RESEARCH NOTE · PUBLIC EDITION

2026 · 11



PROOF OF USEFUL WORK · CRYPTO × SILICON

Compute as Consensus

AI × Crypto × Chips in 2026, starting from Pearl's proof of useful work

Pearl · Useful work · DePIN and agents · ASIC economics · Process nodes · Staged budgets

THE VERDICT

Pearl proves that matrix multiplication can secure a blockchain, but not yet that the work is useful. Across AI × crypto, paid demand is in inference, data and privacy, and an ASIC pays only where the pot is large, the rules are stable and the chip has a second market.

Tan Shuai

Technologist and founder · AI researcher and developer

tanshuai.com · i@tanshuai.com

Data as of November 13, 2026

Money in US dollars

CONTENTS

Contents



◆ **Executive summary** — The theorem holds, the incentive does not — 04

01 **Scope and method** — Scope, method and four tests — 07

02 **Pearl · the mechanism** — Pearl: matrix multiplication as consensus — 08

03 **Pearl · the economics** — Pearl: scale, emission and the usefulness gap — 10

04 **Pearl · the rules** — Pearl: rule risk, numerics and governance — 12

05 **Useful-work chains** — Useful-work chains: neither matrices nor zero knowledge stop ASICs — 13

06 **The wider landscape** — The rest of AI × crypto: real but small demand, and value moving to equity — 16

07 **ASIC economics** — The ASIC question: one formula, three yardsticks — 18

08 **Chip by chip** — Five chip types, five answers — 21

09 **Process nodes** — Process nodes: let the largest lifetime cost choose — 24

10 **A staged budget** — From model to silicon: a staged budget for a first chip — 27

11 **Opportunities** — Where the opportunities are, and what I would do — 29

12 **Signposts** — What to watch, November 2026 to 2027 — 31

13 **Questions to ask** — Questions you did not ask but should have — 32

◆ **About the author** — Tan Shuai — 34

◆ **Sources** — 35

FIGURES

Figures



FIG 01	AI × crypto in six sectors: where paid demand is real	06
FIG 02	On Pearl's integer work, GPUs share an energy ceiling	09
FIG 03	Pearl's first six months: hashrate follows price, rules keep changing	11
FIG 04	The subsidy multiple: how many times outside revenue a network pays	15
FIG 05	Why efficiency stops paying: an ASIC keeps at most 1 − 1/k of revenue	19
FIG 06	Most pots are smaller than one chip design budget	20
FIG 07	Five chip types, five answers	24
FIG 08	A leading-edge wafer costs ten times a mature one	25
FIG 09	Buy evidence before masks: a staged budget for a first chip	28

Tags in the text

- **estimate** — the author's arithmetic or inference from cited data
- **vendor claim** — published by the company that sells the product, not independently verified
- **analysis** — the author's interpretation of the evidence

Data as of November 13, 2026 · Money in US dollars

EXECUTIVE SUMMARY



The theorem holds, the incentive does not

THE VERDICT

Pearl shows that "compute as consensus" works as engineering, but not yet that it is useful.

≈90 EH/s
Pearl's network on November 13, 2026, where one H is one INT7 multiply-accumulate

≈1%
of Pearl blocks carry real AI work (Blockworks, October 28, 2026)

≈$70M
a year of verifiable external revenue at Bittensor, against more than $300 million of emissions

175–180×
Aleo's prover rewards against its fees in the third quarter of 2025

R × (1 – 1/k)
the profit ceiling of an ASIC that is k times cheaper than the marginal miner estimate

≈$1.5M
to reach measured silicon on N7 through a model, an FPGA and an MPW, before a mask set of ≈$3.5M estimate

Pearl

On November 13, 2026, Pearl's network ran at about 90 EH/s. In its consensus code one "hash" is one integer multiply-accumulate, so the network equals the integer throughput of about 122,000 H100s. Yet only about 1% of blocks carry real AI work, and the one public inference partnership is no longer offered on its provider's serverless API. The cause is economic, not cryptographic: a GPU that mines while it serves inference earns only about 1% to 27% of the hashrate of a dedicated miner, so rational miners multiply random matrices.

Paid demand

Across AI × crypto, paid demand that can be verified shows up in three places: selling inference, selling data and selling privacy. Bittensor's verifiable external revenue is about $70 million a year, about four-fifths of it from compute and infrastructure subnets. Grass reports $17 million of data revenue in the first half of 2026. Confidential inference in trusted execution environments is in production. By contrast, payments actually initiated by AI agents come to about $5,000–11,000 a month, and zkML and decentralized training have no disclosed paying demand yet.

Subsidy and value

The subsidy multiple is the most honest health check. Aleo's prover rewards were about 175–180 times its fees; Bittensor's emissions are about 4–11 times its verifiable revenue; Render's market value is about 360 times its on-chain fees; and Pearl's roughly $460 million a year of miner income is mostly block subsidy, almost unconnected to paid AI demand. Subsidy is fuel for a cold start, not a business model. Meanwhile value is moving from tokens to equity: four equity companies that use centralized infrastructure and crypto-native distribution are valued at about $8.5 billion, about the same as the protocol- and application-layer AI tokens together.

ASICs

For chips, the profit ceiling of an ASIC is roughly the revenue per unit of compute times (1 − 1/k), where k is how many times cheaper it is than the marginal hardware. Because that term is concave, efficiency beyond k = 5–10 adds almost no profit; what decides the outcome is the size of the pot, how long the rules last and whether the chip has a second market. My answers by chip type: do not tape out a dedicated INT7 Pearl chip now, and bet instead on an inference-first accelerator that can also mine; stay out of the merchant bitcoin-miner market as a new entrant; do not build a stand-alone training chip as a small team; build inference chips when a customer is signed; hedge on ZK accelerators and position early on FHE.

Nodes and budgets

The process node should be chosen by the largest item in the chip's lifetime cost: energy for chips that run around the clock, packaging and memory for training chips, the SRAM-or-HBM decision for inference chips. And a first chip should be bought in stages, a model, an FPGA, a multi-project wafer and only then a mask set, so that a change in the protocol costs the price of a prototype rather than of a tape-out. The opportunities lie where the fields cross: crypto as the first customer, AI as the second market.

Question	Judgement	Key numbers (sources in the chapters)
① **Does Pearl work, and is its work useful?**	It works: a blockchain secured by matrix multiplication at scale. Not yet useful: almost all mining is on random matrices	About 90 EH/s on November 13, 2026; 1 H = 1 INT7 multiply-accumulate; about 1% of blocks with AI work
② **Where is the real paid demand in AI × crypto?**	Inference, data and privacy. Agent payments, zkML and decentralized training are still pre-demand	Bittensor about $70 million a year of verifiable revenue; Grass $17 million in H1 2026; agent payments about $5,000–11,000 a month
③ **Is an ASIC worth building?**	Only where the pot is large, the rules are stable and the chip has a second market. Today no network offers all three	Profit ceiling ≈ R × (1 − 1/k); bitcoin about $11.5 billion a year of miner revenue, Pearl about $0.46–0.49 billion, Kaspa about $18 million
④ **Which chips should a small team build?**	Inference-first accelerators that can also mine; inference chips for a signed customer; early FHE work. Not merchant bitcoin miners or training chips	Kaspa's best ASIC earns $0.86 a day; 3/2 nm design budgets $60–400 million or more
⑤ **Which process node?**	The largest lifetime cost decides: energy, packaging and memory, or the memory architecture	TSMC N2 about $30,000 a wafer against about $3,000 for N28; HBM and CoWoS about 63% of an H100's cost
⑥ **What does a first chip cost, and how should it be bought?**	In four stages: a model, an FPGA, a multi-project wafer, then masks. Stop before masks if the rules or the buyer are not settled	A commercial MPW slot $50,000–500,000; N7 masks about $3.5 million; a 7/6 nm design budget $15–100 million

FIG 01 **AI × crypto in six sectors: where paid demand is real**

Verifiable paid demand against the subsidy and valuation signals, November 2026

Useful-work chains 01

REAL BUT SMALL

≈$70M a year

Bittensor's verifiable external revenue

Signal: emissions 4–11× revenue; Pearl ≈1% AI blocks

The mechanisms work; usefulness needs payers

Decentralized compute 02

REAL BUT SMALL

$2.35M a year

Render's on-chain fees, the largest verifiable

Signal: market value ≈360× fees

Mostly resale of commodity inference

Training and inference 03

PROFIT IN EQUITY

$100M+ a year

Prime Intellect, an equity company on centralized GPUs

Signal: Gensyn's training mainnet not live

Progress is real; profit goes to equity

Verifiable AI 04

EARLY

398

paying users of Phala's TEE cloud, end of 2025

Signal: no disclosed paid zkML

TEEs in production; zkML early

Agent payments 05

BEFORE DEMAND

$5–11K a month

x402 payments actually initiated by agents

Signal: ≈200M transactions, mostly signalling

The rails came first; demand has not

Data networks 06

REAL BUYERS

$17M

Grass data revenue, first half of 2026 (self-reported)

Signal: market value ≈10× annualized revenue

One of the few sectors with real buyers

Source: DefiLlama, SubConnect (via hoge.gg), TRM Labs, Phala, Dealroom, SiliconANGLE, CoinGecko, Blockworks; the author's synthesis.

CHAPTER 01



01 SCOPE AND METHOD
Scope, method and four tests

Why this report, and why now

In April 2026 a new blockchain, Pearl, put a long-standing idea into production: let miners compete for blocks by doing the matrix multiplications that AI runs on, instead of hashing. Within six months its network reached about 90 EH/s and its token a market value of about $500 million. At the same time, AI compute marketplaces, decentralized training networks, agent payment rails and data networks raised billions, and bitcoin miners began converting their power sites into AI data centers. I wanted to know which of these have real customers, whether custom silicon makes sense anywhere in them, and which process node and which foundry a chip team can actually use.

Four tests for any AI × crypto project

A 2017 security study of Gridcoin argued that useful computation can secure a blockchain only if it is hard to solve, easy to verify and easy to reward. I add a fourth test. Together they decide most of what follows.

01 **Hard to solve.** Honest computation must be the cheapest way to produce a valid proof; there must be no shortcut.

02 **Easy to verify.** Checking the work must cost far less than doing it, or the network pays twice.

03 **Easy to reward.** The protocol must be able to compute the reward from the proof alone, without trusting a third party.

04 **Someone pays for the output.** Unless a customer outside the network pays for the result, the “useful” work is a subsidy with a narrative.

Hash puzzles pass the first three tests and fail the fourth. Real AI tasks pass the fourth but fail the second: proving an AI computation in zero knowledge costs 100 to 10,000 times the computation itself, and proof-of-learning schemes have been shown to be cheaply spoofable. Pearl’s noisy matrix multiplication is the first credible bridge that passes the verification test while staying the same shape as AI compute.

How I worked

This is desk research, done the way I would review a chip or a business plan. I read Pearl’s code, its protocol proposals and its whitepaper; chain data from an independent full node; company filings and earnings releases; papers from cryptography, computer architecture and machine learning; foundry and vendor price data; and trade and technology press. Every number carries its date. Where I derive a number myself, I say so and mark it estimate. Numbers only the seller reports are marked vendor claim; my interpretation of evidence is marked analysis. Money is in US dollars. Status is as of November 13, 2026.

CHAPTER 02

02

PEARL · THE MECHANISM

Pearl: matrix multiplication as consensus

The idea, from theorem to mainnet

In April 2025, Ilan Komargodski and Omri Weinstein published “Proofs of Useful Work from Arbitrary Matrix Multiplication”. They construct a proof of useful work for any matrix product with an overhead of only 1 + o(1) times a naive multiplication; its security rests on the hardness of solving batches of random low-rank linear equations, which is a conjecture, not a proof. The paper names the real difficulty of the field: a proof of useful work with a constant-factor overhead is trivial but useless; the hard part is keeping honest computation the cheapest way to win when miners choose their own inputs.

Pearl’s mainnet started on April 27, 2026, at 09:00 UTC. Miners compute a noisy product (A + E)(B + F), where the low-rank noise is derived from a BLAKE3 commitment, and then strip the noise to recover A·B. The accumulator state of each output tile is hashed into a lottery ticket; a winning tile is opened and proved with a recursive Plonky2 zero-knowledge proof of under 60 KB, inside a block certificate capped at 65 KB. Verification is a small proof of constant size, independent of the matrix size: as cheap to check as a hash puzzle, while the computation itself can be useful.

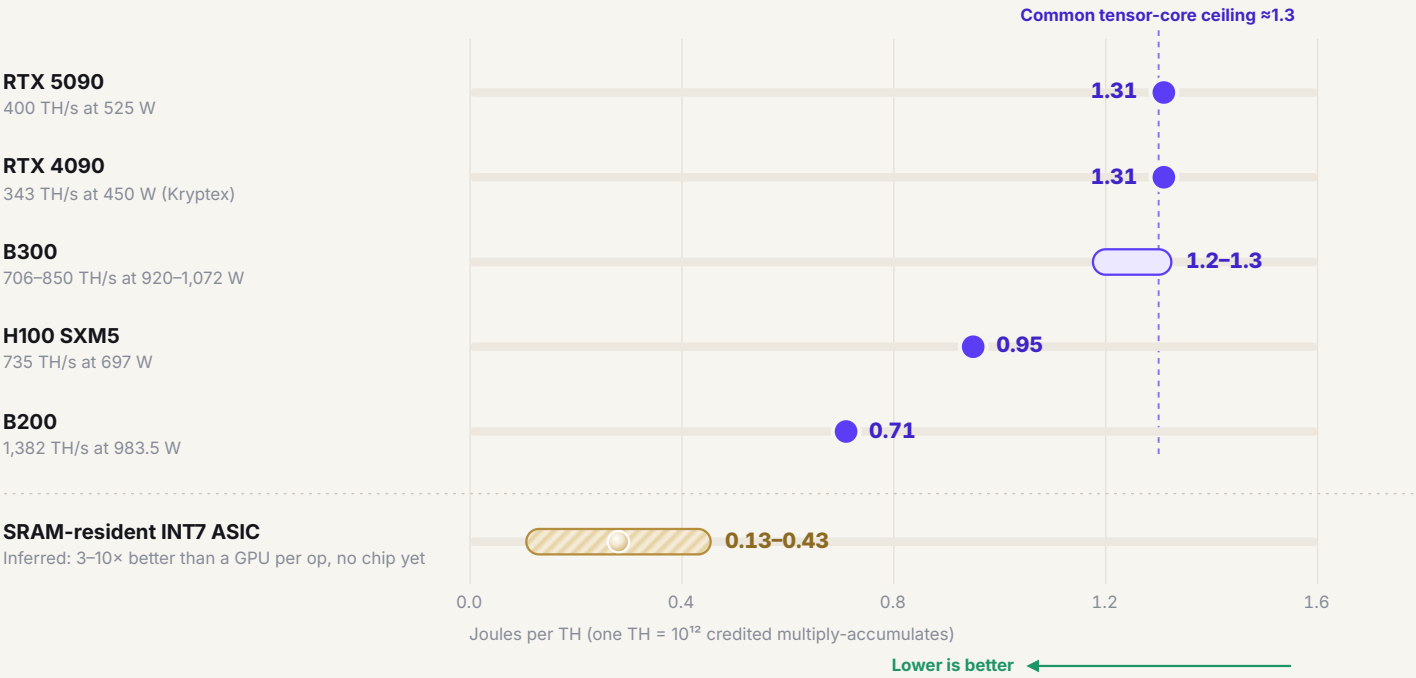
One multiply-accumulate is one “hash”

For a chip designer the most important fact is the unit. Pearl’s consensus code scales each attempt’s target by the tile size times the dot-product length, so **1 H is exactly one credited INT7 × INT7 → INT32 multiply-accumulate (MAC)**. Since block 96,251 a rank penalty requires a noise rank of at least 128; at the minimum rank, 1 H still equals 1 MAC. Pearl’s “EH/s” therefore literally means 10^18 integer multiply-accumulates per second: **for the first time, a blockchain’s security budget is priced in the tensor throughput of AI accelerators.** Measured GPU hashrates fall at 60% to 95% of each card’s dense INT8 multiply-accumulate peak, which fits this reading.

Hardware	Pearl hashrate	Power	Joules per trillion MACs (J/TH)	Source (date)
RTX 5090	400 TH/s	525 W	1.31 estimate	Kryptex (November 12)
RTX 4090	about 305 TH/s (community range 280–330); Kryptex lists 343 TH/s	450 W (Kryptex)	1.31 (Kryptex figures) estimate	prlstats (October 3); Kryptex
H100 SXM5	735 TH/s	697 W	0.95 estimate	PeakMiner, via prlstats (September 19)
B200	1.382 PH/s	983.5 W	0.71 estimate	PeakMiner, via prlstats
B300	706–850 TH/s	920–1,072 W	1.2–1.3 estimate	prlstats; SimeonOnSecurity

FIG 02 **On Pearl's integer work, GPUs share an energy ceiling**

Joules per trillion INT7 multiply-accumulates (J/TH) at measured Pearl hashrate; lower is better



Source: Kryptex (November 12, 2026); PeakMiner and SRBMiner benchmarks via prlstats; SimeonOnSecurity; the ASIC range is the author's inference, not a measurement.

Two details in this table matter to chip designers. First, consumer RTX 4090 and 5090 cards and the data-center B300 converge at about 1.3 J/TH, and even the best card, the B200, reaches only about 0.71 J/TH. A data-center GPU costing tens of thousands of dollars spends nearly as much energy per integer multiply-accumulate as a gaming card, which points to a common ceiling of tensor-core energy per INT8 MAC, and that ceiling is exactly what a dedicated chip could break. Second, the B300 is slower than the B200: NVIDIA cut INT8 throughput sharply in Blackwell Ultra, so the integer work has to take a slower path on that chip. Mainstream AI silicon is moving away from INT8 towards FP8 and FP4. When the number format a consensus chooses diverges from where hardware is going, the protocol is forced to change its rules, and that is the root of the FP8 fork and of my ASIC judgement later in this report.

CHAPTER 03

03 PEARL · THE ECONOMICS
Pearl: scale, emission and the usefulness gap

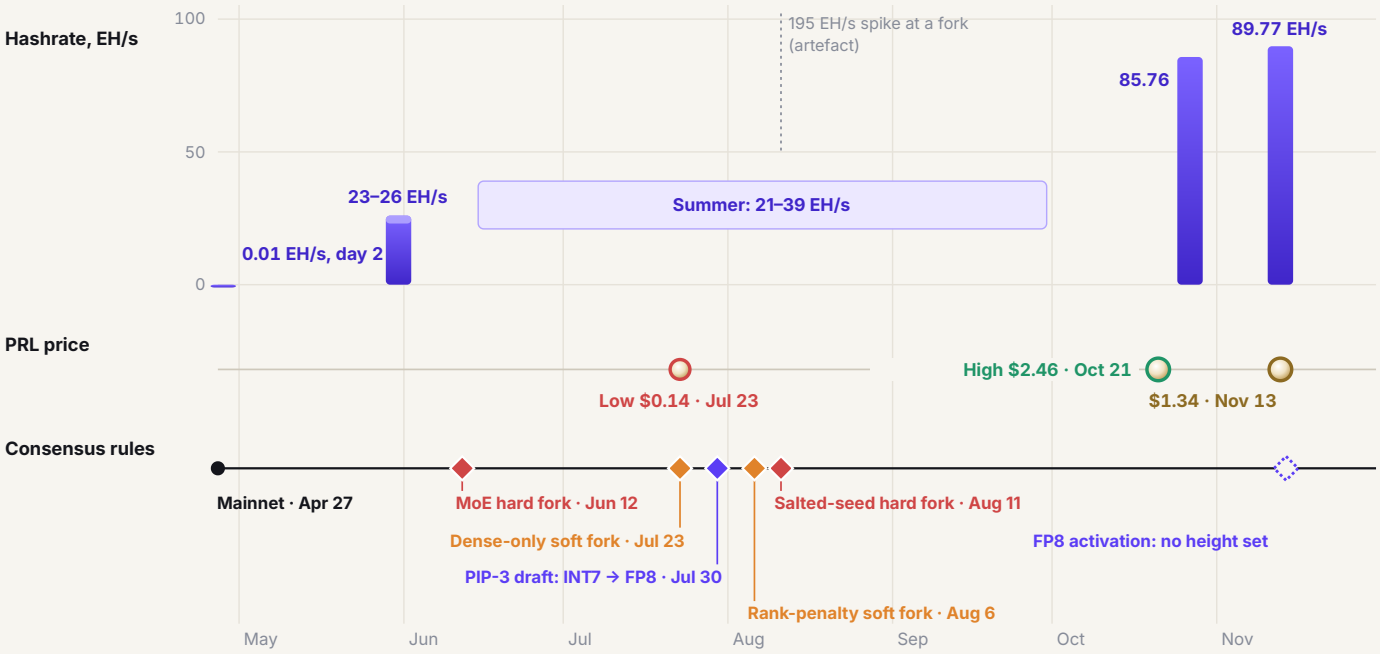
From 0.01 to 90 EH/s in six months

Metric	Value (November 13, 2026)	Basis and source
Network hashrate	About 90 EH/s (baseline 89.77; same-day range 88.20–93.22; intra-day high on November 12 of 110–123)	Independent full node prl-stats; Kryptex
Difficulty and block height	58,841,982 and 142,530	prlstats
Equivalent hardware	About 122,000 H100s or 225,000 RTX 5090s	At 735 and 400 TH/s per card estimate
Price, market value and fully diluted value	$1.32–1.34; about $500 million; about $2.8 billion (CoinGecko rank 101)	CoinGecko
All-time high and low	$2.46 (October 21) and $0.1399 (July 23)	CoinGecko
Coins mined	377.6 million PRL, 17.98% of the 2.1 billion cap	prlstats
Block subsidy and daily emission	About 2,173 PRL and about 0.97–1.03 million PRL	prlstats; Kryptex
New supply in the next 12 months	About 78% of current supply	prlstats
Miner income	About $1.25 million a day (model), about $460 million a year	prlstats
Revenue per unit of hashrate, R	About $14–15 per PH/s per day; an RTX 5090 earns about $5–6 a day	estimate; Kryptex
Blocks carrying real AI work	About 1% (estimate of October 28)	Blockworks
Pool concentration	Kryptex found 46.4% of blocks in the last 24 hours; the top two pools about 62%	prlstats
Holder concentration	Address Gini coefficient 0.969; the top 100 addresses hold 53.0%	prlstats

Pearl's hashrate is close to a lagging function of its price. The day after launch the network ran at about 0.01 EH/s; by the end of May it reached 23–26 EH/s. Over the summer the price fell about 86% from its May high and the hashrate moved sideways at 21–39 EH/s. After the price rose about tenfold in September and October, the hashrate reached 85.76 EH/s on October 27 and about 90 EH/s in November. A spike to 195 EH/s on August 11 coincided with a fork and should be read as an estimation artefact. Energy grew with it: one study estimated about 112 MW for the whole network at 24 EH/s, and a prlstats model puts it at about 190–215 MW today.

FIG 03 **Pearl's first six months: hashrate follows price, rules keep changing**

Network hashrate, three price marks and every consensus change, April 27 to November 13, 2026



Source: prlstats full node and series; Kryptex; CoinGecko; Pearl's params.go and PIP-3. The August 11 spike is an estimation artefact at a fork.

Emission amplifies the reflexivity

The supply cap is 2.1 billion coins and there are no halvings: the block subsidy declines smoothly as E(t) = S·H / ((t + H)(t + H − 1)) with H = 650,226. But emission is counted in blocks, not in time, and the early difficulty started too low and adjusted slowly. **About 44.06 million coins were mined on the first day and about 120 million in the first six days**, about 32% of today's supply estimate. Recent blocks, at about 171–183 seconds, still come faster than the 194-second target. The result is that supply will grow by about 78% over the next 12 months, a dilution that miners, hardware makers and holders all have to face.

Miner economics follow. Network-wide miner income is about $1.25 million a day (model); Kryptex's daily emission times the current price gives about $1.35 million a day, so **revenue per unit of hashrate R ≈ $14–15 per PH/s per day** estimate. An RTX 5090 earns about $5–6 a day. An H100 or H200 earns about $11, while the cheapest public on-demand rental is about $2.09 per card per hour, about $50 a day, so renting cards only to mine loses heavily. R swings widely: by the prlstats series it was about $10 in August and September and briefly about $33 in mid-October estimate. Within weeks of launch, the estimated daily revenue of an RTX 5090 halved from about $33.80 to about $17.19. Renting cloud GPUs to mine paid only in the first weeks and during the October price spike.

The theorem holds; the incentive does not

The real problem is demand, not cryptography. Pearl's only public commercial integration is a Gemma-4-31B-it-pearl endpoint from Together AI, launched on May 15 at more than 25% below the usual price, with the discount funded by the expected future value of the emission. On November 13, the model page showed that it is **not available on Together's serverless API**. In June a census of 8,012 online mining instances, about 21% of the network, concluded that the network was producing no useful AI computation at all (the author calls the evidence suggestive, not conclusive), and showed that shares built from purely random matrices were accepted by pools.

On October 28, Blockworks estimated that **only about 1% of blocks carry real AI work**; without it, it wrote, Pearl is a smaller, less secure, faster-inflating bitcoin.

The reason can be computed. The inference-side overhead is genuinely small: Pearl reports a 5.08% increase in end-to-end time for Llama 70B on four H200s, and an independent test measured a 9.2% throughput loss on one Hopper GPU. The mining side earns too little. The official vLLM plug-in mines only on prefill batches of at least 1,024 tokens and never on decode steps. A card serving inference gets about 27% of a dedicated miner's hashrate on pure prefill traffic, about 5.4% on a chat mix that is 80% decode, and only about 1% on decode-dominated traffic. Without paid inference demand, a rational miner multiplies random matrices; the protocol verifies that the committed matrices were multiplied honestly, not whether they were useful. **The theorem holds; the incentive does not.** Pearl's usefulness is an off-chain commercial problem: someone has to pay for inference.

CHAPTER 04

04 PEARL · THE RULES
Pearl: rule risk, numerics and governance

Four forks in two months

The other face of Pearl is how fast its rules change. From June to August 2026 the lab pushed four consensus changes at hard-coded heights: a mixture-of-experts hard fork (block 71,935, about June 12), a dense-only soft fork (91,630, about July 23), a rank-penalty soft fork (96,251, about August 6) and a salted-seed hard fork (99,000, about August 11). After the salted-seed fork, old miners could produce only invalid shares; the rank penalty cut the effective hashrate of high-rank configurations by 50% to 87.5%.

The next step is PIP-3, still a draft: it moves mining from INT7 to FP8 × FP8 → FP32 and requires bit-exact device profiles, **supporting only two architectures, Hopper and Blackwell**. The document calls hardware dependence a necessary evil of floating-point protocols and warns that any tolerance in verification invites grinding attacks. On November 12, the mainnet parameters of the fp8-scheme branch set no activation height. Under the current draft, the RTX 40 series, the A100 and AMD cards are not admitted, while a low-confidence prlstats model estimates that about 82% of the hashrate comes from consumer cards. Forcing the fork through risks a chain split or a miner backlash.

A number-format trilemma

Pearl faces a trilemma: a number format can be useful for AI (FP8 or FP4), bit-exactly verifiable across hardware, and hardware-neutral, but at most two of the three. The INT7 scheme is verifiable and neutral but diverges from where AI silicon is going. FP8 with a whitelist is useful and verifiable but ties consensus to one vendor. FP8 with a tolerance is useful and neutral but opens the door to grinding attacks. Floating-point arithmetic in tensor cores generally does not follow a numerical standard and is not reproducible across generations; on Hopper, for example, FP8 accumulation keeps only 13 fractional bits and truncates. From a chip designer's point of view, a "Hopper profile" is a fully determined arithmetic that an ASIC could implement bit for bit, perhaps more efficiently than the GPU. But doing so would be business in a grey zone against the protocol's stated intent, with very high governance risk analysis.

Governance

Governance is concentrated. The lab wrote every protocol proposal; forks activate at hard-coded heights with no miner signalling; there is no second client implementation; and as of early October no third-party audit of the node or the zero-knowledge circuits had been published. The Israeli newspaper TheMarker reported that the lab raised about $120–130 million in a stealth round (about $55 million of equity plus a GPU infrastructure fund of about $75 million); Pearl has not confirmed it through official channels. “No investor allocation in the token” and “a company with venture shareholders” can both be true, but the tension remains: AI revenue may flow to equity while dilution falls on token holders.

Taken together, Pearl is the cleanest “compute as consensus” experiment so far: a novel mechanism, verifiable on chain, with a unit that maps one to one onto AI hardware. But usefulness has not arrived, and its security budget, hashrate and miner enthusiasm all hang on a token whose main demand has yet to form and whose supply will grow by about 78% in 12 months. Three signals from Blockworks are worth tracking: whether the upgrade lands smoothly, whether the share of AI-work blocks rises materially, and whether any inference provider passes mining income on to its customers.

CHAPTER 05



USEFUL-WORK CHAINS

05 Useful-work chains: neither matrices nor zero knowledge stop ASICs

Kaspa: a matrix is neither useful nor ASIC-proof

Kaspa is the textbook case. Its kHeavyHash puts a 64 × 64 matrix-vector product over 4-bit values between two hashes, and the product is thrown away. About 16 months after its fair launch in November 2021, IceRiver announced the first Kaspa ASICs on March 14, 2023, and by the end of 2023 GPU mining was finished. Comparing an RTX 4090 at 3.8 GH/s and 300 W with Bitmain’s KS7 at 77 J/TH gives an efficiency gap of about **1,000 times** estimate. After the feast came compression: on November 13, the network ran at 358.9 PH/s, about 64% below its 1 EH/s of August 2024, and the best KS7 earns only **$0.86 a day** at $0.10 per kWh.

Aleo: zero knowledge does not stop special hardware either

Aleo’s prover puzzle is designed to mimic the work of generating a SNARK proof; it does not prove users’ transactions, and the project openly welcomes ZK ASICs. About five months after its September 2024 mainnet, dedicated machines were shipping. In the third quarter of 2025, provers received 31.2 million ALEO (about $7.8 million) while the network collected only about $43,000–44,000 in fees: **a subsidy multiple of about 175–180**. The price fell from about $9 in early trading to $0.22 by the end of that quarter, and ARC-46 now requires a stake of 100,000 to 2.5 million ALEO per solution, turning mining into a stake threshold.

Qubic and Bittensor

Qubic markets “useful proof of work” that trains its Aigarth neural network, but the most visible output of its compute in 2025–2026 was other chains’ blocks. It first mined Monero, raising its share from under 2% in mid-May 2025 to more than 25% by the end of July and claiming 51% in August, which caused a six-block reorganization (another source says 18). On April 1, 2026 it switched to mining Dogecoin with ASICs, using the proceeds for buy-

backs and burns. There is no independent evidence that its AI training is useful, and its market value fell from about $280–300 million to about $50 million on November 13. It also exposed an externality: a chain that pays for general-purpose compute creates a large, rentable, coordinated pool of hashrate that can be turned against CPU-friendly chains.

Bittensor is not proof of work but stake-weighted subjective scoring: validators score miners, and weights are clipped at the stake-weighted median. It is the only network in this group with external revenue at scale. The SubConnect index, which counts only verifiable external customers, shows about **$70 million a year** in early November 2026 (about $35 million on September 1), 78% to 82% of it from compute and infrastructure; but at early-November prices its emissions exceed $300 million a year, so revenue covers only 9% to 23%. Revenue estimates vary widely with the basis, from about $17 million (Artemis) to $172 million annualized from self-reported totals. In April, 67 of 128 subnets were burning more than 99% of their incentives, and when the leading team Covenant left over governance, TAO fell 24% within hours.

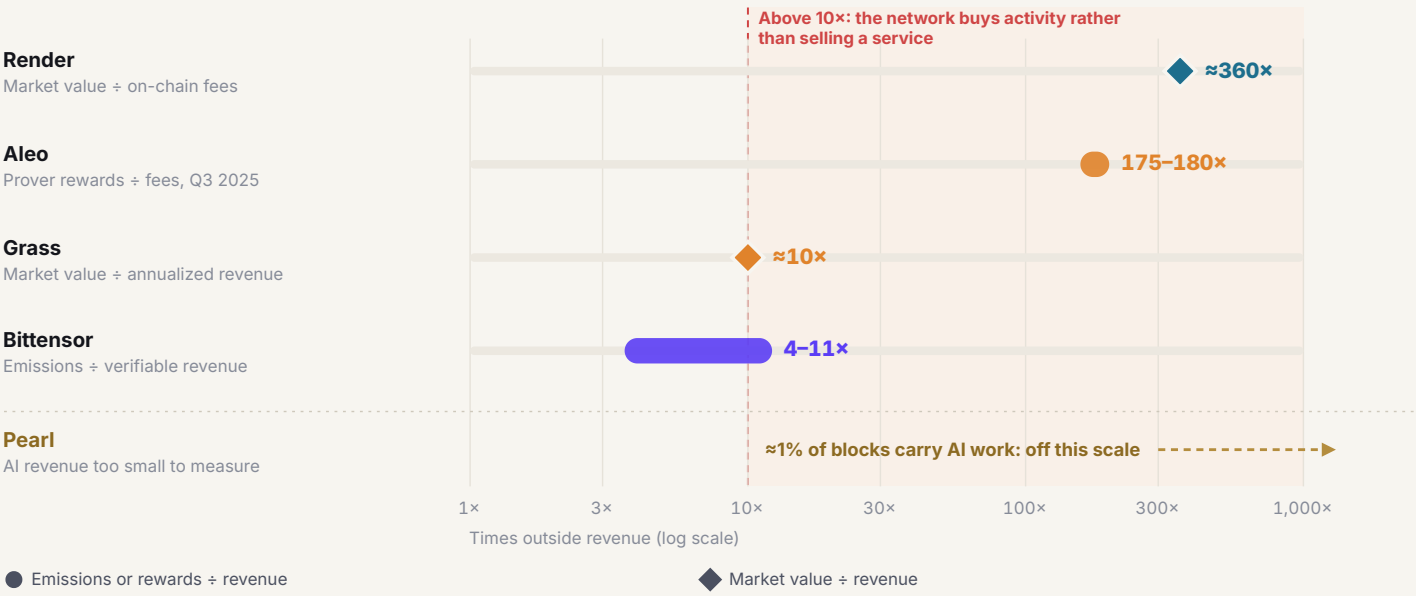
History gives the same answer

Primecoin's prime chains were quick to verify, but nobody needed them. Gridcoin relied on third-party statistics sites to attest BOINC contributions, and researchers forged blocks without doing any computation. Filecoin's storage proofs work technically, but in the third quarter of 2025 utilization was only 36%, and 99.5% of its $793,000 of fees were penalties, with base fees of just $4,100.

Project	Work or consensus	Market value (date)	Hardware path	Paid demand against subsidy	Judgement
Pearl	Noisy matrix multiplication (INT7 MAC plus ZK certificate)	About $500 million (November 13)	GPUs; the FP8 draft would whitelist Hopper and Blackwell	About 1% of blocks with AI work; miner income mostly subsidy	The most credible mechanism; demand unproven
Kaspa	kHeavyHash (a 64 × 64, 4-bit matrix-vector product between hashes)	About $840 million (November 13)	ASICs about 16 months after launch; now all ASIC	Claims no usefulness; miner revenue about $18 million a year estimate	A matrix is neither useful nor ASIC-proof
Aleo	Synthetic SNARK puzzle plus proof of stake	About $118 million (end of Q3 2025)	Dedicated machines after about five months	Prover rewards about 175–180 times fees	Synthetic work, driven by subsidy
Qubic	"uPoW" (claims to train Aigarth)	About $50 million (November 13)	CPU and GPU; ASICs for Dogecoin since April 2026	No independent evidence of usefulness; compute used for outside mining and buybacks	Compute aggregation wrapped in an AI story
Bittensor	Yuma consensus (stake-weighted scoring)	About $2.55 billion (November 11)	General-purpose GPUs; no ASIC path	About $70 million a year of verifiable revenue against more than $300 million of emissions	The only one with external revenue at scale; still a 4–11 times subsidy
Filecoin (reference)	Proofs of replication and spacetime	About $1.5 billion (end of Q3 2025)	—	Utilization 36%; quarterly base fees $4,100	Useful, but short of demand

FIG 04 **The subsidy multiple: how many times outside revenue a network pays**

Emissions or market value divided by revenue paid from outside the network, log scale



Source: Blockworks (Aleo Q3 2025), DefiLlama (Render fees), CoinGecko, hoge.gg (Bittensor), Dealroom (Grass), Blockworks 0xResearch (Pearl); the author's arithmetic.

The lesson for hardware: fixed, regular work paid in a liquid token gets ASICs within about 3 to 16 months, often from challengers before the incumbents. “Useful” claims held only where an outside customer pays for the output. And the main drivers of stranded hardware are a high subsidy relative to revenue and developer control over the work function.

CHAPTER 06

THE WIDER LANDSCAPE

06 The rest of AI × crypto: real but small demand, and value moving to equity

Decentralized compute markets

Decentralized compute has paying customers, but demand is small, concentrated and mostly resale of open-model inference. Fees verifiable on chain are small: DefiLlama shows about $2.35 million of fees for Render over the past year and about $680,000 for Livepeer, and the whole DePIN sector earned only about $72 million of on-chain revenue in 2025. Self-reported numbers are larger but loosely defined: Aethir reports "more than $127.8 million" of revenue for fiscal 2025 without saying whether that is gross billing or protocol take. Token values and revenue have come apart in both directions: Render's market value is about 360 times its on-chain fees, while Aethir's market value of about $87.7 million is below its self-reported revenue estimate.

The claim of being 70% to 85% cheaper holds only against the hyperscalers' on-demand prices. On November 3, the weighted average H100 price on Akash was about **$2.73 per GPU-hour** (range $2.04–3.16), below about $6.88 on AWS and $11.06 on Google Cloud, but above about $1.73 on Vast.ai, $1.99 on RunPod's community cloud and about $2.44 for a one-year reservation (prices on other platforms as read on September 12; Akash, the publisher, is an interested party). Demand is mostly commodity inference distributed through OpenRouter, about 1.7 billion tokens a day on AkashML and more than 4 billion on io.net, competing head-on with every neocloud. The largest "decentralized compute" dollars are recentralizing: Axe Compute, which holds Aethir's token as a treasury asset, signed a five-year contract worth $1.5 billion to rent out its own cluster of 9,200 B300s, against first-quarter 2026 revenue of $35,000.

Decentralized training and inference

Training over the internet has been demonstrated at 8 to 40 billion parameters. Pluralis-8B trained for 40 days on 330 nodes, mostly consumer GPUs, and reached 63% of the efficiency of a centralized H100 baseline, though its paper admits that the run was not decentralized and needed Pluralis's own nodes and an admission authority. Nous Research's Psyche completed the 40-billion-parameter Consilience pretraining run. At frontier scale the work recentralized: Prime Intellect's INTELLECT-3, a 106-billion-parameter mixture-of-experts model, was trained over two months on 512 H200s connected by 400 Gbps InfiniBand. The constraints are plain: nodes need at least 200 Mbit/s of bandwidth and about 100 times communication compression, and trust and verification still depend on the operator's privileged roles.

The money went to equity companies selling centralized products. Prime Intellect had more than $100 million of annualized revenue before raising $130 million at a $1 billion valuation in July; Venice, a private inference app, reports more than $70 million of annualized revenue and profits since the first quarter of 2026 (company figures), raised $65 million at a $1 billion valuation and runs on leased GPUs. On the token side, Gensyn's core training network is still "to be determined" and its token is worth about $17.2 million; the token of Ritual, an AI-native layer-1, fell 72% from its high after mainnet (headline only).

Verifiable AI

Production chose trusted execution environments (TEEs). Hopper's confidential computing adds under 7% overhead to typical LLM queries and nearly nothing for large models; Phala's production overhead is about 0.5% to 5%, but by the end of 2025, 398 of its 10,004 users paid. A TEE's root of trust is the chip vendor: TEE.fail, disclosed in October 2025, used a DDR5 bus interposer costing under $1,000 to extract attestation keys from fully updated Intel TDX and AMD SEV-SNP machines, which in turn undermines NVIDIA's GPU confidential computing, and both vendors say physical attacks are outside their threat model. That is exactly the DePIN setting, where node operators have physical access. zkML can now prove a complete LLM: Lagrange's DeepProve proves GPT-2 at about 174 tokens a minute and Gemma 3 at about 86, orders of magnitude from production inference and with no disclosed paid use.

Agent payments and data

x402 is the best cautionary example. TRM Labs counts about 198.9 million transactions and $52.7 million settled by known facilitators since May 2025, 99.6% in USDC; after removing self-payments, batch flows and sellers with fewer than ten buyers, the share actually initiated by agents is only 0.6% to 7.5%, **about $5,000–11,000 a month**. Artemis judges that more than 95% of the activity is protocol signalling rather than buying and selling. Meanwhile the rails are being standardized by incumbents: the x402 Foundation was set up under the Linux Foundation on July 14, with Visa, Mastercard, Stripe, Google, AWS, Cloudflare and Coinbase as core members, and Tempo, which co-wrote the Machine Payments Protocol with Stripe, launched its mainnet in March at a $5 billion valuation. Agent-token platforms still run on launchpad logic: Virtuals self-reports $479 million of "aGDP", and ElizaOS's founder declared its token dead after a class-action settlement. Value is more likely to be captured by stablecoin issuers and payment incumbents; crypto's distinctive role here may be identity and accountability for agents.

Data is one of the few sub-sectors with real off-chain buyers. Grass reports $17 million of data revenue in 2025 and another $17 million in the first half of 2026, targets $75 million for the year and says it is profitable; GRASS is worth about $328 million, about ten times its annualized first-half revenue estimate. Its moat is the collection scale of distributed residential bandwidth, not crypto itself. Data DAOs and data marketplaces lack visible paid demand: Vana is worth about $21.5 million and Ocean about $27.6 million, and Ocean is in litigation with Fetch.ai. Proof of personhood is gaining real integrations as agents multiply: World says it has verified nearly 18 million people, and Zoom, Docusign and Tinder support parts of World ID.

Capital has voted

Sector	Verifiable paid demand	Subsidy and valuation signals	Judgement
Useful-work chains	Bittensor about $70 million a year of verifiable external revenue; Pearl about 1% of blocks with AI work	Aleo 175–180 times; Bittensor 4–11 times; Pearl mostly subsidy	The mechanisms have value; usefulness needs paying customers
Decentralized compute markets	On-chain fees about $0.68–2.35 million a year (Livepeer, Render); self-reported $4–128 million	Render's market value about 360 times fees	Real but small, mostly resale of commodity inference
Decentralized training and inference	Equity companies: Prime Intellect over $100 million a year, Venice over $70 million (centralized infrastructure)	Gensyn's training mainnet not live; Ritual's token −72% from its high	Technical progress is real; profits go to equity
Verifiable AI	TEEs in production (398 paying users at Phala); no disclosed paid zkML	—	TEEs useful; zkML early

Sector	Verifiable paid demand	Subsidy and valuation signals	Judgement
Agent payments	Agent-initiated flows about $5,000–11,000 a month	About 200 million transactions, mostly signalling and speculation	Rails first, demand later
Data networks	Grass $17 million in H1 2026 (self-reported)	GRASS about 10 times annualized revenue	One of the few sectors with real buyers

Put the six sectors side by side and the conclusion is consistent: **subsidy is fuel for a cold start, not a business model.** On November 13 the CoinGecko AI category was worth about $18.6 billion in total, but about 56% of that is general-purpose layer-1s such as NEAR and ICP, plus consensus-level projects such as TAO, staked TAO (possibly double-counted with TAO) and Pearl; AI tokens at the protocol and application layer total about $8 billion or less estimate. That is about the combined valuation of four equity companies: Tempo ($5 billion), Nous (a reported $1.5 billion), Prime Intellect ($1 billion) and Venice ($1 billion). Crypto venture funding is stuck at about $4.5–5 billion a quarter, with the increase coming from AI-related infrastructure, whose share of disclosed amounts rose from 8.1% to 18.2% in the third quarter of 2026; and 84.7% of the 118 tokens launched in 2025 trade below their launch price. In one sentence: **crypto does distribution and settlement, equity takes the profits**, at least in 2026.

CHAPTER 07

07 ASIC ECONOMICS
The ASIC question: one formula, three yardsticks

One formula: why efficiency stops paying after k = 5

Let R be the daily revenue per unit of compute (network miner income divided by network hashrate), c the daily cost of a unit of compute on the marginal hardware (on Pearl, a GPU), and k how many times cheaper an ASIC is than that marginal hardware ($c_{ASIC} = c/k$). As long as GPUs remain the marginal miners, free entry and exit push R towards c, so an ASIC earns $R - c/k$ per unit of compute:

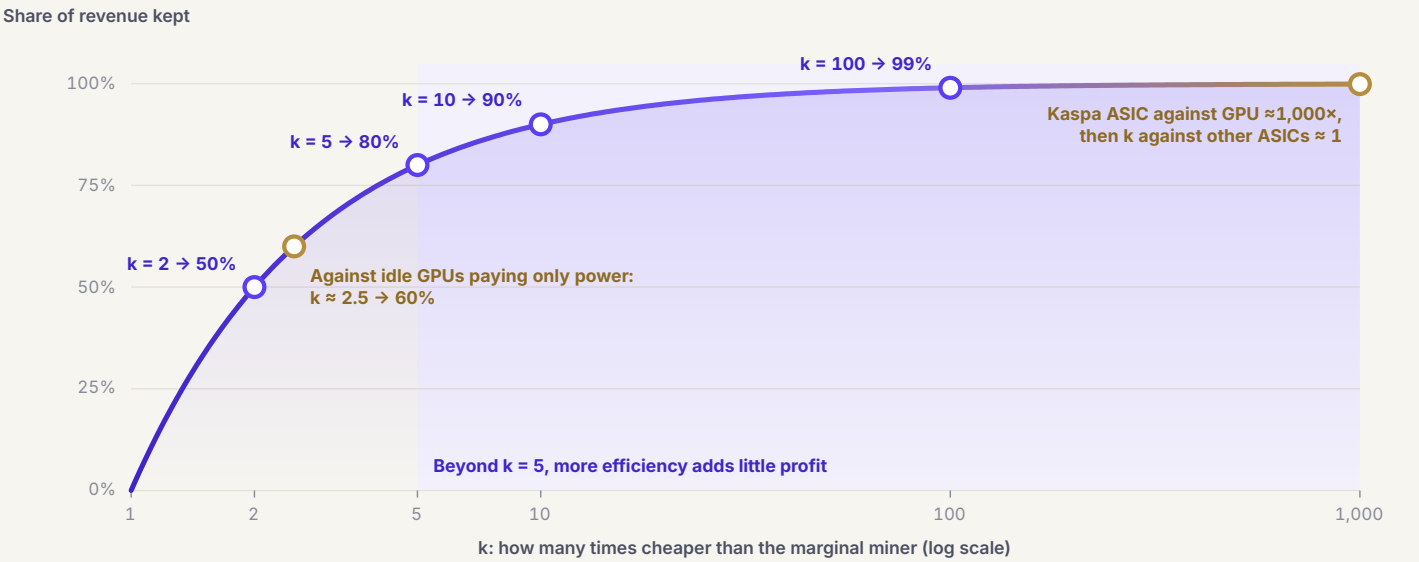
ASIC profit ceiling ≈ R × (1 − 1/k)

Across the network, an ASIC fleet's annual gross profit ceiling is roughly network miner revenue × (1 − 1/k), and only until ASICs replace GPUs as the marginal miners.

The formula has three consequences. First, (1 − 1/k) is concave: at k = 2 an ASIC keeps 50% of revenue, at k = 5 80%, at k = 10 90%, at k = 100 99%. **Beyond k = 5–10, more efficiency adds almost no profit**; the size of R and of the pot decide. Second, the reference resets: once ASICs are the marginal miners, k has to be measured against other ASICs. Kaspa's ASICs had about a 1,000-fold efficiency advantage over GPUs, which bought only a short windfall; today the best KS7 earns $0.86 a day. Third, on a GPU-friendly chain like Pearl, the reference is often not the purchase price of a GPU but the **electricity of an idle GPU**: an owned RTX 5090 at 525 W and $0.10 per kWh costs about $1.26 a day in power and earns about $5–6 estimate. Against GPUs that are sunk costs, even an ASIC that uses one-fifth of the energy has an effective k of only about 2.5 if its depreciation equals its electricity, and a profit ceiling of about 60% of R (an illustrative assumption).

FIG 05 **Why efficiency stops paying: an ASIC keeps at most 1 − 1/k of revenue**

Share of revenue an ASIC can keep against the marginal miner, by its cost advantage k



Source: The author's model: when the marginal hardware sets the price, revenue per unit of compute tends to its cost; Kaspa and RTX 5090 figures from the report.

Three yardsticks: the pot, the payback scale and the rules

The first yardstick is **the pot**. Annualizing about 2,893 BTC of block rewards in the first week of November (about $222 million), bitcoin miners earn about $11.5 billion a year estimate. Pearl's miners earn about $0.46–0.49 billion (at current prices); Aleo's prover rewards, annualized from the third quarter of 2025, about $31 million; and Kaspa's miners, at a block reward of 1.945 KAS and ten blocks a second, about $18 million estimate. **A whole year of Kaspa miner revenue is less than the low end of a single 5/4 nm chip design budget (about $30 million).**

The second yardstick is **the hashrate needed to pay back the NRE**:

h* = NRE ÷ [R × (1 − 1/k) × T × p]

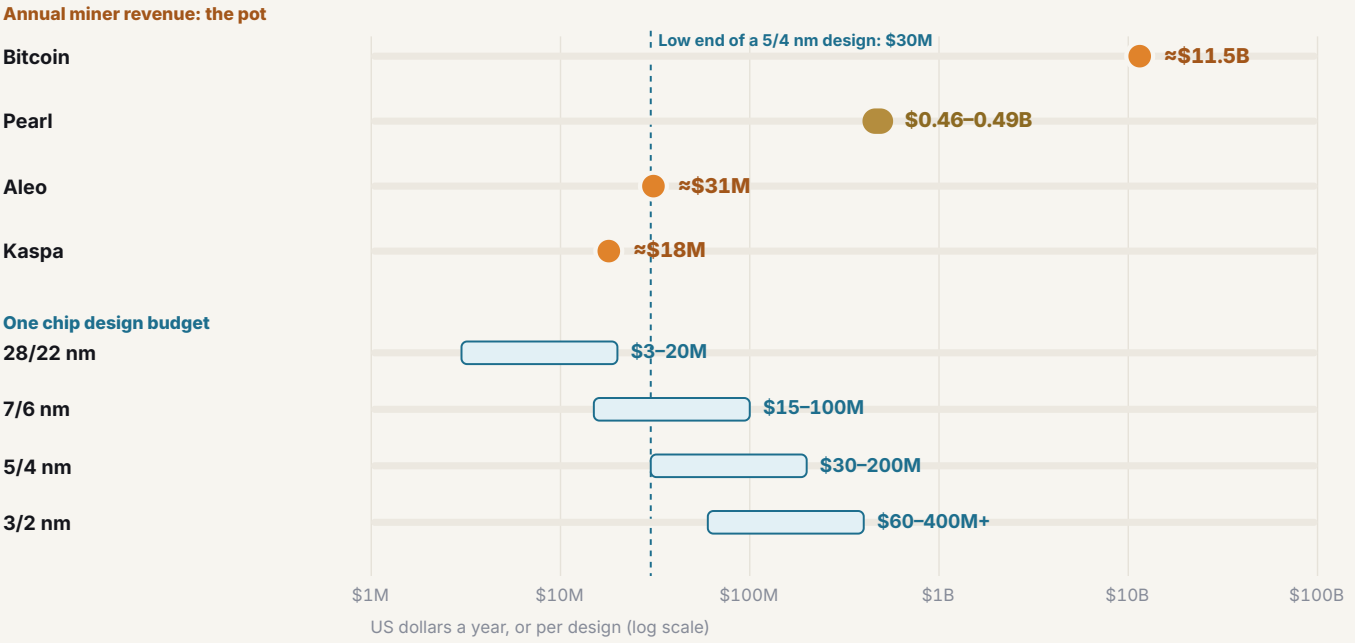
where T is the useful window in days and p the probability that the rules stay unchanged in that window. For Pearl, take R = $14 per PH/s per day, k = 5, T = 365 days and an NRE of $20 million (an N7/N6 repetitive-array chip; at the low end of the ranges of $5–30 million for mid-range nodes from Silicon Analysts and $15–100 million for 7/6 nm compute ASICs from AnySilicon, because a mining chip uses little IP and a repeated structure; my assumption). With p = 1, it needs about **4.9 EH/s** running for a full year, about 5% of the network; with p = 0.5, about **9.8 EH/s**, about 11%; and if R falls back to the summer's level of about $10, even p = 1 needs about 6.9 EH/s estimate. The scale is not out of reach. The problem is p and T.

The third yardstick is **rule risk**, and it is often decisive. Monero's anti-ASIC fork of 2018 stranded Bitmain's X3 and briefly cost the network almost half its hashrate. Sia's fork of the same year bricked Bitmain's and Innosilicon's miners, while the used price of the Obelisk miner, linked to the developers, rose from $2,499 to more than $5,000. A little over a month after Ethereum's Merge, fewer than 20% of former Ethereum miners had found new work. **Whoever controls the work function controls the fate of hardware investors**, and Sia shows that "ASIC resistance" can become the developers' own moat. Hardware with a second market survives: after the Merge, GPUs moved to AI and the cloud, while single-algorithm ASICs crowded into Ethereum Classic. The window is also shrinking: Kaspa's first ASIC took about 16 months, Aleo's about five, and the first to reach the efficiency frontier is often a challenger rather than an incumbent.

Network	Annual miner revenue (about)	Marginal or reference hardware	Rule stability	k available to a new entrant	Judgement
Bitcoin	$11.5 billion	Bitmain S23 XP Hyd (8.9 J/TH, $25.5 per TH)	Very high	≤1	A large pot, but k belongs to the incumbents
Pearl	$0.46–0.49 billion (at current prices)	Owned or idle GPUs (marginal cost ≈ electricity) and data-center GPUs	Very low: four forks in two months; FP8 whitelist pending	Perhaps 3–10 times in efficiency analysis; lower against sunk-cost GPUs	The pot is large enough; rule risk vetoes it
Aleo	About $31 million (Q3 2025, annualized)	IceRiver AE2 and other dedicated machines	Medium: ARC-46 stake threshold	≈1	A small pot, and the threshold is rising
Kaspa	About $18 million	KS7 (77 J/TH)	High	≈1	Annual revenue below the low end of one 5/4 nm design budget

FIG 06 **Most pots are smaller than one chip design budget**

Annual miner revenue of four networks against the design cost of one chip by process node, log scale



Source: Hashrate Index (bitcoin, November 2026), prlstats and Kryptex (Pearl), Blockworks (Aleo, Q3 2025), minerstat (Kaspa), AnySilicon (design budgets, 2026); annualized by the author.

The quadrant of "large pot, stable rules and k > 1" is empty today: bitcoin has the pot and the stable rules, but its k belongs to the incumbents; Pearl's pot is large enough but its rules are the least stable; Kaspa and Aleo have stable enough rules and pots that are too small.

CHAPTER 08

08 CHIP BY CHIP
Five chip types, five answers

Pearl-style PoUW: no dedicated INT7 chip; bet on a miner that can serve inference

The case for an ASIC is real. Only about 1% of blocks carry AI work today, so the whitepaper's economic argument, that a dedicated ASIC loses the ability to do useful work and therefore earns less, is at its weakest right now. The most expensive parts of a GPU, HBM and CoWoS packaging, are about 63% of an H100's manufacturing cost, with the logic die only about 9%. A chip that does nothing but 7-bit integer multiply-accumulates, keeps its operands in on-chip SRAM and runs at low voltage could plausibly reach 3 to 10 times a same-node GPU's energy per operation and a larger advantage in cost per operation [analysis; no measurements yet]. The case against is harder. Foundry lead times at leading nodes can exceed 50 weeks, and a new team needs about 18 to 24 months from start to volume estimate; by the time a chip arrives, Pearl may well have moved to FP8 and then FP4. And because PIP-3 defines the admitted arithmetic through a device whitelist, an INT7 chip would simply stop working once it activates.

So my conclusion is: **do not tape out a dedicated INT7 Pearl chip now.** There are two better bets. One is an inference-first accelerator that can also mine: built around general INT8/FP8 GEMM and SRAM-resident operands, so that if mining access is denied, the value lands in a real inference market. The other is to join the protocol debate and push for a vendor-neutral canonical arithmetic that an ASIC can implement exactly, together with a commitment to algorithm stability, which is the precondition for the "large pot, stable rules, k > 1" quadrant to exist. Until then, an architecture can be validated on FPGAs or mature-node multi-project wafers (MPW), keeping the NRE risk in the range of a million dollars. Chapter 10 turns this into a staged budget.

Bitcoin-style miners: the incumbents own k, and power is the scarce asset

Bitcoin's rule risk is the lowest, but competition pushes a new entrant's k below 1. Bitmain's S23 XP Hyd reaches 8.9 J/TH, sells at $25.5 per TH and ships from November; flagship efficiency fell from 17.5 J/TH for the S21 in 2023 to 8.9 J/TH, about 49% in three years. A new entrant shipping in volume in 2027–2028 would need about 8 J/TH or better to stand estimate. Challengers are leaving: Canaan reported second-quarter 2026 revenue of $31.9 million with a gross loss of $29.3 million and guided only $11–15 million for the third quarter; Bitdeer's third-quarter sales of its own rigs to outside buyers fell to $0.3 million, with a group gross margin of 3.4%; Auradine renamed itself Velaura AI and turned to AI chips; and Core Scientific cancelled an order of about 15 EH/s of 3 nm machines from Block's Proto unit, booking a $41.9 million loss.

Demand is also shrinking. On November 9 the hashprice was about $31.89 per PH/s per day, the forward curve to April 2027 averaged about $32.7, and fees were only 0.72% of block rewards. At $0.045 per kWh, the S23 XP Hyd pays back in about 3.1 years estimate, close to the next halving expected around 2028. More important is the opportunity cost of power: AI and high-performance-computing leases now earn about $1.67–1.9 million per contracted megawatt per year, which at 80% to 85% net operating income is about 1.5 to 2.9 times the gross profit of mining per megawatt (about $0.56–0.9 million) [estimate; the two megawatt bases differ]. Core Scientific will stop mining in the fourth quarter; Bitfarms, now Keel, has shut all its US mining sites; IREN earned $200 million from AI cloud and only $28 million from mining in one quarter. **Power and data-center capacity are the assets; mining machines are consumables.**

The NRE arithmetic is just as unforgiving. Frontier efficiency now needs 3 nm-class processes (by secondary sources, MicroBT uses Samsung SF3, and Auradine's and Proto's chips are 3 nm). Masks cost about $10–20 million and full NRE $30–100 million or more; at $20–25 per TH and a 30% gross margin, each 1 EH/s sold earns about $6–7.5 million of gross profit, so covering annual R&D on Bitdeer's scale, about $110–160 million, requires selling about 20–25 EH/s a year to outside buyers, about 2% of the network estimate. The conclusion: **a new team should not enter the merchant bitcoin-miner market.** Three cases still make sense: vertical integration with cheap power of one's own; niche forms such as home and heating miners (Canaan is also launching home heating products); and taking part in the incumbents' roadmaps through IP and design services.

AI training chips: not as a small team

For training chips, k is hardest to earn. At the same generation and precision, Google's Ironwood leads the B300 in chip-level FP8 efficiency by only about 1.1 to 1.8 times (Ironwood's power is a reporter's estimate) estimate; the 30 to 80 times often quoted from the TPU v1 paper was measured against a K80, which had no tensor cores. The real lever is the number format: moving from FP8 to FP4 usually gives 2 to 3 times the throughput on the same silicon, and NVIDIA moves fastest there: it claims 50 PF of NVFP4 inference for Rubin, in volume in the second half of 2026. Cost and supply thresholds are higher still: 3/2 nm design budgets run $60–400 million or more; CoWoS capacity in 2026 is about 106,000–130,000 wafers a month, and TSMC is reported to be considering price rises of up to 20%; at a contract price of $15.6 per GB for HBM3E, the HBM alone in a 288 GB accelerator costs about $4,500 estimate. This market belongs to Broadcom-style co-development with hyperscalers (chapter 11). The conclusion: **small teams should not build stand-alone training chips**; build IP, chiplets or design services instead.

AI inference chips: trade generality for k, and it pays

Inference is where ASICs have the most room, because k can be bought by giving up generality. Etched claims that its eight-chip server runs Llama-70B at more than 500,000 tokens a second, about 20 times eight H100s; Taalas hard-wires model weights into mask ROM and claims 10 times lower power and 20 times lower manufacturing cost (both vendor claim). SRAM-centred architectures avoid HBM and CoWoS, and their nodes are more manageable: Taalas uses TSMC 6 nm, and Groq moved from 14 nm to Samsung 4 nm. Crypto distribution does not help inference chips, since decentralized compute networks run almost entirely on NVIDIA; the way in is the procurement economics of neoclouds and the large domestic inference market in China, where state-funded data centers must use domestic AI chips, H200 imports are being throttled, and Beijing wants large internet companies to run inference on domestic chips. The conclusion: **worth building**, as long as there is a clear customer and the degree of lock-in to models and architecture is a deliberate choice. It is also the best second market for a PoUW mining chip.

ZK and FHE accelerators: hedge on ZK, position early on FHE

GPUs have already caught up with real-time ZK proving: Succinct's SP1 Hypercube proved 99.7% of Ethereum blocks in under 12 seconds on 16 RTX 5090s. Cysic claims its ZK C1 chip is 10 to 100 times as efficient as GPU provers and will ship in 2026, but its demand comes largely from its own token rewards vendor claim; I found no public evidence that Aleo puzzle ASICs are used for real ZK proving. ZK proving is integer and deterministic, close in shape to what mining chips do, but proof systems change fast (SP1, RISC Zero, Halo2, Plonky2 and more). The conclusion: **wait and hedge**, with programmable MSM, NTT and hash engines, first on FPGAs, then as ASICs.

FHE is the most hardware-bound primitive with the largest upside. Intel's Heracles, on Intel 3 with 197 mm², 176 W and 48 GB of HBM3, runs FHE operations 1,074 to 5,547 times faster than a 24-core Xeon, but Intel has no commercial plans and the project's lead left in December 2025. Niobium is building on Samsung 8LPU with SemiFive

as design partner (a development contract of about $6.86 million) and has raised more than $23 million; Optalysys, with a photonic approach, raised $31 million and says a full commercial chip is still two to three years away; and Zama's open-source FPGA design reaches about 13,000 programmable bootstraps a second at about 200 W, the public baseline any ASIC must clearly beat. The conclusion: **position early and validate with light assets**, for a window around 2027–2029. Demand is more likely to anchor in encrypted AI for enterprises and governments than in cryptocurrency itself, and FHE needs HBM, so its cost structure resembles a training chip's.

Chip type	Where k comes from	Demand and pot	NRE and thresholds	Rule or lock-in risk	Second market	Judgement
Pearl-style PoUW mining chip	Dropping HBM, CoWoS and general-purpose units: perhaps 3–10 times in J/op analysis	Miner income about $0.46–0.49 billion a year, volatile	N7/N6 repetitive array about $10–30 million analysis	Very high: the protocol decides the number format and the device whitelist	AI inference, if built as general INT8/FP8 GEMM	No dedicated INT7 chip; build inference-first and mineable, or wait for canonical arithmetic
Bitcoin SHA-256 mining chip	k ≤ 1 against Bitmain	About $11.5 billion a year, merchant market held by two firms	3 nm-class masks $10–20 million; full NRE $30–100 million or more	Low (stable algorithm); high competition and halving risk	Almost none (single algorithm)	New entrants stay out of the merchant market; only own-power, niche forms or IP
AI training chip	Same-generation systolic arrays only 1.1–1.8 times; the format is the lever	Largest (Broadcom about $58 billion of AI semiconductor revenue in fiscal 2026)	3/2 nm design $60–400 million or more; CoWoS and HBM	NVIDIA's annual cadence	—	Small teams build IP, chiplets and design services, not stand-alone chips
AI inference chip	Giving up generality: 10–20 times claimed vendor claim	Large and fragmented; mainland China's domestic market is large in its own right	SRAM architectures at N6/N7–N4 are manageable	Lock-in to models and architecture	It is the main market	Worth building: sign the customer, then choose the lock-in
ZK accelerator	10–100 times GPUs claimed vendor claim	Small, driven mostly by token rewards	N7/N6 then N5; FPGA first	Proof systems change fast	Limited outside ZK	Wait and hedge: programmable MSM and NTT engines
FHE accelerator	1,074–5,547 times a Xeon (Heracles)	Not yet formed; anchored in enterprise and government	Needs HBM; cost structure close to training chips	Standards and ecosystem unsettled	Encrypted computing, private AI	Position early, validate with light assets; 2027–2029 window

FIG 07 **Five chip types, five answers**

Where the cost advantage k comes from, what the silicon can sell into if the rules change, and the verdict



Pearl-style PoUW mining chip

NOT NOW

k: no HBM or general units: 3–10× per op (inferred)

Second market: AI inference, if general INT8/FP8 GEMM

Bet on an inference-first chip that can mine, not a dedicated INT7 part



Bitcoin SHA-256 mining chip

STAY OUT

k: ≤1 against the incumbents

Second market: almost none

Only with own power, a niche form, or as IP



AI training chip

IP, NOT CHIPS

k: 1.1–1.8× at the same precision

Second market: —

Packaging and HBM decide; build IP, chiplets, services



AI inference chip

BUILD

k: give up generality: 10–20× claimed

Second market: it is the main market

Sign the customer, then choose the lock-in



ZK proving accelerator

WAIT AND HEDGE

k: 10–100× GPUs claimed

Second market: limited outside ZK

Programmable MSM and NTT engines; FPGA first



FHE accelerator

POSITION EARLY

k: 1,074–5,547× a Xeon (Heracles)

Second market: encrypted computing, private AI

A 2027–2029 window; validate with light assets

Source: The author's synthesis of chapters 7 and 8; vendor figures as labelled in the report.

CHAPTER 09

09

PROCESS NODES

Process nodes: let the largest lifetime cost choose

The cost ladder: leading-edge wafers cost ten times mature ones

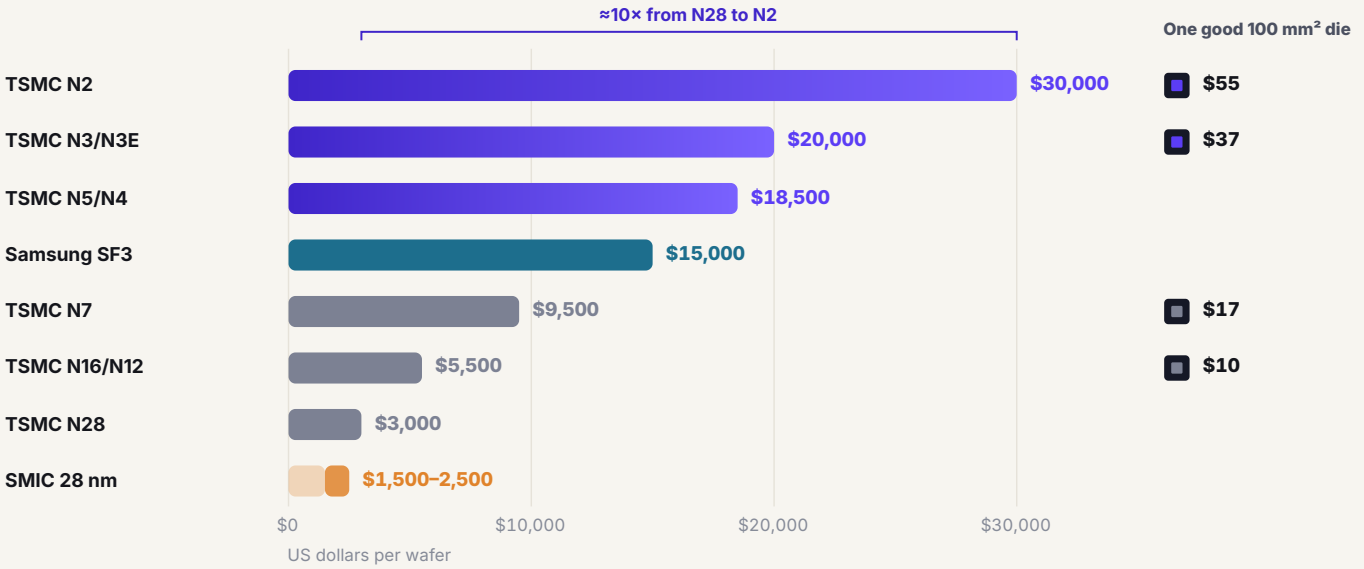
Start with the ledger. A leading-edge wafer now costs about ten times a mature one: TSMC N2 about $30,000, N3 about $20,000, N7 about $9,500 and N28 about $3,000; Samsung SF3 about $15,000, about 25% below N3; and SMIC's 7 nm-class manufacturing cost is reported to be 40% to 50% above TSMC's. TSMC is reported to raise prices by another 3% to 6% from January 2027, most for 2 nm and 3 nm. Masks and design costs differ even more, and sources disagree, so the table shows both.

Node	Wafer price (estimate)	Mask set (model value / guide range)	Design budget (AnySilicon)	MPW (per project)
TSMC N2	About $30,000	—	3/2 nm: $60–400 million or more	—
TSMC N3/N3E	About $20,000 ($18,000–27,000)	$15 million / $10–20 million	As above	—

Node	Wafer price (estimate)	Mask set (model value / guide range)	Design budget (AnySilicon)	MPW (per project)
Samsung SF3	About $15,000 ($13,000–17,000)	$12 million	—	—
TSMC N5/N4	About $18,500 ($16,000–21,300, low confidence)	$6.5 million / $10–20 million	5/4 nm: $30–200 million	—
TSMC N7	About $9,500 ($8,100–10,900)	$3.5 million / $5–10 million	7/6 nm: $15–100 million	Commercial $100,000–500,000
TSMC N16/N12	About $5,500	$1.8 million	16/12 nm: $6–40 million	Academic about $35,000 per mm²
TSMC N28	About $3,000	$1 million / $1–3 million	28/22 nm: $3–20 million	Commercial $50,000–150,000; academic about $12,000 per mm²
SMIC 28 nm	$1,500–2,500	—	—	—
SMIC 7 nm class (N+2/N+3)	40–50% above TSMC	—	—	—

FIG 08 **A leading-edge wafer costs ten times a mature one**

Estimated price per 300 mm wafer by node, and the cost of one good 100 mm² die at 85% yield



Source: Silicon Analysts wafer pricing (estimates, N5 low confidence), Wccftech (SMIC); die costs are the author's arithmetic with an assumed yield.

Then the physics. Horowitz's classic measurements show that a programmable instruction costs about 70 pJ while the arithmetic itself costs a few picojoules: **arithmetic is cheap and overhead is expensive**, which is the reason systolic arrays and dedicated chips exist. Halving the feature size roughly halves the energy per operation, but supply voltage has stalled around 0.9–1.1 V. An N2 wafer costs 50% to 67% more than N3, so unless density rises by more than about 1.5 times, the cost per transistor goes up rather than down estimate: leading nodes have to pay for themselves through efficiency and performance. For a 100 mm² die at 85% yield, the cost of each good die is about $55 at N2, $37 at N3, $29–34 at N5, $17 at N7 and $10 at N16 [estimate; the yield is an assumption].

Choosing a node by chip type and team

My rule is: **let the largest item in the chip's lifetime cost choose the node.** For compute chips that run fully loaded around the clock (mining, ZK), electricity dominates lifetime cost, so once the algorithm is stable they should move to N5/N4 and N3/SF3; while the algorithm is unsettled or money is short, validate on mature-node MPW first. For training chips the bottleneck is packaging and memory (HBM and CoWoS are about 63% of an H100's bill of materials), so the node is secondary. For inference chips the first decision is the memory architecture: SRAM-centred designs can stay at N6/N7 to N4, while HBM designs inherit the cost structure of training chips. FHE is bound by memory bandwidth and cannot avoid HBM.

Chip type	Start-up or small team	Scaled company	Mainland China path	Key points
Matrix-multiplication PoUW mining chip (SRAM-resident, no HBM)	Prototype on 28 nm or 16/12 nm MPW; first volume on N7/N6; N5/N4 once the algorithm and revenue are stable; dies of 50–150 mm² in standard flip-chip packages	N3/N3E or Samsung SF3	SMIC N+1/N+2 is workable (no HBM or CoWoS needed), at about 40–50% more per good die	Energy dominates; protocol rules are the largest risk
SHA-256 mining chip	Do not chase the frontier; niche forms can use mature nodes	3 nm class; incumbents move from TSMC 5 nm to 2 nm and from Samsung SF3 to 2 nm (secondary sources)	Hard to match leading-node efficiency	Power price and the halving cycle decide margins
AI training chip	No stand-alone chip: IP, chiplets, design services	N3P then N2/A16; 2–8 chiplets; secure CoWoS and HBM early	SMIC N+2/N+3 with domestic HBM (the Huawei model); per-chip FP8 about 0.2–0.5 times NVIDIA's, compensated by system scale	Packaging and HBM are the bottleneck
AI inference chip	SRAM-centred or hard-wired, N6/N7 or N4/N5/SF4; INT8, FP8 and FP4 only	N3/N2 with HBM (long context, mixture of experts), or multi-chiplet SRAM	SMIC N+2 with LPDDR or GDDR; HBM-class memory is the choke point	The degree of architectural lock-in is the central decision
ZK accelerator	Start on FPGA or 16/12 nm MPW; then N7/N6 and N5	N5/N3	SMIC N+1/N+2 is workable	Integer and deterministic; easy to verify
FHE accelerator	FPGA (against the open-source baseline) or 16/12 nm MPW; volume needs HBM	N5/N3 with HBM	Limited by access to HBM	Memory bandwidth, not arithmetic, sets the speed

Advanced or mature, four kinds of foundry, and chiplets

Option	Main strengths	Main weaknesses	Best for
Leading nodes (N3/N2, SF3/SF2, Intel 18A)	Each full node roughly halves energy per operation; dense SRAM	Wafers 3–10 times, masks 5–15 times and design budgets 10–20 times those of mature nodes; lead times over 50 weeks; N2 cost per transistor may rise	Compute chips with stable algorithms running around the clock; training chips

Option	Main strengths	Main weaknesses	Best for
Mature and mid-range nodes (N7/N6, N16/12, N28)	Masks in the million-dollar range, cheap MPW, fast iteration, many foundries (TSMC, UMC, GlobalFoundries, SMIC)	Uncompetitive efficiency when fully loaded around the clock	Prototypes, niche markets, edge inference
TSMC	The best yield and ecosystem	The highest prices; N2/N3 capacity tight; long queues for CoWoS	Teams chasing the efficiency frontier with volume commitments
Samsung Foundry	Wafers about 20–25% cheaper (SF3 about $15,000 against about $20,000 for N3); used by Groq's second generation and others	The foundry business is losing money, targeting break-even by the end of 2027	Cost-sensitive compute chips
Intel Foundry	Complex chips such as Heracles proven on Intel 3	No public external pricing or customer data for 18A; the foundry business is losing money	Evaluate ecosystem and support case by case
Foundries in mainland China (SMIC and others)	Close to a large domestic market; 28 nm wafers at $1,500–2,500	7 nm-class capacity about 45,000–60,000 wafers a month, largely absorbed by Huawei; yields about one-third of TSMC's; costs 40–50% higher; DUV only	Inference, mining and ZK chips for the domestic market
Chiplets	Overcome reticle and yield limits; standard for HBM systems (MI355X, Ironwood, Trainium3)	Higher packaging cost and CoWoS dependence; known-good-die testing adds cost	Training and large inference chips; not needed for small mining or ZK chips

CHAPTER 10

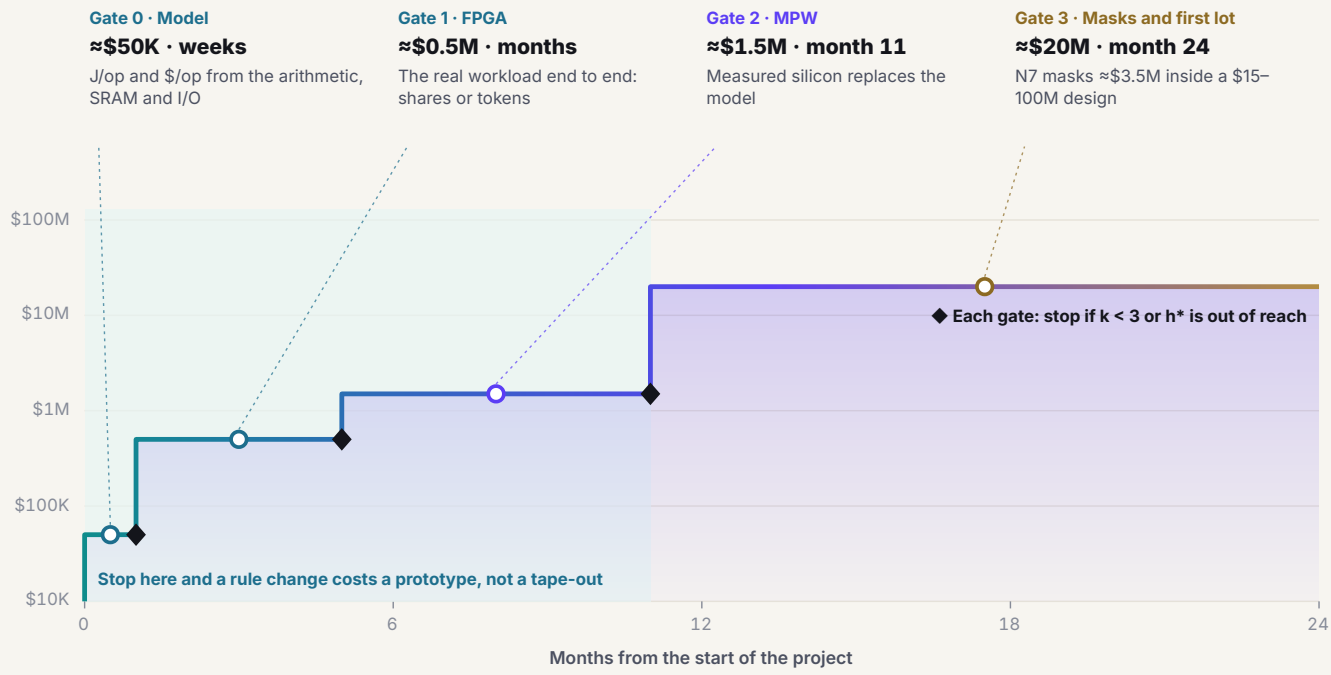
10 A STAGED BUDGET
From model to silicon: a staged budget for a first chip

Buy evidence before masks

A full-custom chip is a series of bets, each larger than the last, and the discipline is to buy evidence before buying masks. I use four gates analysis. Gate 0 is a model: an architecture study that estimates J/op and $/op from the arithmetic, the SRAM and the I/O, against the reference hardware of chapter 7; it costs weeks and tens of thousands of dollars. Gate 1 is an FPGA prototype that runs the real workload end to end, mining shares or inference tokens, and proves the datapath and the protocol integration; it costs a few months and a few hundred thousand dollars. Gate 2 is a multi-project wafer: a commercial slot costs about $50,000–150,000 at N28 and $100,000–500,000 at N7, and measured silicon replaces the model's energy estimates. Gate 3 is the full mask set and the first production lot: masks of about $1 million at N28, $3.5 million at N7 and $15 million at N3, inside a design budget of $15–100 million at 7/6 nm. Each gate needs a number that stops the project, such as k below 3 against the marginal hardware, or a payback hashrate h* that the window cannot reach.

FIG 09 **Buy evidence before masks: a staged budget for a first chip**

Cumulative spend in US dollars of an N7/N6 project through four gates, log scale; each gate has a number that stops the project



Source: Wafer, MPW and mask prices as in chapter 9 (Silicon Analysts, AnySilicon, model values); gate timing and spend are the author's illustrative assumptions.

The staircase is steep. In an N7/N6 project, the first three gates together cost about $1.5 million and about eleven months; the fourth costs more than ten times as much estimate. That is why the order matters more than the node: if the protocol changes its arithmetic, or the buyer walks away, a team that stopped at Gate 2 loses the price of a prototype, not of a tape-out.

What each chip costs, and what has to be true

Project	Gate 2 (prototype)	Gate 3 (node and design budget)	One good 100 mm² die	What has to be true to pay back
Inference-first accelerator that can mine	16/12 nm or N7 MPW, $0.1–0.5 million	N6/N7, about $15–30 million estimate	About $17 at N7, before packaging and memory	A signed inference customer carries most of the NRE; mining income is upside, not the base case
Dedicated Pearl INT7 miner (not recommended now)	28 nm MPW, $0.05–0.15 million	N7/N6, about $20 million	About $17 at N7	About 4.9 EH/s for a full year with p = 1, or 9.8 EH/s with p = 0.5 (chapter 7)
ZK accelerator	FPGA, then 16/12 nm MPW	N7/N6, then N5	About $10 at N16, $17 at N7	Proof systems settle on MSM and NTT kernels that a fixed design can serve
FHE accelerator	FPGA against the open-source baseline	N5/N3 with HBM, $30 million or more	Dominated by HBM and packaging	A buyer for encrypted AI at volume, which does not yet exist

Two commercial points follow. First, the cheapest efficiency is architectural: keeping operands in SRAM and leaving out HBM and CoWoS removes most of a GPU's cost (about 63% of an H100's bill of materials) before any node is chosen. Second, time is a cost: a new team needs about 18 to 24 months from start to volume and leading-edge lead times exceed 50 weeks, so a protocol that changes its arithmetic every few months can make a chip obsolete before Gate 3 pays for itself.

Three rules for a small team

Taken together, these numbers give three rules analysis. First, **stay at Gate 1 or 2 until there is a named buyer or a frozen protocol version**; neither the token price nor the hashrate is a substitute for either. Second, **buy k with architecture first and with the node second**: an SRAM-resident array on N7 can beat a general-purpose part on N3 for one workload, at a fraction of the mask cost. Third, **write down the payback hashrate or volume, with p, before tape-out**, and price milestone payments against a fixed protocol version, so that the customer, not the design house, carries the rule risk.

CHAPTER 11

11 OPPORTUNITIES
Where the opportunities are, and what I would do

Custom chips: demand comes overwhelmingly from AI

In 2026 demand for custom chips comes overwhelmingly from AI. Broadcom's AI semiconductor revenue in the third quarter of its fiscal 2026 was $16.7 billion, up 221% year on year, about $58 billion for the year, 73% of it from XPUs for six customers (from an automated transcript of the earnings call). Global Unichip's second-quarter revenue rose 128% year on year, with turnkey production revenue up 188% and a 21.5% gross margin. Alchip's first-quarter revenue split by node was 21% at 3/2 nm, 63% at 7/5 nm and 16% at 12 nm and above. VeriSilicon's new orders for 2026 reached RMB 8.24 billion (about $1.16 billion) by the end of April, 91.37% of them for AI computing. The market has three tiers: at the top, multi-year co-development worth billions; in the middle, projects from 7 nm to 2 nm with NRE from a few million to over a hundred million dollars ($5–30 million at mid-range nodes, $30–100 million or more at the leading edge); and a long tail of mature-node ASICs with NRE of $0.5–5 million. Crypto and mining customers are a small, cyclical niche, but they are the fastest, most permissionless "first customer" paying for compute today.

The capabilities of a mining-chip team that transfer best, near-threshold low-voltage operation, massively replicated compute arrays, kilowatt-class power delivery and cooling, and fast leading-node tape-outs, are exactly what inference chips and ZK and FHE accelerators need. Capital markets price this: Velaura AI (formerly Auradine) raised $110 million in August at a valuation above $1 billion. The counterexample is Canaan's edge-AI chip unit, which had revenue of $0.9 million in fiscal 2024 while consuming 15% of operating expenses, and was closed in June 2025. The difference lies in an independent route to market, not in reusing mining channels.

AI development: hours are a commodity; the value is in the joints

At the low end, AI development services are commoditized: typical rates on Clutch are $25–49 an hour, and $50–99 in the United States, Canada and similar markets; a blockchain development team costs about $15,000 to more than $90,000 a month. Selling hours has no moat. The value is in the joints between fields: kernels and compilers

for custom chips, mining embedded in inference serving (Pearl's vLLM plug-in is one example), verifiable inference that combines TEEs with ZK spot checks, and payments and identity for agents.

AI × crypto systems: sign the compute buyer before the token

For anyone designing an AI-native chain or compute network, the 2026 verification toolbox has taken shape. In order of cost per unit of work, from lowest to highest: statistical sampling (Hyperbolic's proof of sampling), refereed delegation (Gensyn's Verde, which needs only one honest verifier but requires bit-exact execution), TEEs (near-native speed, but trust in the chip vendor), full redundancy and general zkML. The practical design is layered: TEEs or sampling for low latency, dispute games or ZK spot checks for finality; a December 2025 paper simulated H100 confidential computing with fraud proofs and random ZK spot checks at 99% of centralized throughput and a marginal cost of $0.07 per query. The hardware philosophy decides security and distribution: GPU-native PoUW is protected at first by rentable cloud GPUs, so the cost of attack is tied to rental prices, and Qubic shows that proof of work on general-purpose hardware can be captured wholesale by outside incentives.

Cost and time are not the barrier. Pearl went from paper (April 2025) to mainnet in about 12 months, forked from btcd, with a full node, wallet, light client and miner at launch. Most DeFi audits cost $25,000–100,000, enterprise or multi-chain audits $150,000–250,000 or more, ZK circuits 80% to 120% more, and projects with real value at risk budget $150,000–500,000 a year for security. The real barriers are distribution and liquidity. Founders report that Binance has asked for 0% to 3.5%, and allegedly 8%, of token supply for a listing (Binance denies profiting from listings); all six projects Cointelegraph reviewed that used "token loan plus call option" market-making deals fell, and Movement's market maker sold about $38 million of tokens on listing day. The alternatives are on-chain liquidity, as with Hyperliquid, which passed $10 billion of market value without a Binance listing and whose HIP-3 perpetual markets passed $3.2 billion of open interest in June, and public sales through large exchanges, as when Monad raised $269 million through Coinbase. A high fully diluted value with a low float was the clearest predictor of post-launch decline in 2025–2026.

Token design is part of the system design. Emission should follow paid demand rather than lead it: the networks in this report with the healthiest subsidy multiples tie rewards to work that an outside buyer pays for, and the ones with the worst pay for hashrate or stake with nothing behind it. A practical rule is to publish, before launch, the subsidy multiple the network aims for in its first, second and third year, and to cut emission when it is missed analysis.

Direction	Who buys, and willingness to pay	Window	Key capabilities	Main risk
Inference-first accelerator that can mine	Neoclouds, private deployments, PoUW networks	2027–2028 (depending on the FP8 and FP4 forks)	INT8, FP8 and FP4 GEMM; SRAM architecture; protocol integration	Protocol admission, architectural lock-in
SRAM or hard-wired inference chip	Neoclouds, data centers in mainland China, edge	Now	Compilers and model adaptation; N6–N4 design	Models change too fast
Mining chip (niche or vertically integrated)	Miners with their own power; home and heating miners	Now, bounded by the halving and the hashprice	Low-voltage arrays, efficiency, volume supply chain	Price wars by incumbents, power prices
ZK proving acceleration	Prover networks, rollups	Wait until proof systems converge	Programmable MSM and NTT	Demand depends on token rewards
FHE acceleration	Encrypted AI for enterprises and governments; privacy chains	2027–2029	NTT, HBM systems, cryptographic engineering	Demand not yet formed; long sales cycles

Direction	Who buys, and willingness to pay	Window	Key capabilities	Main risk
TEE hardening and verifiable inference	Confidential inference networks, DePIN	Now	Memory encryption with integrity, independent roots of attestation, ZK spot checks	Dependence on vendor ecosystems
AI-native chain or compute network	Inference buyers, miners, developers	Now	Verification design, emission design, distribution	Post-launch decline, token liquidity

Three paths I would take, each with a first step

If I had to choose, I would take at most three paths, each with a small first step that tests it before money is committed analysis.

- **Path A: an inference-first accelerator that can also mine.** First step: a four-week architecture study that models J/op and $/op for an SRAM-resident INT8/FP8 GEMM array against the RTX 5090 and H100, both on Pearl's work and on transformer inference, followed by FPGA or mature-node MPW validation with an NRE in the range of a million dollars.
- **Path B: custom-silicon services with crypto as the first customer and AI as the second market.** First step: turn this report's NRE and break-even model into a paid feasibility study for compute-chip projects, with milestone payments and acceptance against a fixed protocol version.
- **Path C: an AI-native network that signs its compute buyer before its token.** First step: a design note on layered verification (TEE plus sampling plus ZK spot checks) and on a vendor-neutral canonical arithmetic, written with one named inference buyer before any token design.

CHAPTER 12

12 SIGNPOSTS
What to watch, November 2026 to 2027

When	What to watch	Why it matters
Any time	Pearl's FP8 activation height	Decides whether integer hardware and non-NVIDIA cards stay admitted, and whether the chain splits
Monthly	Pearl's share of blocks carrying AI work	The single best test of whether "useful" is becoming true
Monthly	Whether inference providers pass mining income on to customers	Paid demand meeting emission: the model Pearl needs
Quarterly	Bittensor's verifiable revenue against its emissions	Whether the largest useful-work network is closing its subsidy gap
Quarterly	The share of x402 flows actually initiated by agents	Whether agent payments move from signalling to commerce
January 2027	TSMC's reported 3% to 6% price rise, most for 2 nm and 3 nm	The cost of leading-node tape-outs

When	What to watch	Why it matters
First quarter of 2027	FP4 in Pearl, as reported from the founders' AMA	A further shift of the work towards NVIDIA's newest formats
End of 2027	Samsung Foundry's target to break even	Whether a cheaper second source stays healthy
April 2028	Bitcoin's next halving	Halves the pot for SHA-256 miners

CHAPTER 13

13 QUESTIONS TO ASK
Questions you did not ask but should have

The questions below are the ones I would ask before committing money or silicon to any AI × crypto project.

Who controls the work function? If a small team can change the arithmetic at a hard-coded height, your hardware's life is their decision. Ask for a written stability commitment, or design hardware with a second market.

What is the subsidy multiple? Divide emissions (or market value) by revenue paid from outside the network. A multiple above 10 means the network is buying activity, not selling a service.

What does the next gate cost, and what number stops you? A multi-project wafer costs tens or hundreds of thousands of dollars; a mask set costs millions. Decide the stopping number before you spend.

What is the payback hashrate, with p? Divide the NRE by $R \times (1 - 1/k) \times T \times p$. If the answer is a large share of the network, the chip is a bet on the rules, not on the silicon.

What is the chip's second market? If the protocol forks or the token falls, what else can the silicon sell into? GPUs survived the Ethereum Merge because they had one; single-algorithm ASICs did not.

Does the value land in the token or in the equity? In 2026 the profitable AI × crypto companies are mostly equity companies with crypto distribution; decide which of the two you are building before you raise money.

Who is the first paying customer, by name? Every durable case in this report, from Grass to Bittensor's compute subnets, began with a named buyer outside the network.

These are areas I would like to explore together with others, through technical exchange and technical collaboration: custom chips for mining, AI training and inference, and accelerators; AI development; and AI × crypto systems from chain to node to chip. You can reach me at i@tanshuai.com.

Limitations of this report

This is desk research based on public documents up to November 13, 2026. I did no primary interviews and no independent measurements for this edition. Several key numbers are company or vendor claims, such as Aethir's revenue, Grass's revenue, Etched's and Taalas's performance and Cysic's efficiency; they are labelled as such. Chain metrics for Pearl come from an independent full node and from pools and change by the hour; I used the readings around 07:13 UTC on November 13 and give same-day ranges. Wafer, mask and design costs are third-party estimates and model values that disagree with each other; I show ranges. My ASIC break-even arithmetic and the staged budget use illustrative assumptions (k, T, p, NRE and gate costs) that a reader should replace with their own.



About the author

Tan Shuai is a technologist and founder with more than twenty years of work across chips, hardware and system software, and 20 granted patents. He has many years of fabless chip R&D and management behind him, including full-custom ASIC datapath design for cryptocurrency mining chips and multimedia camera chips, which is the lens of this report: energy per operation, cost per wafer, what a protocol lets you fix in silicon, and who pays. His work has taken him through Silicon Valley, New York, Chicago, Taiwan, the Greater Bay Area (Shenzhen and Hong Kong), Vietnam and Malaysia. Today he works as an AI researcher and developer.

Technical discussion: i@tanshuai.com · tanshuai.com

Notes and disclaimer

This report is independent desk research based on public documents available up to November 13, 2026. It is provided for information and learning only and is not investment or engineering advice. It considers technology and business only. Prices, chain metrics and protocol rules change; verify anything you rely on against the primary source. Figures marked vendor claim, estimate or analysis should be read as such. Crypto-asset prices are highly volatile.

© 2026 Tan Shuai. You may share this report unchanged with attribution.



Sources

All accessed between November 10 and November 13, 2026. Grouped by chapter.

Pearl: protocol, code and chain data

- 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 source code: genesis block: https://github.com/pearl-research-labs/pearl/blob/main/node/chaincfg/genesis.go ; consensus parameters and fork heights: https://github.com/pearl-research-labs/pearl/blob/main/node/chaincfg/params.go ; hashrate accounting: https://github.com/pearl-research-labs/pearl/blob/main/zk-pow/src/api/sanity_checks.rs ; salted-seed fork upgrade guide: https://github.com/pearl-research-labs/pearl/blob/main/docs/salted-seed-fork-upgrade-guide.md ; fp8-scheme branch parameters: https://github.com/pearl-research-labs/pearl/blob/fp8-scheme/node/chaincfg/params.go
- Pearl Improvement Proposals: https://github.com/pearl-research-labs/pips ; PIP-3, floating-point mining: https://github.com/pearl-research-labs/pips/blob/main/PIPS/pip-0003.md
- prlstats (independent full node): network: https://prlstats.com/api/network ; series: https://prlstats.com/api/series ; home and miner income model: https://prlstats.com/api/home ; mined supply: https://prlstats.com/api/mined ; pools: https://prlstats.com/api/pools ; holders: https://prlstats.com/api/holders ; hardware and overhead assumptions: https://prlstats.com/api/assumptions
- Kryptex Pool: Pearl network: https://pool.kryptex.com/prl/about-coin ; hashrate: 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 ; price history: https://api.coingecko.com/api/v3/coins/pearl-2/market_chart?vs_currency=usd&days=365&interval=daily ; "Perle" ticker search: 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", June 2, 2026: 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, partnership with Pearl Research Labs: https://together.ai/blog/together-ai-partners-with-pearl-research-labs ; model page: https://together.ai/models/gemma-4-31b-it-pearl
- Basu, census of Pearl mining instances, arXiv 2606.04819: https://arxiv.org/html/2606.04819v2
- Blockworks, 0xResearch newsletter, October 28, 2026: https://blockworks.com/newsletter/0xresearch/issue/post_daf0d74a-2bf4-41e4-86d3-2a1275475a24
- Datawallet, "Pearl Explained": https://www.datawallet.com/crypto/pearl-explained
- Coingabbar on TheMarker's report of Pearl Research Labs' funding: https://www.coingabbar.com/en/pearl-research-labs-funding-130m-stealth-round-themarker

Useful-work chains and history

- Grothe et al., Gridcoin security analysis, USENIX WOOT 2017: https://www.usenix.org/system/files/conference/woot17/woot17-paper-grothe.pdf
- Proof-of-learning spoofing, arXiv 2208.03567: https://arxiv.org/abs/2208.03567
- hoge.gg, verification and trust in decentralized inference: https://hoge.gg/decentralized-inference-2026-verification-trust/ ; Bittensor revenue and treasury: https://hoge.gg/bittensor-quantum-wallet-revenue-tao-treasury-2026/
- Kaspa: kHeavyHash matrix code: https://github.com/kaspanet/rusty-kaspa/blob/master/consensus/pow/src/matrix.rs ; launch: https://coinranking.com/coin/V8GxkwWow+kaspa-kas/about ; IceRiver KS1/KS2 announcement: https://www.iceriver.io/2023/03/14/1677644580/ ; GPU-to-ASIC history: https://captainmining.com/blogs/crypto-news/the-evolution-of-kaspa-mining-from-gpus-to-asic-powerhouses ; GPU hashrates: https://coincalculators.io/algorithms/kheavyhash ; KS7: https://www.asicminervalue.com/miners/bitmain/antminer-ks7-40th ; network: https://minerstat.com/coin/kas ; 1 EH/s in 2024: https://ourcryptotalk.com/news/kaspa-reaches-1-ehs-in-hashrate ; Crescendo: https://ourcryptotalk.com/news/kaspa-crescendo-live
- Aleo: mainnet FAQ: https://aleo.org/post/aleo-mainnet-faq/ ; decentralized proving: https://aleo.org/post/decentralized-proving-advantages/ ; ASICs: 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 on the Monero reorganization: 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 ; Dogecoin mining: https://hackernoon.com/qubic-is-now-mining-dogecoin-on-its-ai-compute-infrastructure ; CoinGecko: https://www.coingecko.com/en/coins/qubic

- Bittensor consensus documentation: 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
- Monero and ASIC-resistance forks: https://lbry.com/news/forking-for-asic-resistance ; Sia fork: https://hotfix.coindesk.com/markets/2018/11/01/sia-completes-hard-fork-as-bricked-miners-form-rival-blockchain ; Obelisk prices: https://www.cryptoglobe.com/latest/2018/11/obelisk-hardware-skyrockets-in-selling-price-after-sia-hard-fork/ ; miners after the Merge: https://hotfix.coindesk.com/business/2022/11/11/ex-ethereum-miners-token-hop-to-stay-alive-after-the-merge

AI × crypto sectors

- DefiLlama: Render: https://defillama.com/protocol/render-network-bme ; Livepeer: https://defillama.com/protocol/livepeer
- Cointelegraph on Messari's DePIN revenue: https://cointelegraph.com/news/depin-revenues-grow-despite-token-losses
- Aethir, 2025 wrap-up: https://aethir.com/blog-posts/aethirs-2025-wrap-up-decentralized-gpu-cloud-milestones
- CoinGecko, AI category: https://www.coingecko.com/en/categories/artificial-intelligence
- Akash, H100 rental prices: https://akash.network/the-bid/h100-rental-price-2026-cost-per-hour/ ; Q1 2026 report: https://akash.network/blog/akash-network-q1-2026-report/
- io.net, $8 million deal and daily tokens: 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 contract: 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 on INTELLECT-3: https://www.implicator.ai/prime-intellects-intellect-3-open-source-ambition-meets-centralized-reality/
- SiliconANGLE, Prime Intellect funding: 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 funding: https://runtimewire.com/article/nous-research-75-million-round-1-5-billion-valuation
- ownyourmind, Gensyn: https://ownyourmind.ai/projects/gensyn/ ; Phemex, Ritual (headline only): https://phemex.com/news/article/ritual-launches-ai-layer-1-mainnet-and-ritual-token-price-dives-72-from-high-101259
- Confidential computing overhead, arXiv 2409.03992: https://arxiv.org/abs/2409.03992v2 ; Phala 2025 report: 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, measuring AI agent payments: https://www.trmlabs.com/trm-tech-blog/whos-actually-paying-measuring-ai-agent-payments-onchain ; Forkast on the x402 Foundation: 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 mainnet: https://www.theblock.co/amp/post/394131/tempo-mainnet-goes-live-with-machine-payments-protocol-for-agents ; BlockEden on Virtuals' aGDP: https://blockeden.xyz/blog/2026/04/21/agdp-agent-gdp-virtuals-protocol-ai-blockchain-valuation-primitive-tvl-displacement/ ; The Block on ElizaOS: https://theblock.co/post/410774/eliza-labs-native-token-dead
- Dealroom, Grass revenue: https://dealroom.co/news/157794-grass-backed-by-multicoin-projects-75m-in-2026-data-revenue/ ; CryptoSlate, Fetch.ai and Ocean: https://cryptoslate.com/cryptos-flagship-ai-pact-fracture-fetch-sues-ocean-over-263m-fet-community-sales/ ; Fortune on World: https://fortune.com/2026/06/01/internet-according-to-sam-altman-seeing-orb/
- Galaxy, crypto venture capital: https://www.galaxy.com/insights/research/crypto-blockchain-venture-capital-q3 ; Crypto Briefing (headline only): https://cryptobriefing.com/crypto-vc-investment-q3-2026-galaxy/ ; Blockhead on Tiger Research: https://www.blockhead.co/2026/10/14/bitcoin-rose-43-in-q3-crypto-venture-funding-did-not-follow/ ; ForkLog on 2025 token launches: https://forklog.com/en/85-of-tokens-launched-in-2025-fell-below-initial-prices/

ASIC economics and chip types

- Hashrate Index, roundup of November 9, 2026: https://hashrateindex.com/blog/hashrate-index-roundup-november-9-2026/
- AnySilicon, ASIC design cost by node in 2026: https://anysilicon.com/news/how-much-does-asic-design-cost-in-2026-real-price-ranges-by-process-node/ ; Silicon Analysts, fabless tape-out cost guide: https://siliconanalysts.com/analysis/fabless-startup-tapeout-cost-guide ; market data: https://siliconanalysts.com/market-data
- MillionMiner, Bitmain S23 XP Hyd: https://millionminer.com/news/bitmain-wdms-2026-s23-xp-hyd-launch
- Khattak and Mikaitis, FP8 accumulation in tensor cores, arXiv 2512.07004: https://arxiv.org/html/2512.07004v2
- Canaan, second-quarter 2026 results: https://finviz.com/news/389337/canaan-inc-reports-unaudited-second-quarter-2026-financial-results ; Bitdeer, third-quarter 2026 results: 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 and Block: https://theenergymag.com/news/2026-07-29/core-scientific-exit-block-bitcoin-deal-ai ; AI lease economics: https://theenergymag.com/news/2026-08-06/miner-weekly-ai-pumpability-out-fuel
- Blockworks, Cipher and 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, Q1 FY27 results: https://globenewswire.com/news-release/2026/11/05/3397020/0/en/iren-reports-q1-fy27-results.html
- Spark, bitcoin ASIC market concentration: https://www.spark.money/research/bitcoin-asic-manufacturer-market-concentration

- Google, TPU7x documentation: 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 paper, 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 pricing: 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 inference accelerators compared: https://themenonlab.blog/blog/ai-inference-accelerators-compared ; Introl, AI accelerators beyond GPUs: https://introl.com/blog/ai-accelerators-beyond-gpus-tpu-trainium-gaudi-cerebras-2025
- StratNews (Reuters), domestic AI chips in state-funded data centers: https://stratnewsglobal.com/team-sng/china-orders-state-funded-data-centres-to-use-domestic-ai-chips/ ; TNW, H200 imports: https://thenextweb.com/news/nvidia-h200-chips-hong-kong-china-import-approval
- Succinct, real-time proving on 16 GPUs: https://blog.succinct.xyz/real-time-proving-16-gpus/ ; Cysic ZK ASIC products: 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 and SemiFive: https://evertiq.com/design/2026-02-20-semifive-secures-design-win-with-niobium-for-fhe-accelerator ; The Quantum Insider, Niobium funding: 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, HPU on FPGA: https://www.zama.org/post/announcing-hpu-on-fpga-the-first-open-source-hardware-accelerator-for-fhe

Process nodes

- Silicon Analysts, wafer pricing: https://siliconanalysts.com/data/wafer-pricing
- Tom's Hardware, TSMC A16 pricing rumours: 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 pricing: https://wccftech.com/smic-5nm-chip-prices-up-to-50-percent-more-expensive/amp/
- EUROPRACTICE, TSMC MPW schedule 2026: 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, Huawei Ascend roadmap: 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, China's AI chip supply chain: 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

Opportunities and launch economics

- Webull (Benzinga), Broadcom earnings call: https://www.webull.com/news/15511095937917952 ; Quartr, Global Unichip Q2 2026: https://quartr.com/events/global-unichip-corp-3443-q2-2026_F9×6NJQY ; Taiwan News, Alchip: https://www.taiwannews.com.tw/news/6370946 ; Design & Reuse, VeriSilicon: 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, Canaan's AI unit: https://www.theblock.co/post/359250/crypto-miner-canaan-to-shutter-non-core-ai-semiconductor-business-unit-amid-strategic-realignment
- Clutch, AI development pricing: https://clutch.co/developers/artificial-intelligence/pricing ; PixelPlex, blockchain development cost: https://pixelplex.io/blog/blockchain-development-cost/
- Hyperbolic, proof of sampling: 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 ; layered verification, arXiv 2512.20176: https://arxiv.org/abs/2512.20176v1
- Sherlock, audit pricing reference for 2026: https://sherlock.xyz/post/smart-contract-audit-pricing-a-market-reference-for-2026
- Blockworks, Binance listing standards: https://blockworks.com/news/binance-listing-standards-scrutiny ; Cointelegraph, market-maker deals: 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

PROOF OF USEFUL WORK · CRYPTO × SILICON

Compute as Consensus

AI × Crypto × Chips in 2026, starting from Pearl's proof of useful work

Every key number in this report carries its source and date. Where the author derives a number, it is marked as an estimate; where only the seller reports it, it is marked as a vendor claim. It considers technology and business only.

Tan Shuai

Technologist and founder · AI researcher and developer

tanshuai.com · i@tanshuai.com

PUBLIC EDITION · Data as of November 13, 2026

