元
亚马逊 AWS	量产准备	Intel 3+Foveros	Trainium 3 /Inferentia 3 AI 芯片	预付12 亿美元锁定 2027-2028 年产能
谷歌	产能锁定	EMIB-T 先进封装 + Xeon 6	TPU v6/v7 AI 芯片、云服务器 CPU	预付 15 亿美元 + 3-5 年 LTA
英伟达	产能锁定	Foveros/EMIB 先进封装	H200/B100 AI 芯片封装	预付 8 亿美元
高通	稳定量产	Intel 16	射频芯片、物联网芯片	长期供货协议
联发科	稳定量产	Intel 7	电视芯片、中端手机芯片	长期供货协议
博通	小批量生产	Intel 7	部分网络芯片	未披露
特斯拉 / SpaceX/xAI	战略签约	14A	AI 芯片、自动驾驶芯片	特斯拉投入 30 亿美元前期资金

资料来源：公司官网，公司公开发布会，国信证券经济研究所整理

AMD：AI纯度最高，CPU+GPU软硬件生态深度协同

• AMD聚焦高附加值AI算力赛道：

1) 四家同业厂商中AI业务暴露度最高，业绩弹性明显

• 2025年公司数据中心业务（服务器CPU+AI GPU）营收占总营收达50%，增长核心依托高端AI服务器产品、CPU+GPU异构算力协同能力，以及优质的头部客户结构。2025年AMD服务器CPU市场份额升至30%，作为服务器CPU龙头显著受益于板块增长。测算其服务器CPU业务未来五年CAGR可达43%、数据中心GPU未来五年收入CAGR可达52%。

2) CPU+GPU软硬件深度协同，产品性能领先、生态布局完善

① 公司核心壁垒在于CPU+GPU软硬件深度协同：联合Meta、OpenAI推出的Helios机架将EPYC Venice、MI450/MI455X、Pensando网络与开放互联标准结合，形成CPU+GPU+网络的机架级方案，也进一步拉高CPU搭载率。

② 高核心数和SMT线程密度领先。当前Turin已做到192核，Venice将进一步做到256核/512线程。

3) 可能存在产能限制，传统企业渗透较弱

① AMD为Fabless模式，芯片代工完全依赖台积电，先进制程产能成为核心瓶颈。业绩增长将主要依靠高产品均价叠加有限出货量实现。

② 企业级生态积淀不及英特尔，尤其对于对政企、电信等偏保守的客户群体，英特尔在系统稳定性、产品认证上的优势仍十分突出。

表：AMD业务拆分预测（亿美元/%）

亿美元	2025	2026E	2030E	备注
1) 数据中心	166	278	1118	
服务器CPU	96	146	566	五年CAGR 43%，下同
YoY	35%	52%	28%	按照前面章节量价拆分
传统服务器CPU出货量（百万颗）	8	9	12	CAGR 9%
传统服务器CPU ASP（美元）	1050	1050	1202	
AI服务器CPU出货量（百万颗）	0.3	1.2	8	CAGR 90%
AI服务器CPU ASP（美元）	4000	4500	5049	
数据中心GPU	67	133	552	CAGR 52%
YoY	33%	98%	25%	按照AMD GPU产品出货量与单价假设
2) 其他收入	180	183	253	CAGR 7%
YoY	36%	2%	5%	
消费CPU	106	114	140	主要为消费类产品，其中占比较高的消费电子弱周期、汽车与嵌入式场景仍高速增长，假设高个位数小幅增长
游戏CPU	39	30	48	
嵌入式CPU	35	39	65	
总收入	346	461	1371	CAGR 32%
YoY	34%	33%	22%	
毛利率	50%	53%	60%	随着产品高端化，毛利率过去十年稳健上行
经调营业利润率	22%	28%	39%	产品持续高端化，利润率上行，叠加AI强周期，预计突破原有25%-30%区间
经调净利润	68	120	525	CAGR 50%
经调净利率	20%	26%	37%	18-19年收入增速10%时为30%
净利润	43	97	452	CAGR 60%
净利率	13%	21%	33%	
服务器CPU收入占比	27%	29%	40%	
服务器CPU经调营业利润率	23%	33%	44%	考虑服务器CPU营业利润高于消费CPU
服务器CPU 经调净利润	19	41	214	CAGR 62% 假设税率14%

资料来源：公司财报、国信证券经济研究所整理测算 *根据GPU出货、柜内柜外划分、各场景CPU厂商份额和ASP假设

ARM：生态受益于云厂商定制CPU与高效诉求、自研AGI CPU开辟新增量



- ARM依托“IP授权+长期出货提成”的轻资产模式，主要受益于架构生态持续扩容，盈利具备长期稳健弹性。
- 1) 生态规模快速扩张，版税与授权双轮驱动增长
 - ① 版税提成业务：ARM架构服务器CPU收入份额将从2025年的23%（当前主要由英伟达贡献）提升至2030年的34%，**测算2025年到2030年ARM生态相关CPU收入复合增速达58%**，终端规模扩容将持续抬升版税收入。
 - ② IP授权业务：凭借架构开放、灵活度高、低功耗的核心优势，AWS Graviton、Google Axion、Microsoft Cobalt、NVIDIA Grace/Vera均采用了ARM生态。AI芯片技术壁垒更高、产品价值量显著提升，**推动单次IP授权费用高于传统品类**；叠加头部云厂商定制化芯片需求集中爆发，授权业务增长动能充足。
- 2) 自研AGI CPU开辟新增量，产能成为短期约束
- 公司2026年将落地面向AGI场景的自研CPU，产品主要面向端侧推理场景，不会冲击原有合作格局。行业层面整体算力芯片供给偏紧，**产能瓶颈或将成为短期出货限制**。
- 3) 依赖消费电子周期，目前AI业务权重偏低
- 公司业绩深度绑定消费电子行业周期，收入结构呈现明显偏科特征。测算今年ARM 服务器CPU相关收入占比仅17%，营收主力来自消费级CPU如手机、PC/笔记本、物联网（IoT）等消费级终端架构业务。尽管 AI驱动下服务器端ARM生态处于高速增长阶段，但目前该板块收入权重偏低，暂无法成为业绩压舱石。

表：ARM业务拆分预测（亿美元/%）

亿美元	2025	2026E	2030E	备注
1) CPU芯片	0	0	32	三年CAGR 84%
YoY	-	-	86%	假设弱周期收入持平
出货量（百万颗）	-	-	1.3	三年CAGR 69%
ASP（美元）	-	-	2500	
2) Licensing IP授权	18	25	71	五年CAGR 31% 下同
YoY	29%	33%	30%	
服务器CPU IP授权	2	5	23	CAGR 62%
YoY	42%	120%	40%	自研ARM服务器CPU玩家包括云巨头（AWS、谷歌、阿里云、微软）、专业芯片厂商
消费CPU IP授权	16	20	48	
YoY	27%	22%	25%	受益于Arm V9架构，版税率约为v8的2倍
3) Royalty版税	22	27	76	CAGR 29%
YoY	20%	25%	29%	
服务器CPU 版税	3	4	25	CAGR 58%
YoY	33%	61%	28%	假设按照ARM CPU生态收入3.5%抽成
消费CPU 版税	19	23	52	CAGR 22%
YoY	19%	20%	25%	按照ARM消费端（手机、IoT、PC）年均出货量3%*价15%*份额增长3%假设
总收入	40	52	179	CAGR 35%
YoY	24%	29%	37%	
毛利率	97%	98%	98%	普遍95%以上，高增速下维持在高水平
经调营业利润率	47%	47%	56%	普遍50%以上，高增速下提升至更高水平
经调净利润	17	21	83	CAGR 37%
经调净利率	43%	39%	52%	
净利润	8	9	57	CAGR 49%
净利率	20%	18%	32%	股权激励较高
服务器CPU收入占比	12%	17%	50%	
服务器CPU经调营业利润率	43%	43%	52%	需要一定自研投入，营业利润率略低于其他业务
服务器CPU 经调净利润	2	3	21	CAGR 35% ，假设税率14%

资料来源：公司财报、国信证券经济研究所整理测算 *根据GPU出货、柜内柜外划分、各场景CPU厂商份额和

高通：战略转型AI赛道，聚焦低功耗、低延迟的AI推理



高通高通全面布局AI领域，主要路径分为三层：**AI服务器CPU、AI推理加速器（AI200/AI250）、定制芯片开发**。作为服务器赛道新晋厂商，中长期具备差异化增长弹性。

1) 跨界技术复用，低功耗与定制化能力强

① **依托移动端ARM生态成熟的低功耗AI、推理加速技术跨界复用**，大幅降低服务器芯片研发成本，低功耗与LP DDR大容量路径差异化。这使其更适成本敏感的推理场景，而非训练峰值性能竞争。

② 定制化能力增强。收购Alphawave补强高速互连、收购Ventana补强RISC-V/CPU能力。

2) 把消费级CPU锁定的台积电先进产能、存储产能转向高价值CPU/加速器

作为消费级CPU龙头，台积电代工产能和存储产能优先级较高，当前**消费端的投入产出比持续恶化，转投服务器CPU产品将大幅提升单位产能价值、产品均价与毛利率**，对冲消费电子周期压力，从手机周期股转型AI算力基础设施股。

3) 可提供芯片、软件、场景优化一体化全栈方案

此前业务聚焦手机SoC（骁龙系列）、面向安卓高端旗舰机。全球5G基带市占55%+，长期垄断高端手机基带，射频前端/连接芯片。AI产品可配套手机WiFi、蓝牙、GPS、射频收发，提供一体化解决方案。

4) 已获头部AI科技厂商意向订单

与**META、字节、亚马逊均有意向合作订单**（26Q1业绩会与财联社等信源披露），定制化开发ASIC CPU或采购其CPU芯片。HUMAIN计划部署200MW的AI200/AI250。

5) 服务器生态与客户基础最弱，业绩弹性较弱

服务器生态积累薄弱，缺乏大规模量产落地经验与稳定订单，短期难以贡献业绩。同时AI推理赛道竞争激烈，业绩放量存在明显天花板。

表：高通业务拆分预测（亿美元/%）

亿美元	2025	2026E	2030E	备注
1) QCT芯片/硬件业务	384	375	684	五年CAGR 12%
YoY	16%	-2%	20%	
AI相关终端	0	10	174	四年CAGR 104%
YoY	-	-	35%	新增业务
服务器CPU/ASIC等外售	0	0	94	三年CAGR 84%
YoY	-	-	46%	
出货量（百万颗）	-	-	3.1	三年CAGR 44%
ASP（美元）	-	-	3000	端侧推理芯片为主
服务器CPU/ASIC定制	-	10	80	四年CAGR 68%
YoY	-	-	23%	未来三年将落地合作假设
手机等硬件终端	278	238	260	五年CAGR -1%，下同
YoY	12%	-14%	13%	消费电子弱周期，假设收入持平
汽车、IOT和其他终端	106	127	250	五年CAGR 19%
YoY	2%	39%	36%	汽车发展快速，按照过去6年增速均值假设
2) QTL专利授权业务	56	51	60	CAGR 1%
YoY	55%	34%	17%	按照过去6年情况假设收入持平
总收入	440	426	744	CAGR 11%
YoY	13%	-3%	19%	
毛利率	55%	54%	57%	过去十年毛利率在55%-60%左右
经调营业利润率	35%	33%	35%	过去20年稳定在25-35%区间，高收入增速周期假设上行
经调净利润	133	132	231	CAGR 12%
经调净利率	30%	31%	31%	
净利润	55	128	156	CAGR 23%
净利率	13%	30%	21%	
服务器CPU收入占比	0%	2%	23%	
服务器CPU经调营业利润率	-	20%	30%	需要一定自研投入，营业利润率略低于其他业务
服务器CPU经调净利润	0	2	45	四年CAGR 126%，假设税率14%

资料来源：公司财报、国信证券经济研究所整理测算 *根据GPU出货、柜内柜外划分、各场景CPU厂商份额和ASP

- 高通作为消费级CPU赛道龙头，其技术从“买Arm公版再做修改”升级到“只买Arm指令集、自研完整CPU微架构”。主要是2021年收购Nuvia，形成Oryon CPU路线，核心壁垒集中在前端宽度、缓存层级、乱序执行深度、访存系统与预取机制几个层面，而这些恰恰决定了高性能CPU的上限。
- 高通不仅是Arm IP下游客户，更升级为Arm生态“架构级定制者+跨终端复用者+潜在云侧定制方案商”。其优势在于：
 - ① 可以摆脱Arm公版内核迭代偏慢、性能折中明显的限制；
 - ② 能把手机、PC、汽车积累的低功耗高性能CPU/IP整合能力，迁移到服务器CPU与AI ASIC；
 - ③ 叠加NPU和Alphawave的SerDes/UCIe/Chiplet能力后，高通具备了从CPU核、NPU、互连到定制ASIC的相对完整拼图。

表：高通在消费级CPU赛道的能力演进

阶段	技术路径	代表情况	壁垒变化
2008年：早期自研阶段	基于Arm指令集自研	Scorpion架构	有架构基因，但未形成连续壁垒
2013:公版核修改阶段	Arm公版CPU核+高通自研GPU/NPU/ISP等	多代骁龙旗舰SoC	壁垒主要在SoC整合，不在CPU本体。CPU差异化弱，易被联发科追近
2021年至今：Nuvia/Oryon阶段	Arm ISA授权 + 自研完整微架构Oryon	骁龙X、骁龙8至尊版	CPU本体成为核心壁垒

资料来源：钛媒体、华尔街新闻、国信证券经济研究所整理

表：高通转型服务器CPU的基础条件

基础条件	已有资产	对服务器CPU的意义
高性能自研CPU核	Oryon / Nuvia Phoenix系能力	Oryon自研核已在手机/PC/汽车验证，可提升核心性能不够
低功耗SoC整合经验	手机、PC、汽车多年量产	适合AI头节点、推理控制面、云侧节能部署
AI/异构协同经验	Hexagon NPU（AI200/AI250）、GPU、ISP、系统软件	能把CPU定义为异构系统的一部分，而非孤立CPU
高速互连/Chiplet能力	Alphawave带来SerDes、UCIe、chiplet、定制硅	补齐数据中心必需的I/O与扩展短板

资料来源：钛媒体、华尔街新闻、国信证券经济研究所整理

Intel、AMD、ARM、高通核心业务测算对比



- 收入端：整体收入增速ARM（35%）、AMD（32%）收入CAGR领先，高通较慢；服务器CPU为增长核心，其占比整体收入体量来看，AMD与Intel领先，2030年四家该业务收入占比将达 57%/40%/50%/23%，成为收入主引擎。
- 盈利端：Intel经营杠杆显著、经调营业利率改善最明显（6%→38%），其余三家维持高盈利。平均利润贡献体量Intel、AMD较为领先。

表：各家公司产品2025年和2030年业绩收入与服务器CPU贡献对比

亿美元/%	英特尔（Intel）		超威半导体（AMD）		安谋控股（ARM）		高通（Qualcomm）	
	2025	2030E	2025	2030E	2025	2030E	2025	2030E
总收入	529	1626	346	1371	40	179	447	744
五年CAGR	25%		32%		35%		11%	
毛利率	35%	63%	50%	60%	97%	98%	55%	57%
经调营业利润率	6%	38%	22%	39%	47%	56%	35%	35%
经调净利润	19	496	68	525	17	83	133	231
五年CAGR	92%		50%		37%		12%	
经调净利率	4%	35%	20%	37%	43%	52%	30%	31%
服务器CPU收入	146	801	95	566	5	80	0	174
服务器CPU收入占比总收入	27%	57%	27%	40%	12%	50%	-	23%
服务器CPU营业利润率测算	27%	42%	23%	44%	43%	52%	-	30%
服务器CPU净利润测算	39	297	19	214	2	21	0	45
2027年经调整口径PE估值*	30x(扣掉IFS亏损)		35x		125x		20x	

资料来源：公司官网，各公司业绩会，国信证券
请务必阅读正文之后的免责声明及其项下所有

1. **AI进展不及预期风险：**AI智能体落地不及预期，AI商业化进展不及预期。
2. **传统CPU增长不及预期风险：**传统CPU受到消费电子周期影响恢复放缓、ARPU受竞争压制或新产品推广不及预期。
3. **竞争加剧风险：**CPU市场新玩家入局，带来产品和价格竞争激烈的风险。
4. **技术验证与失败风险：**新CPU产品迭代不及预期，先进制程良率爬坡不及预期。

国信证券投资评级			
投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后6到12个月内的相对市场表现，也即报告发布日后的6到12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数（HSI.HI）作为基准；美国市场以标普500指数（SPX.GI）或纳斯达克指数（IXIC.GI）为基准。	股票投资评级	优于大市	股价表现优于市场代表性指数10%以上
		中性	股价表现介于市场代表性指数±10%之间
		弱于大市	股价表现弱于市场代表性指数10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业投资评级	优于大市	行业指数表现优于市场代表性指数10%以上
		中性	行业指数表现介于市场代表性指数±10%之间
		弱于大市	行业指数表现弱于市场代表性指数10%以上

分析师承诺

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

重要声明

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

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

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

证券投资咨询业务的说明

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

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。



国信证券
GUOSEN SECURITIES

国信证券经济研究所

深圳

深圳市福田区福华一路125号国信金融大厦36层

邮编：518046 总机：0755-82130833

上海

上海浦东民生路1199弄证大五道口广场1号楼12楼

邮编：200135

北京

北京西城区金融大街兴盛街6号国信证券9层

邮编：100032