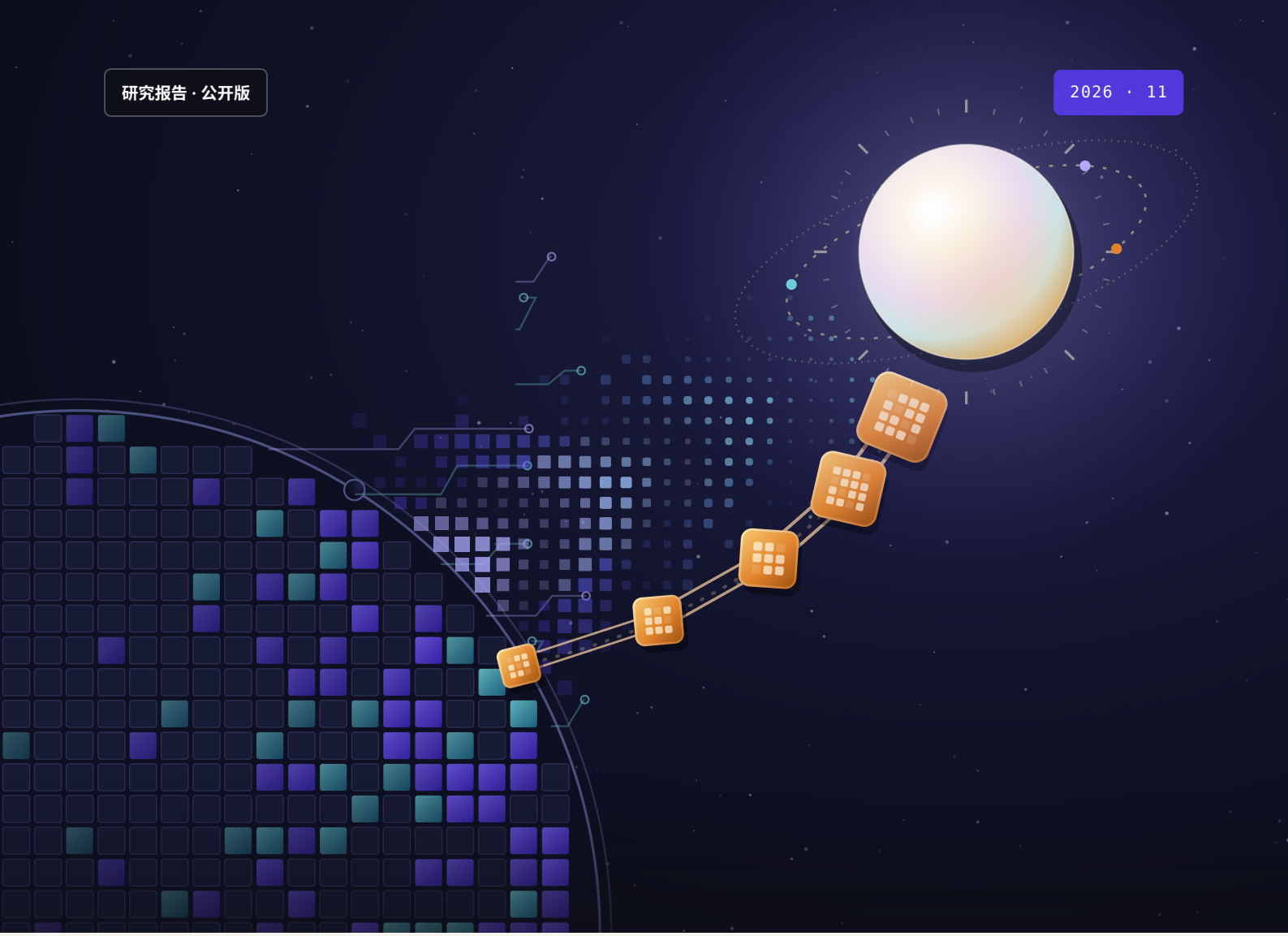
研究报告 · 公开版

2026 · 11



有用工作量证明 · 加密 × 芯片

算力即共识

AI × Crypto × 芯片全景研究（2026）：从 Pearl 的有用工作量证明谈起

Pearl · 有用工作量 · DePIN 与智能体 · ASIC 经济学 · 工艺节点 · 分阶段预算

结论

Pearl 证明了矩阵乘法可以保护一条公链，但还没有证明这些工作有用。在 AI × Crypto 领域，付费需求集中在推理、数据与隐私；ASIC 只在盘子大、规则稳、芯片有第二市场时才值得做。

谭帅

技术专家与创业者 · AI 研发者

tanshuai.com · i@tanshuai.com

数据截至 2026 年 11 月 13 日

金额单位：美元

目录

目录



◆ **执行摘要** 04
定理成立，激励未成立

01 **范围与方法** 06
范围、方法与四项检验

02 **Pearl · 机制** 07
Pearl：以矩阵乘法为共识

03 **Pearl · 经济学** 09
Pearl：规模、排放与有用性缺口

04 **Pearl · 规则** 11
Pearl：规则风险、数值格式与治理

05 **有用工作量类链** 12
有用工作量类链：矩阵与零知识都挡不住 ASIC

06 **更广的版图** 14
AI × Crypto 的其他赛道：需求真实但很小，价值流向股权

07 **ASIC 经济学** 16
ASIC 之问：一个公式，三把尺子

08 **逐类芯片** 18
五类芯片，五个答案

09 **工艺节点** 21
工艺节点：让终生成本的最大项来选

10 **分阶段预算** 24
从模型到硅片：第一颗芯片的分阶段预算

11 **机会** 25
机会在哪里，我会怎么做

12 **观察信号** 27
2026 年 11 月至 2027 年值得关注的信号

13 **该问的问题** 28
你没有问、但应该问的问题

◆ **关于作者** 29
谭帅

◆ **来源** 30

图表



图 01 AI × Crypto 六条赛道：付费需求真实在哪里 05
图 02 在 Pearl 的整数工作上，GPU 共享一道能效天花板 08
图 03 Pearl 的头六个月：算力追随币价，规则不断变化 10
图 04 补贴倍数：网络支付的是外部收入的多少倍 13
图 05 效率为何不再值钱：ASIC 最多留下 1 − 1/k 的收益 16
图 06 多数盘子比一次芯片设计预算还小 18
图 07 五类芯片，五个答案 21
图 08 领先节点晶圆的价格是成熟节点的十倍 22
图 09 先买证据，再买掩模：第一颗芯片的分阶段预算 24

文中标注

笔者测算	笔者依据所引数据所做的计算或推断
厂商宣称	由销售该产品的公司发布，未经独立验证
笔者分析	笔者对证据的解读

数据截至 2026 年 11 月 13 日 · 金额单位：美元

定理成立，激励未成立

结论

Pearl 证明了“算力即共识”在工程上可行，但还没有证明它“有用”。

≈90 EH/s
2026 年 11 月 13 日的 Pearl 全网算力，其中 1 H 即一次 INT7 乘加

≈1%
的 Pearl 区块承载真实 AI 工作（Blockworks，2026 年 10 月 28 日）

≈7,000 万美元
Bittensor 每年可验证的外部收入，而排放超过 3 亿美元

175–180×
2025 年第三季度 Aleo 证明者奖励相对其手续费的倍数

R × (1 − 1/k)
成本比边际矿工低 k 倍的 ASIC 的利润上限
笔者测算

≈150 万美元
在 N7 上经由模型、FPGA 与 MPW 拿到实测硅片，之后才是 ≈350 万美元的掩模组
笔者测算

Pearl

2026 年 11 月 13 日，Pearl 全网算力约 90 EH/s。按其共识代码，一个“哈希”就是一次整数乘加，全网相当于约 12.2 万张 H100 的整数算力。但只有约 1% 的区块承载真实 AI 工作，唯一公开的推理合作也已不在其服务商的无服务器 API 上提供。症结在经济而不在密码学：边推理边挖矿的 GPU，只能拿到专用矿卡约 1%–27% 的算力，因此理性矿工都去算随机矩阵。

付费需求

纵观 AI × Crypto，可验证的付费需求集中在三处：卖推理、卖数据、卖隐私。Bittensor 可验证的外部收入年化约 7,000 万美元，约八成来自算力与基础设施类子网。Grass 自报 2026 年上半年数据收入 1,700 万美元。可信执行环境中的机密推理已进入生产。相比之下，真正由 AI 智能体发起的支付每月只有约 5,000–11,000 美元，zkML 与去中心化训练尚无披露的付费需求。

补贴与价值

补贴倍数是最诚实的体检指标。Aleo 的证明者奖励约为其手续费的 175–180 倍；Bittensor 的排放约为其可验证收入的 4–11 倍；Render 的市值约为其链上费用的 360 倍；Pearl 年化约 4.6 亿美元的矿工收入以区块补贴为主，与付费 AI 需求几乎脱钩。补贴是冷启动的燃料，不是商业模式。与此同时，价值正在从代币迁移到股权：四家“用中心化基础设施、做加密原生分发”的股权公司估值合计约 85 亿美元，大致相当于协议层与应用层 AI 代币的总和。

ASIC

在芯片方面，ASIC 的利润上限约等于单位算力收益 ×(1 − 1/k)，其中 k 是它相对边际硬件的成本优势倍数。由于这一项是凹函数，k 超过 5–10 之后，效率再提升几乎不增加利润；决定成败的是盘子大小、规则能维持多久，以及芯片有没有第二市场。我按芯片类型给出的答案是：现在不流片专用的 INT7 Pearl 芯片，转而押注“推理优先、可挖矿”的加速器；新进入者不进入比特币商户矿机市场；小团队不做独立训练芯片；签下客户后再做推理芯片；ZK 加速器观望对冲，FHE 早布局。

节点与预算

工艺节点应由芯片终生成本中最大的一项决定：全天候运行的芯片看能耗，训练芯片看封装与存储，推理芯片看 SRAM 还是 HBM。第一颗芯片应分阶段购买——先模型，再 FPGA，再多项目晶圆，最后才

是掩模组——这样协议一旦变化，损失的是一次原型的钱，而不是一次流片的钱。机会在各领域的交叉处：以 Crypto 为第一个客户，以 AI 为第二市场。

问题	判断	关键数字（来源见各章）
① **Pearl 可行吗？它的工作有用吗？**	可行：一条以矩阵乘法规模化保护的公链。尚未有用：几乎所有挖矿都在算随机矩阵	2026 年 11 月 13 日约 90 EH/s；1 H = 1 次 INT7 乘加；约 1% 的区块含 AI 工作
② **AI × Crypto 的真实付费需求在哪里？**	推理、数据与隐私。Agent 支付、zkML 与去中心化训练仍处于需求之前	Bittensor 可验证收入年化约 7,000 万美元；Grass 2026 年上半年 1,700 万美元；Agent 支付每月约 5,000–11,000 美元
③ **ASIC 值得做吗？**	只在盘子大、规则稳、芯片有第二市场时才值得。今天没有任何网络三者兼具	利润上限 ≈ R × (1 − 1/k)；比特币矿工年收入约 115 亿美元，Pearl 约 4.6–4.9 亿美元，Kaspa 约 1,800 万美元
④ **小团队该做哪些芯片？**	推理优先、可挖矿的加速器；面向已签约客户的推理芯片；早期 FHE 工作。不做商户比特币矿机或训练芯片	Kaspa 最好的 ASIC 每天赚 0.86 美元；3/2nm 设计预算 6,000 万至 4 亿美元以上
⑤ **选哪个工艺节点？**	由终生成本最大项决定：能耗、封装与存储，或存储架构	台积电 N2 每片晶圆约 30,000 美元，N28 约 3,000 美元；HBM 与 CoWoS 约占 H100 成本的 63%
⑥ **第一颗芯片要花多少钱，该怎么买？**	分四个阶段：模型、FPGA、多项目晶圆，最后才是掩模。若规则或买家尚未落定，就停在掩模之前	商用 MPW 名额 5 万–50 万美元；N7 掩模约 350 万美元；7/6nm 设计预算 1,500 万–1 亿美元

图 01 **AI × Crypto 六条赛道：付费需求真实在哪里**

可验证的付费需求与补贴、估值信号对照，2026 年 11 月



来源： DefiLlama、SubConnect（经 hoge.gg）、TRM Labs、Phala、Dealroom、SiliconANGLE、CoinGecko、Blockworks；笔者综合。

第 01 章

01 范围、方法与四项检验

范围与方法

为什么写这份报告，为什么是现在

2026 年 4 月，一条名为 Pearl 的新公链把一个由来已久的想法投入了生产：让矿工用 AI 赖以运行的矩阵乘法、而不是哈希来竞争出块。六个月内，它的全网算力达到约 90 EH/s，代币市值约 5 亿美元。与此同时，AI 算力市场、去中心化训练网络、Agent 支付通道与数据网络融资数十亿美元，比特币矿企开始把电力站点改造成 AI 数据中心。我想弄清楚：这些方向里哪些有真实客户，定制芯片在其中哪里说得通，一个芯片团队实际能用上哪个工艺节点、哪家代工厂。

检验任何 AI × Crypto 项目的四项标准

2017 年一项针对 Gridcoin 的安全研究认为，有用计算要能保护一条链，必须难以求解、易于验证、奖励易于计算。我再加上第四项。这四项决定了后文的大部分判断。

01 **难以求解。** 诚实计算必须是产出有效证明的最便宜方式，不能有捷径。

02 **易于验证。** 检查工作的成本必须远低于完成工作的成本，否则网络要付两次钱。

03 **奖励易于计算。** 协议必须能仅凭证明本身算出奖励，而不必信任第三方。

04 **输出有人付费。** 除非网络之外有客户为结果付费，否则所谓“有用”的工作只是一笔带着叙事的补贴。

哈希谜题满足前三项而不满足第四项。真实 AI 任务满足第四项，却过不了第二项：用零知识证明一次 AI 计算，成本是计算本身的 100 到 10,000 倍，学习证明（proof-of-learning）方案也已被证明可以低成本伪造。Pearl 的带噪矩阵乘法，是第一座可信地跨过验证关、又与 AI 算力同构的桥。

我的研究方法

这是一项案头研究，做法与我评审一颗芯片或一份商业计划书相同。我阅读了 Pearl 的代码、协议提案与白皮书；来自独立全节点的链上数据；公司公告与财报；密码学、计算机体系结构与机器学习方面的论文；代工厂与厂商的价格数据；以及行业与科技媒体的报道。每个数字都注明日期。凡是我自己推算的数字，都会说明并标注 笔者测算。只有卖方自报的数字标注 厂商宣称；我对证据的解读标注 笔者分析。金额单位为美元。状态截至 2026 年 11 月 13 日。

第 02 章

02

Pearl · 机制

Pearl：以矩阵乘法为共识

从定理到主网

2025 年 4 月，Ilan Komargodski 与 Omri Weinstein 发表论文《Proofs of Useful Work from Arbitrary Matrix Multiplication》。他们为任意矩阵乘积构造了有用工作量证明，额外开销仅为朴素矩阵乘法的 1 + o(1) 倍；其安全性建立在“求解一批随机低秩线性方程”的困难性之上，这是猜想而非证明。论文点破了这个领域的真正难点：常数倍开销的有用工作量证明“平凡但无用”，难的是在矿工自选输入的前提下，让诚实计算依然是最便宜的胜出方式。

Pearl 主网于 2026 年 4 月 27 日 09:00 UTC 启动。矿工计算带噪乘积 (A + E)(B + F)，其中低秩噪声由 BLAKE3 承诺导出，再剥离噪声得到 A·B。每个输出分块的累加器状态被哈希成一张“彩票”；中奖分块被打开，并用递归 Plonky2 零知识证明压缩到 60 KB 以内，放进上限为 65 KB 的区块证书。验证是一个常数大小的小证明，与矩阵规模无关：它和哈希谜题一样便宜可验，计算本身却可以有用。

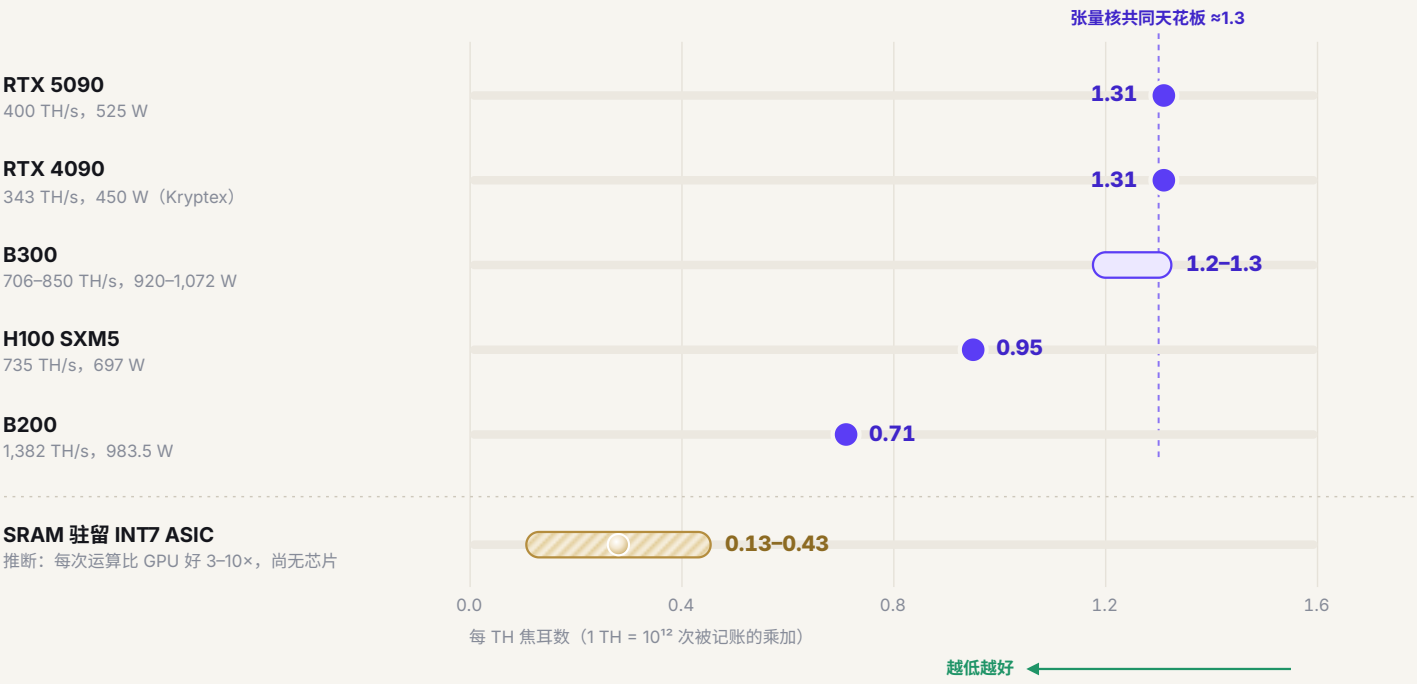
一次乘加就是一个“哈希”

对芯片设计者而言，最重要的事实是计量单位。Pearl 共识代码按“分块大小 × 点积长度”缩放每次尝试的目标值，因此 **1 H 恰好等于一次被记账的 INT7 × INT7 → INT32 乘加（MAC）**。自区块 96,251 起，秩惩罚要求噪声秩至少为 128；在最低秩时，1 H 仍等于 1 MAC。因此 Pearl 的“EH/s”字面意义上就是每秒 10^18 次整数乘加：**区块链的安全预算第一次以 AI 加速器的张量吞吐来计价。** 各类 GPU 的实测算力落在其稠密 INT8 乘加峰值的 60% 到 95%，与这一解读吻合。

硬件	Pearl 算力	功耗	每万亿次乘加焦耳数（J/TH）	来源（日期）
RTX 5090	400 TH/s	525 W	1.31 笔者测算	Kryptex（11 月 12 日）
RTX 4090	约 305 TH/s（社区区间 280–330）；Kryptex 列 343 TH/s	450 W（Kryptex）	1.31（按 Kryptex 读数）笔者测算	prlstats（10 月 3 日）；Kryptex
H100 SXM5	735 TH/s	697 W	0.95 笔者测算	PeakMiner，经 prlstats（9 月 19 日）
B200	1.382 PH/s	983.5 W	0.71 笔者测算	PeakMiner，经 prlstats
B300	706–850 TH/s	920–1,072 W	1.2–1.3 笔者测算	prlstats；SimeonOnSecurity

图 02 **在 Pearl 的整数工作上，GPU 共享一道能效天花板**

按 Pearl 实测算力计，每万亿次 INT7 乘加的焦耳数（J/TH）；越低越好



来源: Kryptex（2026 年 11 月 12 日）；PeakMiner 与 SRBMiner 基准，经 prlstats；SimeonOnSecurity；ASIC 区间为笔者推断，非实测。

这张表里有两个对芯片设计者很重要的细节。第一，消费级 RTX 4090/5090 与数据中心 B300 的能效几乎重合在约 1.3 J/TH，即便最好的 B200 也只有约 0.71 J/TH。一张数万美元的数据中心 GPU，在每次整数乘加上的能耗与游戏卡相差无几，说明张量核每次 INT8 乘加的能耗存在共同的天花板，而这个天花板正是专用芯片可以突破的地方。第二，B300 反而比 B200 慢：英伟达在 Blackwell Ultra 上大幅削减了 INT8 吞吐，整数运算在这颗芯片上只能走更慢的路径。主流 AI 硅片正在离开 INT8、走向 FP8 和 FP4。共识选定的数值格式一旦与硬件演进方向背离，协议就会被迫改规则，这正是后文 FP8 分叉以及我对 ASIC 判断的根源。

第 03 章

03 Pearl · 经济学
Pearl：规模、排放与有用性缺口

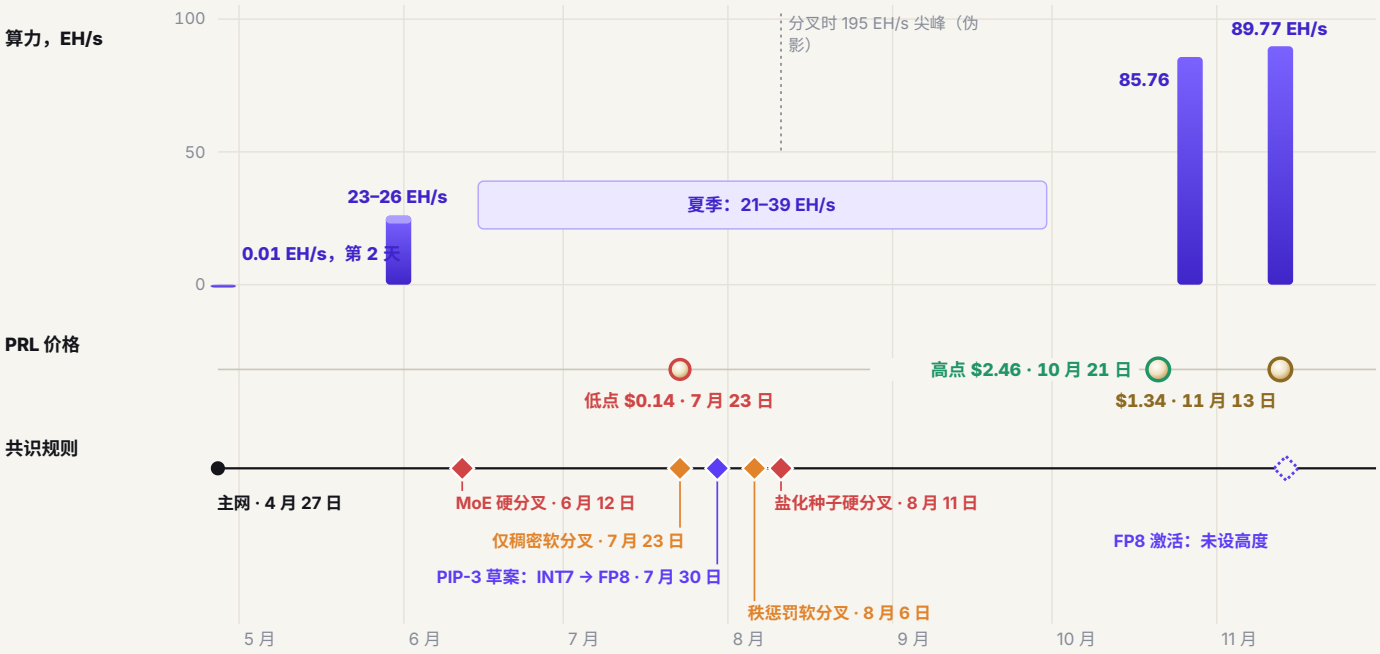
半年从 0.01 到 90 EH/s

指标	数值（2026 年 11 月 13 日）	口径与来源
全网算力	约 90 EH/s（基准 89.77；同日区间 88.20–93.22；11 月 12 日日内高点 110–123）	独立全节点 prlstats；Kryptex
难度与区块高度	58,841,982 与 142,530	prlstats
等效硬件	约 12.2 万张 H100 或 22.5 万张 RTX 5090	按每卡 735 与 400 TH/s 计 笔者测算
价格、市值与完全稀释估值	1.32–1.34 美元；约 5 亿美元；约 28 亿美元（CoinGecko 第 101 名）	CoinGecko
历史高点与低点	2.46 美元（10 月 21 日）与 0.1399 美元（7 月 23 日）	CoinGecko
已挖供应	3.776 亿 PRL，占 21 亿上限的 17.98%	prlstats
区块补贴与日排放	约 2,173 PRL 与约 97–103 万 PRL	prlstats；Kryptex
未来 12 个月新增供应	约为当前供应的 78%	prlstats
矿工收入	约 125 万美元/天（模型值），年化约 4.6 亿美元	prlstats
单位算力收益 R	约 14–15 美元/PH/s·天；一张 RTX 5090 每天约 5–6 美元	笔者测算；Kryptex
承载真实 AI 工作的区块	约 1%（10 月 28 日估计）	Blockworks
矿池集中度	Kryptex 占近 24 小时出块的 46.4%；前两大矿池合计约 62%	prlstats
持币集中度	地址基尼系数 0.969；前 100 地址持有 53.0%	prlstats

Pearl 的算力几乎是币价的滞后函数。上线次日全网约 0.01 EH/s；到 5 月底升至 23–26 EH/s。夏季币价较 5 月高点回撤约 86%，算力在 21–39 EH/s 横盘。9–10 月币价约十倍反弹后，算力在 10 月 27 日升至 85.76 EH/s，11 月站上约 90 EH/s。8 月 11 日那次 195 EH/s 的尖峰恰逢分叉，应视为估算伪影。能耗随之放大：一项研究估计 24 EH/s 时全网约 112 MW，prlstats 模型估计如今约 190–215 MW。

图 03 **Pearl 的头六个月：算力追随币价，规则不断变化**

全网算力、三个价格节点与每次共识变更，2026 年 4 月 27 日至 11 月 13 日



来源: prlstats 全节点与序列；Kryptex；CoinGecko；Pearl 的 params.go 与 PIP-3。8 月 11 日的尖峰是分叉时的估算伪影。

排放放大了反身性

供应上限为 21 亿枚，没有减半：区块补贴按 E(t) = S·H / ((t + H)(t + H − 1)) 平滑递减，其中 H = 650,226。但排放按区块而非时间计算，而早期难度起点过低、调整缓慢。**首日即挖出约 4,406 万枚，前六天约 1.2 亿枚**，约占今日供应的 32% 笔者测算。近期出块约 171–183 秒，仍快于 194 秒的目标。结果是未来 12 个月供应还将增长约 78%，这是矿工、矿机厂商和持币者都必须正视的稀释。

矿工经济学由此可算。全网矿工收入约 125 万美元/天（模型值）；以 Kryptex 的日排放乘以现价，约为 135 万美元/天，因此**单位算力收益 R ≈ 14–15 美元/PH/s·天** 笔者测算。一张 RTX 5090 每天约赚 5–6 美元。H100 或 H200 每天约 11 美元，而最便宜的公开按需租价约为每卡每小时 2.09 美元，约合每天 50 美元，纯为挖矿租卡会深度亏损。R 的波动极大：按 prlstats 序列，8–9 月约 10 美元，10 月中旬一度约 33 美元 笔者测算。上线后数周内，RTX 5090 的估计日收益就从约 33.80 美元腰斩至约 17.19 美元。租用云 GPU 挖矿只在上线初期几周和 10 月的价格尖峰时赚过钱。

定理成立，激励未成立

真正的问题在需求，不在密码学。Pearl 唯一公开的商业集成，是 Together AI 于 5 月 15 日推出的 Gemma-4-31B-it-pearl 端点：价格较常规低 25% 以上，折扣由排放的预期未来价值补贴。11 月 13 日，该模型页面显示其**不在 Together 的无服务器 API 上提供**。6 月，一项对 8,012 个在线挖矿实例（约占全网 21%）的普查得出结论：网络完全没有产生任何有用的 AI 计算（作者称这些证据属提示性而非结论性），并证明用纯随机矩阵构造的份额会被矿池接受。10 月 28 日，Blockworks 估计**只有约 1% 的区块承载真实 AI 工作**；它写道，没有这部分工作，Pearl 只是一个更小、更不安全、通胀更快的比特币。

原因可以算出来。推理侧的额外开销确实很小：Pearl 自报在 4 张 H200 上运行 Llama 70B，端到端时间增加 5.08%，一项独立测试在单张 Hopper GPU 上测得 9.2% 的吞吐损失。但挖矿侧赚得太少。官方 vLLM 插件只在至少 1,024 个 token 的预填充批次上挖矿，解码步骤从不挖矿。一张在线推理的卡，在纯预填充流量下能拿到专用矿卡约 27% 的算力，在 80%

为解码的对话流量下约 5.4%，在解码主导的流量下仅约 1%。没有付费推理需求，理性矿工就会去算随机矩阵；协议验证的是承诺过的矩阵被诚实地乘过，而不是这些矩阵有没有用。**定理成立，激励未成立。** Pearl 的“有用”是一个链下的商业问题：必须有人为推理付费。

第 04 章

04 Pearl · 规则
Pearl：规则风险、数值格式与治理

两个月四次分叉

Pearl 的另一张面孔是规则变化之快。2026 年 6–8 月，实验室以硬编码高度连推四次共识变更：混合专家（MoE）硬分叉（区块 71,935，约 6 月 12 日）、仅稠密软分叉（91,630，约 7 月 23 日）、秩惩罚软分叉（96,251，约 8 月 6 日）和盐化种子硬分叉（99,000，约 8 月 11 日）。盐化种子分叉之后，旧矿工只能产出无效份额；秩惩罚则让高秩配置的有效算力损失 50% 到 87.5%。

下一步是仍处草案状态的 PIP-3：把挖矿从 INT7 改为 FP8 × FP8 → FP32，并要求逐位精确的设备配置文件，**只支持 Hopper 与 Blackwell 两种架构**。文件称硬件依赖是浮点协议的必要之恶，并警告验证中的任何容差都会招致研磨攻击。截至 11 月 12 日，fp8-scheme 分支的主网参数尚未设置激活高度。按现行草案，RTX 40 系、A100 与 AMD 显卡都不在准入之列，而 prlstats 一个低置信度的模型估计约 82% 的算力来自消费级显卡。强推分叉，有链分裂或矿工反弹的风险。

数值格式三难

Pearl 面对一个三难：数值格式可以对 AI 有用（FP8 或 FP4）、跨硬件逐位可验证、硬件中立，但三者最多得其二。INT7 方案可验证且中立，却与 AI 硅片的演进方向背离。FP8 加白名单有用且可验证，却把共识绑在单一厂商身上。FP8 加容差有用且中立，却打开了研磨攻击之门。张量核中的浮点运算通常不遵循数值标准，跨代结果不可复现；以 Hopper 为例，FP8 累加只保留 13 位小数位并采用截断。从芯片设计者的角度看，一份“Hopper 配置文件”就是一套完全确定的算术，ASIC 可以逐位实现，甚至可能比 GPU 更省电。但这样做是在违背协议明示意图的灰区里做生意，治理风险极高 笔者分析。

治理

治理高度集中。所有协议提案都由实验室撰写；分叉一律以硬编码高度激活，不设矿工信号；没有第二套客户端实现；截至 10 月初，也未见对节点或零知识电路的第三方审计公布。以色列《TheMarker》报道，实验室在一轮秘密融资中筹得约 1.2–1.3 亿美元（约 5,500 万美元股权，外加约 7,500 万美元的 GPU 基础设施基金）；Pearl 未在官方渠道确认。“代币没有投资人份额”与“公司有风投股东”可以同时成立，但张力始终存在：AI 收入可能流向股权，稀释却由持币者承担。

综合来看，Pearl 是迄今最干净的“算力即共识”实验：机制新颖、链上可验证，计量单位与 AI 硬件一一对应。但“有用”尚未到来，而它的安全预算、算力与矿工热情，都系于一个主要需求尚未成形、12 个月内供应还要增长约 78% 的代币。Blockworks 提出的三个信号值得跟踪：升级能否平稳落地，AI 工作区块的占比能否实质上升，是否有推理服务商把挖矿收益让渡给客户。

第 05 章

05 有用工作量类链：矩阵与零知识都挡不住 ASIC

Kaspa：矩阵既不等于有用，也不等于抗 ASIC

Kaspa 是教科书式的案例。它的 kHeavyHash 在两次哈希之间夹了一个 64 × 64、4 位元素的矩阵向量乘法，乘积随即被丢弃。2021 年 11 月公平启动后约 16 个月，IceRiver 于 2023 年 3 月 14 日发布首批 Kaspa ASIC，到 2023 年底 GPU 挖矿基本出局。按 RTX 4090 的 3.8 GH/s（300 W）与比特大陆 KS7 的 77 J/TH 对比，能效差距约 **1,000 倍** 笔者测算。盛宴之后是压缩：11 月 13 日全网算力为 358.9 PH/s，较 2024 年 8 月的 1 EH/s 下降约 64%，最先进的 KS7 在每千瓦时 0.10 美元的电价下每天仅赚 **0.86 美元**。

Aleo：零知识同样挡不住专用硬件

Aleo 的证明者谜题被设计为模仿生成 SNARK 证明的工作，并不证明用户的交易，项目方还公开欢迎 ZK ASIC。2024 年 9 月主网上线后约五个月，专用矿机即已出货。2025 年第三季度，证明者获得 3,120 万枚 ALEO（约 780 万美元），而全网手续费只有约 43,000–44,000 美元：**补贴倍数约 175–180**。币价从早期交易的约 9 美元跌至该季度末的 0.22 美元，ARC-46 如今又要求每个解质押 10 万至 250 万枚 ALEO，挖矿事实上变成了权益门槛。

Qubic 与 Bittensor

Qubic 以训练其 Aigarth 神经网络的“有用工作量证明”为卖点，但其算力在 2025–2026 年最可见的产出是别的链的区块。它先挖门罗币，份额从 2025 年 5 月中旬的不足 2% 升至 7 月底的 25% 以上，8 月宣称拿下 51%，引发了 6 个区块深的重组（另有来源称 18 个）。2026 年 4 月 1 日，它转向用 ASIC 挖狗狗币，所得用于回购与销毁。其 AI 训练的有用性没有独立证据，市值从约 2.8–3 亿美元降至 11 月 13 日的约 5,000 万美元。它还暴露了一种外部性：为通用算力付费的链，会制造出一个庞大、可租用、可协调的算力池，并能反过来攻击 CPU 友好型链。

Bittensor 不是工作量证明，而是质押加权的主观打分：验证者给矿工评分，权重在质押加权中位数处被截断。它是这一组里唯一拥有规模化外部收入的网络。只统计可验证外部客户的 SubConnect 指数显示，2026 年 11 月初年化约 **7,000 万美元**（9 月 1 日约 3,500 万美元），其中 78% 到 82% 来自算力与基础设施；但按 11 月初的币价，其年排放超过 3 亿美元，收入只覆盖 9% 到 23%。收入估算随口径差异很大，从约 1,700 万美元（Artemis）到按自报总量年化的 1.72 亿美元不等。4 月时，128 个子网中有 67 个销毁了 99% 以上的激励；头部团队 Covenant 因治理问题出走时，TAO 数小时内下跌 24%。

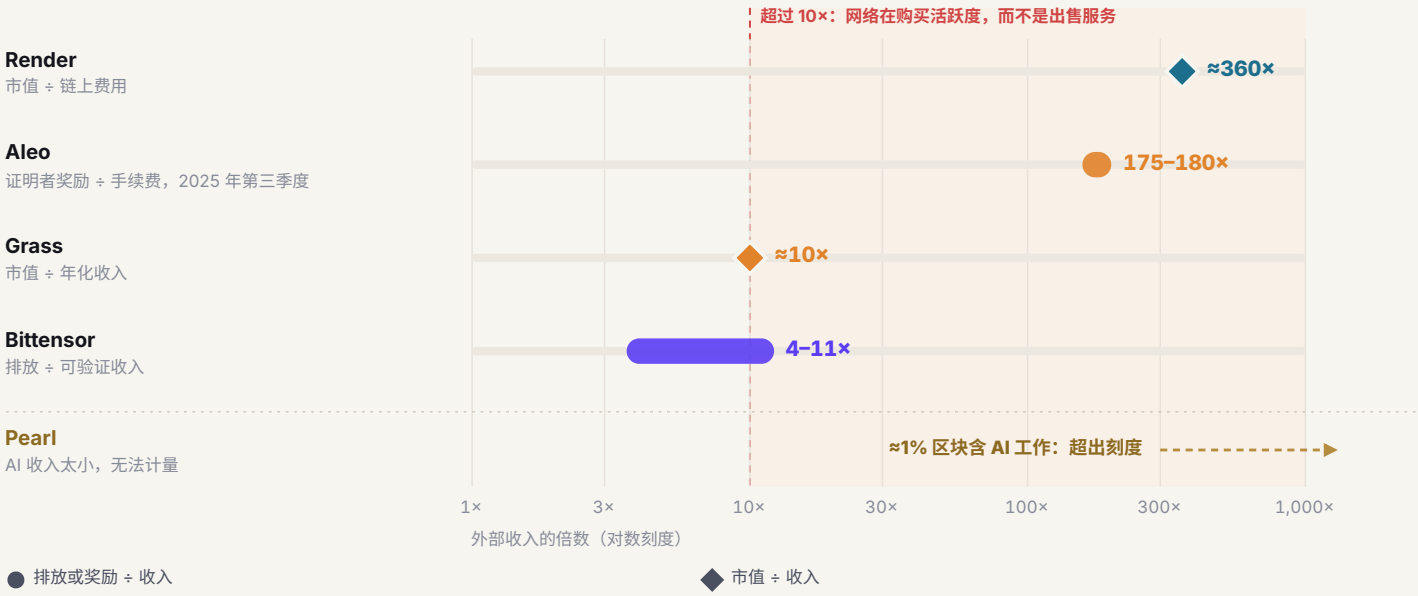
历史给出同样的答案

Primecoin 的素数链可以快速验证，却无人需要。Gridcoin 依赖第三方统计网站证明 BOINC 贡献，研究者在不做任何计算的情况下就伪造出块。Filecoin 的存储证明在技术上成立，但 2025 年第三季度利用率仅 36%，79.3 万美元手续费中 99.5% 是罚金，基础费只有 4,100 美元。

项目	工作 / 共识	市值（日期）	硬件走向	付费需求 vs 补贴	判断
Pearl	带噪矩阵乘法（INT7 乘加 + ZK 证书）	约 5 亿美元（11 月 13 日）	GPU；FP8 草案拟设 Hopper 与 Blackwell 白名单	约 1% 区块含 AI 工作；矿工收入以补贴为主	机制最可信，需求待证
Kaspa	kHeavyHash（哈希之间夹 64 × 64、4 位矩阵向量乘）	约 8.4 亿美元（11 月 13 日）	上线约 16 个月后出 ASIC；现全部为 ASIC	不主张“有用”；矿工年收入约 1,800 万美元 笔者测算	矩阵既不等于有用，也不等于抗 ASIC
Aleo	合成 SNARK 谜题 + 权益证明	约 1.18 亿美元（2025 年第三季度末）	约五个月后出专用机	证明者奖励约为手续费的 175–180 倍	合成工作，补贴驱动
Qubic	“uPoW”（宣称训练 Aigarth）	约 5,000 万美元（11 月 13 日）	CPU 与 GPU；2026 年 4 月起以 ASIC 挖狗狗币	有用性无独立证据；算力用于外部挖矿与回购	AI 叙事包装的算力聚合
Bittensor	Yuma 共识（质押加权打分）	约 25.5 亿美元（11 月 11 日）	通用 GPU；无 ASIC 路径	可验证收入年化约 7,000 万美元 vs 排放逾 3 亿美元	唯一有规模外部收入者；仍是 4–11 倍补贴
Filecoin（参照）	复制证明与时空证明	约 15 亿美元（2025 年第三季度末）	—	利用率 36%；季度基础费 4,100 美元	有用，但需求不足

图 04 **补贴倍数：网络支付的是外部收入的多少倍**

排放或市值除以来自网络外部的付费收入，对数刻度



来源: Blockworks（Aleo 2025 年第三季度）、DefiLlama（Render 费用）、CoinGecko、hoge.gg（Bittensor）、Dealroom（Grass）、Blockworks 0xResearch（Pearl）；笔者计算。

对硬件的启示：以流动性代币支付的固定、规则化工作，大约 3 到 16 个月就会迎来 ASIC，率先抵达的往往是挑战者而不是在位者。“有用”的说法只在有外部客户为输出付费的地方成立。而让硬件搁浅的主要因素，是相对收入过高的补贴，以及开发方对工作函数的控制。

第 06 章

06 更广的版图
AI × Crypto 的其他赛道：需求真实但很小，价值流向股权

去中心化算力市场

去中心化算力确有付费客户，但需求小、集中，而且主要是开源模型推理的转售。可在链上核验的费用很少：DefiLlama 显示 Render 过去一年费用约 235 万美元，Livepeer 约 68 万美元，整个 DePIN 板块 2025 年链上收入仅约 7,200 万美元。自报数字更大，但口径宽松：Aethir 称 2025 财年收入“1.278 亿美元以上”，却未说明是总账单还是协议抽成。代币估值与收入在两个方向上同时脱钩：Render 的市值约为其链上费用的 360 倍，而 Aethir 约 8,770 万美元的市值却低于其自报收入 笔者测算。

“便宜 70% 到 85%”的说法只对云巨头的按需价成立。11 月 3 日，Akash 上 H100 的加权均价约为**每卡每小时 2.73 美元**（区间 2.04–3.16 美元），低于 AWS 的约 6.88 美元与 Google Cloud 的 11.06 美元，却高于 Vast.ai 的约 1.73 美元、RunPod 社区云的 1.99 美元，以及一年期预留价的约 2.44 美元（其他平台价格为 9 月 12 日读数；发布方 Akash 是利益相关方）。需求主要是经由 OpenRouter 分发的商品化推理，AkashML 每天约 17 亿 token，io.net 每天逾 40 亿 token，与所有新兴云正面竞争。最大的“去中心化算力”美元正在再中心化：以 Aethir 代币为金库资产的 Axe Compute 签下一份 5 年期、价值 15 亿美元的合同，出租的是自有的 9,200 张 B300 集群，而其 2026 年第一季度收入仅 35,000 美元。

去中心化训练与推理

互联网上的训练已在 80 亿至 400 亿参数级得到验证。Pluralis-8B 在 330 个以消费级 GPU 为主的节点上训练 40 天，达到集中式 H100 基线效率的 63%，但论文坦言此次运行并非去中心化，需要 Pluralis 自己的节点和一个准入授权方。Nous Research 的 Psyche 完成了 400 亿参数的 Consilience 预训练。到了前沿规模，工作重新集中化：Prime Intellect 的 INTELLECT-3 是一个 1,060 亿参数的混合专家模型，在 512 张以 400 Gbps InfiniBand 互连的 H200 上训练了两个月。约束清晰可见：节点需要至少 200 Mbit/s 的带宽和约 100 倍的通信压缩，信任与验证仍依赖运营方的特权角色。

钱流向了卖中心化产品的股权公司。Prime Intellect 在 7 月以 10 亿美元估值融资 1.3 亿美元之前，年化收入已超过 1 亿美元；私密推理应用 Venice 自报年化收入超过 7,000 万美元、自 2026 年第一季度起盈利（公司口径），以 10 亿美元估值融资 6,500 万美元，运行在租用的 GPU 上。代币一侧，Gensyn 的核心训练网络仍“待定”，其代币市值约 1,720 万美元；AI 原生 L1 Ritual 主网上线后，代币较高点下跌 72%（仅见标题）。

可验证 AI

生产系统选择了可信执行环境（TEE）。Hopper 机密计算对典型 LLM 查询的额外开销低于 7%，对大模型几乎为零；Phala 生产环境的开销约为 0.5% 到 5%，但到 2025 年底，其 10,004 名用户中付费者只有 398 人。TEE 的信任根在芯片厂商：2025 年 10 月披露的 TEE.fail，用一个成本不足 1,000 美元的 DDR5 总线插入器，从完全更新的 Intel TDX 与 AMD SEV-SNP 机器中提取出证明密钥，进而危及英伟达 GPU 机密计算，而两家厂商都表示物理攻击不在其威胁模型之内。这恰恰是节点运营者能物理接触机器的 DePIN 场景。zkML 已能证明完整的 LLM：Lagrange 的 DeepProve 证明 GPT-2 约每分钟 174 个 token、Gemma 3 约 86 个，距离生产推理仍差几个数量级，也没有披露付费使用。

Agent 支付与数据

x402 是最好的反面教材。TRM Labs 统计，自 2025 年 5 月以来，已知结算方共处理约 1.989 亿笔交易、5,270 万美元，99.6% 为 USDC；剔除自付、批量流转与买家少于十个的卖家之后，真正由智能体发起的份额只有 0.6% 到 7.5%，**约每月 5,000–11,000 美元**。Artemis 认为 95% 以上的活动是协议信号而非买卖。与此同时，通道正被在位巨头标准化：x402 基金会于 7 月 14 日在 Linux 基金会下成立，Visa、Mastercard、Stripe、Google、AWS、Cloudflare 与 Coinbase 均为核心成员；与 Stripe 共同制定机器支付协议（Machine Payments Protocol）的 Tempo 于 3 月上线主网，估值 50 亿美元。智能体代币平台仍是发射台逻辑：Virtuals 自报“aGDP”4.79 亿美元，ElizaOS 创始人在一宗集体诉讼和解后宣布其代币“已死”。价值更可能被稳定币发行方与支付巨头捕获；加密在这里的独特角色，或许在于智能体的身份与问责。

数据是少数有链下真实买家的子赛道之一。Grass 自报 2025 年数据收入 1,700 万美元、2026 年上半年再收 1,700 万美元，全年目标 7,500 万美元，并称已经盈利；GRASS 市值约 3.28 亿美元，约为其上半年年化收入的十倍 笔者测算。它的护城河是分布式住宅带宽带来的采集规模，而不是加密本身。数据 DAO 与数据市场缺乏可见的付费需求：Vana 市值约 2,150 万美元，Ocean 约 2,760 万美元，且 Ocean 正与 Fetch.ai 陷入诉讼。人格证明则随着智能体增多而获得真实集成：World 称已验证近 1,800 万人，Zoom、Docusign 与 Tinder 已支持 World ID 的部分功能。

资本已经投票

赛道	可验证的付费需求	补贴与估值信号	判断
有用工作量类链	Bittensor 可验证外部收入年化约 7,000 万美元；Pearl 约 1% 区块含 AI 工作	Aleo 175–180 倍；Bittensor 4–11 倍；Pearl 以补贴为主	机制有价值；“有用”要靠付费客户证明
去中心化算力市场	链上费用每年约 68 万–235 万美元（Livepeer、Render）；自报 400 万–1.28 亿美元	Render 市值约为费用的 360 倍	真实但小，以商品化推理转售为主
去中心化训练与推理	股权公司：Prime Intellect 年化逾 1 亿美元，Venice 逾 7,000 万美元（中心化基础设施）	Gensyn 训练主网未上线；Ritual 代币较高点 −72%	技术进步真实；利润归股权
可验证 AI	TEE 进入生产（Phala 398 名付费用户）；zkML 未披露付费使用	—	TEE 有用；zkML 尚早
Agent 支付	智能体发起的流量每月约 5,000–11,000 美元	约 2 亿笔交易，多为信号与投机	通道先行，需求未至
数据网络	Grass 2026 年上半年 1,700 万美元（自报）	GRASS 约为年化收入的 10 倍	少数有真实买家的赛道之一

把六条赛道放在一起，结论非常一致：**补贴是冷启动的燃料，不是商业模式。** 11 月 13 日，CoinGecko AI 板块总市值约 186 亿美元，但其中约 56% 是 NEAR、ICP 这类通用 L1，以及 TAO、质押 TAO（可能与 TAO 重复计算）和 Pearl 这类共识层项目；真正属于协议层与应用层的 AI 代币合计约 80 亿美元或更少 笔者测算。这大致相当于四家股权公司的估值之和：Tempo（50 亿美元）、Nous（报道值 15 亿美元）、Prime Intellect（10 亿美元）与 Venice（10 亿美元）。加密风投的季度规模停留在约 45–50 亿美元，增量来自 AI 相关基础设施，其在披露金额中的占比于 2026 年第三季度从 8.1% 升至 18.2%；而 2025 年发行的 118 个代币中，84.7% 跌破了发行价。一句话：**加密负责分发与结算，股权收走利润**，至少在 2026 年如此。

第 07 章

07 ASIC 经济学
ASIC 之问：一个公式，三把尺子

一个公式：为什么 k 超过 5 之后效率不再值钱

设 R 为单位算力的日收益（全网矿工收入 ÷ 全网算力），c 为边际硬件（在 Pearl 上是 GPU）提供单位算力的日成本，k 为 ASIC 相对边际硬件的成本优势倍数（c_ASIC = c/k）。只要 GPU 仍是边际矿工，自由进出就会把 R 压向 c，于是 ASIC 每单位算力赚 R − c/k：

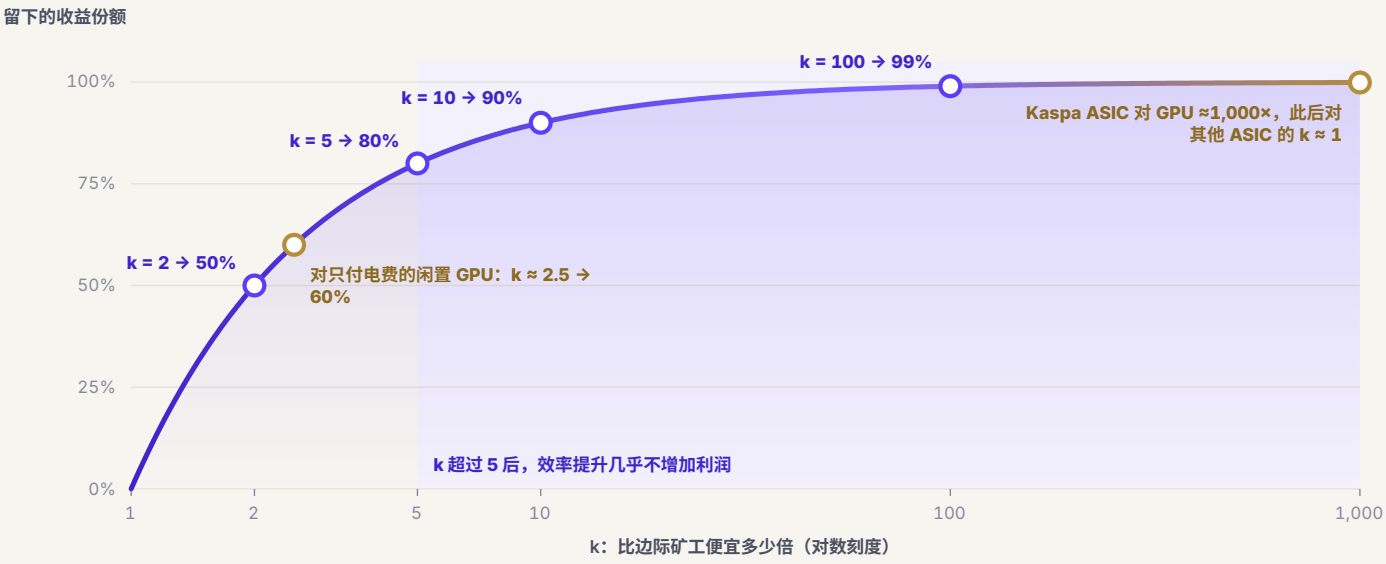
ASIC 利润上限 ≈ R × (1 − 1/k)

放大到全网，ASIC 机群的年毛利上限约为全网矿工年收入 × (1 − 1/k)，而且只在 ASIC 取代 GPU 成为边际矿工之前成立。

这个公式有三层含义。其一，(1 − 1/k) 是凹函数：k = 2 时 ASIC 留下收益的 50%，k = 5 时 80%，k = 10 时 90%，k = 100 时 99%。**k 超过 5–10 以后，效率再提升几乎不增加利润**；决定成败的是 R 与盘子的大小。其二，参照系会重置：一旦 ASIC 成为边际矿工，k 就必须相对其他 ASIC 计算。Kaspa 上 ASIC 对 GPU 约 1,000 倍的能效优势只换来一段短暂红利；如今最先进的 KS7 每天只赚 0.86 美元。其三，在 Pearl 这类 GPU 友好的链上，参照往往不是 GPU 的购置价，而是**闲置 GPU 的电费**：一张自有 RTX 5090 按 525 W、每千瓦时 0.10 美元计，电费每天约 1.26 美元，收益约 5–6 美元 笔者测算。与沉没成本的 GPU 竞争时，即便 ASIC 的能耗只有其五分之一，只要 ASIC 的折旧与电费相当，有效 k 也只有约 2.5，利润上限约为 R 的 60%（示意性假设）。

图 05　效率为何不再值钱：ASIC 最多留下 1 − 1/k 的收益

ASIC 相对边际矿工能留下的收益份额，按其成本优势倍数 k



来源： 笔者模型：边际硬件定价时，单位算力收益趋向其成本；Kaspa 与 RTX 5090 数据见报告正文。

三把尺子：盘子、回本规模与规则

第一把尺子是**盘子**。按 11 月第一周约 2,893 BTC 的区块奖励（约 2.22 亿美元）年化，比特币矿工年收入约 115 亿美元 笔者测算。Pearl 矿工约 4.6–4.9 亿美元（按现价）；Aleo 证明者奖励按 2025 年第三季度年化约 3,100 万美元；Kaspa 矿工按 1.945 KAS 的区块奖励、每秒十个区块计，约 1,800 万美元 笔者测算。**Kaspa 全网矿工一年的收入，还不到一次 5/4nm 芯片设计预算的下限（约 3,000 万美元）。**

第二把尺子是 **NRE 回本所需的算力规模**：

h* = NRE ÷ [R × (1 − 1/k) × T × p]

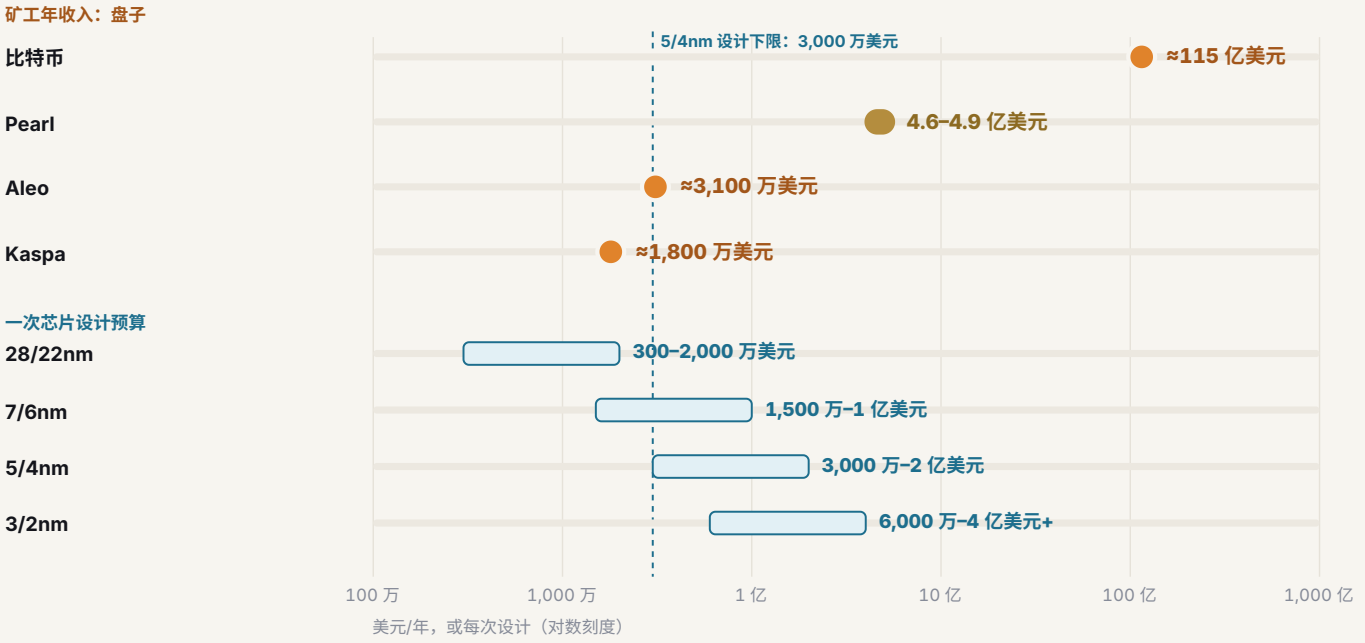
其中 T 为有效窗口天数，p 为窗口内规则不变的概率。以 Pearl 为例，取 R = 14 美元/PH/s·天、k = 5、T = 365 天、NRE 为 2,000 万美元（N7/N6 重复阵列型芯片；处于 Silicon Analysts 中端节点 500 万–3,000 万美元区间与 AnySilicon 7/6nm 计算类 ASIC 1,500 万–1 亿美元区间的偏低端，因为矿机芯片用的 IP 少、结构重复；属笔者假设）。p = 1 时需要全年稳定运行约 **4.9 EH/s**，约占全网 5%；p = 0.5 时约 **9.8 EH/s**，约 11%；若 R 回落到夏季约 10 美元的水平，即便 p = 1 也需约 6.9 EH/s 笔者测算。规模本身并非遥不可及。问题出在 p 与 T。

第三把尺子是**规则风险**，它往往一票否决。门罗币 2018 年的反 ASIC 分叉让比特大陆 X3 搁浅，全网算力一度损失近一半。Sia 同年的分叉让比特大陆与芯动的矿机变砖，而与开发方关联的 Obelisk 矿机二手价从 2,499 美元涨到 5,000 美元以上。以太坊合并一个多月后，前以太坊矿工中只有不到 20% 找到了新工作。**谁控制工作函数，谁就控制硬件投资者的命运**；Sia 的案例还说明，“抗 ASIC”可以变成开发方自己的护城河。有第二市场的硬件才能活下来：合并后 GPU 流向 AI 与云，单一算法 ASIC 则挤向以太坊经典（ETC）。窗口也在缩短：Kaspa 的首款 ASIC 用了约 16 个月，Aleo 约五个月，率先抵达效率前沿的往往是挑战者而不是在位者。

网络	全网矿工年收入（约）	边际 / 参照硬件	规则稳定性	新进入者可得的 k	判断
比特币	115 亿美元	比特大陆 S23 XP Hyd（8.9 J/TH，每 TH 25.5 美元）	极高	≤1	盘子大，但 k 归在位者
Pearl	4.6–4.9 亿美元（按现价）	自有或闲置 GPU（边际成本 ≈ 电费）与数据中心 GPU	极低：两个月四次分叉；FP8 白名单待定	能效或可达 3–10 倍 笔者分析；对沉没成本 GPU 更低	盘子够大；规则风险一票否决
Aleo	约 3,100 万美元（2025 年第三季度年化）	IceRiver AE2 等专用机	中：ARC-46 质押门槛	≈1	盘子小，门槛还在升
Kaspa	约 1,800 万美元	KS7（77 J/TH）	高	≈1	年收入低于一次 5/4nm 设计预算的下限

图 06 **多数盘子比一次芯片设计预算还小**

四个网络的矿工年收入与按工艺节点划分的单颗芯片设计成本，对数刻度



来源: Hashrate Index（比特币，2026 年 11 月）、prlstats 与 Kryptex（Pearl）、Blockworks（Aleo，2025 年第三季度）、minerstat（Kaspa）、AnySilicon（设计预算，2026）；笔者年化。

“盘子大、规则稳、k > 1”三者兼具的象限，今天是空的：比特币有盘子、规则也稳，但 k 已归在位者；Pearl 的盘子够大，规则却最不稳；Kaspa 与 Aleo 规则算稳，盘子又太小。

第 08 章

08 逐类芯片 五类芯片，五个答案

Pearl 类 PoUW：不做专用 INT7 芯片，押注“能做推理的矿机”

支持做 ASIC 的理由是真实的。今天只有约 1% 的区块承载 AI 工作，因此白皮书“专用 ASIC 失去做有用工作的能力、因而更不赚钱”这一经济学论证，恰在此时最弱。GPU 上最贵的部分——HBM 与 CoWoS 封装——约占 H100 制造成本的 63%，逻辑裸片仅约 9%。一颗只做 7 位整数乘加、操作数驻留片上 SRAM、低压运行的芯片，相对同节点 GPU 取得 3 到 10 倍的每次运算能效优势、以及更大的每次运算成本优势，在原理上是说得通的 [笔者分析；尚无实测]。反对的理由更硬。领先节点的代工交期可超过 50 周，新团队从立项到量产大约需要 18 到 24 个月 笔者测算；芯片抵达时，Pearl 很可能已迁移到 FP8，乃至 FP4。而且由于 PIP-3 以设备白名单定义准入的算术，一旦激活，INT7 芯片将直接失效。

所以我的结论是：**现在不流片专用的 INT7 Pearl 芯片。** 更好的押注有两种。一是“推理优先、可挖矿”的加速器：以通用 INT8/FP8 GEMM 和 SRAM 驻留操作数为核心，即使挖矿准入落空，价值也会落在真实存在的推理市场上。二是参与协议层讨论，推动一种厂商中立、ASIC 能精确实现的规范算术，以及算法稳定承诺，这是“盘子大、规则稳、k > 1”象限得以

出现的前提。在此之前，可以先用 FPGA 或成熟节点多项目晶圆（MPW）验证架构，把 NRE 风险控制在百万美元量级。第 10 章把这一思路落成分阶段的预算。

BTC 类矿机：k 归在位者，电力才是稀缺资产

比特币的规则风险最低，但竞争把新进入者的 k 压到 1 以下。比特大陆的 S23 XP Hyd 能效 8.9 J/TH、售价每 TH 25.5 美元，11 月起发货；旗舰能效从 2023 年 S21 的 17.5 J/TH 降到 8.9 J/TH，三年约降 49%。2027–2028 年才量产的新进入者，大约需要 8 J/TH 或更好才能立足 笔者测算。挑战者纷纷退场：嘉楠 2026 年第二季度收入 3,190 万美元、毛亏 2,930 万美元，第三季度指引仅 1,100–1,500 万美元；比特小鹿第三季度自有矿机对外销售跌至 30 万美元，集团毛利率 3.4%；Auradine 更名 Velaura AI 转向 AI 芯片；Core Scientific 取消了 Block 旗下 Proto 部门约 15 EH/s 的 3nm 矿机订单，计提 4,190 万美元损失。

需求端也在收缩。11 月 9 日哈希价格约为 31.89 美元/PH/s·天，至 2027 年 4 月的远期曲线均值约 32.7 美元，手续费只占区块奖励的 0.72%。按每千瓦时 0.045 美元计，S23 XP Hyd 约 3.1 年回本 笔者测算，已逼近预计在 2028 年前后的下一次减半。更关键的是电力的机会成本：AI 与高性能计算租约如今每合同兆瓦每年约收入 167–190 万美元，按 80% 到 85% 的净营业收入率计，约为挖矿每兆瓦毛利（约 56–90 万美元）的 1.5 到 2.9 倍 [笔者测算；两者的兆瓦口径不同]。Core Scientific 将在第四季度停止挖矿；Bitfarms（现名 Keel）已关停全部美国矿场；IREN 一个季度 AI 云收入 2 亿美元，挖矿仅 2,800 万美元。**电力与机房是资产，矿机是耗材。**

NRE 的账同样严苛。前沿能效如今需要 3nm 级工艺（据二手资料，MicroBT 采用三星 SF3，Auradine 与 Proto 的芯片均为 3nm）。掩模约 1,000–2,000 万美元，全口径 NRE 3,000 万至 1 亿美元以上；按每 TH 20–25 美元、30% 毛利计，每对外销售 1 EH/s 约得毛利 600–750 万美元，要覆盖比特小鹿量级每年约 1.1–1.6 亿美元的研发投入，需要每年对外销售约 20–25 EH/s，约占全网的 2% 笔者测算。结论：**新团队不应进入比特币商户矿机市场。** 仍说得通的只有三种情形：自带低价电力的垂直整合；家用与供暖矿机等细分形态（嘉楠也在推出家用供暖产品）；以 IP 与设计服务的方式参与在位者的产品路线。

AI 训练芯片：小团队不做

训练芯片的 k 最难获得。同代、同精度下，谷歌 Ironwood 相对 B300 的芯片级 FP8 能效优势只有约 1.1 到 1.8 倍（Ironwood 功耗为记者估计） 笔者测算；常被引用的 TPU v1 论文中的 30 到 80 倍，比较对象是没有张量核的 K80。真正的杠杆是数值格式：从 FP8 到 FP4，在同一硅片上通常带来 2 到 3 倍的吞吐，而英伟达在这条路上走得最快：它宣称 Rubin 有 50 PF 的 NVFP4 推理算力，2026 年下半年量产。成本与供给门槛更高：3/2nm 设计预算为 6,000 万至 4 亿美元以上；2026 年 CoWoS 产能约为每月 10.6–13 万片晶圆，据报台积电正考虑最高 20% 的涨价；按 HBM3E 每 GB 15.6 美元的合约价，一颗 288 GB 加速器光 HBM 就约 4,500 美元 笔者测算。这个市场属于博通式的超大规模云厂商协同开发（见第 11 章）。结论：**小团队不做独立训练芯片**，转而做 IP、chiplet 或设计服务。

AI 推理芯片：用放弃通用性换取 k，值得做

推理是 ASIC 最有空间的地方，因为 k 可以用放弃通用性来换。Etched 宣称其 8 芯片服务器运行 Llama-70B 可达每秒 50 万 token 以上，约为 8 张 H100 的 20 倍；Taalas 把模型权重硬连进掩模 ROM，宣称功耗低 10 倍、制造成本低 20 倍（均为 厂商宣称）。以 SRAM 为中心的架构绕开了 HBM 与 CoWoS，节点也更可控：Taalas 用台积电 6nm，Groq 则从 14nm 迁到三星 4nm。加密分发帮不了推理芯片，因为去中心化算力网络几乎清一色是英伟达；入口在新兴云的采购经济学，以及中国庞大的本土推理市场：国资数据中心须使用国产 AI 芯片，H200 进口被限流，北京还希望大型互联网公司在国产芯片上运行推理。结论：**值得做**，前提是有明确的客户，并主动选择对模型与架构的锁定程度。它也是 PoUW 矿机芯片最好的第二市场。

ZK 与 FHE 加速器：ZK 观望对冲，FHE 早布局

GPU 已经追上了实时 ZK 证明：Succinct 的 SP1 Hypercube 用 16 张 RTX 5090，在 12 秒内证明了 99.7% 的以太坊区块。Cysic 宣称其 ZK C1 芯片效率为 GPU 证明者的 10 到 100 倍、2026 年出货，但其需求很大程度上来自它自己的代币奖励 厂商宣称；我没有找到 Aleo 谜题 ASIC 被用于真实 ZK 证明的公开证据。ZK 证明是整数、确定性的运算，与矿机芯片所做的工作高度同构，但证明系统迭代很快（SP1、RISC Zero、Halo2、Plonky2 等）。结论：**观望与对冲**，做可编程的 MSM、NTT 与哈希引擎，先 FPGA，后 ASIC。

FHE 是最受硬件制约、上行空间最大的原语。英特尔的 Heracles 在 Intel 3 上以 197 mm²、176 W 和 48 GB HBM3，运行 FHE 运算比 24 核至强快 1,074 到 5,547 倍，但英特尔没有商业化计划，项目负责人已于 2025 年 12 月离开。Niobium 在三星 8LPU 上开发，由 SemiFive 担任设计伙伴（开发合同约 686 万美元），融资逾 2,300 万美元；采用光子方案的 Optalysys 融资 3,100 万美元，并称完整的商用芯片仍需两到三年；Zama 开源的 FPGA 方案在约 200 W 下每秒约 13,000 次可编程自举，是任何 ASIC 都必须明显超越的公开基线。结论：**早布局、轻资产验证**，窗口约在 2027–2029 年。需求更可能锚定在企业与政府的加密 AI，而非加密货币本身；而且 FHE 需要 HBM，成本结构接近训练芯片。

芯片类型	k 的来源	需求与盘子	NRE 与门槛	规则或锁定风险	第二市场	结论
Pearl 类 PoUW 矿机芯片	去掉 HBM、CoWoS 与通用单元：J/op 或达 3–10 倍 笔者分析	矿工收入每年约 4.6–4.9 亿美元，波动大	N7/N6 重复阵列约 1,000–3,000 万美元 笔者分析	极高：数值格式与设备白名单由协议方决定	AI 推理（若做成通用 INT8/FP8 GEMM）	不做专用 INT7；做“推理优先、可挖矿”，或等待规范算术
比特币 SHA-256 矿机芯片	相对比特大陆 k ≤ 1	每年约 115 亿美元，商户市场集中于两家	3nm 级掩模 1,000–2,000 万美元；全口径 NRE 3,000 万–1 亿美元以上	低（算法稳定）；竞争与减半风险高	几乎没有（单一算法）	新进入者不做商户市场；只做自带电力、细分形态或 IP
AI 训练芯片	同代脉动阵列仅 1.1–1.8 倍；格式才是杠杆	最大（博通 2026 财年 AI 半导体收入约 580 亿美元）	3/2nm 设计 6,000 万–4 亿美元以上；CoWoS 与 HBM	英伟达的年度迭代节奏	—	小团队做 IP、chiplet 与设计服务，不做独立芯片
AI 推理芯片	放弃通用性：宣称 10–20 倍 厂商宣称	大且分散；中国内地本土市场本身就很大	N6/N7–N4 的 SRAM 架构可控	模型与架构锁定	自身即主市场	值得做：先签客户，再选锁定程度
ZK 加速器	宣称为 GPU 的 10–100 倍 厂商宣称	小，主要靠代币奖励驱动	N7/N6 再到 N5；先 FPGA	证明系统迭代快	ZK 之外有限	观望对冲：可编程 MSM 与 NTT 引擎
FHE 加速器	对至强 1,074–5,547 倍（Heracles）	尚未成形；锚在企业与政府	需要 HBM；成本结构接近训练芯片	标准与生态未定	加密计算、隐私 AI	早布局、轻资产验证；2027–2029 窗口

图 07 **五类芯片，五个答案**

成本优势 k 从何而来，规则变化时硅片还能卖到哪里，以及结论



Pearl 类 PoUW 矿机芯片

暂不做

k： 无 HBM、无通用单元：每次运算 3–10×（推断）

第二市场： AI 推理（若做通用 INT8/FP8 GEMM）

押注推理优先、可挖矿的芯片，而非专用 INT7



比特币 SHA-256 矿机芯片

不进入

k： 相对在位者 ≤1

第二市场： 几乎没有

只做自带电力、细分形态或 IP



AI 训练芯片

做 IP，不做芯片

k： 同精度下 1.1–1.8×

第二市场： —

封装与 HBM 决定成败；做 IP、chiplet、服务



AI 推理芯片

值得做

k： 放弃通用性：宣称 10–20×

第二市场： 自身即主市场

先签客户，再选锁定程度



ZK 证明加速器

观望对冲

k： 宣称为 GPU 的 10–100×

第二市场： ZK 之外有限

可编程 MSM 与 NTT 引擎；先 FPGA



FHE 加速器

早布局

k： 对至强 1,074–5,547×（Heracles）

第二市场： 加密计算、隐私 AI

2027–2029 窗口；轻资产验证

来源： 笔者对第 7、8 章的综合；厂商数据按报告中的标注。

第 09 章

09

工艺节点

工艺节点：让终生成本的最大项来选

成本阶梯：领先节点晶圆已是成熟节点的十倍

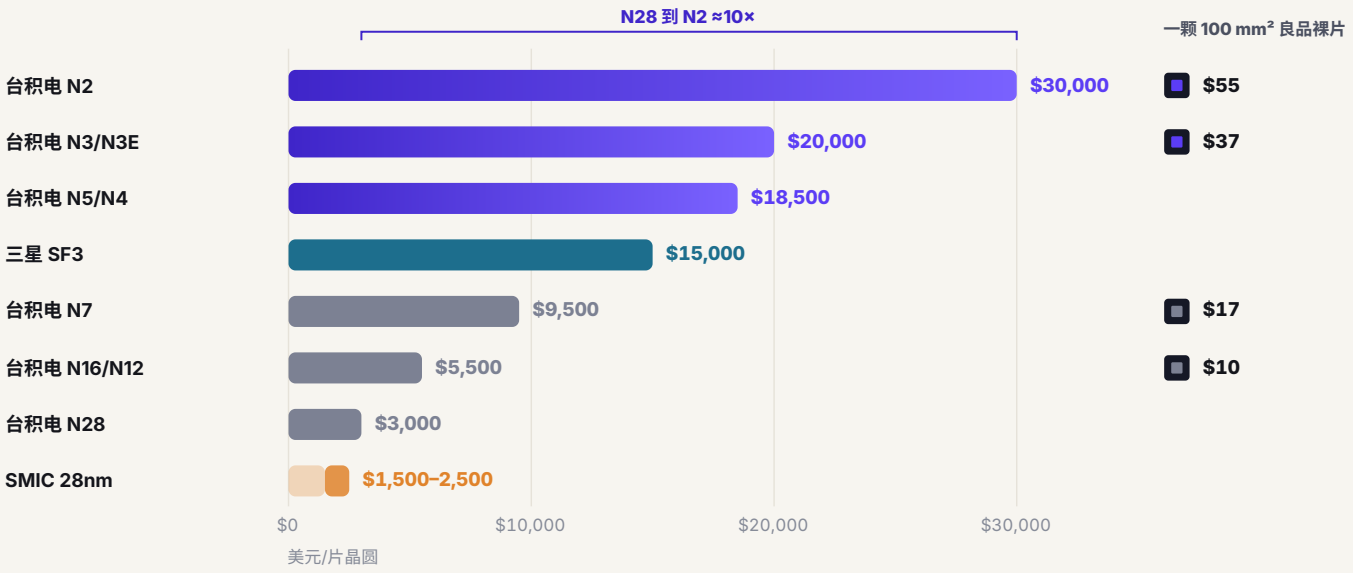
先看账本。领先节点的晶圆如今约为成熟节点的十倍：台积电 N2 约 30,000 美元、N3 约 20,000 美元、N7 约 9,500 美元、N28 约 3,000 美元；三星 SF3 约 15,000 美元，比 N3 便宜约 25%；SMIC 7nm 级的制造成本据报比台积电高 40% 到 50%。据报台积电将自 2027 年 1 月起再涨价 3% 到 6%，2nm 与 3nm 涨幅最大。掩模与设计成本的差距更大，而且不同来源互有出入，下表同时列出。

节点	晶圆价格（估算）	掩模组（模型值 / 指南区间）	设计预算（AnySilicon）	MPW（每项目）
台积电 N2	约 30,000 美元	—	3/2nm：6,000 万–4 亿美元以上	—
台积电 N3/N3E	约 20,000 美元（18,000–27,000 美元）	1,500 万美元 / 1,000–2,000 万美元	同上	—
三星 SF3	约 15,000 美元（13,000–17,000 美元）	1,200 万美元	—	—
台积电 N5/N4	约 18,500 美元（16,000–21,300 美元，低置信度）	650 万美元 / 1,000–2,000 万美元	5/4nm：3,000 万–2 亿美元	—

节点	晶圆价格（估算）	掩模组（模型值 / 指南区间）	设计预算（AnySilicon）	MPW（每项目）
台积电 N7	约 9,500 美元（8,100–10,900 美元）	350 万美元 / 500–1,000 万美元	7/6nm：1,500 万–1 亿美元	商用 10–50 万美元
台积电 N16/N12	约 5,500 美元	180 万美元	16/12nm：600 万–4,000 万美元	学术约每 mm² 35,000 美元
台积电 N28	约 3,000 美元	100 万美元 / 100–300 万美元	28/22nm：300 万–2,000 万美元	商用 5–15 万美元；学术约每 mm² 12,000 美元
SMIC 28nm	1,500–2,500 美元	—	—	—
SMIC 7nm 级（N+2/N+3）	较台积电高 40%–50%	—	—	—

图 08 **领先节点晶圆的价格是成熟节点的十倍**

各节点 300 mm 晶圆估算价格，以及 85% 良率下一颗 100 mm² 良品裸片的成本



来源: Silicon Analysts 晶圆价格（估算，N5 低置信度）、Wccftech（SMIC）；裸片成本为笔者按假设良率计算。

再看物理。Horowitz 的经典测算表明，一条可编程指令约耗 70 pJ，而运算本身只要几皮焦：**算术便宜，开销昂贵**，这是脉动阵列与专用芯片存在的根本理由。特征尺寸缩小一半，每次运算的能耗大约减半，但供电电压已停滞在 0.9–1.1 V 附近。N2 晶圆比 N3 贵 50% 到 67%，除非密度提升超过约 1.5 倍，单位晶体管成本将不降反升 笔者测算：领先节点只能靠能效与性能来买单。以 100 mm² 裸片、85% 良率计，每颗良品裸片的成本约为：N2 55 美元、N3 37 美元、N5 29–34 美元、N7 17 美元、N16 10 美元 [笔者测算；良率为假设]。

按芯片类型与团队情况选节点

我的原则是：**让芯片终生成本中最大的那一项来选节点。** 7×24 满载运行的算力芯片（挖矿、ZK）由电费主导终生成本，算法一旦稳定，就应向 N5/N4 和 N3/SF3 推进；算法未定或资金有限时，先用成熟节点 MPW 验证。训练芯片的瓶颈是封装与存储（HBM 与 CoWoS 约占 H100 物料成本的 63%），节点反而是次要问题。推理芯片的首要决策是存储架构：以 SRAM 为中心的设计可以留在 N6/N7 至 N4，HBM 方案则继承训练芯片的成本结构。FHE 受内存带宽约束，绕不开 HBM。

芯片类型	初创 / 小团队	规模化公司	中国内地路径	要点
矩阵乘法 PoUW 矿机芯片（SRAM 驻留、无 HBM）	28nm 或 16/12nm MPW 原型；首批量产 N7/N6；算法与收益稳定后上 N5/N4；裸片 50–150 mm²，常规倒装封装	N3/N3E 或三星 SF3	SMIC N+1/N+2 可行（不需要 HBM 或 CoWoS），每颗良品裸片成本约高 40%–50%	能效主导；协议规则是最大风险
SHA-256 矿机芯片	不追前沿；细分形态可用成熟节点	3nm 级；在位者从台积电 5nm 走向 2nm、从三星 SF3 走向 2nm（二手资料）	能效难以对标领先节点	电价与减半周期决定利润率
AI 训练芯片	不做独立芯片：IP、chiplet、设计服务	N3P 再到 N2/A16；2–8 颗 chiplet；提前锁定 CoWoS 与 HBM	SMIC N+2/N+3 配国产 HBM（华为模式）；单芯片 FP8 约为英伟达的 0.2–0.5 倍，以系统规模补偿	封装与 HBM 是瓶颈
AI 推理芯片	SRAM 为中心或硬连线，N6/N7 或 N4/N5/SF4；只做 INT8、FP8 与 FP4	N3/N2 配 HBM（长上下文、混合专家），或多 chiplet SRAM	SMIC N+2 配 LPDDR 或 GDDR；HBM 级内存是卡点	架构锁定程度是核心决策
ZK 加速器	从 FPGA 或 16/12nm MPW 起步；再到 N7/N6 与 N5	N5/N3	SMIC N+1/N+2 可行	整数、确定性；易于验证
FHE 加速器	FPGA（对标开源基线）或 16/12nm MPW；量产需要 HBM	配 HBM 的 N5/N3	受 HBM 获取限制	决定速度的是内存带宽，而非算术

先进还是成熟、四类代工厂与 chiplet

方案	主要优点	主要缺点	更适合
领先节点（N3/N2、SF3/SF2、Intel 18A）	每个全节点约让每次运算能耗减半；SRAM 密度高	晶圆 3–10 倍、掩模 5–15 倍、设计预算 10–20 倍于成熟节点；交期逾 50 周；N2 单位晶体管成本或上升	算法稳定、全天候运行的算力芯片；训练芯片
成熟与中端节点（N7/N6、N16/12、N28）	掩模在百万美元量级，MPW 便宜，迭代快，代工厂选择多（台积电、联电、格芯、SMIC）	全天候满载时能效缺乏竞争力	原型验证、细分市场、边缘推理
台积电	良率与生态最佳	价格最高；N2/N3 产能紧张；CoWoS 排队时间长	追求能效前沿、有量产承诺的团队
三星代工	晶圆便宜约 20%–25%（SF3 约 15,000 美元 vs N3 约 20,000 美元）；Groq 第二代等已采用	代工业务亏损，目标 2027 年底盈亏平衡	成本敏感的算力芯片
英特尔代工	Intel 3 已有 Heracles 这类复杂芯片验证	18A 未公开对外报价与客户数据；代工业务亏损	需逐项评估生态与支持
中国内地代工厂（SMIC 等）	贴近庞大的本土市场；28nm 晶圆 1,500–2,500 美元	7nm 级产能约每月 45,000–60,000 片，多被华为吸收；良率约为台积电的三分之一；成本高 40%–50%；仅 DUV	面向本土市场的推理、矿机与 ZK 芯片
chiplet	突破光罩与良率限制；是 HBM 系统的标配（MI355X、Ironwood、Trainium3）	增加封装成本与 CoWoS 依赖；已知良品裸片（KGD）测试增加成本	训练与大型推理芯片；小型矿机或 ZK 芯片不需要

第 10 章

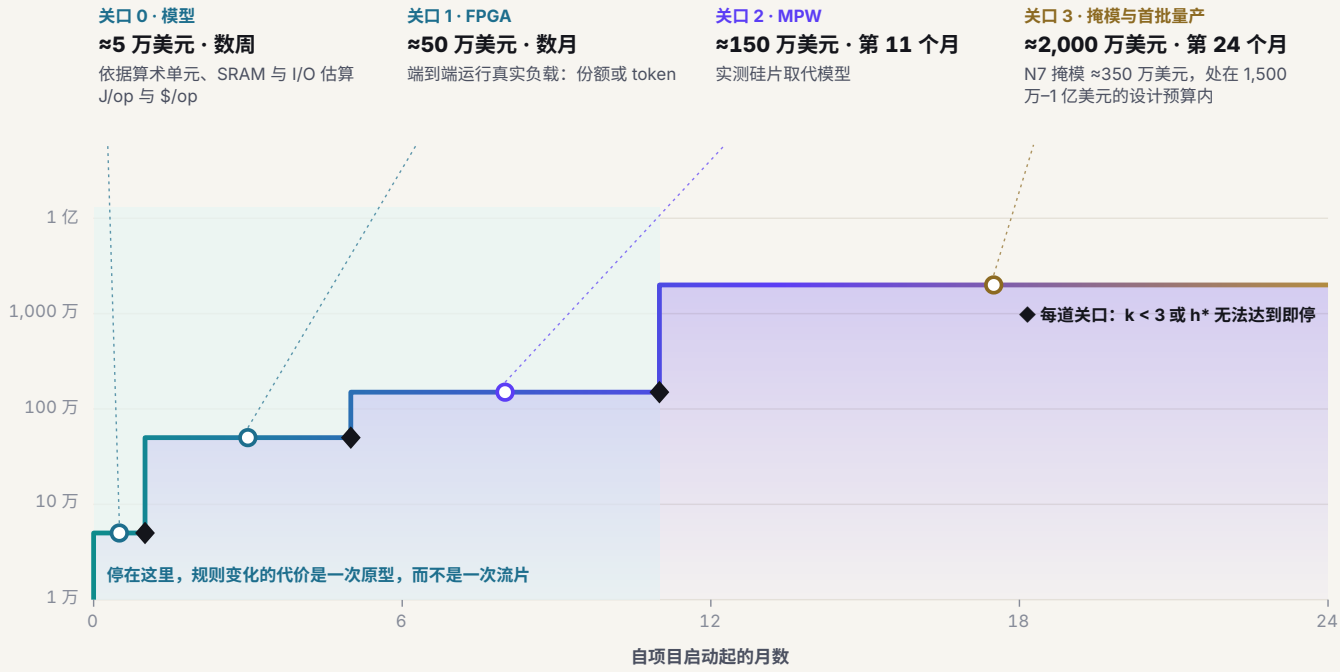
10 分阶段预算 从模型到硅片：第一颗芯片的分阶段预算

先买证据，再买掩模

一颗全定制芯片是一连串赌注，一注比一注大，纪律在于先买证据、再买掩模。我用四道关口 笔者分析。关口 0 是模型：一项架构研究，依据算术单元、SRAM 与 I/O 估算 J/op 与 $/op，并与第 7 章的参照硬件对比；耗时数周，花费数万美元。关口 1 是 FPGA 原型：端到端运行真实负载，产出挖矿份额或推理 token，验证数据通路与协议对接；耗时数月，花费数十万美元。关口 2 是多项目晶圆：商用名额在 N28 约 5 万–15 万美元，在 N7 约 10 万–50 万美元，实测硅片取代模型的能耗估算。关口 3 是全套掩模与首批量产：掩模在 N28 约 100 万美元、N7 约 350 万美元、N3 约 1,500 万美元，处在 7/6nm 1,500 万–1 亿美元的设计预算之内。每道关口都需要一个让项目停下的数字，比如相对边际硬件的 k 低于 3，或者窗口内无法达到的回本算力 h*。

图 09 **先买证据，再买掩模：第一颗芯片的分阶段预算**

一个 N7/N6 项目经过四道关口的累计支出（美元，对数刻度）；每道关口都有一个让项目停下的数字



来源: 晶圆、MPW 与掩模价格同第 9 章（Silicon Analysts、AnySilicon、模型值）；各关口的时间与支出为笔者的示意性假设。

这道阶梯很陡。在一个 N7/N6 项目里，前三道关口合计约 150 万美元、约十一个月；第四道的花费是其十倍以上 笔者测算。所以顺序比节点更重要：如果协议改了算术，或者买家走了，停在关口 2 的团队损失的是一次原型的钱，而不是一次流片的钱。

每类芯片要花多少钱，什么条件必须成立

项目	关口 2（原型）	关口 3（节点与设计预算）	一颗 100 mm² 良品裸片	回本必须成立的条件
推理优先、可挖矿的加速器	16/12nm 或 N7 MPW，10 万–50 万美元	N6/N7，约 1,500 万–3,000 万美元 笔者测算	N7 约 17 美元，不含封装与存储	已签约的推理客户承担大部分 NRE；挖矿收入是上行空间，而非基本盘
专用 Pearl INT7 矿机（目前不推荐）	28nm MPW，5 万–15 万美元	N7/N6，约 2,000 万美元	N7 约 17 美元	p = 1 时全年约 4.9 EH/s，p = 0.5 时 9.8 EH/s（见第 7 章）
ZK 加速器	FPGA，再到 16/12nm MPW	N7/N6，再到 N5	N16 约 10 美元，N7 约 17 美元	证明系统收敛到固定设计能服务的 MSM 与 NTT 内核
FHE 加速器	对标开源基线的 FPGA	配 HBM 的 N5/N3，3,000 万美元以上	由 HBM 与封装主导	出现规模化的加密 AI 买家，而这样的买家尚不存在

由此可得两点商业结论。第一，最便宜的能效来自架构：让操作数驻留 SRAM、去掉 HBM 与 CoWoS，在选择任何节点之前就已省掉 GPU 的大部分成本（约占 H100 物料成本的 63%）。第二，时间就是成本：新团队从立项到量产大约需要 18 到 24 个月，领先节点交期超过 50 周，一个每隔几个月就改算术的协议，可能让芯片在关口 3 回本之前就已过时。

小团队的三条规则

综合这些数字，可以得到三条规则 笔者分析。第一，**在出现具名买家或冻结的协议版本之前，停在关口 1 或 2**；代币价格和全网算力都替代不了这两者中的任何一个。第二，**先用架构买 k，再用节点买 k**：针对单一负载，N7 上的 SRAM 驻留阵列可以胜过 N3 上的通用芯片，掩模成本却只是零头。第三，**流片之前写下带 p 的回本算力或销量**，并以固定的协议版本为里程碑付款定价，让客户而不是设计公司承担规则风险。

第 11 章

11 机会在哪里，我会怎么做

定制芯片：需求压倒性地来自 AI

2026 年的定制芯片需求压倒性地来自 AI。博通 2026 财年第三季度 AI 半导体收入 167 亿美元，同比增长 221%，全年约 580 亿美元，其中 73% 来自面向六家客户的 XPU（据财报电话会的自动转写稿）。创意电子第二季度收入同比增长 128%，量产交付业务收入增长 188%，毛利率 21.5%。世芯第一季度收入的节点结构为 3/2nm 21%、7/5nm 63%、12nm 及以上 16%。芯原 2026 年新签订单至 4 月底已达 82.4 亿元人民币（约 11.6 亿美元），其中 91.37% 为 AI 算力。市场分三层：顶层是价值数十亿美元的多年期协同开发；中层是 7nm 至 2nm 的项目，NRE 从数百万到上亿美元不等（中端节点 500 万–3,000 万美元，领先节点 3,000 万–1 亿美元以上）；长尾是 NRE 50 万–500 万美元的成熟节点 ASIC。加密与挖矿客户只是一个小而周期性的利基，但它是今天为算力付费的、最快也最无需许可的“第一个客户”。

矿机芯片团队最可迁移的能力——近阈值低压运行、海量复制的运算阵列、千瓦级供电与散热、快速的领先节点流片——恰是推理芯片与 ZK、FHE 加速器所需要的。资本市场愿意为此定价：Velaura AI（原 Auradine）8 月以超过 10 亿美元的估值融资 1.1 亿美元。反面教材是嘉楠的边缘 AI 芯片部门：2024 财年收入 90 万美元，却消耗了 15% 的运营开支，于 2025 年 6 月关闭。两者的差别在于独立的市场进入路径，而不是复用挖矿渠道。

AI 开发：工时已商品化，价值在结合部

AI 开发服务在低端已经商品化：Clutch 上的典型费率为每小时 25–49 美元，美国、加拿大等市场为 50–99 美元；一支区块链开发团队每月开支约 15,000 美元至 90,000 美元以上。卖工时没有护城河。价值在各领域的结合部：为定制芯片写算子与编译器，把挖矿嵌进推理服务（Pearl 的 vLLM 插件就是一例），把 TEE 与 ZK 抽检组合成可验证推理，为智能体接入支付与身份。

AI × Crypto 系统：先签算力买家，再发币

如果要设计一条 AI 原生的链或一张算力网络，2026 年的验证工具箱已经成形。按单位工作的成本由低到高，依次是统计抽样（Hyperbolic 的抽样证明）、裁判式委托（Gensyn 的 Verde：只需一名诚实验证者，但要求逐位确定的执行）、TEE（接近原生速度，但要信任芯片厂商）、全量冗余，以及通用 zkML。务实的设计是分层组合：TEE 或抽样负责低延迟，争议博弈或 ZK 抽检负责最终性；一篇 2025 年 12 月的论文在模拟中以 H100 机密计算加欺诈证明与随机 ZK 抽检，达到集中式吞吐的 99%，每次查询边际成本 0.07 美元。硬件哲学决定安全性与分发：GPU 原生的 PoUW 在起步阶段由可租用的云 GPU 保护，攻击成本与租价挂钩；Qubic 则说明，通用硬件上的工作量证明可能被外部激励整体收编。

成本与周期并不是门槛。Pearl 从论文（2025 年 4 月）到主网约 12 个月，基于 btcd 分叉，上线即提供全节点、钱包、轻客户端与矿工程序。多数 DeFi 审计费用为 25,000–100,000 美元，企业级或多链审计为 150,000–250,000 美元以上，ZK 电路贵 80% 到 120%，有实质资金风险的项目每年安全预算为 150,000–500,000 美元。真正的门槛是分发与流动性。有创始人反映，币安上币曾要求 0% 至 3.5%、甚至据称 8% 的代币供应量（币安否认从上币中获利）；Cointelegraph 审视的六个采用“代币借贷 + 看涨期权”做市协议的项目全部下跌，Movement 的做市商在上币当日卖出约 3,800 万美元代币。替代路径有两条：一是链上流动性，比如 Hyperliquid，未上币安即突破 100 亿美元市值，其 HIP-3 永续合约市场的未平仓量在 6 月突破 32 亿美元；二是通过大型交易所公开发售，比如 Monad 经 Coinbase 募得 2.69 亿美元。高 FDV、低流通，是 2025–2026 年发行后回撤最清晰的预测因子。

代币设计是系统设计的一部分。排放应当跟随付费需求，而不是走在需求前面：本报告中补贴倍数最健康的网络，把奖励与外部买家付费的工作挂钩；最差的网络，则为背后空无一物的算力或质押付钱。一条务实的规则是：在上线前公布网络在第一、二、三年的目标补贴倍数，未达标就削减排放 笔者分析。

方向	需求方与付费意愿	时间窗口	关键能力	主要风险
推理优先、可挖矿的加速器	新兴云、私有部署、PoUW 网络	2027–2028（视 FP8 与 FP4 分叉而定）	INT8、FP8 与 FP4 GEMM；SRAM 架构；协议对接	协议准入、架构锁定
SRAM 或硬连线推理芯片	新兴云、中国内地数据中心、边缘	当前	编译器与模型适配；N6–N4 设计	模型迭代过快
矿机芯片（细分或垂直整合）	自有电力的矿企；家用与供暖矿机	当前，受减半与哈希价格约束	低压阵列、能效、量产供应链	在位者价格战、电价
ZK 证明加速	证明者网络、rollup	观望至证明系统收敛	可编程 MSM 与 NTT	需求依赖代币奖励
FHE 加速	企业与政府的加密 AI；隐私链	2027–2029	NTT、HBM 系统、密码工程	需求未成形；销售周期长
TEE 加固与可验证推理	机密推理网络、DePIN	当前	带完整性的内存加密、独立的证明根、ZK 抽检	依赖厂商生态
AI 原生链或算力网络	推理买家、矿工、开发者	当前	验证机制、排放设计、分发	发行后回撤、代币流动性

我会走的三条路，各有第一步

如果必须选择，我最多走三条路，每条都先用一个小的第一步在投入资金之前加以检验 笔者分析。

- **路径 A：推理优先、可挖矿的加速器。** 第一步：用四周做一次架构研究，针对 Pearl 的工作和 Transformer 推理两种负载，对比 SRAM 驻留的 INT8/FP8 GEMM 阵列与 RTX 5090、H100 的 J/op 与 $/op；随后以 FPGA 或成熟节点 MPW 验证，NRE 控制在百万美元量级。
- **路径 B：以 Crypto 为第一个客户、以 AI 为第二市场的定制芯片服务。** 第一步：把本报告的 NRE 与盈亏平衡模型做成面向算力芯片项目的付费可行性研究，按里程碑付款，并以固定的协议版本作为验收依据。
- **路径 C：先签算力买家、再发币的 AI 原生网络。** 第一步：在做任何代币设计之前，与一家具名的推理买家共同撰写一份设计说明，内容涵盖分层验证（TEE + 抽样 + ZK 抽检）与厂商中立的规范算术。

第 12 章

12 观察信号 2026 年 11 月至 2027 年值得关注的信号

时间	关注什么	为什么重要
随时	Pearl 的 FP8 激活高度	决定整数硬件与非英伟达显卡能否继续准入，以及链是否分裂
每月	Pearl 承载 AI 工作的区块占比	检验“有用”是否正在成真的最佳单一指标
每月	推理服务商是否向客户让渡挖矿收益	付费需求与排放相遇：Pearl 需要的模式
每季度	Bittensor 可验证收入相对其排放	最大的有用工作量网络是否在缩小补贴缺口
每季度	x402 中真正由智能体发起的流量占比	Agent 支付能否从信号走向商业
2027 年 1 月	据报台积电涨价 3% 到 6%，2nm 与 3nm 涨幅最大	领先节点流片的成本
2027 年第一季度	Pearl 引入 FP4（据创始人 AMA 的报道）	工作进一步转向英伟达最新的数值格式
2027 年底	三星代工的盈亏平衡目标	更便宜的第二货源能否保持健康
2028 年 4 月	比特币下一次减半	SHA-256 矿机的盘子减半

第 13 章

13 该问的问题
你没有问、但应该问的问题

下面这些问题，是我在把资金或硅片投入任何 AI × Crypto 项目之前会问的。

谁控制工作函数？ 如果一个小团队可以在硬编码高度改变算术，你的硬件寿命就由他们决定。要么索取书面的稳定承诺，要么设计有第二市场的硬件。

补贴倍数是多少？ 用排放（或市值）除以网络外部支付的收入。倍数超过 10，意味着网络在购买活跃度，而不是在出售服务。

下一道关口要花多少钱，哪个数字会让你停下？ 一次多项目晶圆要花数万到数十万美元；一套掩模要花数百万美元。花钱之前就定好止损数字。

回本算力是多少，p 取多少？ 用 NRE 除以 R × (1 − 1/k) × T × p。如果结果占全网的很大份额，这颗芯片押的是规则，而不是硅片。

芯片的第二市场是什么？ 如果协议分叉或代币下跌，这块硅还能卖到哪里？GPU 熬过了以太坊合并，因为它们有第二市场；单一算法 ASIC 没有。

价值落在代币里还是股权里？ 2026 年赚钱的 AI × Crypto 公司，多是带加密分发的股权公司；在融资之前，先决定你在建的是哪一种。

第一个付费客户是谁，叫什么名字？ 本报告中每一个持久的案例，从 Grass 到 Bittensor 的算力子网，都始于网络之外一个具名的买家。

这些是我希望与大家通过技术交流与技术合作共同探索的方向：面向挖矿、AI 训练与推理的定制芯片及各类加速器；AI 研发；以及从链到节点再到芯片的 AI × Crypto 系统。欢迎通过 i@tanshuai.com 与我联系。

本报告的局限

这是一项基于截至 2026 年 11 月 13 日公开文件的案头研究。本版没有做一手访谈，也没有做独立测量。若干关键数字是公司或厂商的自报，如 Aethir 的收入、Grass 的收入、Etched 与 Taalas 的性能以及 Cysic 的效率；文中均已注明。Pearl 的链上指标来自独立全节点与矿池，按小时变化；我采用 11 月 13 日 07:13 UTC 前后的读数，并给出同日区间。晶圆、掩模与设计成本是第三方估算与模型值，彼此并不一致；我给出的是区间。我的 ASIC 盈亏平衡测算与分阶段预算使用了示意性假设（k、T、p、NRE 与各关口成本），读者应换成自己的数值。

关于作者

谭帅是一名技术专家与创业者，在芯片、硬件与系统软件领域工作二十余年，拥有 20 项已授权专利。他有多年无晶圆厂芯片研发与管理经验，包括加密货币矿机芯片与多媒体摄像头芯片的全定制 ASIC 数据通路设计，这也是本报告的视角：每次运算的能耗、每片晶圆的成本、协议允许你把什么固化进硅片，以及谁来付钱。他的工作经历遍及硅谷、纽约、芝加哥、台湾地区、大湾区（深圳与香港）、越南和马来西亚。如今他是一名 AI 研发者。

技术交流：i@tanshuai.com · tanshuai.com

说明与免责声明

本报告是基于截至 2026 年 11 月 13 日可得公开文件的独立案头研究，仅供信息与学习之用，不构成投资或工程建议。本报告只讨论技术与商业。价格、链上指标与协议规则都会变化；凡需依赖的信息，请以原始来源核实。标注为 厂商宣称 、 笔者测算 或 笔者分析 的数字，请按其标注理解。加密资产价格波动剧烈。

© 2026 谭帅。欢迎在注明出处的前提下原样分享本报告。

来源

全部访问于 2026 年 11 月 10 日至 11 月 13 日之间，按章节分组。

Pearl：协议、代码与链上数据

- Komargodski and Weinstein, "Proofs of Useful Work from Arbitrary Matrix Multiplication", arXiv 2504.09971: https://arxiv.org/abs/2504.09971
- Pearl Research Labs, INT whitepaper: https://pearlresearch.ai/research/int-whitepaper
- Pearl 源代码：创世区块：https://github.com/pearl-research-labs/pearl/blob/main/node/chaincfg/genesis.go ；共识参数与分叉高度：https://github.com/pearl-research-labs/pearl/blob/main/node/chaincfg/params.go ；算力计量：https://github.com/pearl-research-labs/pearl/blob/main/zk-pow/src/api/sanity_checks.rs ；盐化种子分叉升级指南：https://github.com/pearl-research-labs/pearl/blob/main/docs/salted-seed-fork-upgrade-guide.md ；fp8-scheme 分支参数：https://github.com/pearl-research-labs/pearl/blob/fp8-scheme/node/chaincfg/params.go
- Pearl 改进提案（PIP）：https://github.com/pearl-research-labs/pips ；PIP-3，浮点挖矿：https://github.com/pearl-research-labs/pips/blob/main/PIPS/pip-0003.md
- prlstats（独立全节点）：网络：https://prlstats.com/api/network ；序列：https://prlstats.com/api/series ；首页与矿工收入模型：https://prlstats.com/api/home ；已挖供应：https://prlstats.com/api/mined ；矿池：https://prlstats.com/api/pools ；持币地址：https://prlstats.com/api/holders ；硬件与开销假设：https://prlstats.com/api/assumptions
- Kryptex Pool：Pearl 网络：https://pool.kryptex.com/prl/about-coin ；算力：https://pool.kryptex.com/prl/hashrate ；RTX 5090：https://pool.kryptex.com/de/device/gpu/nvidia/nvidia-geforce-rtx-5090 ；RTX 4090：https://pool.kryptex.com/device/gpu/nvidia/nvidia-geforce-rtx-4090 ；"Pearl FP8 fork"：https://pool.kryptex.com/articles/pearl-fp8-fork-en
- CoinGecko，Pearl：https://api.coingecko.com/api/v3/coins/pearl-2 ；https://www.coingecko.com/en/coins/pearl-research ；价格历史：https://api.coingecko.com/api/v3/coins/pearl-2/market_chart?vs_currency=usd&days=365&interval=daily ；“Perle”代号搜索：https://api.coingecko.com/api/v3/search?query=perle
- SimeonOnSecurity, "Mining Pearl (PRL) in 2026": https://simeononsecurity.com/articles/pearl-prl-mining-guide-2026/
- Hashrate Index, "Pearl (PRL): Inside the AI-Compute Cryptocurrency"，2026 年 6 月 2 日：https://hashrateindex.com/blog/pearl-prl-ai-compute-cryptocurrency/
- Tom's Hardware, "New AI compute cryptocurrency Pearl sparks a GPU mining rush, but profitability is sliding": https://www.tomshardware.com/tech-industry/cryptomining/new-ai-compute-cryptocurrency-pearl-sparks-a-gpu-mining-rush-but-profitability-is-sliding
- Together AI，与 Pearl Research Labs 的合作：https://together.ai/blog/together-ai-partners-with-pearl-research-labs ；模型页面：https://together.ai/models/gemma-4-31b-it-pearl
- Basu，Pearl 挖矿实例普查，arXiv 2606.04819：https://arxiv.org/html/2606.04819v2
- Blockworks，0xResearch 通讯，2026 年 10 月 28 日：https://blockworks.com/newsletter/0xresearch/issue/post_daf0d74a-2bf4-41e4-86d3-2a1275475a24
- Datawallet, "Pearl Explained": https://www.datawallet.com/crypto/pearl-explained
- Coingabbar 关于 TheMarker 报道 Pearl Research Labs 融资：https://www.coingabbar.com/en/pearl-research-labs-funding-130m-stealth-round-themarker

有用工作量类链与历史

- Grothe et al.，Gridcoin 安全分析，USENIX WOOT 2017：https://www.usenix.org/system/files/conference/woot17/woot17-paper-grothe.pdf
- 学习证明伪造，arXiv 2208.03567：https://arxiv.org/abs/2208.03567
- hoge.gg，去中心化推理中的验证与信任：https://hoge.gg/decentralized-inference-2026-verification-trust/ ；Bittensor 收入与金库：https://hoge.gg/bittensor-quantum-wallet-revenue-tao-treasury-2026/
- Kaspa：kHeavyHash 矩阵代码：https://github.com/kaspanet/rusty-kaspa/blob/master/consensus/pow/src/matrix.rs ；启动：https://coinranking.com/coin/V8GxkwWow+kaspa-kas/about ；IceRiver KS1/KS2 发布：https://www.iceriver.io/2023/03/14/1677644580/ ；从 GPU 到 ASIC 的历史：https://captainmining.com/blogs/crypto-news/the-evolution-of-kaspa-mining-from-gpus-to-asic-powerhouses ；GPU 算力：https://coincalculators.io/algorithms/kheavyhash ；KS7：https://www.asicminervalue.com/miners/bitmain/antminer-ks7-40th ；网络：https://minerstat.com/coin/kas ；2024 年的 1 EH/s：https://ourcryptotalk.com/news/kaspa-reaches-1-ehs-in-hashrate ；Crescendo：https://ourcryptotalk.com/news/kaspa-crescendo-live
- Aleo：主网 FAQ：https://aleo.org/post/aleo-mainnet-faq/ ；去中心化证明：https://aleo.org/post/decentralized-proving-advantages/ ；ASIC：https://whattomine.com/coins/432-aleo-zksnark/asics ；Blockworks, "State of Aleo Q3 2025"：https://app.blockworks.com/report/state-of-aleo-q3-2025 ；ARC-46：https://chaincatcher.com/en/article/2144326
- Qubic：The Block 关于门罗币重组：https://www.theblock.co/post/366535/monero-faces-chain-reorganization-fears-after-qubic-says-it-controls-51-of-hashrate ；Halborn：https://www.halborn.com/blog/post/explained-the-monero-51-percent-attack-august-2025 ；狗狗币挖矿：https://hackernoon.com/qubic-is-now-mining-dogecoin-on-its-ai-compute-infrastructure ；CoinGecko：https://www.coingecko.com/en/coins/qubic
- Bittensor 共识文档：https://www.bittensor.com/llms.mdx/docs/internals/consensus/content.md ；FalconX, "State of Bittensor"：https://www.falconx.io/newsroom/state-of-bittensor-subnet-adoption-trends-network-mechanics-and-covenants-departure

- Primecoin：https://en.wikipedia.org/wiki/Primecoin ；Blockworks, "State of Filecoin Q3 2025"：https://app.blockworks.com/report/state-of-filecoin-q3-2025
- 门罗币与抗 ASIC 分叉：https://lbry.com/news/forking-for-asic-resistance ；Sia 分叉：https://hotfix.coindesk.com/markets/2018/11/01/sia-completes-hard-fork-as-bricked-miners-form-rival-blockchain ；Obelisk 价格：https://www.cryptoglobe.com/latest/2018/11/obelisk-hardware-skyrockets-in-selling-price-after-sia-hard-fork/ ；合并后的矿工：https://hotfix.coindesk.com/business/2022/11/11/ex-ethereum-miners-token-hop-to-stay-alive-after-the-merge

AI × Crypto 各赛道

- DefiLlama：Render：https://defillama.com/protocol/render-network-bme ；Livepeer：https://defillama.com/protocol/livepeer
- Cointelegraph 关于 Messari 的 DePIN 收入数据：https://cointelegraph.com/news/depin-revenues-grow-despite-token-losses
- Aethir，2025 年回顾：https://aethir.com/blog-posts/aethirs-2025-wrap-up-decentralized-gpu-cloud-milestones
- CoinGecko，AI 板块：https://www.coingecko.com/en/categories/artificial-intelligence
- Akash，H100 租赁价格：https://akash.network/the-bid/h100-rental-price-2026-cost-per-hour/ ；2026 年第一季度报告：https://akash.network/blog/akash-network-q1-2026-report/
- io.net，800 万美元合同与每日 token 量：https://lifestyle.kbew98country.com/story/271004/io-net-to-burn-up-to-12m-tokens-as-network-closes-8m-deal-and-hits-4-billion-daily-ai-tokens/
- DCD，Axe Compute 合同：https://www.datacenterdynamics.com/en/news/axe-compute-secures-15bn-compute-contract-with-unnamed-customer/
- Pluralis-8B：https://www.alphaxiv.org/abs/2607.13332.md ；Nous Research，Psyche：https://nousresearch.com/the-next-phase-of-psyche ；Implicator 关于 INTELLECT-3：https://www.implicator.ai/prime-intellects-intellect-3-open-source-ambition-meets-centralized-reality/
- SiliconANGLE，Prime Intellect 融资：https://siliconangle.com/2026/07/08/prime-intellect-raises-130m-1b-valuation-ai-training-platform/ ；The Block，Venice：https://www.theblock.co/post/406934 ；RuntimeWire，Nous 融资：https://runtimewire.com/article/nous-research-75-million-round-1-5-billion-valuation
- ownyourmind，Gensyn：https://ownyourmind.ai/projects/gensyn/ ；Phemex，Ritual（仅见标题）：https://phemex.com/news/article/ritual-launches-ai-layer-1-mainnet-and-ritual-token-price-dives-72-from-high-101259
- 机密计算开销，arXiv 2409.03992：https://arxiv.org/abs/2409.03992v2 ；Phala 2025 年报告：https://phala.com/posts/phala-2025-report ；TEE.fail：https://thehackernews.com/2025/10/new-teefail-side-channel-attack.html ；DeepProve，IACR ePrint 2026/1112：https://eprint.iacr.org/2026/1112
- TRM Labs，AI 智能体支付测量：https://www.trmlabs.com/trm-tech-blog/whos-actually-paying-measuring-ai-agent-payments-onchain ；Forkast 关于 x402 基金会：https://forkast.news/the-x402-foundation-activated-a-27-year-old-internet-code-200-million-transactions-later-the-real-volume-is-still-tiny/ ；The Block，Tempo 主网：https://www.theblock.co/amp/post/394131/tempo-mainnet-goes-live-with-machine-payments-protocol-for-agents ；BlockEden 关于 Virtuals 的 aGDP：https://blockeden.xyz/blog/2026/04/21/agdp-agent-gdp-virtuals-protocol-ai-blockchain-valuation-primitive-tvl-displacement/ ；The Block 关于 ElizaOS：https://theblock.co/post/410774/eliza-labs-native-token-dead
- Dealroom，Grass 收入：https://dealroom.co/news/157794-grass-backed-by-multicoin-projects-75m-in-2026-data-revenue/ ；CryptoSlate，Fetch.ai 与 Ocean：https://cryptoslate.com/cryptos-flagship-ai-pact-fracture-fetch-sues-ocean-over-263m-fet-community-sales/ ；Fortune 关于 World：https://fortune.com/2026/06/01/internet-according-to-sam-altman-seeing-orb/
- Galaxy，加密风险投资：https://www.galaxy.com/insights/research/crypto-blockchain-venture-capital-q3 ；Crypto Briefing（仅见标题）：https://cryptobriefing.com/crypto-vc-investment-q3-2026-galaxy/ ；Blockhead 关于 Tiger Research：https://www.blockhead.co/2026/10/14/bitcoin-rose-43-in-q3-crypto-venture-funding-did-not-follow/ ；ForkLog 关于 2025 年代币发行：https://forklog.com/en/85-of-tokens-launched-in-2025-fell-below-initial-prices/

ASIC 经济学与芯片类型

- Hashrate Index，2026 年 11 月 9 日综述：https://hashrateindex.com/blog/hashrate-index-roundup-november-9-2026/
- AnySilicon，2026 年各节点 ASIC 设计成本：https://anysilicon.com/news/how-much-does-asic-design-cost-in-2026-real-price-ranges-by-process-node/ ；Silicon Analysts，无晶圆厂流片成本指南：https://siliconanalysts.com/analysis/fabless-startup-tapeout-cost-guide ；市场数据：https://siliconanalysts.com/market-data
- MillionMiner，比特大陆 S23 XP Hyd：https://millionminer.com/news/bitmain-wdms-2026-s23-xp-hyd-launch
- Khattak and Mikaitis，张量核中的 FP8 累加，arXiv 2512.07004：https://arxiv.org/html/2512.07004v2
- 嘉楠，2026 年第二季度业绩：https://finviz.com/news/389337/canaan-inc-reports-unaudited-second-quarter-2026-financial-results ；比特小鹿，2026 年第三季度业绩：https://greenstocknews.com/news/nasdaq/btdr/bitdeer-reports-unaudited-third-quarter-2026-results-record-revenue-of-us-320-7-million-adjusted-ebitda-up-173-to-us-69-6-million
- TheEnergyMag：Velaura AI：https://theenergymag.com/news/market-news/velaura-ai-series-a-bitcoin-pivot ；Core Scientific 与 Block：https://theenergymag.com/news/2026-07-29/core-scientific-exit-block-bitcoin-deal-ai ；AI 租约经济学：https://theenergymag.com/news/2026-08-06/miner-weekly-ai-pumpability-out-fuel
- Blockworks，Cipher 与 Fluidstack：https://blockworks.co/news/cipher-mining-inks-3b-ai-hosting-deal-with-fluidstack ；Cointelegraph，Core Scientific：https://cointelegraph.com/news/core-scientific-to-quit-bitcoin-mining-by-year-end-as-ai-business-surges ；ForkLog，Bitfarms：https://forklog.com/en/former-bitfarms-shuts-down-all-mining-sites-in-the-us/ ；IREN，2027 财年第一季度业绩：https://globenewswire.com/news-release/2026/11/05/3397020/0/en/iren-reports-q1-fy27-results.html
- Spark，比特币 ASIC 市场集中度：https://www.spark.money/research/bitcoin-asic-manufacturer-market-concentration

- Google，TPU7x 文档：https://docs.cloud.google.com/tpu/docs/tpu7x?hl=es ；The Register，Ironwood：https://www.theregister.com/2025/04/10/googles_7thgen_ironwood_tpus_debut/ ；SemiAnalysis，B300：https://inferencex.semianalysis.com/chips/b300 ；TPU v1 论文，arXiv 1704.04760：https://arxiv.org/abs/1704.04760 ；Glenn Lockwood，MI355X：https://glennklockwood.com/garden/processors/mi355x
- Tom's Hardware，Vera Rubin：https://www.tomshardware.com/pc-components/gpus/nvidia-launches-vera-rubin-nvl72-ai-supercomputer-at-ces-promises-up-to-5x-greater-inference-performance-and-10x-lower-cost-per-token-than-blackwell-coming-2h-2026 ；TrendForce，CoWoS 定价：https://trendforce.com/news/2026/10/27/news-tsmc-reportedly-eyes-20-cowos-price-hike-skips-equipment-price-renegotiations-amid-ai-demand-surge
- Menon Lab，AI 推理加速器对比：https://themenonlab.blog/blog/ai-inference-accelerators-compared ；Introl，GPU 之外的 AI 加速器：https://introl.com/blog/ai-accelerators-beyond-gpus-tpu-trainium-gaudi-cerebras-2025
- StratNews（Reuters），国资数据中心使用国产 AI 芯片：https://stratnewsglobal.com/team-sng/china-orders-state-funded-data-centres-to-use-domestic-ai-chips/ ；TNW，H200 进口：https://thenextweb.com/news/nvidia-h200-chips-hong-kong-china-import-approval
- Succinct，16 张 GPU 上的实时证明：https://blog.succinct.xyz/real-time-proving-16-gpus/ ；Cysic ZK ASIC 产品：https://cysic-docs.pages.dev/hardware-products/zk-asic-products
- Tom's Hardware，Intel Heracles：https://www.tomshardware.com/tech-industry/cyber-security/intels-heracles-chip-computes-fully-encrypted-data-without-decrypting-it-chip-is-1-074-to-5-547-times-faster-than-a-24-core-intel-xeon-in-fhe-math-operations ；IEEE Spectrum：https://spectrum.ieee.org/fhe-intel ；Evertiq，Niobium 与 SemiFive：https://evertiq.com/design/2026-02-20-semifive-secures-design-win-with-niobium-for-fhe-accelerator ；The Quantum Insider，Niobium 融资：https://thequantuminsider.com/2025/12/03/niobium-23m-fhe-funding/ ；Optica OPN，Optalysys：https://www.optica-opn.org/home/industry/2026/february/optalysys_raises_us_$31_million_to_push_photonics_computing_toward_cloud_infrastructure/ ；Zama，FPGA 上的 HPU：https://www.zama.org/post/announcing-hpu-on-fpga-the-first-open-source-hardware-accelerator-for-fhe

工艺节点

- Silicon Analysts，晶圆价格：https://siliconanalysts.com/data/wafer-pricing
- Tom's Hardware，台积电 A16 定价传闻：https://www.tomshardware.com/tech-industry/semiconductors/tsmc-could-charge-up-to-usd45-000-for-1-6nm-wafers-rumors-allege-a-50-percent-increase-in-pricing-over-prior-gen-wafers ；Wccftech，SMIC 定价：https://wccftech.com/smic-5nm-chip-prices-up-to-50-percent-more-expensive/amp/
- EUROPRACTICE，2026 年台积电 MPW 时间表：https://europractice-ic.com/wp-content/uploads/2025/11/TSMC-Europractice-MPW-schedule-2026.pdf
- Horowitz, "Computing's Energy Problem", ISSCC 2014: https://www.Gwern.net/doc/cs/hardware/2014-horowitz-2.pdf
- Tom's Hardware，华为昇腾路线图：https://www.tomshardware.com/tech-industry/artificial-intelligence/huawei-ascend-npu-roadmap-examined-company-targets-4-zettaflops-fp4-performance-by-2028-amid-manufacturing-constraints ；The Substrate，中国 AI 芯片供应链：https://www.the-substrate.net/p/where-chinas-ai-chip-supply-chain ；Tom's Hardware，Trainium3：https://www.tomshardware.com/tech-industry/artificial-intelligence/amazon-launches-trainium3-ai-accelerator-competing-directly-against-blackwell-ultra-in-fp8-performance-new-trn3-gen2-ultraserver-takes-vertical-scaling-notes-from-nvidias-playbook

机会与发行经济学

- Webull（Benzinga），博通财报电话会：https://www.webull.com/news/15511095937917952 ；Quartr，创意电子 2026 年第二季度：https://quartr.com/events/global-unichip-corp-3443-q2-2026_F9×6NJQY ；Taiwan News，世芯：https://www.taiwannews.com.tw/news/6370946 ；Design & Reuse，芯原：https://www.design-reuse.com/news/202530457-surge-in-custom-chip-business-propels-verisilicon-s-q1-revenue-to-double-with-new-orders-skyrocketing-to-8-2-billion-yuan-increased-r-d-investment-leads-to-wider-net-loss-earnings-insight/ ；The Block，嘉楠的 AI 部门：https://www.theblock.co/post/359250/crypto-miner-canaan-to-shutter-non-core-ai-semiconductor-business-unit-amid-strategic-realignment
- Clutch，AI 开发定价：https://clutch.co/developers/artificial-intelligence/pricing ；PixelPlex，区块链开发成本：https://pixelplex.io/blog/blockchain-development-cost/
- Hyperbolic，抽样证明：https://hyperbolic.ai/blog/deep-dive-into-hyperbolic-proof-of-sampling ；Gensyn，Verde：https://www.gensyn.ai/research/verde-a-verification-system-for-machine-learning-over-untrusted-nodes ；分层验证，arXiv 2512.20176：https://arxiv.org/abs/2512.20176v1
- Sherlock，2026 年审计定价参考：https://sherlock.xyz/post/smart-contract-audit-pricing-a-market-reference-for-2026
- Blockworks，币安上币标准：https://blockworks.com/news/binance-listing-standards-scrutiny ；Cointelegraph，做市协议：https://cointelegraph.com/magazine/market-maker-deals-quietly-killing-crypto-projects ；Blockworks，Movement：https://blockworks.co/news/movement-rushi-manche-suspension-market-maker-web3port ；Nansen，HIP-3：https://nansen.ai/post/what-is-hip-3-hyperliquid ；The Block，Monad：https://www.theblock.co/post/380094/monad-mainnet-launches

有用工作量证明 · 加密 × 芯片

算力即共识

AI × Crypto × 芯片全景研究（2026）：从 Pearl 的有用工作量证明谈起

本报告的每个关键数字都注明来源与日期。笔者自行推算的数字标注为“笔者测算”；只有卖方自报的数字标注为“厂商宣称”。本报告只讨论技术与商业。

谭帅

技术专家与创业者 · AI 研发者

tanshuai.com · i@tanshuai.com

公开版 · 数据截至 2026 年 11 月 13 日

