

# 高通研报：布局未来，低估当下

作者：大宇 @BTCdayu

发布时间：2026-05-04 23:03:43 UTC

来源：<https://x.com/BTCdayu/status/2051437646426370309>

文章链接：<https://x.com/i/article/2051437438049161216>

浏览 182573 · 点赞 149 · 转发 39 · 回复 4

26年4月29日发布

## 26年4月29日发布

高通今晚美股盘后，将发布新一期财报，这会是一个挺值得关注的事。一方面，目前端侧AI拐点已现，高通是不是主角？另一方面高通自身的手机业务缩水已经是共识，尤其是与苹果的多年虐恋终于分手影响很大。

而且从一季度财报发布后的情况来看，市场对手机业务下滑很敏感，盘后股价还是跌了 9%，而本季度从体感上也预计手机业务不会出现大的增长，很可能是数据下滑+指引进一步下降。

当前，23家券商目标价从 到216，跨度极宽，市场对高通的看法已经分成了两派：一派觉得这是手机业务缩水估值仍然过高；一派觉得这是被低估的端侧 AI 标的。

但研究过后，发现情况比这还要复杂有趣，甚至也有一些待判断的机会。

我关注高通是从推理的研究开始的，推理会成为未来AI世界的主流，而且很多的推理会在端侧发生，那**端侧模型到底解决什么问题？**高通又什么什么位置？现在市场对高通的悲观定价怎么看？

这是本人《AI投资地图》第11篇文章，长1.8万字，建议先转发收藏。

## 一、端侧 AI 到底是什么

### 1.1 对话不需要端侧模型

ChatGPT、Claude 等任何AI，几乎所有计算都在云端完成。手机只是一个收发文字的终端，云端的AI比任何能塞进手机的模型都更强。所以"端侧 AI"如果定义成"在手机上跑一个 LLM 和你对话"，云端能力碾压。

但端侧 AI 肯定不是这个意思。它解决的是**那些云端做不到、做不好、或者做了不划算的问题。**

### 1.2 一个真实事件：豆包 + 努比亚 M153 与 2025 年 12 月的"全网封堵"

2025 年 12 月 1 日，字节跳动发布"豆包手机助手技术预览版"，搭载于与中兴合作的努比亚 M153 工程机上。这台手机使用 Snapdragon 8 至尊版（即 Snapdragon 8 Elite Gen 5）芯片，将豆包大模型深度集成到 OS 底层——它能在不同 App 之间自动跳转，完成"打开美团搜索→对比价格→淘宝下单→截图发微信"这种**跨应用任务自动化**的工作。

3499 元、首批仅 3 万台，几小时内售罄，闲鱼溢价至 7000+。

## 12月3日开始，连锁反应出现：

- 微信检测到异常操作后被动下线，腾讯风控触发，导致登录失败
- 农行、建行等金融 App 启动针对 AI 屏幕共享的监测，要求关闭 AI 助手才能登录
- 阿里系（淘宝、闲鱼、大麦）开始拒绝“豆包手机”登录

12月5日，豆包发布公告，主动下线对金融类、游戏类等敏感场景的 AI 操作能力。

这件事的戏剧性在于：**它用一种非常现实的方式，证明了端侧 AI 的必要性。**让 AI Agent 跨越应用边界、读屏、自动化操作，是个真实的需求；但要做这件事，AI 必须**在本地完成所有敏感数据处理**——读取的屏幕内容、输入的密码、跨 App 抓取的信息，绝不能离开设备。否则，要么触发各 App 的安全风控（这次发生了），要么引发不可承受的隐私和监管问题。

努比亚 M153 的官方描述里有一句：“**数据处理采用端云协同架构，本地端侧模型处理高度隐私数据，云端传输实施 AES-256 加密保护。**”这一行字，清楚地说明了端侧模型在这个产品里要做什么。

第二代豆包 AI 手机预计 2026 年底出货（《智能涌现》/ 新浪科技口径）。

### 1.3 端侧 AI 真正在解决的几类问题

端侧 AI 不是“在手机上跑 LLM 对话”，而是云端 LLM 不擅长或者不可能处理的另一半智能任务：

**操作系统级智能。** Apple Intelligence 在 iOS 26 已经把多步骤的 agentic 行为放在本地运行——请求拆分、工具选择、响应生成都在端侧，只有需要联网的具体动作（查天气、查地图）才上云。延迟从云端的 2-3 秒降到本地的 200-400ms，对话才感觉自然。Apple 的 Foundation Models 框架（WWDC 2025 发布）首次向第三方 App 开放离线调用端侧模型的能力，使用的是约 30 亿参数的多语言多模态基础模型。Galaxy AI 同理——三星 Galaxy S26 系列的 Live Translate 在通话过程中本地翻译双方语言，对方甚至可以用座机。

**Agent 操作设备本身。**当 AI 要替你点外卖、订车、改设置、发消息时，它需要直接读屏、点按钮、跨 App。这些 UI 操作的反馈循环每一步都要 100ms 以内，云端做不了。这就是 2026 年所有手机厂商都在卷的“agentic AI”。豆包 + 努比亚 M153 是这个方向最戏剧性的中国本土实证；荣耀 YOYO Agent 已获中国信通院 L3 级智能体认证，支持 3000+ 场景自动执行；OPPO Find X8 支持 128K 端侧超长上下文。

**实时翻译和字幕。**这个场景如果走云端，延迟和断网都会让产品报废——电话不能等 2 秒再说话，地铁里没信号你也得让它工作。Gemini Nano 在这一类任务上的延迟控制在 100ms 以内。

**相机和影像。**现在的旗舰手机相机背后跑着一堆神经网络——HDR 合成、夜景降噪、人像分割、场景识别、动作捕捉。这些必须在按下快门后毫秒级完成，根本没有时间联网。

**隐私敏感场景。**面部解锁、健康数据分析、私人照片识别、密码管理。Apple Intelligence 的核心论点就是“数据不离开设备”，这是产品定义而非性能取舍。GDPR Article 25 的“设计隐私”原则和中国《个人信息保护法》从合规层面为端侧处理提供了制度推力。

**必须永远在线的可穿戴和眼镜。**智能眼镜上的 AI 助手不可能让用户每说一句话都等服务器。高通 AR1+ Gen 1 这块芯片专门为智能眼镜设计，关键升级是允许 1B 参数 AI 模型完全在本地运行。Ray-Ban Meta 用的是基础 AR1，下一代将不再需要联回 Meta 服务器。这一类形态注定是端侧的，因为眼镜的体积、电池、隐私顾虑都不允许走云端。

**机器人和汽车的安全关键场景。** 自动驾驶决策延迟超过 100ms 就可能撞车。机器人抓取物体、避开人类同样如此。这种场景必须本地，且对算力效率要求极高。

把这几类场景放在一起看，**端侧 AI 不是云端 LLM 的竞争者，而是补集**——它处理的是云端不擅长或者不可能处理的另一半智能任务。

## 1.4 云端成本的反向逼迫

除了延迟和隐私这两个常被讨论的维度，还有一个更少被提及但更重要的论点——**云端 AI 的网络出口成本 (egress cost) 在规模化之后会成为不可承受之重。**

具体来看：传统聊天（微信、Telegram）的上行速率通常低于 3 KB/s，但带有语音、图像、持续屏幕识别的 AI Agent 应用产生的是持续双向高带宽负载，动辄达到 10 KB/s 以上。AWS、Azure 等云服务商的网络出口数据费率约为一个中型应用每天产生推理输出，仅网络传输费每月就是 2,600-3,600。

当用户规模达到上亿级别时，这种纯云端模式将导致服务商面临不可承受的财务亏损。这是为何**必须将大量推理任务卸载 (Offload) 到端侧设备的底层经济逻辑**——不是技术品味问题，是商业模式问题。

这个论点在企业级部署中得到更强的实证。AI Vanguard 对欧洲物流运营商的部署分析显示：月度推理量 42M 的任务，纯云方案月度成本 ，平均延迟；混合方案（6,800，平均延迟降到 70ms——月度成本降低 52%，3 年 TCO 节省 45%，硬件投资 \$35,982 的回收期仅 4.9 个月。

更极端的零售案例：一家拥有 250 家门店的零售商部署店内计算机视觉，纯云方案因为每天 48M 图像、每月 2,880TB 数据上传产生的网络转移费就高达；边缘方案月度总成本只要 33,682，节省 89%。

把上面拼到一起，揭示了决策规则：**月度推理量在 500 万-1,000 万附近，是端侧/混合架构的盈亏平衡点。**低于此阈值，云端的变动成本模式更灵活；高于此阈值，端侧的固定资本投资可通过规模效应摊薄。

## 1.5 一个重要的技术趋势：3B 模型成为手机端侧"甜点"

2025-2026 年发生了一件事，让端侧 AI 从概念变成可行：**专业化小模型在垂直任务上已经够用，并且已经能在主流旗舰机上流畅跑起来。**

vivo AI 全球研究院院长周围在 2025 年开发者大会上的实测数据提供了关键实证：

- **7B 模型**：旗舰手机出词速度仅 10.9 token/s；Prefill（首词处理）阶段优化前仅 19.7 token/s——1,000 token 输入需要超过 50 秒首词延迟；持续推理功耗高达 2A+，是重载游戏（《原神》峰值约 1A）的两倍以上；内存占用 3.6GB，对 12GB RAM 主流旗舰机已是负担
- **3B 模型**：通过 2.8bit 量化后内存占用仅 2GB，极限出词速度超过 200 tokens/s，日常待机功耗 750mA
- **1B 模型**：可 24 小时在线实时监测和记忆

这一实证直接推动行业从 7B 向 3B 的迁移。**3B 参数模型成为 2025-2026 年手机端侧的"甜点"**——它既够用（基础对话、摘要、视觉理解、翻译），又能在主流旗舰机的内存、功耗、散热预算内持续运行。

NVIDIA 自家的研究团队在 2025 年发表的立场论文从另一个角度论证了同一件事：服务一个 7B 参数的 SLM 在延迟、能耗、FLOPs 上比 70-175B 的 LLM 便宜 10-30 倍，而对于 agent 任务这种"重复性、专业化、低多样性"的工作负载，小模型经过微调后表现已经不输大模型。MobileLLM 研究则揭示了一个反直觉的发现：在小参数规模下，架构设计比参数数量更重要。

这意味着什么？端侧 AI 不再是"等技术成熟"的故事，技术已经成熟了。现在的瓶颈在系统集成、模型优化、应用生态。这正好是高通这种平台型公司的主场。

## 1.6 一个隐性的差异化战场：低功耗持续推理

峰值算力竞赛是当前市场的显性竞争维度，但有一个更具差异化的隐性维度正在浮现：低功耗持续推理。

高通 Sensing Hub 的设计使得常开 AI (Always-on AI) 功耗 < 1mA。这一指标看似微小，但对比 vivo 实测的 7B 模型端侧运行功耗 2A+——这是"峰值推理"与"持续感知"之间的数量级差异。如果未来个人 AI Agent 需要"像呼吸一样持续存在"，它不可能每次呼吸都调用云端 API (成本、延迟、隐私均不可接受)，而必须在端侧以 mW 级功耗持续运行。

这个方向的竞争格局与峰值算力竞争截然不同：

- Apple 在能效比上保持领先 (M4/M5 在轻薄设备中维持极低功耗)
- 高通的差异化在于把移动端积累的功耗管理经验向 PC 和汽车场景迁移
- 骁龙 X2 Elite 宣称在同功耗下 CPU 性能提升 31%，NPU 性能提升 37% 且功耗降低 16%

低功耗决定了端侧 AI 能否从"功能演示"走向"日常刚需"。当前端侧 AI 在消费端的痛点不是模型不够大，而是电池不够耐用——本地运行 LLM 30 分钟可消耗智能手机 15-25% 的电池电量。如果谁能在下一代芯片中实现"全天候 Agent"而电池续航不显著下降，将创造真正的产品差异化。

## 1.7 市场规模有多大？

端侧市场分散在手机、PC、汽车、眼镜、机器人、IoT 一堆碎片化的终端形态里，每个形态都要单独的芯片、软件栈、客户工程支持。这种"碎片化但每片都需要深度集成"的市场结构，对哪类公司有利？大概率不是云端 GPU 公司——是已经在多终端形态都有商用平台的公司。

**2025 年规模数字非常集中：5 家研究机构落在 \$24-30B 区间。**其中 Edge AI 的数字是 \$26.14B，这对高通特别有价值，因为这个口径就是芯片级 + SoC NPU 的实际价值，不含整机也不含软件服务。这一细分口径预计 2025-2030 年 CAGR 17.6%，到 2030 年达到 **\$58.90B**。

当前，Edge AI 硬件的 5 大全球玩家之一（其余为华为、三星、苹果、联发科）。

## 1.8 中国市场是端侧 AI 全球最活跃的试验田

把视角从全球转到中国：

- **IDC 2025 年 12 月发布的预测：**2026 年中国新一代 AI 手机出货量将达到 1.47 亿台，同比增长 31.6%，**占整体市场 53%——首次过半。**这是端侧 AI 在中国从"早期采用"走向"主流标配"的关键拐点
- **2025 年上半年，**全球 23 家端侧 AI 芯片上市公司合计营收 2,937.76 亿元 (+12.3%)，归母净利润 622.03 亿元 (+9%)
- **中国端侧 AI 市场：**从 2018 到 2023 年，CAGR 高达 116.3%，2023 年规模已达 1,939 亿元
- **Counterpoint 数据：**2025 年全球支持生成式 AI 的智能手机 SoC 出货量同比增长 74%

中国市场的"端云结合"趋势尤为突出——手机厂商更多调用外部云侧大模型，同时将自身资源集中于端侧多模态模型的轻量化与深度推理能力建设。这恰恰是高通的主场——它是 **Android 阵营 GenAI SoC 市场份额 57% 的供应商**，全球 GenAI SoC 市场份额 35%（仅次于 Apple 的 46%）。

## 二、特斯拉不用高通，其他呢？

### 2.1 特斯拉是大户，但它不用

特斯拉的所有自动驾驶硬件——HW3、HW4、即将出货的 **HW5/AI5**——全部是特斯拉自研，台积电和三星代工，与高通无关。

- **HW3**：特斯拉与三星 14nm 联合设计，2019 年部署，单芯片约 144 TOPS
- **HW4 (FSD Computer 2)**：三星 7nm 制造，定制 SoC，单芯片约 73 TOPS（双芯片 146 TOPS）
- **HW5/AI5**：由 TSMC 和三星 双源代工。Musk 宣称单颗 AI5 算力相当于 Nvidia H100、比 AI4 快约 40 倍。首批样品 2026 年底，高量产推迟到 2027 年中后期——Cybercab Q2 2026 量产仍将使用 AI4。AI6 计划由 Samsung 2nm 制造；AI6.5 用 TSMC 2nm。

特斯拉走的是"垂直整合到底"的路线——硬件、软件、模型、数据、车都自己做。这条路的优势是端到端优化，可以让模型架构和芯片架构匹配；劣势是研发成本极高，且每个产品线都要重新走一遍。

2.2 但 2025 年 8 月发生了一件加固高通投资逻辑的事——特斯拉关闭了 Dojo

### 2025 年 8 月 7 日，特斯拉解散了 Dojo 团队。

- Dojo 团队负责人 Peter Bannon 离职
- 约 20 名核心工程师离职去了新创立的 DensityAI
- 剩余团队被重新分配到其他数据中心和算力项目
- **特斯拉转向 NVIDIA、AMD、Samsung 三方采购**：训练用 Nvidia 和 AMD GPU，AI5/AI6 自研芯片由三星德州工厂代工（\$165 亿、8 年合同）

这件事意味着什么？**连特斯拉自己都关闭了训练侧的全自研路线**。即使是垂直整合的标杆，也发现"训练芯片 + 推理芯片 + 软件 + 模型 + 车"这条路全部自研做到底是不经济的，必须回到产业分工。

### 2.3 全球绝大多数车企走不了特斯拉的路

**但全球绝大多数车企不可能像特斯拉那样自研**。汽车行业的现实是：

- 自研一颗符合车规、ASIL-D 安全等级的 SoC 大约要 5-7 年和 \$20-30 亿研发投入
- 算法团队、芯片团队、车端软件团队是三套独立的能力栈
- 每代车型 18-24 个月一个周期，跟不上自研芯片的迭代节奏

黄仁勋对车企自研的评论一针见血："你不是在跟公司竞争，你在跟团队竞争。世界上没有那么多能造出这种复杂产品的优秀团队。"

严格意义上的"全栈自研"（芯片 + 算法 + 软件 + 整车），全球**只有特斯拉一家车企做到**（2025 年交付 163.6 万辆，同比 -8.6%，连续两年下滑）。华为虽然不直接造车，但通过鸿蒙智行 + HI 模式深度参与车企，自研智驾芯

片 + 鸿蒙座舱 OS + ADS 算法的全栈方案；2025 年鸿蒙智行五界累计交付 58.91 万辆（同比 +32%），仅次于零跑（59.66 万）位列新势力第二；加上 HI 模式合作车型，华为智驾解决方案搭载车辆规模更大。

中国新势力中：**蔚来神玁 NX9031**（5nm，全球首颗量产 5nm 智驾芯片）首发于 2025 年 3 月 ET9 量产，**截至 2026 年 4 月累计出货超 20 万套**；2025 年 6 月蔚来芯片业务剥离独立成立“安徽神玁”，已获 22 亿元融资。

**小鹏图灵芯片**（7nm，单颗 750 TOPS）2025 年 6 月在 G7 Ultra 量产首发，已洽谈向大众供应；2026 款 P7+/G6/G9 已全系搭载（顶配 3 颗 2,250 TOPS）。

**理想自研马赫 100 芯片采用数据流架构**（行业唯一，不同于传统 GPGPU），5nm + Chiplet 工艺，**单颗 1,280 TOPS、双芯片 2,560 TOPS**，2026 Q2 在全新 L9 Livis 量产首发，相关论文已被 2026 ISCA Industry Track 录用——为汽车行业首家。

**比亚迪销量全球最大**（460 万辆/2025），**自研智驾芯片“玄机”已于 2026 年初流片成功**，算力 80-100 TOPS，2026 下半年装车，主要替代天神之眼 C 方案中外采的英伟达 Orin N 和地平线 J6 系列。

**小米玄戒 O1** 已出货超 100 万颗，雷军 4 月 27 日明确表态“后续自研芯片将用于小米汽车”，玄戒 O2 预计 2026 年 8 月发布、将推广至平板/汽车/电脑。

**海外车企没有任何一家在做车规级 AI 芯片自研。**

**这些自研力量并没有蚕食高通。**高通汽车业务的真正护城河是**智能座舱**。**2025 年中国座舱域控 SoC 市场，高通装机量份额超过 70%**（与非网/盖世汽车数据），是全球高端市场的事实标准。

但智驾芯片又有不同。2025 年 1-11 月中国乘用车智驾域控 SoC 市场（佐思汽研，总装机量 796.0 万颗），**英伟达 + 特斯拉合计占 63.5%**；剔除特斯拉自用，外部市场由英伟达主导。国产厂商（华为海思、地平线、黑芝麻智能等）合计 20.7%，同比 +4.6 个百分点。具体到城区 NOA 这一最有增长性的细分领域，**英伟达 + 华为 + 地平线合计 90%**，英伟达领先华为约 26 个百分点。**高通在中国智驾芯片市场至今未能进入主流玩家序列。**

这意味着：特斯拉、华为、蔚来、小鹏、理想自研芯片冲击的主要是英伟达 Orin/Thor 的智驾市场（以及部分蚕食地平线征程系列的中算力市场），不是高通的座舱市场。高通在汽车的真正战役不是“抢英伟达”，而是“用座舱优势作为锚点，通过舱驾一体（SA8775/SA8797）把英伟达从智驾领域挤出来”——这是高通和零跑、宝马合作 SA8797 双芯片方案的战略意义所在。

对高通真正的威胁，其实是联发科——联发科已与英伟达合作进军座舱市场，正面挑战高通。

## 2.4 高通的客户群有多大

特斯拉证明了“自动驾驶 + 智能座舱 + 端到端 AI”这件事在技术上是可行的。证明完之后，剩下 95% 的车企面对一个选择——要么自研（5 年+ \$20 亿+，且未必做得比特斯拉好），要么找一个能提供完整方案的伙伴。

高通能发力的区域：

**指标**

**数据**

**时点/来源**

**累计搭载任意 Snapdragon Digital Chassis 解决方案的车辆 3.5 亿+ 辆**

2024 年公司披露

**搭载集成 AI 的 Snapdragon Cockpit Platforms 的车辆 7,500 万+ 辆**

2025 年 6 月数据

**Snapdragon Ride (ADAS) SoC 累计出货近 100 万颗**

CES 2026 数据

**中国智能座舱域控制器市场份额75.2%**

2025 Q1, gasgoo 数据

**全球智能座舱/车载连接领域总份额~55%**

Counterpoint 数据

**设计中标 pipeline (design-win, 未出货 lifetime revenue) \$450 亿**

公司披露, 其中 ADAS 占 \$150 亿

**FY2026 Q1 汽车营收\$1.10B**

+15% YoY

**FY2025 全年汽车营收\$39.6 亿**

+36% YoY

**FY2029 汽车营收目标\$80 亿**

投资者日, 其中 80% 已被现有合同锁定

最后一行最关键。**\$80 亿 FY2029 目标里有 80% 已经被现有合同锁定**——这意味着这条增长曲线不是"故事", 是"合同"。这个可见性在所有半导体公司里都是少有的。具体客户清单(部分): 宝马、奔驰、大众集团、丰田、Stellantis、Hyundai、Volvo、雷诺, 加上中国的理想、零跑、小鹏、蔚来、吉利、长城、奇瑞、东风等。

## 2.5 新催化剂: Wayve

2026 年 3 月 10 日, 高通和英国自动驾驶初创公司 Wayve 宣布合作, 在 **Snapdragon Ride 平台上预集成端到端 AI 驾驶系统**。Wayve 的"AI Driver"软件栈采用端到端神经网络(与特斯拉 FSD 路径相似——直接从摄像头数据学习驾驶行为, 不依赖 HD map), 现在成为基于 Snapdragon Ride 的车企的可选 AI 驾驶软件方案。

紧接着 4 月 15 日, **QCOM Ventures + AMD + Arm 联合向 Wayve Series D 追加 \$6,000 万投资**。Wayve 总估值在前一轮(2026 年 2 月, 亿)达 86 亿, 投资方包括 NVIDIA、Microsoft、Uber、Mercedes-Benz、Nissan、Stellantis; 客户包括 Nissan (驾驶辅助+与 Uber 合作的伦敦/东京 robotaxi 项目)。

战略意义: **这是高通对抗 NVIDIA Drive Thor 的重要软件武器**。NVIDIA 在 ADAS 高端市场的优势不只是 Thor 算力, 更是 CUDA + Omniverse 全栈软件。高通过去缺少的就是端到端 AI 驾驶软件栈——现在通过 Wayve, 它把这块短板补上了, 且 Wayve 还是平台中立的(同时会跑在 NVIDIA Drive Hyperion 上)。

这意味着**特斯拉那条端到端路线, 正在被泛化成产业标准**: 不是每家车企都要自己复制特斯拉, 而是可以通过 Snapdragon Ride + Wayve AI Driver 这样的预集成方案, 获得"端到端 AI 驾驶"的能力。这反过来加固了高通的位置。

## 2.6 特斯拉反而印证了高通的位置

特斯拉是个孤例。它走自己的路, 与高通无竞争关系。但它的存在反而让其他车企更确信"汽车需要变成轮子上的智能终端", 而这件事 95% 的车企不可能每家都自研。**高通的汽车业务本质上是在收割"我做不了特斯拉, 但我也得变成软件定义汽车"那个市场的钱。**

## 三、高通在端侧 AI 中的位置

### 3.1 业务结构现状 (FY2026 Q1)

先看 FY2026 Q1 (2025 年 12 月 28 日 ending) 实际数字:

**总计\$12.25B+5%100%业务**

**营收**

**YoY**

**占比**

QCT 手机

\$7.82B

+3%

64%

QCT 汽车

\$1.10B

+15%

9%

QCT IoT

\$1.69B

+9%

14%

QTL 许可

\$1.59B

+4%

13%

来源: 公司 8-K 财报披露。Non-GAAP EPS \$3.50 (创纪录), QCT EBT 利润率 31%, QTL EBT 利润率 77%。

需要意识到的几件事:

第一, **手机仍然是最大的业务, 占比近三分之二**。任何关于"高通已经成功转型"的判断都为时过早。手机的命运还是高通的命运。

第二, **QTL 是利润奶牛**。77% 的 EBT 利润率意味着每收 就有 1.23B 流到税前利润。这部分钱来自对全球 5G 标准必要专利 (SEP) 的许可——华为持有的 5G 专利数量是高通的 6.5 倍, 但高通 QTL 授权收入是华为的 13 倍。这说明高通的专利集中在设备端无线空口 (波束成形、调制、频谱效率) 等核心领域, 技术上难以绕开。这是高通最稳定、最高利润、也最被市场低估的业务。

第三, **汽车 + IoT 合计 \$2.79B, 已经接近手机的 36%**。在 FY2024 时这个比例还是 25% 上下。多元化在数字上是真的发生了。

放到 FY2025 全年来看:

- **总营收 \$44.28B, +13.7% YoY**
- **Non-GAAP EPS \$12.03, +18% YoY**

- **自由现金流 \$12.8B** (创纪录)
- **non-Apple QCT 营收增长 +18% YoY** (这个数字很重要——剔除 Apple 后核心业务的增长水平)
- **资本回报**: FY2025 回购 股息 3.8B = \$12.6B, 相当于 100% 自由现金流返还股东
- **2026 年 3 月新批 \$200 亿股票回购授权** (约市值 14%), 季度股息从提升到 0.92

### 3.2 七个终端形态的产品矩阵

端侧 AI 的硬件竞赛是以 NPU 算力 (TOPS, 每秒万亿次操作) 为核心指标的。看 2026 年的实际产品:

**手机端: Snapdragon 8 Elite Gen 5** 2025 年 9 月发布, 是当前所有 2026 年安卓旗舰的标配, 算力约 100 TOPS (INT8 行业估算), 能在端侧以 220 tokens/s 速度运行 30 亿参数 LLM、支持 32K 上下文。覆盖客户: 荣耀、IQOO、努比亚、一加、OPPO、POCO、realme、红米、红魔、华硕 ROG、三星、索尼、vivo、小米、ZTE。

**PC 端: Snapdragon X2 Elite Extreme** 2025 年 9 月发布, 2026 年 H1 出货: 18 核, 单双核加速 5.0 GHz (首款 ARM Windows 处理器达此频率), 全核 4.4 GHz, CES 2026 发布了 **Snapdragon X2 Plus**——把 80 TOPS 的能力下沉到 \$800 价位的主流笔记本。Geekbench AI 88,615 分, 远超 Apple M4 的 52,193 和 Intel Core Ultra X9 的 15,628)。预计 2026 年 150+ 设计出货。

**汽车: Snapdragon Ride (ADAS) 和 Snapdragon Cockpit (座舱)** 平台覆盖从入门到 L4 自动驾驶的全段位。**Ride Flex (SA8775)** 进一步把座舱和 ADAS 整合到一颗芯片上。**Ride Elite / Cockpit Elite (SA8797)**——Qualcomm 官方在 Ride Elite 发布时主动淡化具体 TOPS 数字, 强调"NPU 性能较前代座舱平台提升 12 倍"和"配备 transformer 加速器、向量引擎、混合精度支持", 行业分析 (AnandTech 等) 估算其 dense INT8 算力约 320 TOPS。2026 年 BMW iX3 是 Snapdragon Ride Pilot 首发车型。

**XR / 智能眼镜: AR1+ Gen 1** 比基础版功耗低 7%, 体积缩小 28%, 可在本地运行最大 1B 参数 AI 模型。智能眼镜市场目前是高通的天下: Ray-Ban Meta (基础 AR1 Gen 1)、Oakley Meta、Meta Ray-Ban Display 全部使用高通平台, 2025 年 6 月 AWE 发布的 AR1+ Gen 1 业内普遍认为对应 Meta 下一代高端眼镜 Hypernova; 三星 2026 H2 上市的 AI 智能眼镜采用 AR1+ Gen 1; Snap Spectacles 也用高通, Snap 子公司 Specs 与高通签署多年战略合作协议。

**但 Meta 同时在准备替代方案**: 2023 年 11 月起, Meta 与联发科已正式合作共同开发**独家定制 AR 眼镜 SoC**——Meta 主导设计、联发科代工流片, 类似 Google Pixel Tensor 的"半自研"模式 (Meta 之前完全自研的 FAST 团队已经解散)。**2024 年 Meta Orion AR 眼镜原型机已用上这颗 Meta + 联发科定制芯片**。这意味着 Meta 在真正的 AR 形态 (带显示、空间感知) 走向量产时, 会逐步降低对高通的依赖。智能眼镜不会是高通的独占市场。

**机器人**: CES 2026 推出 **Dragonwing IQ10** (18-core CPU), 用于工业 AMR 和全尺寸人形机器人。合作伙伴包括 **Figure** (人形机器人头部)、Booster、VinMotion、Kuka Robotics、Advantech。VinMotion Motion 2 用 IQ9 系列。

**数据中心: AI200** (2026 商用) 和 **AI250** (2027 商用) 是高通进入数据中心 AI 推理市场的产品。AI200 每张 PCIe 卡支持 768GB LPDDR; AI250 用 near-memory computing 架构, 宣称内存带宽提升 10 倍。沙特 PIF (公共投资基金) 旗下的国家级 AI 公司 HUMAIN 计划 2026 年起部署 200 兆瓦的 AI200/AI250 系统——Wolfe Research 估算这个合同价值约 \$10 亿, **Wells Fargo 估算最高可达 \$20 亿**。

**IoT / 工业：Dragonwing IQ 系列覆盖从 1 TOPS 到 700 TOPS 的算力区间。2025 年高通完成了两次关键收购整合到这个生态：**

- **Edge Impulse**（2025 年 3 月收购）：边缘 AI 开发平台，25 万+ 活跃开发者
- **Arduino**（2025 年 10 月收购）：全球最大开源创客社区，3,000-3,200 万用户

加上 Snapdragon AI Hub（250+ 预优化模型），高通的开发者生态总规模达数千万级别。

### 3.3 这个产品矩阵的战略意义

**高通是目前唯一一家在所有七种主要终端形态——手机、PC、汽车、智能眼镜、可穿戴、机器人、数据中心推理——都有商用 NPU 平台的公司。**

苹果的产品矩阵只覆盖自己的设备，不外卖。英伟达在数据中心和机器人有强势地位，但手机、PC、汽车都不是它的主场。联发科主要在手机和 PC 入门段位。三星只做手机和家电。Intel 在 PC 占据存量市场但 NPU 性能落后。

这种全终端覆盖的意义不只是"业务多元化"，而是**软件栈和开发者生态可以复用**。Qualcomm 自定义的 Oryon CPU 核心已经从手机用到 PC 用到汽车——同一个 IP 跨产品线复用，给 OEM 提供端到端解决方案，给高通提供巨大的内部杠杆。开发者为 Snapdragon 8 Elite Gen 5 优化的端侧 AI 模型，在 X2 Elite 笔记本和 Snapdragon Ride 汽车上都能跑。

这是 2010 年代英伟达靠 CUDA 在数据中心做到的事——硬件标准化 + 软件复用 + 生态绑定。高通在端侧的位置类似（虽然规模和盈利能力差很远）。

### 3.4 护城河评估

**护城河维度**

**强度**

**核心支撑**

**主要威胁**

**专利 (QTL) ★★★★★**

5G eMBB 领域约 20% 声明专利；EBT 利润率 77%

华为协议已到期；被许可方选择仅授权蜂窝 SEP；监管压力

**芯片平台整合★★★★☆**

Oryon CPU 全自研，Hexagon NPU6 领先，跨平台复用

联发科价格竞争；Apple 自研；x86 生态壁垒

**OEM/运营商关系★★★★☆**

全球 20+ 旗舰 OEM 合作，Android 旗舰"默认选择"

OEM 垂直整合（Samsung Exynos、Google Tensor）

**边缘 AI 能力★★★★☆**

100 TOPS 移动 NPU、80 TOPS PC NPU、舱驾融合

NVIDIA 数据中心优势、专用 AI 芯片竞争

**软件 / 开发者生态★★★★☆☆**

Qualcomm AI Hub、Edge Impulse、Arduino

工具链碎片化、CUDA 生态锁定、文档滞后

**这个评估的结论是：**高通的专利护城河最深但面临天花板，芯片平台护城河正在加固但存在性能差距，软件生态护城河最弱但是估值弹性最大的变量。

具体表现为：

- **正面反馈：**高通的 INT8 量化能力极为先进；Qualcomm AI Hub 大幅简化部署流程，"从 Hugging Face 模型到手机运行不到一小时"
- **负面反馈：**SNPE/QNN/QAIRT/Genie 等多套 SDK 并存，工具链碎片化；文档更新滞后；SDK 优先面向企业客户导致社区开发者体验不佳
- **与 NVIDIA CUDA 对比：**CUDA 已形成网络效应——数千名熟悉 CUDA 编程模型的开发者、主流 AI 框架最成熟的 GPU 后端、围绕 CUDA 构建的模型和工具生态。高通距离 CUDA 级别的生态锁定仍有相当距离

高通正在通过几个动作补这块短板：

- 支持行业标准框架（PyTorch、TensorFlow、ONNX）
- 集成 vLLM 推理引擎
- 开源 AIMET 量化工具和 Hexagon-MLIR 编译器堆栈
- 收购 Edge Impulse + Arduino，在工业 IoT 和创客社区两端布局
- 2026 年 2 月在印度承诺 \$1.5 亿 AI 风投基金，触达新兴市场开发者

**若软件生态构建成功，将突破手机芯片的周期性估值陷阱，使估值逻辑从"硬件周期股"转变为"端侧 AI 基础设施平台"。**

### 3.5 客户集中度需要正视

- **来自 Apple、Samsung、小米的收入各超 10%**——三家合计可能超过 35-40%
- **中国大陆+香港占总营收 45.9%**（FY2025 10-K，从 FY2023 的 37% 上升）
- **中国市场绝对营收约 \$17.8B**

高通对几个关键客户的依赖度，在地缘政治持续紧张的背景下，是结构性风险。

---

## 四、几个关键的业务变量——逐项拆解

### 4.1 Apple modem 切换

与苹果的分手，是高通未来两年最大的利空，不过股价也一定程度上反映了这一点。

时间线清楚：

- **C1（2025）：**iPhone 16e 首发，仅 sub-6 GHz，无 mmWave
- **C1X（2025 秋）：**iPhone Air 用，仍仅 sub-6
- **C2 Ganymede（2026 秋）：**iPhone 18 Pro 全系，加入 mmWave 支持

- **C3 Prometheus** (2027) : 与高通授权协议同年到期

CEO Cristiano Amon 多次在公开场合给出的精确节奏:

- 2025 年 (iPhone 17 系列) : 高通供应约 **70%** 的 iPhone modem
- 2026 年 (iPhone 18 系列) : 降到 **20%**
- 2027 年 (iPhone 19 系列) : **0%**

### 数量上的影响

- 来自 Apple 的 modem 销售: \$57-59 亿/年
- 来自 RF 组件和其他子系统: \$16-19 亿/年
- 合计: \$73-78 亿/年

完全退出将造成约 **16-18% 总营收的缺口**, 全面体现于2027财年, 但2026 已开始承压, 预计影响 \$20-30 亿 modem 收入。

但有几件事需要拆开看:

**第一, 专利授权 (QTL) 和 modem 供应是两件事。** 即便 Apple 完全不用高通 modem, 2027 年 3 月之前 Apple 仍要付 5G 标准必要专利费——这部分进 QTL, 是 77% EBT 利润率的部分。2027 年之后协议续不续是开放问题, 但即便 Apple 拒绝, 高通可以走法庭路线 (2017-2019 年已走过一次, 最后是 Apple 和解付钱)。

**第二, 高通的 modem 卖给 Apple 的毛利率本来就低。** Apple 长期是高通最难谈判的客户, 所以 modem 的实际边际利润流失, 比 \$73-78 亿营收的影响小一些。

**第三, Apple modem 切换的时间点错开。** 2026 年只是 Pro 系列切, 标准版 iPhone 18 仍可能用 Snapdragon X80 modem。这意味着 calendar 2026 财年 (高通 FY2027) 才是真正的拐点。

## 4.2 一个催化剂: OpenAI 智能手机芯片合作

4 月 27 日才爆出。TF International Securities 分析师 Ming-Chi Kuo 在 X 上披露: **OpenAI 与 Qualcomm 和 MediaTek 共同开发智能手机处理器, 由 Luxshare 独家代工, 预计 2028 年量产。**

具体细节:

- **目标年出货量 3-4 亿台**——这个量级与 iPhone 当前出货量相当 (实际兑现难度很大, OpenAI 几乎没有手机生态基础, 需要从零搭建消费品供应链和分销网络)
- **设备形态:** app-free 设计, AI Agent 替代传统应用图标
- **架构:** 端云协同——本地处理用户实时状态、内存层级管理、低延迟响应; 高强度任务上云
- **2026 年底-2027 Q1 完成芯片规格定型**

战略意义有几层:

**一是直接补 Apple 流失。** 如果 2028 年这个项目真出货 3 亿台, 对应的处理器订单可能在 \$50-100 亿/年级别——这相当于直接对冲 Apple modem 退出的 \$73-78 亿缺口。这是市场目前还没有定价的反向催化剂。

**二是 OpenAI 对架构的要求与高通核心能力高度契合：**低功耗持续推理、agentic AI 处理、端侧多模态理解、内存层级管理——这些正是 Hexagon NPU6 和 Sensing Hub 设计的方向。

**三是供应链多元化保险。**OpenAI 选择 Qualcomm + MediaTek 双源（而非单一供应商），降低 Qualcomm 依赖，但同时也意味着高通至少能拿到一半订单。

**风险也要看清：**这个合作目前没有 OpenAI 官方确认，2028 年量产时间表意味着至少还有 2 年时间走完芯片设计、流片、车厂认证。OpenAI 还要解决一个根本问题——没有 Google Play Services 生态，AI Agent 怎么访问应用功能？这件事在 OpenAI 那边的不确定性，目前比在高通那边的不确定性大。

市场对 4 月 27 日的反应（QCOM 单日 +12%，盘中触及 \$160）说明投资者也意识到这一点。

### 4.3 内存周期：最被低估的短期变量

这是市场目前真正担心的问题。

高通 CEO 在 Q1 FY2026 财报电话会议上说：**整个财年的手机出货规模由内存供应决定。**DRAM 制造商正在把产能从手机内存转向 HBM 用于 AI 数据中心，分析师警告 DRAM 短缺可能持续到 2028 年。

具体到数字：**Q2 FY2026 指引营收 \$10.2-11.0B**（vs Q1 实际），手机营收预计约 6.0B，比 Q1 的 \$7.8B 下降 23%。指引的主要原因就是手机厂商客户在内存短缺的不确定性下削减库存。

**这是一个真实的、但本质上短期的、且对高通是双刃剑的问题。**

- **短期看：**手机厂商砍单 → 高通手机芯片营收下降 → calendar 2026 业绩压力
- **中期看：**内存价格上涨可能推高高通 BOM 成本（虽然高通自己不直接买内存，但客户接受涨价能力下降）
- **反向看：**DRAM 转向 HBM 的根本原因是数据中心 AI 需求爆发，这恰恰是高通 AI200/AI250 即将切入的市场

最坏情况下，**手机业务季度营收**回落到 \$60 亿出头（指引下沿）水平，FY2026 全年手机营收同比可能下滑 10-15%。基础情况下，FY2027 应该恢复正常节奏。这个周期波动不是结构性问题。

### 4.4 三星 Galaxy S26 开始不用骁龙

Samsung Galaxy S25 系列在 2025 年是全球 100% 使用高通 Snapdragon。但 **2026 年 2 月发布的 Galaxy S26 系列改变了规则：**

- **Galaxy S26 Ultra：**全球 100% 使用 Snapdragon 8 Elite Gen 5 for Galaxy
- **Galaxy S26 / S26+：**
- 北美（美国 + 加拿大）、中国、日本：使用 Snapdragon 8 Elite Gen 5
- **欧洲、印度、韩国及其他市场：**使用三星自研 Exynos 2600（首款 2nm GAA 智能手机芯片）

测试数据：Snapdragon 版本续航 9 小时 26 分钟，比 Exynos 版本（6 小时 48 分钟）**长 39%**（多 2 小时 38 分钟），单核 CPU 性能高 19%。

**对高通的意义：**高通在三星旗舰系列的份额，从 S25 时代的 100%，**降到 S26 时代估计的 60-70%**（Ultra 100% + S26/S26+ 部分市场用高通）。这是 2026 年高通手机业务一个新增的逆风——**它和 Apple 切换是叠加的。**

更长期看，Samsung Exynos 2600 的成功（即使只是部分成功）意味着三星正在重新建立自己的芯片设计能力。如果未来几年三星扩大 Exynos 在 S 系列的覆盖比例，高通会面临持续的份额损失。

#### 4.5 联发科的真实威胁：高端旗舰份额

看一组对比数据，2026 年 OPPO Find X9 系列同时推出搭载高通 Snapdragon 8 Elite Gen 5 的 Ultra 版本和搭载联发科（MediaTek）Dimensity 9500 的 Pro 版本——同一品牌、同一时间、配置接近，相当于一次纯硅之争。结果是联发科已经在 GPU 能效、ray tracing、温控上有优势，单核 CPU 高通仍领先。

更早的市场份额数据：MediaTek 在 2026 年预计以 **34% 份额**保持智能手机 SoC 出货量第一，Qualcomm **24.7%**，Samsung 升至 **12.1%**（Counterpoint 数据）。

这个 close-to-parity 的局面意味着什么？

意味着**高通在高端旗舰的定价权可能被压缩**。过去高通靠"性能旗舰"溢价，OEM 不得不付，因为联发科只是中端选项。现在 OEM 可以让两家芯片厂商比价。

高通在高端旗舰仍然是首选，但联发科正在持续扩展存在感。这个变量短期看不出来，但对 QCT 手机的长期 ASP（平均售价）和利润率是真实的压力。

#### 4.6 汽车业务：最实在的增长引擎

把数字放在一起看节奏：

##### 80% 已被现有合同锁定指标

##### 数据

##### 来源

FY2025 全年汽车营收

\$39.6 亿

+36% YoY

FY2026 Q1 汽车营收

亿 (1.1B)

+15% YoY，连续两季超 \$10 亿

Q2 FY2026 指引

>35% YoY 增长

财报电话会

Design-win pipeline

\$450 亿

公司披露，其中 ADAS 占 \$150 亿

FY2029 营收目标

\$80 亿

如果 FY2029 真做到 亿汽车营收（全年39.6 亿），就是 4 年大致翻 1 倍。从一个低基数起步、且 80% 已签约，这个增速并不夸张。

#### 4.7 PC 业务：势头好但份额仍小

2025 年 AI PC 占笔记本出货 44% (IDC 预测 2027 年达 60%)，但 ARM 在 PC 整体营收只占 13% (x86 87%)。Snapdragon X 系列从 2024 年中第一次出货，2025 年扩展产品线，2026 年 H1 X2 Elite 出货。

份额的实际曲线：

- 2024 Q3：仅 0.8% (约 72 万台) ——市场上很多人忘了起步有多惨
- 2024 年 12 月：在美国 \$800+ 价位 Windows 笔记本零售市场达到 10%
- 2026 年预计：150+ 设计商业化
- 2027 年底 (IDC 预测)：ARM 笔记本可能占 Windows 市场 20-25%

最关键的两个观察：

- **Snapdragon X2 Elite 预计 2026 年覆盖 150+ 设计**，比第一代显著扩张
- **X2 Plus 把 80 TOPS NPU 下沉到 \$800 主流笔记本**——这是从“高端尝鲜”走向“主流标配”的关键一步

但 ARM PC 的关键瓶颈不在性能而在软件兼容。Microsoft 的 Prism 模拟层在改善 (2025 年底新增 AVX/AVX2 指令集支持)，但游戏 (约 10% 因反作弊软件不兼容)、专业软件 (Adobe Premiere Pro 等)、企业 IT 部署的迁移成本仍然存在。

**FY2029 PC 业务目标 \$40 亿** (投资者日数字)。当前营收基数很小，要做到 \$40 亿要求 2026-2029 年保持极高增速。

#### 4.8 数据中心：野心大，落地存疑

高通 2025 年 10 月公布 AI200/AI250：

- **AI200**：2026 年商用，每 PCIe 卡 768GB LPDDR
- **AI250**：2027 年商用，near-memory computing 架构，宣称内存带宽提升 10 倍

挑战 NVIDIA 在 AI 推理 (不是训练) 市场的主导地位，主打低 TCO 和能效。

这件事的特点：

**第一，重要客户已经签了，但高度集中**：HUMAIN 是沙特政府背书的 AI 创业公司，200MW 部署起步，目标是为沙特和全球提供 AI 推理服务。这一个客户就足以让 AI200 项目跑起来，但也意味着客户多元化是后面要解决的问题。

**第二，Alphawave 收购的意义比表面看更大**。\$2.4B 收购完成于 2025 年 12 月，Alphawave 的 PAM4 SerDes 和 chiplet 互连 IP 是数据中心 AI 系统的关键组件，前 Alphawave CEO Tony Pialis 现在领导 Qualcomm 数据中心业务。这是一次“买技术 + 买团队 + 买市场入场券”的整合。

**第三，机会和门槛都很大**。NVIDIA 在 AI 数据中心 80%+ 份额。即便高通在推理细分市场 (不和 H100/B200 训练正面冲突) 能拿到 5-10% 份额，几年后也可能是 \$20-40 亿/年的业务。但若执行不到位，这个数字可能是几亿。

数据中心业务对 2026-2027 年的财报贡献会很有限。公司自己说 fiscal 2027 才会开始看到数据中心 AI 芯片的营收贡献。Barclays 期待高通在 2026 年 6 月分析师日 (analyst day) 披露更清晰的数据中心计划——这是个具体的时间窗口可以观察。

需要正视的还有：高通过去两次数据中心尝试均失败（2017 年 Centriq 服务器 CPU、2022 年 Nuvia 服务器项目）。第三次进攻执行风险不容忽视。此外，Nuvia 核心创始人 Gerard Williams 和 John Bruno 于 2026 年 1 月离职，对高通的数据中心 CPU 路线图构成直接威胁。

#### 4.9 机器人和工业 IoT：故事很大，离营收很远

高通认为 physical AI（机器人）市场到 2040 年达 \$1 万亿。Dragonwing IQ10 + Arduino（已收购）+ Edge Impulse 的全栈布局是认真的。Figure（人形机器人头部）这个合作客户也很有分量。

当下 humanoid 机器人销量按千台计，距离百万台量产还有很长的路。这块业务在 2026-2027 年财报上几乎看不到，2028 年开始有点贡献，2030 年之后才可能有意义的数字。

**FY2029 工业 IoT 目标 \$40 亿**（投资者日数字），与 PC 业务并列为 IoT 大类下的两个 \$40 亿子目标。

把它当成一个免费看涨期权——成了，是巨大的增长来源；不成，影响今天的估值不大。

#### 4.10 把所有变量叠加起来——FY2029 多元化的全景

最后，把所有非手机增长引擎拼到一起，看 FY2029 高通投资者日给出的目标：

更新后的表格（FY2025 实际数据起点 + 4 年 CAGR）

**~20%FY2029 业务**

**目标营收**

**起点 (FY2025 实际)**

**4 年隐含 CAGR**

**汽车**\$80 亿

\$39.6 亿

~19%

**PC**\$40 亿

极低基数

极高

**工业 IoT**\$40 亿

~\$25 亿 (估算)

~12%

**XR**\$20 亿+

~\$8 亿 (估算)

~26%

**其他 IoT**\$40 亿

~\$33 亿 (估算)

~5%

**IoT 合计**\$140 亿

\$66.2 亿（官方）

~21%汽车 + IoT 总计\$220 亿\$105.8 亿（官方）

两个关键观察：

第一，FY2025 起点 亿220 亿，4 年隐含 CAGR ~20%。这不是"高估的故事"，而是**首年已超前于目标曲线**的可量化目标。其中汽车 80% 已签约。

第二，加上 QTL（高通技术许可业务）（FY2025 亿，预计微增至60 亿）和数据中心新业务（公司未公开 FY2029 目标，分析师估 亿），非手机非业务在可能合计达到330-340 亿。

这里值得专门拎出来说说 QTL（高通技术许可业务）——它在 FY2029 估值故事里的位置，比 11% 的营收占比看起来重要得多。

QTL 是和 QCT 完全独立的业务线——QCT 卖芯片，QTL 卖专利使用许可，把高通在 5G/4G 标准必要专利组合的授权卖给所有做无线设备的厂商。**关键差别在于：QCT 收入取决于客户买不买高通芯片，QTL 收入取决于客户做不做 5G 设备**——这两件事可以分开。

Apple 2020 年起切自研 modem、2027 年完成替代——QCT 这部分会失去苹果几乎全部 modem 营收（约 \$73-78 亿/年）。但只要 Apple 还做 5G iPhone，苹果就要继续付 QTL 专利费——除非 Apple 同时挑战专利授权协议本身。这件事 2017 年发生过一次，2019 年和解，最终 Apple 认输付钱。

利润结构上，QTL 是软件公司级别的——FY2026 Q1 EBT 利润率 77%，FY2025 全年 72%。每年 ~亿营收能产出约43-45 亿税前利润，几乎不受芯片业务周期影响。**叠加进 FY2029 全公司展望：QTL 营收占比 11%，但贡献的税前利润占比可能接近 30%**——QCT 周期波动可以蚕食很多东西，但很难蚕食到 QTL 这块利润底。

这是为什么 forward PE 12-14x 看起来便宜——**下行最大支柱不是 \$200 亿回购，是 QTL 这条稳定现金流**。只要 5G 还存在、5G 标准必要专利还有效，这条现金流就在。

## 五、估值

高通是否被低估，本质上是问：**四个非手机引擎能否在 Apple 切换的窗口期（2026-2027）以足够速度兑现，覆盖每年 \$73-78 亿的苹果业务流失。**

这是一个**赔率高于胜率的对赌**——不是说"高通一定会赢"，而是说当前估值（Forward PE 12-14x）已经定价了非常悲观的情景，而正面催化剂（OpenAI 智能手机、Wayve、HUMAIN、汽车 80% 签约可见性、Arm 反诉胜诉、\$200 亿回购）的兑现概率，似乎被市场低估了。

### 5.1 当前估值数据点

截至 2026 年 4 月 28 日：

- **股价区间**：4 月 24 日 +11.12% 至（财报拉动），月日因合作传闻盘前160 一度，4 月 28 日收盘 **\$150.04**
- **市值**：约 亿（160B）
- **52 周区间**：（年月日）205.95（2025 年 10 月 27 日）。从高点已下跌约 34%
- **TTM PE**：~27（GAAP；Yahoo Finance 数据）
- **Forward PE**：~12-14（基于 FY2026 共识 EPS 11.20）

- TTM 营收: \$44.87B
- TTM 净利润: \$5.37B (GAAP)
- 经营现金流 (FY2025) : \$14.0B
- 自由现金流 (FY2025) : \$12.8B (+14.86% YoY)
- FCF Yield: ~8.5% (在大型科技股中接近"价值股"水平)
- 净有息负债: 约 (11.8B 现金 vs \$14.8B 长期债务)

历史对比:

- 10 年平均 PE: 21; 过去 5 年平均: 18.5; 过去 3 年平均: 20.6
- 2018 年 Apple modem 争端时期, PE 也曾跌至 12x 以下——**当前估值水平与那次结构性危机时期高度相似**
- 2022 年库存危机期间最低季度 PE 为 10.5 倍

## 5.2 同业横向比较

把高通和半导体板块主要公司放在一起对比 forward PE:

公司

Forward PE

TTM PE

EV/EBITDA

FCF Yield

LTM 营收增长

Qualcomm~12-14~27

~11

~8.5%

+5-11%

Broadcom (AVGO)

~36

~35

~46

~4%

+68%

Marvell (MRVL)

~40+

—

~40

~3%

+18%

NVIDIA

~25-30

~42

~36

~2%

+215%

AMD

~30

~133

~44

~3%

+25%

Intel

~25

—

—

—

-5%

Texas Instruments

~36

~47

~25

~4%

-5%

Micron

~10

—

—

—

—

半导体行业平均~46~64~38~3%+65%

观察：

- 高通的 Forward PE 和 EV/EBITDA 在大型半导体里都是地板水平（除了 Micron 这种典型的内存周期股）
- 与同样面临"成长性争议"的标的相比，估值折扣可能过大——但这个判断的前提是认为高通能完成转型
- **FCF Yield 8.5% 在大型科技股中接近"价值股"水平**——这反映出市场对其增长前景的悲观预期

### 5.3 分析师集体调低未来盈利预测（EPS 共识的下修）

时点

**FY2026 EPS 共识**

**FY2027 EPS 共识**

3 个月前（约 2026 年 1 月）

\$12.10

\$12.50 (估算)

60 天前 (约 2026 年 2 月底)

\$11.20 \$11.50 当前 (2026 年 4 月)

\$11.03 \$10.99 累计下修幅度

-8.8%~-12%

EPS 的持续下修表明，低 Forward PE 并非单纯的"市场错误定价"，而是盈利预期下调后的"被动降低"。这一动态与 2018 年 Apple modem 争端时期高度相似——当时高通 PE 也曾跌至 12x 以下，伴随的是盈利预期的同步下修。

这个路径是市场目前的"中性情景"——不假设转型大成功，也不假设彻底失败。

#### 5.4 4 月 14-22 日的"降级潮"——情绪定价已经发生

仅 4 月中旬一周内，高通就吃到了三家大行的下调：

- **JPMorgan** (4 月 16 日)：从 Overweight 下调至 Neutral，目标价从 下调至 140；分析师 Samik Chatterjee 引述"非手机业务多元化进展缓慢、缺乏近期催化剂"
- **BNP Paribas** (4 月 17 日)：从 Outperform 下调至 Neutral，目标价从 下调至 120；引述"smartphone woes 看不到尽头"
- **Barclays** (4 月 22 日)：以 **Underweight** 重新覆盖，目标价 \$130；分析师 Thomas O'Malley 警告高通对内存短缺敞口大，2026 年手机业务可能两位数下滑

加上 Bank of America 2026 年 3 月给出的 Underperform 评级 (预计高通 FY2025-FY2028 营收年化增长仅 2%、EPS 年化增长仅 1%)，**2026 年初至今高通已吃到至少 8 次降级**——分析师情绪降到近 20 年最低点。

但这种情绪定价已经发生了。对 4 月 27 日 OpenAI 合作消息的反应 (单日 +12%) 说明，**市场对正面催化剂的反应弹性也很大**——只要叙事方向逆转，悲观定价可以快速反转。

#### 5.5 三档情景分析 (带概率权重)

基于上述所有信息，构建三档情景：

情景

核心假设

FY2029 营收

EPS (FY2029)

Forward PE

**乐观** 汽车 140 亿 + PC 亿数据中心 50 亿+ 兑现；OpenAI 合作于 2028 年量产；Apple QTL 续约

\$55-60 亿

\$15-17

18-20x

**中性** 多元化部分成功，汽车 + IoT 合计 \$180-200 亿；OpenAI 合作推迟或缩水至小规模；Apple QTL 协议谈判艰难但保留

\$44-48 亿

\$11-12

13-15x

**悲观**苹果缺口无法填补；手机业务持续下滑；联发科和华为继续蚕食中国市场；数据中心 AI200 销量低于预期；OpenAI 合作未实现

\$38-42 亿

\$9-10

9-11x

**这个分布的关键特征是非对称性**——上行空间大于下行空间。在 forward PE 12-14x 的水平，下行有"\$200 亿回购 + 当下盈利能力 + FCF yield 8.5%"支撑；上行如果转型故事兑现，有重估空间。

## 5.6 估值观察

把上面拼到一起看，几个观察：

**当下定价反映的是什么：**市场基本上 100% 定价了苹果完全退出 + 智能手机周期下行 + Galaxy S26 部分 Exynos 回归。但似乎没有给汽车（80% 签约可见性）、PC（X2 平台爬坡）、数据中心（HUMAIN 已签）、OpenAI 合作（4 月 27 日新催化剂）、Wayve 战略整合任何溢价。

**估值的不对称性：**在当前估值水平上（市盈率约 12-14 倍——大型半导体公司里的地板价），股价继续往下跌的空间有三重托底：高通刚批准的 \$200 亿股票回购计划、QTL 等稳定业务带来的盈利的底、年自由现金流相当于公司市值 8.5%（这个比例放在大型科技股里属于偏高水平，意味着公司每年实打实赚到的现金多）；上行如果转型故事兑现，重估空间大。

**最大的判断点：**你需要相信"端侧 AI + 软件定义汽车"是真实的、长期的趋势，且高通在其中的位置不会被颠覆。这个前提如果成立，当下估值看起来便宜；如果不成立（比如 Apple 模式可复制 + 自研芯片成为车企标配），那高通可能被困在一摊持续萎缩的生意里——客户一个接一个走、新业务长不出来、规模像冰一样慢慢化掉。

到 216 的巨大估值差异——不是一个数字模型问题，是一个产业判断问题。

---

## 六、风险

按重要性排：

**第一，智能手机超长期下行 + 内存供应延长。** 当下市场最大的担忧。BNP Paribas 认为"smartphone woes 看不到尽头"，对应的场景是：手机出货量在 11-12 亿台/年附近停滞甚至下行 + 内存周期延长 + 中国厂商加速自研芯片。如果这个组合成立，QCT 手机营收可能在 \$250 亿长期徘徊。Counterpoint 已预测 2026 年高端手机 SoC 出货量下降 7%。

**第二，苹果对 QTL 授权的最后一战。** 现有协议 2027 年 3 月到期。Apple 替换完 modem 之后，会不会再次挑战 5G 标准必要专利的费率？历史上 Apple 已经打过一次，最后认输。但 2027 年的产业格局和当时不同——苹果有更强的自研基带后，谈判筹码不一样。如果 QTL 失去 Apple 这个客户，每年净损失大约 \$7-10 亿利润（高利润率部分）。

**第三，数据中心未必能站稳。** NVIDIA 在数据中心 AI 占 80%+ 份额，AMD 缓慢蚕食，Intel 也在做 Crescent Island 推理芯片。AI200/AI250 的硬件参数好看，但没有 CUDA 生态意味着每一个客户都要重新做软件适配。

HUMAIN 之后的下一个大客户在哪？OpenAI、Anthropic、Google、Meta 这些大厂都在自研或绑定 NVIDIA。Nuvia 创始人 Gerard Williams 和 John Bruno 于 2026 年 1 月离职是个负面信号。

**第四，三星 Galaxy S26 Exynos 部分回归**——一个新增加但容易被忽视的逆风。Galaxy S25 时代 100% 高通，S26 估计降到 60-70%。如果三星持续推进 Exynos 在 S 系列的覆盖范围，未来几年会成为持续的份额损失。

**第五，机器人/智能眼镜不是高通独占。** 英伟达 Jetson Thor 在机器人有先发优势和 CUDA 生态。Meta 与联发科合作做 AR 眼镜定制 SoC，意味着 Meta 不会让高通独占这个市场。这两个长期增长点的竞争比汽车更激烈。

**第六，中国市场的双重风险。** 中国 OEM（小米、OPPO、vivo、荣耀）是高通手机芯片的核心客户。但中国自研芯片的进展，加上潜在的中美科技政策摩擦，可能在 2-3 年后蚕食这块市场。2025 年 10 月中国国家市场监督管理总局对高通收购 Autotalks 发起反垄断调查，是个值得关注的信号。

**第七，OpenAI 合作的执行风险。** 4 月 27 日 Ming-Chi Kuo 披露的合作目前没有 OpenAI 官方确认。2028 年量产时间表意味着至少还有 2 年走完芯片设计、流片、车厂认证。OpenAI 还需要解决无 Google Play Services 生态下 AI Agent 如何访问应用功能的根本问题。这件事在 OpenAI 那边的不确定性，目前比在高通那边的不确定性大。即便最终方案缩水或推迟，对高通来说不是大灾难（因为还没定价）；但如果完全不发生，意味着失去一个反向催化剂。

**第八，关税和供应链。** 高通无晶圆，依赖 TSMC 3nm（旗舰）和三星 4nm（部分汽车/IoT）。任何台海局势、美国对华出口管制升级都直接影响生产。Apple 已锁定 50%+ 台积电首批 2nm 产能，意味着高通在 2026-2027 年面临先进制程结构性紧张。

**第九，估值陷阱。** PE 12-14x 便宜是事实，但便宜的东西可以更便宜。如果 calendar 2026 财报持续 miss，估值可能进一步压缩到 PE 10x（对应股价 \$110 区间）。

一个已经从风险变成正面催化剂的事件：Arm 诉讼终局

值得单独说明的是 Arm 诉讼——这个之前被市场视为悬而未决的风险，现在状态已经反转：

- **2024 年 12 月：** Delaware 联邦陪审团判定高通胜诉（没有违反 ALA 或 TLA）
- **2025 年 9 月 30 日：** Delaware 法官 Maryellen Noreika 驳回 Arm 全部剩余主张，**完全胜诉，full and final judgment**
- **2025 年 10 月：** Arm 提起上诉
- **2026 年 3 月：** 高通的反诉案（高通诉 Arm 干扰客户关系、阻碍创新、计划自己做芯片）开庭审理

**所以这件事现在的状态是：**主案高通已经完胜，且现在反过来在告 Arm。这从“风险”变成了“潜在的赔偿收入 + 战略筹码”。Counterpoint Research VP Neil Shah 评论这次胜诉对高通“是 monumental 的”——清除了 Snapdragon X PC、汽车、AI 服务器多个领域 Nuvia-based CPU 部署的法律不确定性。

## 七、回到最初的问题

回到这份研报开头的几个核心困惑：

**Q：端侧模型市场规模有多大？** 2025 年 亿50 亿是合理的共识区间（Edge AI 整体口径）。预计 2030 年硬件 亿、年整体1180 亿。但比规模更重要的是结构——这个市场分散在手机、PC、汽车、眼镜、机器人多个终端

形态，每个形态都需要专门的硅 + 软件 + 客户深度集成。中国是全球最活跃的试验田，IDC 预测 2026 年中国 AI 手机出货 1.47 亿台、占比 53% 首次过半。

### Q：未来 AI 在端侧会怎么演变？

1. **3B 模型成为手机端侧"甜点"** —— vivo 实测 7B 模型功耗 2A+ 难以普及，3B 模型 200+ tokens/s 内存仅 2GB，已是当前最优解
2. **agentic AI 让设备替你执行任务** —— 这必须低延迟必须本地。豆包 + 努比亚 M153 案例是中国本土最戏剧化的实证
3. **OS 厂商把 AI 内嵌到系统级别** —— Apple Intelligence、Galaxy AI 已经做了，会变成所有手机、PC 的标配
4. **智能眼镜和机器人这两类新形态本质上必须端侧** —— 体积、电池、隐私顾虑都不允许走云端
5. **云端 egress cost 的反向逼迫** —— 大规模 AI 应用走纯云端在经济学上不可持续，混合架构在月推理量 500-1000 万附近达到盈亏平衡
6. **低功耗持续推理成为新战场** —— 不是峰值算力，而是 mW 级 always-on 的能力，决定 7×24 Agent 的可行性

**Q：高通在"人手一个 AI"场景中的核心竞争力？** 实事求是地说：**高通是目前唯一一家在所有主要端侧形态都有商用 NPU 平台的公司。**这意味着开发者、OEM、汽车厂、机器人公司可以选择"全栈用高通"，跨产品线复用软件和优化。这是一个真实的护城河，但不是一个无敌的护城河——苹果在自有生态完全不需要它，联发科在手机端追得很紧，英伟达在数据中心和机器人有先发优势。

它的位置是"端侧的安卓阵营 + 汽车 + PC 的事实标准供应商"，不是"端侧的 NVIDIA"。

\* \* \*

[英伟达份额大降，AI革命新阶段机会在哪？](#)

[曦智科技研报：热度仅次于阿里和快手上手，今天开盘怎么看？](#)

### [错过20倍，我找到了AI投资的笨办法](#)

如果觉得有帮助请点赞转发、点击“在看”等，您的支持是我持续分享的最大动力，谢谢！

**关注我，在别人看热闹的地方，看到机会。更多内容，访问 [dayu.xyz](http://dayu.xyz)**

以上数据仅供参考，不构成任何投资建议。投资有风险，决策需谨慎。

由 Hermes 根据公开 X Article 导出为中文原版 PDF。原文版权归作者所有；内容仅作参考阅读归档，不构成投资建议。