FORSKNINGSNOTAT · OFFENTLIG UTGÅVA

2026 · 11



BEVIS PÅ NYTTIGT ARBETE · KRYPTO × KISEL

# Beräkningskraft som konsensus

## AI × krypto × chip 2026, med Pearls bevis på nyttigt arbete som utgångspunkt

Pearl · Nyttigt arbete · DePIN och agenter · ASIC-ekonomi · Processnoder · Budgetar i steg

**SLUTSATSEN**

**Pearl visar att matrismultiplikation kan säkra en blockkedja, men ännu inte att arbetet är nyttigt. Inom AI × krypto finns den betalda efterfrågan i inferens, data och integritet, och en ASIC lönar sig bara där potten är stor, reglerna stabila och chipet har en andra marknad.**


**Tan Shuai**

Teknolog och grundare · AI-forskare och AI-utvecklare

tanshuai.com · i@tanshuai.com


Data per 13 november 2026

Belopp i US-dollar

INNEHÅLL

# Innehåll


◆ **Sammanfattning** — Teoremet håller, incitamentet gör det inte — 04

01 **Omfattning och metod** — Omfattning, metod och fyra test — 07

02 **Pearl · mekanismen** — Pearl: matrismultiplikation som konsensus — 08

03 **Pearl · ekonomin** — Pearl: skala, emission och nyttoglappet — 10

04 **Pearl · reglerna** — Pearl: regelrisk, numerik och styrning — 12

05 **Kedjor för nyttigt arbete** — Kedjor för nyttigt arbete: varken matriser eller nollkunskap stoppar ASIC:er — 13

06 **Det bredare landskapet** — Resten av AI × krypto: verklig men liten efterfrågan, och värdet flyttar till aktierna — 16

07 **ASIC-ekonomi** — ASIC-frågan: en formel, tre måttstockar — 18

08 **Chip för chip** — Fem chiptyper, fem svar — 21

09 **Processnoder** — Processnoder: låt den största livstidskostnaden välja — 24

10 **En budget i steg** — Från modell till kisel: en budget i steg för ett första chip — 27

11 **Möjligheter** — Var möjligheterna finns, och vad jag skulle göra — 29

12 **Signaler att följa** — Vad man ska följa, november 2026 till 2027 — 32

13 **Frågor att ställa** — Frågor du inte ställde men borde ha ställt — 33

◆ **Om författaren** — Tan Shuai — 34

◆ **Källor** — 35

FIGURER

# Figurer


| | | |
|---|---|---|
| FIG. 01 | AI × krypto i sex sektorer: var betald efterfrågan är verklig | 06 |
| FIG. 02 | På Pearls heltalsarbete delar GPU:er samma energitak | 09 |
| FIG. 03 | Pearls första sex månader: hashkraften följer priset, reglerna ändras | 11 |
| FIG. 04 | Subventionsmultipeln: hur många gånger de externa intäkterna ett nätverk betalar | 15 |
| FIG. 05 | Varför effektivitet slutar löna sig: en ASIC behåller högst 1 − 1/k av intäkterna | 19 |
| FIG. 06 | De flesta potter är mindre än en enda chipbudget | 20 |
| FIG. 07 | Fem chiptyper, fem svar | 24 |
| FIG. 08 | En wafer i spetsnod kostar tio gånger en mogen | 25 |
| FIG. 09 | Köp belägg före masker: en budget i steg för ett första chip | 28 |


**Märkningar i texten**

- **uppskattning** författarens beräkning eller slutsats utifrån citerade data
- **leverantörsuppgift** publicerat av företaget som säljer produkten, inte oberoende verifierat
- **analys** författarens tolkning av underlaget

Detta är en översättning av den engelska utgåvan; där de skiljer sig åt gäller den engelska utgåvan.

Data per 13 november 2026 · Belopp i US-dollar

SAMMANFATTNING

# Teoremet håller, incitamentet gör det inte

SLUTSATSEN

**Pearl visar att ”beräkningskraft som konsensus” fungerar som teknik, men ännu inte att det är nyttigt.**

**≈90 EH/s**
Pearls nätverk den 13 november 2026, där en H är en INT7-multiplikation med ackumulering

**≈1 %**
av Pearls block bär verkligt AI-arbete (Blockworks, 28 oktober 2026)

**≈70 mn $**
om året i verifierbara externa intäkter hos Bittensor, mot mer än 300 miljoner dollar i emissioner

**175–180×**
Aleos belöningar till bevisare jämfört med avgifterna under tredje kvartalet 2025

**R × (1 – 1/k)**
vinsttaket för en ASIC som är k gånger billigare än den marginella minern
uppskattning

**≈1,5 mn $**
för att nå uppmätt kisel på N7 via en modell, en FPGA och en MPW, före ett maskset på ≈3,5 mn $ uppskattning

**Pearl**

Den 13 november 2026 körde Pearls nätverk med omkring 90 EH/s. I konsensuskoden är en ”hash” en heltalsmultiplikation med ackumulering, så nätverket motsvarar heltalskapaciteten hos omkring 122 000 H100. Ändå bär bara omkring 1 % av blocken verkligt AI-arbete, och det enda offentliga inferenspartnerskapet erbjuds inte längre i leverantörens serverlösa API. Orsaken är ekonomisk, inte kryptografisk: en GPU som utvinner medan den kör inferens får bara omkring 1 % till 27 % av hashkraften hos en dedikerad miner, så rationella miners multiplicerar slumpmatriser.

**Betald efterfrågan**

Inom AI × krypto syns betald efterfrågan som går att verifiera på tre ställen: försäljning av inferens, försäljning av data och försäljning av integritet. Bittensors verifierbara externa intäkter är omkring 70 miljoner dollar om året, varav omkring fyra femtedelar från subnät för beräkning och infrastruktur. Grass redovisar 17 miljoner dollar i dataintäkter under första halvåret 2026. Konfidentiell inferens i betrodda exekveringsmiljöer är i produktion. Betalningar som faktiskt initieras av AI-agenter uppgår däremot till omkring 5 000–11 000 dollar i månaden, och zkML och decentraliserad träning har ännu ingen redovisad betalande efterfrågan.

**Subvention och värde**

Subventionsmultipeln är det ärligaste hälsotestet. Aleos belöningar till bevisare var omkring 175–180 gånger avgifterna; Bittensors emissioner är omkring 4–11 gånger de verifierbara intäkterna; Renders marknadsvärde är omkring 360 gånger avgifterna i kedjan; och Pearls miners intäkter på ungefär 460 miljoner dollar om året är till största delen blocksubvention, nästan utan koppling till betald AI-efterfrågan. Subventioner är bränsle för en kallstart, inte en affärsmodell. Samtidigt flyttar värdet från tokens till aktier: fyra aktiebolag som använder centraliserad infrastruktur och kryptonativ distribution värderas till omkring 8,5 miljarder dollar, ungefär lika mycket som AI-tokens på protokoll- och applikationsnivå tillsammans.

**ASIC:er**

För chip är vinsttaket för en ASIC ungefär intäkten per beräkningsenhet gånger (1 − 1/k), där k anger hur många gånger billigare den är än den marginella hårdvaran. Eftersom termen är konkav ger effektivitet bortom k = 5–10 nästan ingen ytterligare vinst; det som avgör utfallet är pottens storlek, hur länge reglerna består och om chipet har en andra marknad. Mina svar per chiptyp: gör ingen tape-out av ett dedikerat INT7-chip för Pearl nu, utan satsa i stället på en inferensaccelerator som också kan utvinna; håll dig som nykomling borta från marknaden för bitcoin-miners till försäljning; bygg inget fristående träningschip som litet team; bygg inferenschip när en kund är tecknad; gardera med ZK-acceleratorer och positionera dig tidigt inom FHE.

**Noder och budgetar**

Processnoden bör väljas efter den största posten i chipets livstidskostnad: energi för chip som går dygnet runt, kapsling och minne för träningschip, valet mellan SRAM och HBM för inferenschip. Och ett första chip bör köpas i steg, en modell, en FPGA, en multiprojektwafer och först därefter ett maskset, så att en ändring i protokollet kostar priset för en prototyp snarare än för en tape-out. Möjligheterna finns där fälten korsas: kryptot som första kund, AI som andra marknad.

| Fråga | Bedömning | Nyckeltal (källor i kapitlen) |
|---|---|---|
| ① **Fungerar Pearl, och är arbetet nyttigt?** | Det fungerar: en blockkedja som säkras av matrismultiplikation i stor skala. Ännu inte nyttigt: nästan all mining sker på slumpmatriser | Omkring 90 EH/s den 13 november 2026; 1 H = 1 INT7-MAC; omkring 1 % av blocken med AI-arbete |
| ② **Var finns den verkliga betalda efterfrågan inom AI × krypto?** | Inferens, data och integritet. Agentbetalningar, zkML och decentraliserad träning är fortfarande före efterfrågan | Bittensor omkring 70 miljoner dollar om året i verifierbara intäkter; Grass 17 miljoner dollar första halvåret 2026; agentbetalningar omkring 5 000–11 000 dollar i månaden |
| ③ **Är en ASIC värd att bygga?** | Bara där potten är stor, reglerna stabila och chipet har en andra marknad. I dag erbjuder inget nätverk alla tre | Vinsttak ≈ R × (1 − 1/k); bitcoin omkring 11,5 miljarder dollar om året i miningintäkter, Pearl omkring 0,46–0,49 miljarder, Kaspa omkring 18 miljoner |
| ④ **Vilka chip bör ett litet team bygga?** | Inferensacceleratorer som också kan utvinna; inferenschip för en tecknad kund; tidigt FHE-arbete. Inte bitcoin-miners till försäljning eller träningschip | Kaspas bästa ASIC tjänar 0,86 dollar om dagen; designbudgetar för 3/2 nm 60–400 miljoner dollar eller mer |
| ⑤ **Vilken processnod?** | Den största livstidskostnaden avgör: energi, kapsling och minne, eller minnesarkitekturen | TSMC N2 omkring 30 000 dollar per wafer mot omkring 3 000 dollar för N28; HBM och CoWoS omkring 63 % av kostnaden för en H100 |
| ⑥ **Vad kostar ett första chip, och hur bör det köpas?** | I fyra steg: en modell, en FPGA, en multiprojektwafer, sedan masker. Stanna före maskerna om reglerna eller köparen inte är klara | En kommersiell MPW-plats 50 000–500 000 dollar; N7-masker omkring 3,5 miljoner dollar; en designbudget för 7/6 nm 15–100 miljoner dollar |

**FIG. 01** **AI × krypto i sex sektorer: var betald efterfrågan är verklig**

Verifierbar betald efterfrågan mot signaler om subvention och värdering, november 2026

### Kedjor för nyttigt arbete 01

VERKLIG MEN LITEN

**≈70 mn $ om året**

Bittensors verifierbara externa intäkter

**Signal:** emissioner 4–11× intäkterna; Pearl ≈1 % AI-block

**Mekanismerna fungerar; nyttan kräver betalare**

### Decentraliserad beräkning 02

VERKLIG MEN LITEN

**2,35 mn $ om året**

Renders avgifter i kedjan, de största verifierbara

**Signal:** marknadsvärde ≈360× avgifterna

**Mest vidareförsäljning av standardinferens**

### Träning och inferens 03

VINST I BOLAGEN

**över 100 mn $ om året**

Prime Intellect, ett aktiebolag på centraliserade GPU:er

**Signal:** Gensyns huvudnät för träning ej i drift

**Framstegen är verkliga; vinsten går till ägarna**

### Verifierbar AI 04

TIDIGT

**398**

betalande användare av Phalas TEE-moln, slutet av 2025

**Signal:** ingen redovisad betald zkML

**TEE i produktion; zkML tidigt**

### Agentbetalningar 05

FÖRE EFTERFRÅGAN

**5 000–11 000 $/mån**

x402-betalningar som agenter faktiskt initierat

**Signal:** ≈200 mn transaktioner, mest signalering

**Rälsen kom först; efterfrågan har inte kommit**

### Datanätverk 06

VERKLIGA KÖPARE

**17 mn $**

Grass dataintäkter, första halvåret 2026 (egen uppgift)

**Signal:** marknadsvärde ≈10× årstakten

**En av få sektorer med verkliga köpare**

**Källa:** DefiLlama, SubConnect (via hoge.gg), TRM Labs, Phala, Dealroom, SiliconANGLE, CoinGecko, Blockworks; författarens sammanställning.



# 01 OMFATTNING OCH METOD
# Omfattning, metod och fyra test

## Varför denna rapport, och varför nu

I april 2026 satte en ny blockkedja, Pearl, en gammal idé i produktion: låt miners tävla om block genom att utföra de matrismultiplikationer som AI bygger på, i stället för att hasha. Inom sex månader nådde nätverket omkring 90 EH/s och dess token ett marknadsvärde på omkring 500 miljoner dollar. Samtidigt tog marknadsplatser för AI-beräkning, nätverk för decentraliserad träning, betalräls för agenter och datanätverk in miljarder, och bitcoin-miners började bygga om sina kraftanläggningar till AI-datacenter. Jag ville veta vilka av dessa som har verkliga kunder, om kundanpassat kisel är meningsfullt någonstans bland dem, och vilken processnod och vilken foundry ett chip-team faktiskt kan använda.

## Fyra test för varje projekt inom AI × krypto

En säkerhetsstudie av Gridcoin från 2017 hävdade att nyttig beräkning kan säkra en blockkedja bara om den är svår att lösa, lätt att verifiera och lätt att belöna. Jag lägger till ett fjärde test. Tillsammans avgör de det mesta av det som följer.

01 **Svår att lösa.** Ärlig beräkning måste vara det billigaste sättet att ta fram ett giltigt bevis; det får inte finnas någon genväg.

02 **Lätt att verifiera.** Att kontrollera arbetet måste kosta långt mindre än att utföra det, annars betalar nätverket två gånger.

03 **Lätt att belöna.** Protokollet måste kunna beräkna belöningen enbart utifrån beviset, utan att lita på en tredje part.

04 **Någon betalar för resultatet.** Om inte en kund utanför nätverket betalar för resultatet är det ”nyttiga” arbetet en subvention med en berättelse.

Hashpussel klarar de tre första testen och faller på det fjärde. Verkliga AI-uppgifter klarar det fjärde men faller på det andra: att bevisa en AI-beräkning med nollkunskap kostar 100 till 10 000 gånger själva beräkningen, och system för proof of learning har visat sig gå att förfalska billigt. Pearls brusiga matrismultiplikation är den första trovärdiga bryggan som klarar verifieringstestet och samtidigt har samma form som AI-beräkning.

## Hur jag arbetade

Det här är skrivbordsforskning, gjord på samma sätt som jag skulle granska ett chip eller en affärsplan. Jag läste Pearls kod, dess protokollförslag och dess vitbok; kedjedata från en oberoende fullnod; företagsrapporter och resultatmeddelanden; artiklar inom kryptografi, datorarkitektur och maskininlärning; prisdata från foundries och leverantörer; samt bransch- och teknikpress. Varje siffra anges med datum. Där jag själv härleder en siffra säger jag det och märker den uppskattning. Siffror som bara säljaren uppger märks leverantörsuppgift; min tolkning av underlaget märks analys. Belopp anges i US-dollar. Läget avser 13 november 2026.

KAPITEL 02

# 02 PEARL · MEKANISMEN
# Pearl: matrismultiplikation som konsensus

## Idén, från teorem till huvudnät

I april 2025 publicerade Ilan Komargodski och Omri Weinstein ”Proofs of Useful Work from Arbitrary Matrix Multiplication”. De konstruerar ett bevis på nyttigt arbete för varje matrisprodukt med en merkostnad på bara 1 + o(1) gånger en naiv multiplikation; säkerheten vilar på att det är svårt att lösa satser av slumpmässiga linjära ekvationer med låg rang, vilket är en förmodan, inte ett bevis. Artikeln pekar ut fältets verkliga svårighet: ett bevis på nyttigt arbete med en merkostnad som är en konstant faktor är trivialt men oanvändbart; det svåra är att hålla ärlig beräkning som det billigaste sättet att vinna när miners själva väljer sina indata.

Pearls huvudnät startade den 27 april 2026 kl. 09:00 UTC. Miners beräknar en brusig produkt (A + E)(B + F), där bruset med låg rang härleds ur ett BLAKE3-åtagande, och tar sedan bort bruset för att återfå A·B. Ackumulatortillståndet i varje utdataruta hashas till en lott; en vinnande ruta öppnas och bevisas med ett rekursivt Plonky2-nollkunskapsbevis under 60 KB, inuti ett blockcertifikat med ett tak på 65 KB. Verifieringen är ett litet bevis av konstant storlek, oberoende av matrisstorleken: lika billigt att kontrollera som ett hashpussel, medan själva beräkningen kan vara nyttig.

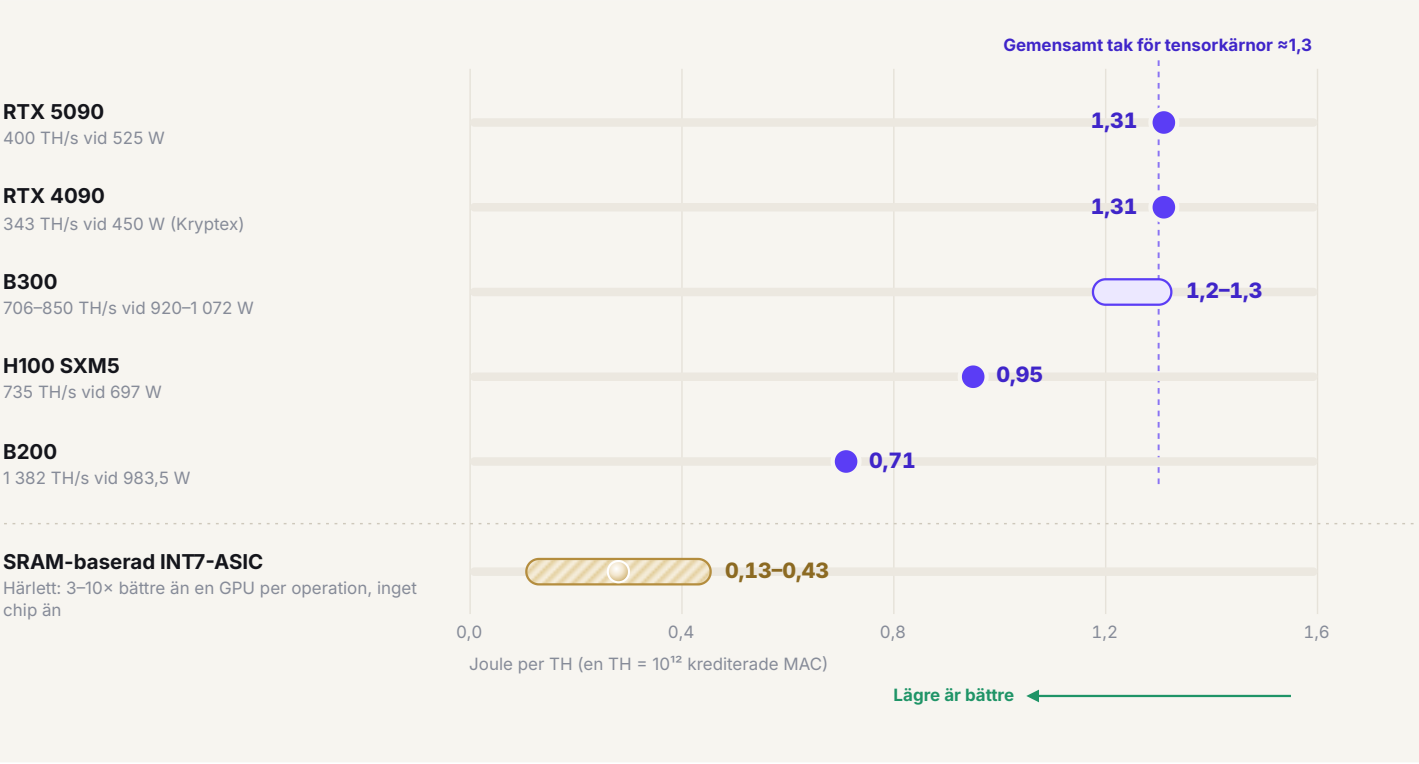
## En MAC-operation är en ”hash”

För en chipdesigner är det viktigaste faktumet enheten. Pearls konsensuskod skalar varje försöks mål med rutstorleken gånger skalärproduktens längd, så **1 H är exakt en krediterad INT7 × INT7 → INT32-multiplikation med ackumulering (MAC)**. Sedan block 96 251 kräver ett rangstraff en brusrang på minst 128; vid minsta rang är 1 H fortfarande lika med 1 MAC. Pearls ”EH/s” betyder alltså bokstavligen 10^18 heltals-MAC per sekund: **för första gången prissätts en blockkedjas säkerhetsbudget i AI-acceleratorers tensorgenomströmning.** Uppmätt hashkraft på GPU:er ligger på 60 % till 95 % av varje korts toppvärde för täta INT8-MAC, vilket stämmer med denna läsning.

| Hårdvara | Hashkraft på Pearl | Effekt | Joule per biljon MAC (J/TH) | Källa (datum) |
|---|---|---|---|---|
| **RTX 5090** | 400 TH/s | 525 W | 1,31 uppskattning | Kryptex (12 november) |
| **RTX 4090** | omkring 305 TH/s (communityns intervall 280–330); Kryptex anger 343 TH/s | 450 W (Kryptex) | 1,31 (Kryptex siffror) uppskattning | prlstats (3 oktober); Kryptex |
| **H100 SXM5** | 735 TH/s | 697 W | 0,95 uppskattning | PeakMiner, via prlstats (19 september) |
| **B200** | 1,382 PH/s | 983,5 W | 0,71 uppskattning | PeakMiner, via prlstats |
| **B300** | 706–850 TH/s | 920–1 072 W | 1,2–1,3 uppskattning | prlstats; SimeonOnSecurity |

**FIG. 02** **På Pearls heltalsarbete delar GPU:er samma energitak**

Joule per biljon INT7-MAC (J/TH) vid uppmätt hashkraft på Pearl; lägre är bättre




**Källa:** Kryptex (12 november 2026); PeakMiner- och SRBMiner-mätningar via prlstats; SimeonOnSecurity; ASIC-intervallet är författarens slutsats, ingen mätning.

Två detaljer i tabellen är viktiga för chipdesigners. För det första möts konsumentkorten RTX 4090 och 5090 och datacenterkortet B300 vid omkring 1,3 J/TH, och även det bästa kortet, B200, når bara omkring 0,71 J/TH. En datacenter-GPU som kostar tiotusentals dollar gör av med nästan lika mycket energi per heltals-MAC som ett spelkort, vilket pekar på ett gemensamt tak för tensorkärnornas energi per INT8-MAC, och just det taket är vad ett dedikerat chip skulle kunna bryta. För det andra är B300 långsammare än B200: NVIDIA skar kraftigt i INT8-genomströmningen i Blackwell Ultra, så heltalsarbetet måste ta en långsammare väg på det chipet. Den gängse AI-kiseln rör sig bort från INT8 mot FP8 och FP4. När det talformat en konsensus väljer avviker från vart hårdvaran är på väg tvingas protokollet ändra sina regler, och det är roten till FP8-forken och till min ASIC-bedömning längre fram i rapporten.

KAPITEL 03

# 03 PEARL · EKONOMIN
# Pearl: skala, emission och nyttoglappet

## Från 0,01 till 90 EH/s på sex månader

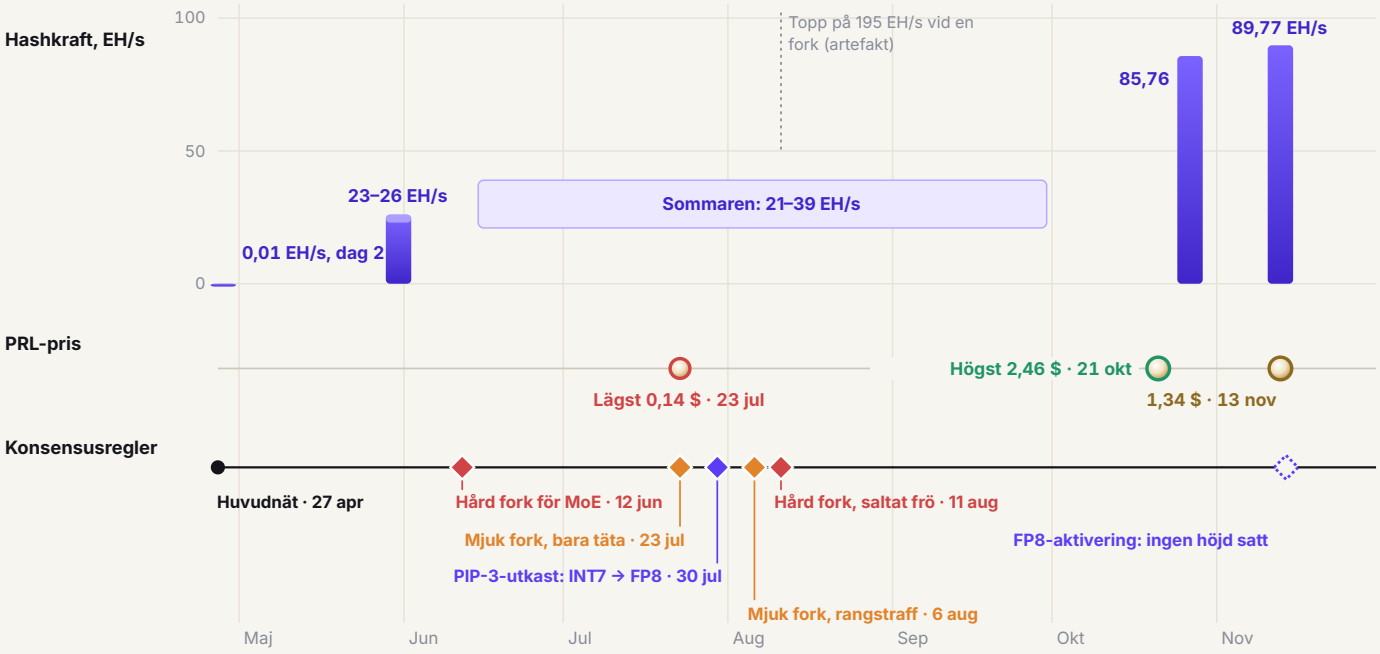
| Mått | Värde (13 november 2026) | Underlag och källa |
|---|---|---|
| **Nätverkets hashkraft** | Omkring 90 EH/s (basvärde 89,77; intervall samma dag 88,20–93,22; dagens högsta den 12 november 110–123) | Oberoende fullnod prlstats; Kryptex |
| **Svårighetsgrad och blockhöjd** | 58 841 982 och 142 530 | prlstats |
| **Motsvarande hårdvara** | Omkring 122 000 H100 eller 225 000 RTX 5090 | Vid 735 respektive 400 TH/s per kort uppskattning |
| **Pris, marknadsvärde och fullt utspätt värde** | 1,32–1,34 dollar; omkring 500 miljoner dollar; omkring 2,8 miljarder dollar (plats 101 på CoinGecko) | CoinGecko |
| **Högsta och lägsta notering någonsin** | 2,46 dollar (21 oktober) och 0,1399 dollar (23 juli) | CoinGecko |
| **Utvunna mynt** | 377,6 miljoner PRL, 17,98 % av taket på 2,1 miljarder | prlstats |
| **Blocksubvention och daglig emission** | Omkring 2 173 PRL och omkring 0,97–1,03 miljoner PRL | prlstats; Kryptex |
| **Nytt utbud under de kommande 12 månaderna** | Omkring 78 % av nuvarande utbud | prlstats |
| **Minerintäkter** | Omkring 1,25 miljoner dollar om dagen (modell), omkring 460 miljoner dollar om året | prlstats |
| **Intäkt per enhet hashkraft, R** | Omkring 14–15 dollar per PH/s och dag; ett RTX 5090 tjänar omkring 5–6 dollar om dagen | uppskattning ; Kryptex |
| **Block som bär verkligt AI-arbete** | Omkring 1 % (uppskattning från 28 oktober) | Blockworks |
| **Koncentration bland pooler** | Kryptex hittade 46,4 % av blocken under de senaste 24 timmarna; de två största poolerna omkring 62 % | prlstats |
| **Koncentration bland innehavare** | Gini-koefficient för adresser 0,969; de 100 största adresserna äger 53,0 % | prlstats |

Pearls hashkraft är nära en eftersläpande funktion av priset. Dagen efter lanseringen körde nätverket med omkring 0,01 EH/s; i slutet av maj nådde det 23–26 EH/s. Under sommaren föll priset omkring 86 % från toppen i maj och hashkraften rörde sig i sidled på 21–39 EH/s. Sedan priset stigit ungefär tiofaldigt i september och oktober nådde hashkraften 85,76 EH/s den 27 oktober och omkring 90 EH/s i november. En topp på 195 EH/s den 11 augusti sammanföll med en fork och bör läsas som en skattningsartefakt. Energiförbrukningen växte i takt: en studie

uppskattade omkring 112 MW för hela nätverket vid 24 EH/s, och en modell från prlstats anger omkring 190–215 MW i dag.

**FIG. 03** **Pearls första sex månader: hashkraften följer priset, reglerna ändras**

Nätverkets hashkraft, tre prisnivåer och varje konsensusändring, 27 april–13 november 2026




**Källa:** prlstats fullnod och serier; Kryptex; CoinGecko; Pearls params.go och PIP-3. Toppen den 11 augusti är en skattningsartefakt vid en fork.

## Emissionen förstärker reflexiviteten

Utbudstaket är 2,1 miljarder mynt och det finns inga halveringar: blocksubventionen minskar jämnt enligt E(t) = S·H / ((t + H)(t + H − 1)) med H = 650 226. Men emissionen räknas i block, inte i tid, och den tidiga svårighetsgraden startade för lågt och justerades långsamt. **Omkring 44,06 miljoner mynt utvanns den första dagen och omkring 120 miljoner under de första sex dagarna**, omkring 32 % av dagens utbud uppskattning. De senaste blocken, på omkring 171–183 sekunder, kommer fortfarande snabbare än målet på 194 sekunder. Resultatet är att utbudet växer med omkring 78 % under de kommande 12 månaderna, en utspädning som miners, hårdvarutillverkare och innehavare alla måste hantera.

Minerekonomin följer av detta. Minerintäkterna i hela nätverket är omkring 1,25 miljoner dollar om dagen (modell); Kryptex dagliga emission gånger aktuellt pris ger omkring 1,35 miljoner dollar om dagen, så **intäkten per enhet hashkraft R ≈ 14–15 dollar per PH/s och dag** uppskattning. Ett RTX 5090 tjänar omkring 5–6 dollar om dagen. En H100 eller H200 tjänar omkring 11 dollar, medan den billigaste offentliga hyran på begäran är omkring 2,09 dollar per kort och timme, omkring 50 dollar om dagen, så att hyra kort bara för att utvinna ger stora förluster. R svänger kraftigt: enligt prlstats serie låg den på omkring 10 dollar i augusti och september och kortvarigt på omkring 33 dollar i mitten av oktober uppskattning. Inom några veckor efter lanseringen halverades den uppskattade dagsintäkten för ett RTX 5090 från omkring 33,80 dollar till omkring 17,19 dollar. Att hyra GPU:er i molnet för att utvinna lönade sig bara under de första veckorna och under pristoppen i oktober.

## Teoremet håller; incitamentet gör det inte

Det verkliga problemet är efterfrågan, inte kryptografi. Pearls enda offentliga kommersiella integration är en Gemma-4-31B-it-pearl-endpoint från Together AI, lanserad den 15 maj till mer än 25 % under ordinarie pris, där

rabatten finansieras av emissionens förväntade framtida värde. Den 13 november visade modellsidan att den **inte är tillgänglig i Togethers serverlösa API**. I juni drog en inventering av 8 012 aktiva mininginstanser, omkring 21 % av nätverket, slutsatsen att nätverket inte producerade någon nyttig AI-beräkning alls (författaren kallar belägget antydande, inte avgörande), och visade att andelar byggda av rena slumpmatriser accepterades av pooler. Den 28 oktober uppskattade Blockworks att **bara omkring 1 % av blocken bär verkligt AI-arbete**; utan det, skrev Blockworks, är Pearl en mindre, mindre säker och snabbare inflaterande bitcoin.

Orsaken går att räkna ut. Merkostnaden på inferenssidan är verkligen liten: Pearl rapporterar 5,08 % längre total tid för Llama 70B på fyra H200, och ett oberoende test mätte 9,2 % förlust i genomströmning på en Hopper-GPU. Miningsidan tjänar för lite. Det officiella vLLM-insticksprogrammet utvinner bara på prefill-satser om minst 1 024 token och aldrig på avkodningssteg. Ett kort som kör inferens får omkring 27 % av en dedikerad miners hashkraft vid ren prefill-trafik, omkring 5,4 % vid en chattmix som till 80 % är avkodning, och bara omkring 1 % vid trafik som domineras av avkodning. Utan betald efterfrågan på inferens multiplicerar en rationell miner slumpmatriser; protokollet verifierar att de utlovade matriserna multiplicerades ärligt, inte att de var nyttiga. **Teoremet håller; incitamentet gör det inte.** Pearls nytta är ett kommersiellt problem utanför kedjan: någon måste betala för inferensen.

KAPITEL 04

# 04 PEARL · REGLERNA
# Pearl: regelrisk, numerik och styrning

## Fyra forkar på två månader

Pearls andra sida är hur snabbt reglerna ändras. Från juni till augusti 2026 drev labbet igenom fyra konsensusändringar vid hårdkodade höjder: en hård fork för mixture-of-experts (block 71 935, omkring 12 juni), en mjuk fork för enbart täta modeller (91 630, omkring 23 juli), en mjuk fork med rangstraff (96 251, omkring 6 augusti) och en hård fork med saltat frö (99 000, omkring 11 augusti). Efter forken med saltat frö kunde gamla miners bara ge ogiltiga andelar; rangstraffet sänkte den effektiva hashkraften för konfigurationer med hög rang med 50 % till 87,5 %.

Nästa steg är PIP-3, fortfarande ett utkast: det flyttar mining från INT7 till FP8 × FP8 → FP32 och kräver bitexakta enhetsprofiler, **med stöd för bara två arkitekturer, Hopper och Blackwell**. Dokumentet kallar hårdvaruberoende ett nödvändigt ont för flyttalsprotokoll och varnar för att varje tolerans i verifieringen bjuder in till grinding-attacker. Den 12 november angav huvudnätsparametrarna i grenen fp8-scheme ingen aktiveringshöjd. Enligt det nuvarande utkastet godkänns inte RTX 40-serien, A100 eller AMD-kort, medan en modell med låg säkerhet från prlstats uppskattar att omkring 82 % av hashkraften kommer från konsumentkort. Att tvinga igenom forken riskerar en kedjedelning eller en motreaktion bland miners.

## Ett trilemma för talformat

Pearl står inför ett trilemma: ett talformat kan vara nyttigt för AI (FP8 eller FP4), bitexakt verifierbart på olika hårdvara och hårdvaruneutralt, men högst två av de tre. INT7-schemat är verifierbart och neutralt men avviker från vart AI-kislet är på väg. FP8 med vitlista är nyttigt och verifierbart men binder konsensus till en leverantör. FP8 med tolerans är nyttigt och neutralt men öppnar för grinding-attacker. Flyttalsaritmetik i tensorkärnor följer i allmänhet ingen numerisk standard och är inte reproducerbar mellan generationer; på Hopper behåller till exempel FP8-ackumuleringen bara 13 bråkbitar och trunkerar. Ur en chipdesigners synvinkel är en ”Hopper-profil” en helt

bestämd aritmetik som en ASIC skulle kunna implementera bit för bit, kanske effektivare än GPU:n. Men att göra det vore affärer i en gråzon mot protokollets uttalade avsikt, med mycket hög styrningsrisk analys.

### Styrning

Styrningen är koncentrerad. Labbet skrev varje protokollförslag; forkar aktiveras vid hårdkodade höjder utan signalering från miners; det finns ingen andra klientimplementation; och i början av oktober hade ingen tredjeparts-granskning av noden eller nollkunskapskretsarna publicerats. Den israeliska tidningen TheMarker rapporterade att labbet tagit in omkring 120–130 miljoner dollar i en hemlig runda (omkring 55 miljoner dollar i eget kapital plus en fond för GPU-infrastruktur på omkring 75 miljoner dollar); Pearl har inte bekräftat detta i officiella kanaler. ”Ingen tilldelning till investerare i token” och ”ett bolag med riskkapitalägare” kan båda vara sanna, men spänningen kvarstår: AI-intäkter kan gå till aktieägarna medan utspädningen drabbar tokeninnehavarna.

Sammantaget är Pearl det renaste experimentet med ”beräkningskraft som konsensus” hittills: en ny mekanism, verifierbar i kedjan, med en enhet som motsvarar AI-hårdvara ett till ett. Men nyttan har inte kommit, och dess säkerhetsbudget, hashkraft och minerentusiasm hänger alla på en token vars huvudsakliga efterfrågan ännu inte har uppstått och vars utbud växer med omkring 78 % på 12 månader. Tre signaler från Blockworks är värda att följa: om uppgraderingen går smidigt, om andelen block med AI-arbete stiger väsentligt, och om någon inferensleverantör för vidare miningintäkterna till sina kunder.

KAPITEL 05

KEDJOR FÖR NYTTIGT ARBETE

# 05 Kedjor för nyttigt arbete: varken matriser eller nollkunskap stoppar ASIC:er

### Kaspa: en matris är varken nyttig eller ASIC-säker

Kaspa är skolexemplet. Dess kHeavyHash lägger en 64 × 64-produkt mellan matris och vektor över 4-bitarsvärden mellan två hashar, och produkten kastas bort. Omkring 16 månader efter den rättvisa lanseringen i november 2021 tillkännagav IceRiver de första ASIC:erna för Kaspa den 14 mars 2023, och i slutet av 2023 var GPU-mining över. En jämförelse av ett RTX 4090 med 3,8 GH/s och 300 W mot Bitmains KS7 med 77 J/TH ger ett effektivitetsgap på omkring **1 000 gånger** uppskattning. Efter festen kom kompressionen: den 13 november körde nätverket med 358,9 PH/s, omkring 64 % under 1 EH/s i augusti 2024, och den bästa KS7 tjänar bara **0,86 dollar om dagen** vid 0,10 dollar per kWh.

### Aleo: inte heller nollkunskap stoppar specialhårdvara

Aleos bevisarpussel är utformat för att efterlikna arbetet med att ta fram ett SNARK-bevis; det bevisar inte användarnas transaktioner, och projektet välkomnar öppet ZK-ASIC:er. Omkring fem månader efter huvudnätet i september 2024 levererades dedikerade maskiner. Under tredje kvartalet 2025 fick bevisarna 31,2 miljoner ALEO (omkring 7,8 miljoner dollar) medan nätverket bara samlade in omkring 43 000–44 000 dollar i avgifter: **en subventionsmultipel på omkring 175–180**. Priset föll från omkring 9 dollar i den tidiga handeln till 0,22 dollar i slutet av det kvartalet, och ARC-46 kräver nu en insats på 100 000 till 2,5 miljoner ALEO per lösning, vilket gör mining till en insatströskel.

## Qubic och Bittensor

Qubic marknadsför ”useful proof of work” som tränar dess neurala nätverk Aigarth, men det mest synliga resultatet av dess beräkningskraft 2025–2026 var andra kedjors block. Först utvann det Monero, ökade sin andel från under 2 % i mitten av maj 2025 till mer än 25 % i slutet av juli och gjorde anspråk på 51 % i augusti, vilket orsakade en omorganisation på sex block (en annan källa säger 18). Den 1 april 2026 gick det över till att utvinna Dogecoin med ASIC:er och använda intäkterna till återköp och bränning. Det finns inga oberoende belägg för att dess AI-träning är nyttig, och marknadsvärdet föll från omkring 280–300 miljoner dollar till omkring 50 miljoner dollar den 13 november. Det blottlade också en extern effekt: en kedja som betalar för generell beräkning skapar en stor, hyrbar, samordnad pool av hashkraft som kan vändas mot CPU-vänliga kedjor.

Bittensor är inte proof of work utan insatsviktad subjektiv poängsättning: validerare poängsätter miners, och vikterna kapas vid den insatsviktade medianen. Det är det enda nätverket i gruppen med externa intäkter i stor skala. SubConnect-indexet, som bara räknar verifierbara externa kunder, visar omkring **70 miljoner dollar om året** i början av november 2026 (omkring 35 miljoner dollar den 1 september), varav 78 % till 82 % från beräkning och infrastruktur; men vid priserna i början av november överstiger emissionerna 300 miljoner dollar om året, så intäkterna täcker bara 9 % till 23 %. Intäktsuppskattningarna varierar kraftigt med underlaget, från omkring 17 miljoner dollar (Artemis) till 172 miljoner dollar i årstakt utifrån egenrapporterade summor. I april brände 67 av 128 subnät mer än 99 % av sina incitament, och när det ledande teamet Covenant lämnade efter en styrningskonflikt föll TAO 24 % inom några timmar.

## Historien ger samma svar

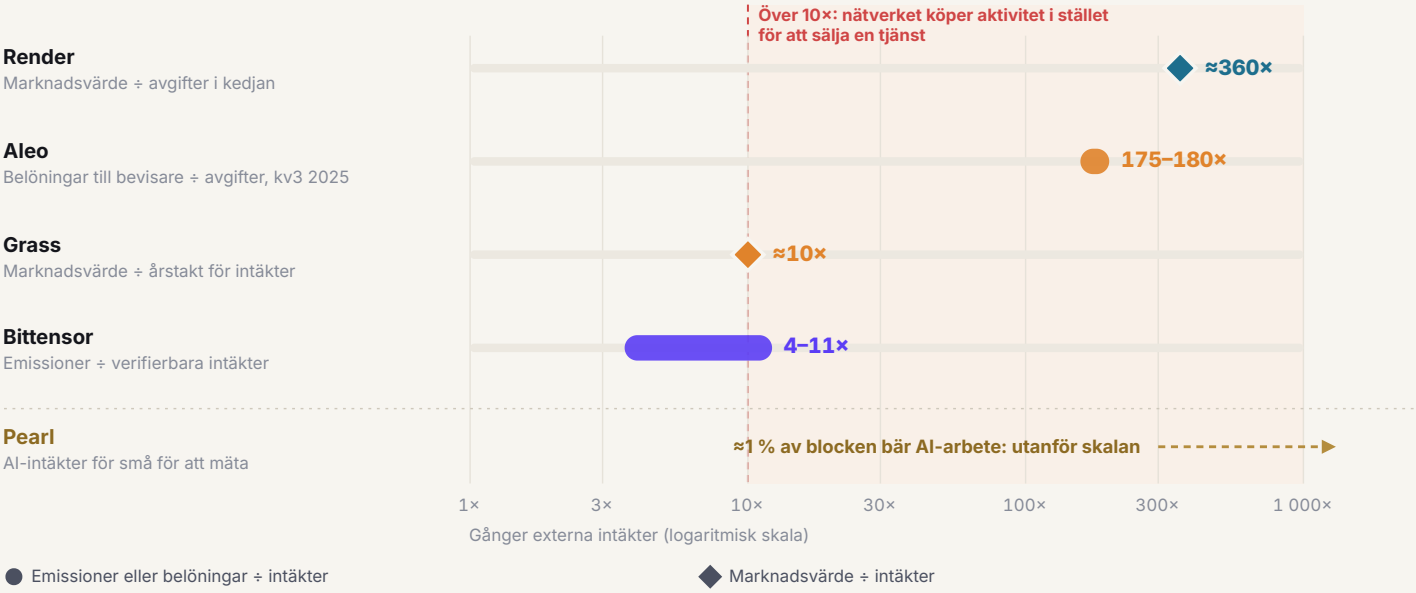
Primecoins primtalskedjor gick snabbt att verifiera, men ingen behövde dem. Gridcoin förlitade sig på statistiksajter från tredje part för att intyga BOINC-bidrag, och forskare förfalskade block utan att utföra någon beräkning. Filecoins lagringsbevis fungerar tekniskt, men under tredje kvartalet 2025 var utnyttjandet bara 36 %, och 99,5 % av avgifterna på 793 000 dollar var straffavgifter, med basavgifter på bara 4 100 dollar.

| Projekt | Arbete eller konsensus | Marknadsvärde (datum) | Hårdvaruväg | Betald efterfrågan mot subvention | Bedömning |
|---|---|---|---|---|---|
| **Pearl** | Brusig matrismultiplikation (INT7-MAC plus ZK-certifikat) | Omkring 500 miljoner dollar (13 november) | GPU:er; FP8-utkastet skulle vitlista Hopper och Blackwell | Omkring 1 % av blocken med AI-arbete; minerintäkter mest subvention | Den mest trovärdiga mekanismen; efterfrågan obevisad |
| **Kaspa** | kHeavyHash (en 64 × 64-produkt mellan matris och vektor med 4 bitar mellan hashar) | Omkring 840 miljoner dollar (13 november) | ASIC:er omkring 16 månader efter lanseringen; nu helt ASIC | Gör inga anspråk på nytta; miningintäkter omkring 18 miljoner dollar om året uppskattning | En matris är varken nyttig eller ASIC-säker |
| **Aleo** | Syntetiskt SNARK-pussel plus proof of stake | Omkring 118 miljoner dollar (slutet av kv3 2025) | Dedikerade maskiner efter omkring fem månader | Belöningar till bevisare omkring 175–180 gånger avgifterna | Syntetiskt arbete, drivet av subventioner |
| **Qubic** | ”uPoW” (uppger att det tränar Aigarth) | Omkring 50 miljoner dollar (13 november) | CPU och GPU; ASIC:er för Dogecoin sedan april 2026 | Inga oberoende belägg för nytta; beräkningskraften används till extern mining och återköp | Aggregering av beräkningskraft inslagen i en AI-berättelse |
| **Bittensor** | Yuma-konsensus (insatsviktad poängsättning) | Omkring 2,55 miljarder dollar (11 november) | Generella GPU:er; ingen ASIC-väg | Omkring 70 miljoner dollar om året i verifierbara intäkter mot mer än 300 miljoner dollar i emissioner | Den enda med externa intäkter i stor skala; ändå en subvention på 4–11 gånger |

| Projekt | Arbete eller konsensus | Marknadsvärde (datum) | Hårdvaruväg | Betald efterfrågan mot subvention | Bedömning |
|---|---|---|---|---|---|
| **Filecoin (referens)** | Bevis på replikering och rumtid | Omkring 1,5 miljarder dollar (slutet av kv3 2025) | — | Utnyttjande 36 %; basavgifter per kvartal 4 100 dollar | Nyttigt, men efterfrågan saknas |

**FIG. 04** **Subventionsmultipeln: hur många gånger de externa intäkterna ett nätverk betalar**

Emissioner eller marknadsvärde delat med intäkter betalda utifrån, logaritmisk skala




**Källa:** Blockworks (Aleo kv3 2025), DefiLlama (Renders avgifter), CoinGecko, hoge.gg (Bittensor), Dealroom (Grass), Blockworks 0xResearch (Pearl); författarens beräkning.

Lärdomen för hårdvara: fast, regelbundet arbete som betalas i en likvid token får ASIC:er inom omkring 3 till 16 månader, ofta från utmanare före de etablerade. Anspråk på ”nytta” höll bara där en extern kund betalar för resultatet. Och de främsta orsakerna till strandad hårdvara är hög subvention i förhållande till intäkterna och utvecklarnas kontroll över arbetsfunktionen.

KAPITEL 06

DET BREDARE LANDSKAPET

# 06 Resten av AI × krypto: verklig men liten efterfrågan, och värdet flyttar till aktierna

## Decentraliserade beräkningsmarknader

Decentraliserad beräkning har betalande kunder, men efterfrågan är liten, koncentrerad och består mest av vidareförsäljning av inferens på öppna modeller. Avgifterna som går att verifiera i kedjan är små: DefiLlama visar omkring 2,35 miljoner dollar i avgifter för Render det senaste året och omkring 680 000 dollar för Livepeer, och hela DePIN-sektorn tjänade bara omkring 72 miljoner dollar i intäkter i kedjan 2025. Egenrapporterade siffror är större men löst definierade: Aethir redovisar ”mer än 127,8 miljoner dollar” i intäkter för räkenskapsåret 2025 utan att säga om det är bruttofakturering eller protokollets andel. Tokenvärden och intäkter har glidit isär åt båda hållen: Renders marknadsvärde är omkring 360 gånger avgifterna i kedjan, medan Aethirs marknadsvärde på omkring 87,7 miljoner dollar ligger under dess egenrapporterade intäkter uppskattning.

Påståendet om att vara 70 % till 85 % billigare håller bara mot hyperskalarnas priser på begäran. Den 3 november var det viktade genomsnittspriset för H100 på Akash omkring **2,73 dollar per GPU-timme** (intervall 2,04–3,16 dollar), under omkring 6,88 dollar hos AWS och 11,06 dollar hos Google Cloud, men över omkring 1,73 dollar hos Vast.ai, 1,99 dollar i RunPods community-moln och omkring 2,44 dollar för en ettårig reservation (priser på andra plattformar avlästa den 12 september; Akash, som publicerade siffrorna, är en part i saken). Efterfrågan är mest standardinferens som distribueras via OpenRouter, omkring 1,7 miljarder token om dagen på AkashML och mer än 4 miljarder på io.net, i direkt konkurrens med varje neocloud. De största dollarbeloppen inom ”decentraliserad beräkning” centraliseras på nytt: Axe Compute, som håller Aethirs token som kassatillgång, tecknade ett femårskontrakt värt 1,5 miljarder dollar för att hyra ut sitt eget kluster med 9 200 B300, mot intäkter på 35 000 dollar under första kvartalet 2026.

## Decentraliserad träning och inferens

Träning över internet har visats med 8 till 40 miljarder parametrar. Pluralis-8B tränades i 40 dagar på 330 noder, mest konsument-GPU:er, och nådde 63 % av effektiviteten hos en centraliserad H100-referens, även om artikeln medger att körningen inte var decentraliserad och krävde Pluralis egna noder och en antagningsinstans. Nous Researchs Psyche genomförde förträningskörningen Consilience med 40 miljarder parametrar. I frontskala centraliserades arbetet igen: Prime Intellects INTELLECT-3, en mixture-of-experts-modell med 106 miljarder parametrar, tränades under två månader på 512 H200 sammankopplade med 400 Gbps InfiniBand. Begränsningarna är tydliga: noderna behöver minst 200 Mbit/s bandbredd och omkring 100 gångers kompression av kommunikationen, och förtroende och verifiering beror fortfarande på operatörens privilegierade roller.

Pengarna gick till aktiebolag som säljer centraliserade produkter. Prime Intellect hade mer än 100 miljoner dollar i intäkter i årstakt innan bolaget tog in 130 miljoner dollar till en värdering på 1 miljard dollar i juli; Venice, en privat inferensapp, redovisar mer än 70 miljoner dollar i intäkter i årstakt och vinst sedan första kvartalet 2026 (bolagets siffror), tog in 65 miljoner dollar till en värdering på 1 miljard dollar och kör på hyrda GPU:er. På tokensidan är Gensyns centrala träningsnätverk fortfarande ”ej fastställt” och dess token är värd omkring 17,2 miljoner dollar; token för Ritual, ett AI-nativt lager 1, föll 72 % från toppen efter huvudnätet (endast rubrikuppgift).

## Verifierbar AI

Produktionen valde betrodda exekveringsmiljöer (TEE). Hoppers konfidentiella beräkning lägger till under 7 % merkostnad för typiska LLM-frågor och nästan ingenting för stora modeller; Phalas merkostnad i produktion är omkring 0,5 % till 5 %, men i slutet av 2025 betalade 398 av dess 10 004 användare. En TEE:s förtroendeankare är chipleverantören: TEE.fail, offentliggjort i oktober 2025, använde en mellanlänk på DDR5-bussen som kostade under 1 000 dollar för att extrahera attesteringsnycklar från fullt uppdaterade maskiner med Intel TDX och AMD SEV-SNP, vilket i sin tur undergräver NVIDIAs konfidentiella beräkning på GPU, och båda leverantörerna säger att fysiska attacker ligger utanför deras hotmodell. Det är precis DePIN-situationen, där nodoperatörer har fysisk åtkomst. zkML kan nu bevisa en komplett LLM: Lagranges DeepProve bevisar GPT-2 med omkring 174 token i minuten och Gemma 3 med omkring 86, flera storleksordningar från produktionsinferens och utan redovisad betald användning.

## Agentbetalningar och data

x402 är det bästa varnande exemplet. TRM Labs räknar omkring 198,9 miljoner transaktioner och 52,7 miljoner dollar avvecklade av kända förmedlare sedan maj 2025, 99,6 % i USDC; efter att självbetalningar, satsflöden och säljare med färre än tio köpare rensats bort är andelen som faktiskt initierats av agenter bara 0,6 % till 7,5 %, **omkring 5 000–11 000 dollar i månaden**. Artemis bedömer att mer än 95 % av aktiviteten är protokollsignalering snarare än köp och försäljning. Samtidigt standardiseras rälsen av de etablerade aktörerna: x402 Foundation inrättades under Linux Foundation den 14 juli, med Visa, Mastercard, Stripe, Google, AWS, Cloudflare och Coinbase som kärnmedlemmar, och Tempo, som skrev Machine Payments Protocol tillsammans med Stripe, lanserade sitt huvudnät i mars till en värdering på 5 miljarder dollar. Plattformar för agenttokens drivs fortfarande av lanseringsplattformarnas logik: Virtuals redovisar själv 479 miljoner dollar i ”aGDP”, och ElizaOS grundare förklarade dess token död efter en förlikning i en grupptalan. Värdet fångas troligare av utgivare av stablecoins och etablerade betalningsaktörer; kryptots särskilda roll här kan vara identitet och ansvarsskyldighet för agenter.

Data är en av få delsektorer med verkliga köpare utanför kedjan. Grass redovisar 17 miljoner dollar i dataintäkter 2025 och ytterligare 17 miljoner dollar under första halvåret 2026, siktar på 75 miljoner dollar för året och säger att bolaget är lönsamt; GRASS är värd omkring 328 miljoner dollar, omkring tio gånger första halvårets intäkter i årstakt uppskattning. Vallgraven är insamlingsskalan i distribuerad bandbredd från hushåll, inte kryptot i sig. Data-DAO:er och datamarknadsplatser saknar synlig betald efterfrågan: Vana är värd omkring 21,5 miljoner dollar och Ocean omkring 27,6 miljoner dollar, och Ocean är i tvist med Fetch.ai. Bevis på personskap får verkliga integrationer i takt med att agenterna blir fler: World säger att det har verifierat nästan 18 miljoner människor, och Zoom, Docusign och Tinder stöder delar av World ID.

## Kapitalet har röstat

| Sektor | Verifierbar betald efterfrågan | Signaler om subvention och värdering | Bedömning |
|---|---|---|---|
| **Kedjor för nyttigt arbete** | Bittensor omkring 70 miljoner dollar om året i verifierbara externa intäkter; Pearl omkring 1 % av blocken med AI-arbete | Aleo 175–180 gånger; Bittensor 4–11 gånger; Pearl mest subvention | Mekanismerna har värde; nyttan kräver betalande kunder |
| **Decentraliserade beräkningsmarknader** | Avgifter i kedjan omkring 0,68–2,35 miljoner dollar om året (Livepeer, Render); egenrapporterat 4–128 miljoner dollar | Renders marknadsvärde omkring 360 gånger avgifterna | Verklig men liten, mest vidareförsäljning av standardinferens |
| **Decentraliserad träning och inferens** | Aktiebolag: Prime Intellect över 100 miljoner dollar om året, Venice över 70 miljoner dollar (centraliserad infrastruktur) | Gensyns huvudnät för träning ej i drift; Ritual-token −72 % från toppen | De tekniska framstegen är verkliga; vinsterna går till aktierna |

| Sektor | Verifierbar betald efterfrågan | Signaler om subvention och värdering | Bedömning |
|---|---|---|---|
| **Verifierbar AI** | TEE i produktion (398 betalande användare hos Phala); ingen redovisad betald zkML | — | TEE nyttiga; zkML tidigt |
| **Agentbetalningar** | Flöden initierade av agenter omkring 5 000–11 000 dollar i månaden | Omkring 200 miljoner transaktioner, mest signalering och spekulation | Rälsen först, efterfrågan senare |
| **Datanätverk** | Grass 17 miljoner dollar första halvåret 2026 (egenrapporterat) | GRASS omkring 10 gånger intäkterna i årstakt | En av få sektorer med verkliga köpare |

Lägg de sex sektorerna bredvid varandra så blir slutsatsen enhetlig: **subventioner är bränsle för en kallstart, inte en affärsmodell.** Den 13 november var CoinGeckos AI-kategori värd omkring 18,6 miljarder dollar totalt, men omkring 56 % av det är generella lager 1-kedjor som NEAR och ICP, plus projekt på konsensusnivå som TAO, stakad TAO (möjligen dubbelräknad med TAO) och Pearl; AI-tokens på protokoll- och applikationsnivå uppgår till omkring 8 miljarder dollar eller mindre uppskattning. Det är ungefär den sammanlagda värderingen av fyra aktiebolag: Tempo (5 miljarder dollar), Nous (enligt uppgift 1,5 miljarder dollar), Prime Intellect (1 miljard dollar) och Venice (1 miljard dollar). Riskkapitalet till krypto ligger fast på omkring 4,5–5 miljarder dollar per kvartal, med ökningen från AI-relaterad infrastruktur, vars andel av redovisade belopp steg från 8,1 % till 18,2 % under tredje kvartalet 2026; och 84,7 % av de 118 tokens som lanserades 2025 handlas under lanseringspriset. I en mening: **kryptot sköter distribution och avräkning, aktierna tar vinsterna**, åtminstone 2026.

KAPITEL 07

07

ASIC-EKONOMI

# ASIC-frågan: en formel, tre måttstockar

## En formel: varför effektivitet slutar löna sig efter k = 5

Låt R vara den dagliga intäkten per beräkningsenhet (nätverkets minerintäkter delat med nätverkets hashkraft), c den dagliga kostnaden för en beräkningsenhet på den marginella hårdvaran (på Pearl en GPU), och k hur många gånger billigare en ASIC är än den marginella hårdvaran ($c_{ASIC} = c/k$). Så länge GPU:er förblir de marginella minerna driver fritt inträde och utträde R mot c, så en ASIC tjänar $R - c/k$ per beräkningsenhet:

**Vinsttak för en ASIC ≈ R × (1 − 1/k)**

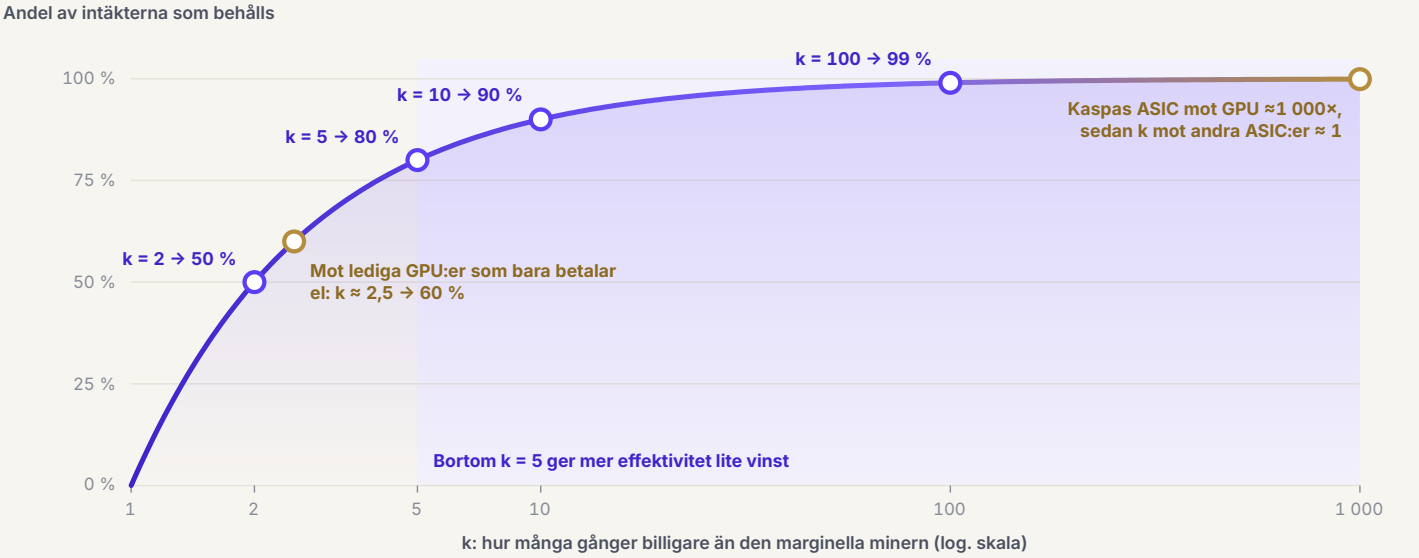
Över hela nätverket är en ASIC-flottas tak för årlig bruttovinst ungefär nätverkets miningintäkter × (1 − 1/k), och bara tills ASIC:er ersätter GPU:er som de marginella minerna.

Formeln har tre följder. För det första är (1 − 1/k) konkav: vid k = 2 behåller en ASIC 50 % av intäkterna, vid k = 5 80 %, vid k = 10 90 %, vid k = 100 99 %. **Bortom k = 5–10 ger mer effektivitet nästan ingen ytterligare vinst**; storleken på R och på potten avgör. För det andra nollställs referensen: när ASIC:er väl är de marginella minerna måste k mätas mot andra ASIC:er. Kaspas ASIC:er hade omkring 1 000 gångers effektivitetsfördel mot GPU:er, vilket bara köpte en kort bonanza; i dag tjänar den bästa KS7 0,86 dollar om dagen. För det tredje är referensen på en GPU-vänlig kedja som Pearl ofta inte inköpspriset för en GPU utan **elen för en ledig GPU**: ett eget RTX 5090 på 525 W vid 0,10 dollar per kWh kostar omkring 1,26 dollar om dagen i el och tjänar omkring 5–6 dollar uppskattning. Mot GPU:er som är sänkta kostnader har även en ASIC som använder en femtedel av energin ett ef-

fektivt k på bara omkring 2,5 om dess avskrivning är lika stor som dess el, och ett vinsttak på omkring 60 % av R (ett illustrativt antagande).

**FIG. 05** **Varför effektivitet slutar löna sig: en ASIC behåller högst 1 – 1/k av intäkterna**

Andel av intäkterna en ASIC kan behålla mot den marginella minern, efter kostnadsfördelen k




**Källa:** Författarens modell: när den marginella hårdvaran sätter priset går intäkten per beräkningsenhet mot dess kostnad; siffror för Kaspa och RTX 5090 från rapporten.

## Tre måttstockar: potten, återbetalningsskalan och reglerna

Den första måttstocken är **potten**. Räknar man om omkring 2 893 BTC i blockbelöningar under första veckan i november (omkring 222 miljoner dollar) till årstakt tjänar bitcoin-miners omkring 11,5 miljarder dollar om året uppskattning. Pearls miners tjänar omkring 0,46–0,49 miljarder dollar (till nuvarande priser); Aleos belöningar till bevisare, i årstakt utifrån tredje kvartalet 2025, omkring 31 miljoner dollar; och Kaspas miners, vid en blockbelöning på 1,945 KAS och tio block per sekund, omkring 18 miljoner dollar uppskattning. **Ett helt års miningintäkter i Kaspa är mindre än den nedre gränsen för en enda chipdesignbudget på 5/4 nm (omkring 30 miljoner dollar).**

Den andra måttstocken är **den hashkraft som krävs för att betala tillbaka NRE**:

**h* = NRE ÷ [R × (1 – 1/k) × T × p]**

där T är det användbara fönstret i dagar och p sannolikheten att reglerna förblir oförändrade under det fönstret. För Pearl, ta R = 14 dollar per PH/s och dag, k = 5, T = 365 dagar och en NRE på 20 miljoner dollar (ett N7/N6-chip med upprepad matris; i nedre delen av intervallen 5–30 miljoner dollar för mellannoder från Silicon Analysts och 15–100 miljoner dollar för beräknings-ASIC:er på 7/6 nm från AnySilicon, eftersom ett miningchip använder lite IP och en upprepad struktur; mitt antagande). Med p = 1 krävs omkring **4,9 EH/s** som går ett helt år, omkring 5 % av nätverket; med p = 0,5 omkring **9,8 EH/s**, omkring 11 %; och om R faller tillbaka till sommarens nivå på omkring 10 dollar krävs även med p = 1 omkring 6,9 EH/s uppskattning. Skalan är inte utom räckhåll. Problemet är p och T.

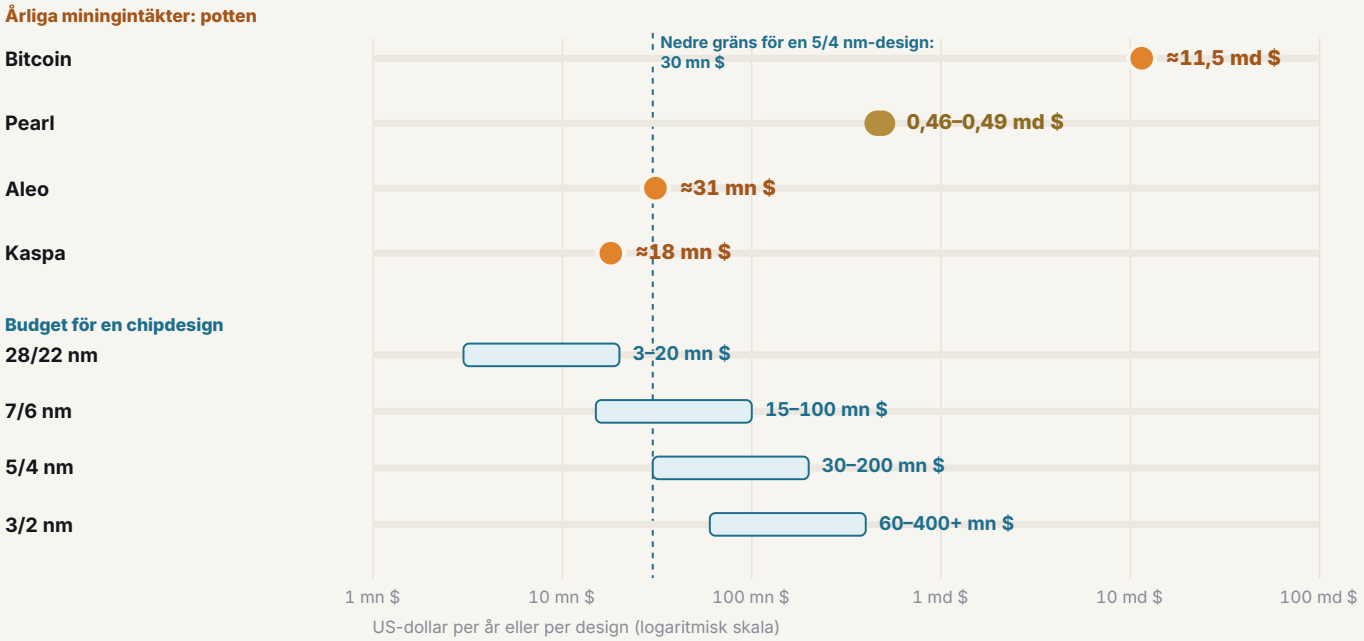
Den tredje måttstocken är **regelrisken**, och den är ofta avgörande. Moneros anti-ASIC-fork 2018 strandade Bitmains X3 och kostade kortvarigt nätverket nästan halva hashkraften. Sias fork samma år gjorde Bitmains och Innosilicons miners obrukbara, medan andrahandspriset för Obelisk-minern, som var knuten till utvecklarna, steg från 2 499 dollar till mer än 5 000 dollar. Lite mer än en månad efter Ethereums Merge hade färre än 20 % av de tidigare Ethereum-minerna hittat nytt arbete. **Den som styr arbetsfunktionen styr hårdvaruinvesterarnas öde**,

och Sia visar att ”ASIC-resistens” kan bli utvecklarnas egen vallgrav. Hårdvara med en andra marknad överlever: efter the Merge flyttade GPU:erna till AI och molnet, medan ASIC:er för en enda algoritm trängdes in i Ethereum Classic. Fönstret krymper också: Kaspas första ASIC tog omkring 16 månader, Aleos omkring fem, och den som först når effektivitetsgränsen är ofta en utmanare snarare än en etablerad aktör.

| Nätverk | Årliga miningintäkter (omkring) | Marginell hårdvara eller referenshårdvara | Regelstabilitet | k tillgängligt för en nykomling | Bedömning |
|---|---|---|---|---|---|
| **Bitcoin** | 11,5 miljarder dollar | Bitmain S23 XP Hyd (8,9 J/TH, 25,5 dollar per TH) | Mycket hög | ≤1 | En stor pott, men k tillhör de etablerade |
| **Pearl** | 0,46–0,49 miljarder dollar (till nuvarande priser) | Egna eller lediga GPU:er (marginalkostnad ≈ el) och datacenter-GPU:er | Mycket låg: fyra forkar på två månader; FP8-vitlista väntar | Kanske 3–10 gånger i effektivitet analys ; lägre mot GPU:er som är sänkta kostnader | Potten är stor nog; regelrisken lägger in sitt veto |
| **Aleo** | Omkring 31 miljoner dollar (kv3 2025, årstakt) | IceRiver AE2 och andra dedikerade maskiner | Medel: insatströskel enligt ARC-46 | ≈1 | En liten pott, och tröskeln stiger |
| **Kaspa** | Omkring 18 miljoner dollar | KS7 (77 J/TH) | Hög | ≈1 | Årsintäkter under nedre gränsen för en designbudget på 5/4 nm |

**FIG. 06** **De flesta potter är mindre än en enda chipbudget**

Årliga miningintäkter i fyra nätverk mot designkostnaden för ett chip per processnod, logaritmisk skala




**Källa:** Hashrate Index (bitcoin, november 2026), prlstats och Kryptex (Pearl), Blockworks (Aleo, kv3 2025), minerstat (Kaspa), AnySilicon (designbudgetar, 2026); omräknat till årstakt av författaren.

Kvadranten ”stor pott, stabila regler och k > 1” är tom i dag: bitcoin har potten och de stabila reglerna, men dess k tillhör de etablerade; Pearls pott är stor nog men dess regler är minst stabila; Kaspa och Aleo har tillräckligt sta-

bila regler och för små potter.

KAPITEL 08

08

CHIP FÖR CHIP

# Fem chiptyper, fem svar

## PoUW i Pearls stil: inget dedikerat INT7-chip; satsa på en miner som kan köra inferens

Argumenten för en ASIC är verkliga. Bara omkring 1 % av blocken bär AI-arbete i dag, så vitbokens ekonomiska argument, att en dedikerad ASIC förlorar förmågan att utföra nyttigt arbete och därför tjänar mindre, är som svagast just nu. De dyraste delarna av en GPU, HBM och CoWoS-kapsling, utgör omkring 63 % av tillverkningskostnaden för en H100, medan logikdien bara står för omkring 9 %. Ett chip som inte gör något annat än 7-bitars heltals-MAC, håller sina operander i SRAM på chipet och går på låg spänning skulle rimligen kunna nå 3 till 10 gånger bättre energi per operation än en GPU på samma nod och en större fördel i kostnad per operation [analys; inga mätningar än]. Argumenten emot väger tyngre. Ledtiderna hos foundries i de främsta noderna kan överstiga 50 veckor, och ett nytt team behöver omkring 18 till 24 månader från start till volym uppskattning; när ett chip väl kommer kan Pearl mycket väl ha gått över till FP8 och sedan FP4. Och eftersom PIP-3 definierar den godkända aritmetiken genom en vitlista över enheter skulle ett INT7-chip helt enkelt sluta fungera när den aktiveras.

Min slutsats är därför: **gör ingen tape-out av ett dedikerat INT7-chip för Pearl nu.** Det finns två bättre satsningar. Den ena är en inferensaccelerator som också kan utvinna: byggd kring generell INT8/FP8-GEMM och operander i SRAM, så att värdet hamnar på en verklig inferensmarknad om tillgången till mining stängs. Den andra är att delta i protokolldebatten och driva på för en leverantörsneutral kanonisk aritmetik som en ASIC kan implementera exakt, tillsammans med ett åtagande om algoritmstabilitet, vilket är förutsättningen för att kvadranten ”stor pott, stabila regler, k > 1” ska kunna finnas. Fram till dess kan en arkitektur valideras på FPGA:er eller på multiprojektwafers (MPW) i mogna noder, så att NRE-risken hålls i storleksordningen en miljon dollar. Kapitel 10 omsätter detta i en budget i steg.

## Miners i bitcoins stil: de etablerade äger k, och el är den knappa tillgången

Bitcoins regelrisk är lägst, men konkurrensen trycker en nykomlings k under 1. Bitmains S23 XP Hyd når 8,9 J/TH, säljs för 25,5 dollar per TH och levereras från november; effektiviteten i toppmodellerna har gått från 17,5 J/TH för S21 2023 till 8,9 J/TH, omkring 49 % på tre år. En nykomling som levererar i volym 2027–2028 skulle behöva omkring 8 J/TH eller bättre för att stå sig uppskattning. Utmanarna lämnar: Canaan redovisade intäkter på 31,9 miljoner dollar för andra kvartalet 2026 med en bruttoförlust på 29,3 miljoner dollar och prognostiserade bara 11–15 miljoner dollar för tredje kvartalet; Bitdeers försäljning av egna maskiner till externa köpare under tredje kvartalet föll till 0,3 miljoner dollar, med en bruttomarginal för koncernen på 3,4 %; Auradine bytte namn till Velaura AI och gick över till AI-chip; och Core Scientific avbeställde en order på omkring 15 EH/s 3 nm-maskiner från Blocks enhet Proto och bokförde en förlust på 41,9 miljoner dollar.

Efterfrågan krymper också. Den 9 november var hashpriset omkring 31,89 dollar per PH/s och dag, terminskurvan till april 2027 låg i genomsnitt på omkring 32,7 dollar, och avgifterna var bara 0,72 % av blockbelöningarna. Vid 0,045 dollar per kWh betalar sig S23 XP Hyd på omkring 3,1 år uppskattning, nära nästa halvering som väntas omkring 2028. Viktigare är elens alternativkostnad: leasing för AI och högprestandaberäkning ger nu omkring 1,67–1,9 miljoner dollar per kontrakterad megawatt och år, vilket vid 80 % till 85 % i rörelseöverskott är omkring 1,5 till 2,9 gånger miningens bruttovinst per megawatt (omkring 0,56–0,9 miljoner dollar) [uppskattning; de två

megawattbaserna skiljer sig åt]. Core Scientific slutar utvinna under fjärde kvartalet; Bitfarms, numera Keel, har stängt alla sina miningsajter i USA; IREN tjänade 200 miljoner dollar på AI-moln och bara 28 miljoner dollar på mining under ett kvartal. **El och datacenterkapacitet är tillgångarna; miningmaskiner är förbrukningsvaror.**

NRE-kalkylen är lika obarmhärtig. Toppeffektivitet kräver nu processer i 3 nm-klassen (enligt sekundära källor använder MicroBT Samsung SF3, och Auradines och Protos chip är 3 nm). Masker kostar omkring 10–20 miljoner dollar och full NRE 30–100 miljoner dollar eller mer; vid 20–25 dollar per TH och 30 % bruttomarginal ger varje såld 1 EH/s omkring 6–7,5 miljoner dollar i bruttovinst, så att täcka en årlig FoU i Bitdeers skala, omkring 110–160 miljoner dollar, kräver försäljning av omkring 20–25 EH/s om året till externa köpare, omkring 2 % av nätverket uppskattning. Slutsatsen: **ett nytt team bör inte gå in på marknaden för bitcoin-miners till försäljning.** Tre fall är fortfarande rimliga: vertikal integration med egen billig el; nischformer som miners för hem och uppvärmning (Canaan lanserar också produkter för hemuppvärmning); och att delta i de etablerade aktörernas färdplaner genom IP och designtjänster.

## AI-träningschip: inte som litet team

För träningschip är k svårast att förtjäna. Vid samma generation och precision leder Googles Ironwood över B300 i FP8-effektivitet på chipnivå med bara omkring 1,1 till 1,8 gånger (Ironwoods effekt är en journalists uppskattning) uppskattning; de 30 till 80 gånger som ofta citeras från artikeln om TPU v1 mättes mot en K80, som saknade tensorkärnor. Den verkliga hävstången är talformatet: att gå från FP8 till FP4 ger vanligen 2 till 3 gånger genomströmningen på samma kisel, och NVIDIA rör sig snabbast där: bolaget uppger 50 PF NVFP4-inferens för Rubin, i volym under andra halvåret 2026. Trösklarna för kostnad och tillgång är ännu högre: designbudgetar för 3/2 nm ligger på 60–400 miljoner dollar eller mer; CoWoS-kapaciteten 2026 är omkring 106 000–130 000 wafers i månaden, och TSMC uppges överväga prishöjningar på upp till 20 %; vid ett kontraktspris på 15,6 dollar per GB för HBM3E kostar enbart HBM-minnet i en accelerator med 288 GB omkring 4 500 dollar uppskattning. Den här marknaden tillhör samutveckling med hyperskalare i Broadcoms stil (kapitel 11). Slutsatsen: **små team bör inte bygga fristående träningschip**; bygg i stället IP, chiplets eller designtjänster.

## AI-inferenschip: byt generalitet mot k, och det lönar sig

Inferens är där ASIC:er har störst utrymme, eftersom k kan köpas genom att ge upp generalitet. Etched uppger att dess server med åtta chip kör Llama-70B med mer än 500 000 token per sekund, omkring 20 gånger åtta H100; Taalas hårdkodar modellvikter i mask-ROM och uppger 10 gånger lägre effekt och 20 gånger lägre tillverkningskostnad (båda leverantörsuppgift). SRAM-centrerade arkitekturer undviker HBM och CoWoS, och deras noder är mer hanterbara: Taalas använder TSMC 6 nm, och Groq gick från 14 nm till Samsung 4 nm. Kryptodistribution hjälper inte inferenschip, eftersom decentraliserade beräkningsnätverk nästan helt körs på NVIDIA; vägen in är neoclouderna inköpsekonomi och den stora inhemska inferensmarknaden i Kina, där statligt finansierade datacenter måste använda inhemska AI-chip, importen av H200 stryps och Peking vill att stora internetbolag kör inferens på inhemska chip. Slutsatsen: **värt att bygga**, så länge det finns en tydlig kund och graden av inlåsning till modeller och arkitektur är ett medvetet val. Det är också den bästa andra marknaden för ett PoUW-miningchip.

## ZK- och FHE-acceleratorer: gardera med ZK, positionera dig tidigt inom FHE

GPU:er har redan kommit ikapp med ZK-bevisning i realtid: Succincts SP1 Hypercube bevisade 99,7 % av Ethereums block på under 12 sekunder med 16 RTX 5090. Cysic uppger att dess chip ZK C1 är 10 till 100 gånger så effektivt som GPU-bevisare och levereras 2026, men efterfrågan kommer till stor del från dess egna tokenbelöningar leverantörsuppgift; jag hittade inga offentliga belägg för att ASIC:er för Aleos pussel används för verklig ZK-bevisning. ZK-bevisning är heltalsbaserad och deterministisk, till formen nära det miningchip gör, men bevissystemen ändras snabbt (SP1, RISC Zero, Halo2, Plonky2 med flera). Slutsatsen: **vänta och gardera**, med programmerbara MSM-, NTT- och hashmotorer, först på FPGA:er, sedan som ASIC:er.

FHE är den mest hårdvarubundna primitiven med störst uppsida. Intels Heracles, på Intel 3 med 197 mm², 176 W och 48 GB HBM3, kör FHE-operationer 1 074 till 5 547 gånger snabbare än en Xeon med 24 kärnor, men Intel har inga kommersiella planer och projektets ledare slutade i december 2025. Niobium bygger på Samsung 8LPU med SemiFive som designpartner (ett utvecklingskontrakt på omkring 6,86 miljoner dollar) och har tagit in mer än 23 miljoner dollar; Optalysys, med en fotonisk metod, tog in 31 miljoner dollar och säger att ett fullständigt kommersiellt chip fortfarande ligger två till tre år bort; och Zamas FPGA-design med öppen källkod når omkring 13 000 programmerbara bootstraps per sekund vid omkring 200 W, den offentliga referens som varje ASIC tydligt måste slå. Slutsatsen: **positionera dig tidigt och validera med lätta tillgångar**, för ett fönster omkring 2027–2029. Efterfrågan förankras troligare i krypterad AI för företag och myndigheter än i kryptovalutor i sig, och FHE behöver HBM, så kostnadsstrukturen liknar ett träningschips.

| Chiptyp | Var k kommer ifrån | Efterfrågan och pott | NRE och trösklar | Regel- eller inlåsningsrisk | Andra marknad | Bedömning |
|---|---|---|---|---|---|---|
| **PoUW-mining-chip i Pearls stil** | Att slopa HBM, CoWoS och generella enheter: kanske 3–10 gånger i J/op analys | Minerintäkter omkring 0,46–0,49 miljarder dollar om året, volatila | Upprepad matris på N7/N6 omkring 10–30 miljoner dollar analys | Mycket hög: protokollet bestämmer talformatet och vitlistan över enheter | AI-inferens, om det byggs som generell INT8/FP8-GEMM | Inget dedikerat INT7-chip; bygg inferensinriktat och utvinningsbart, eller vänta på kanonisk aritmetik |
| **SHA-256-miningchip för bitcoin** | k ≤ 1 mot Bitmain | Omkring 11,5 miljarder dollar om året, försäljningsmarknaden hålls av två företag | Masker i 3 nm-klassen 10–20 miljoner dollar; full NRE 30–100 miljoner dollar eller mer | Låg (stabil algoritm); hög risk för konkurrens och halvering | Nästan ingen (en enda algoritm) | Nykomlingar avstår från försäljningsmarknaden; bara egen el, nischformer eller IP |
| **AI-träningschip** | Systoliska matriser av samma generation bara 1,1–1,8 gånger; formatet är hävstången | Störst (Broadcom omkring 58 miljarder dollar i intäkter från AI-halvledare räkenskapsåret 2026) | Design på 3/2 nm 60–400 miljoner dollar eller mer; CoWoS och HBM | NVIDIAs årliga takt | — | Små team bygger IP, chiplets och designtjänster, inte fristående chip |
| **AI-inferenschip** | Att ge upp generalitet: 10–20 gånger enligt uppgift leverantörsuppgift | Stor och splittrad; Fastlandskinas inhemska marknad är stor i sig | SRAM-arkitekturer på N6/N7–N4 är hanterbara | Inlåsning till modeller och arkitektur | Det är huvudmarknaden | Värt att bygga: teckna kunden, välj sedan inlåsningen |
| **ZK-accelerator** | 10–100 gånger GPU:er enligt uppgift leverantörsuppgift | Liten, driven mest av tokenbelöningar | N7/N6, sedan N5; FPGA först | Bevissystemen ändras snabbt | Begränsad utanför ZK | Vänta och gardera: programmerbara MSM- och NTT-motorer |
| **FHE-accelerator** | 1 074–5 547 gånger en Xeon (Heracles) | Ännu inte formad; förankrad i företag och myndigheter | Kräver HBM; kostnadsstruktur nära träningschip | Standarder och ekosystem oklara | Krypterad beräkning, privat AI | Positionera tidigt, validera med lätta tillgångar; fönster 2027–2029 |

**FIG. 07** **Fem chiptyper, fem svar**

Var kostnadsfördelen k kommer ifrån, vad kislet kan säljas till om reglerna ändras, och slutsatsen



### PoUW-miningchip i Pearls stil

INTE NU

**k:** utan HBM och generella enheter: 3–10× per operation (härlett)

**Andra marknad:** AI-inferens, om generell INT8/FP8-GEMM

**Satsa på ett inferenschip som kan utvinna, inte på en dedikerad INT7-krets**



### SHA-256-miningchip för bitcoin

AVSTÅ

**k:** ≤1 mot de etablerade

**Andra marknad:** nästan ingen

**Bara med egen el, en nischform eller som IP**



### AI-träningschip

IP, INTE CHIP

**k:** 1,1–1,8× vid samma precision

**Andra marknad:** —

**Kapsling och HBM avgör; bygg IP, chiplets, tjänster**



### AI-inferenschip

BYGG

**k:** avstå generalitet: 10–20× enligt uppgift

**Andra marknad:** det är huvudmarknaden

**Teckna kunden, välj sedan inlåsningen**



### Accelerator för ZK-bevis

VÄNTA OCH GARDERA

**k:** 10–100× GPU:er enligt uppgift

**Andra marknad:** begränsad utanför ZK

**Programmerbara MSM- och NTT-motorer; FPGA först**



### FHE-accelerator

POSITIONERA TIDIGT

**k:** 1 074–5 547× en Xeon (Heracles)

**Andra marknad:** krypterad beräkning, privat AI

**Ett fönster 2027–2029; validera med lätta tillgångar**

**Källa:** Författarens sammanställning av kapitel 7 och 8; leverantörssiffror märkta som i rapporten.

KAPITEL 09

09

PROCESSNODER

# Processnoder: låt den största livstidskostnaden välja

## Kostnadsstegen: wafers i spetsnoder kostar tio gånger mogna

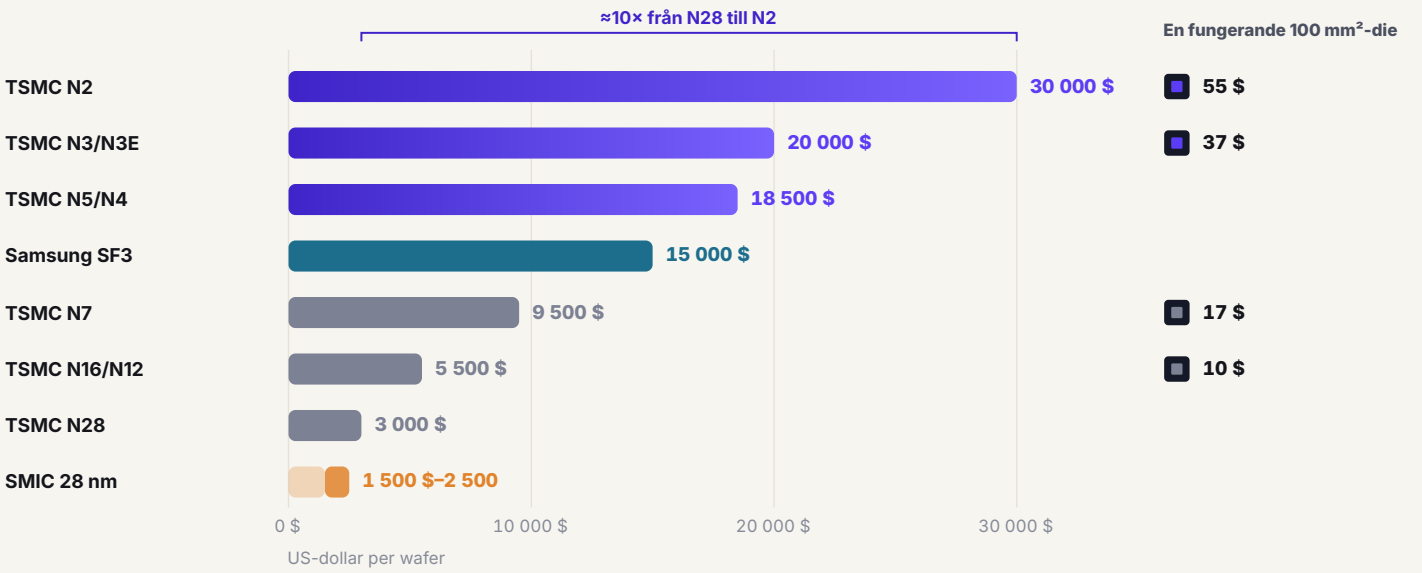
Börja med kalkylen. En wafer i spetsnod kostar nu omkring tio gånger en mogen: TSMC N2 omkring 30 000 dollar, N3 omkring 20 000 dollar, N7 omkring 9 500 dollar och N28 omkring 3 000 dollar; Samsung SF3 omkring 15 000 dollar, omkring 25 % under N3; och SMIC:s tillverkningskostnad i 7 nm-klassen uppges ligga 40 % till 50 % över TSMC:s. TSMC uppges höja priserna med ytterligare 3 % till 6 % från januari 2027, mest för 2 nm och 3 nm. Masker och designkostnader skiljer sig ännu mer, och källorna är oense, så tabellen visar båda.

| Nod | Waferpris (uppskattning) | Maskset (modellvärde / vägledande intervall) | Designbudget (AnySilicon) | MPW (per projekt) |
|---|---|---|---|---|
| **TSMC N2** | Omkring 30 000 dollar | — | 3/2 nm: 60–400 miljoner dollar eller mer | — |
| **TSMC N3/N3E** | Omkring 20 000 dollar (18 000–27 000 dollar) | 15 miljoner dollar / 10–20 miljoner dollar | Som ovan | — |

| Nod | Waferpris (uppskattning) | Maskset (modellvärde / vägledande intervall) | Designbudget (AnySilicon) | MPW (per projekt) |
|---|---|---|---|---|
| **Samsung SF3** | Omkring 15 000 dollar (13 000–17 000 dollar) | 12 miljoner dollar | — | — |
| **TSMC N5/N4** | Omkring 18 500 dollar (16 000–21 300 dollar, låg säkerhet) | 6,5 miljoner dollar / 10–20 miljoner dollar | 5/4 nm: 30–200 miljoner dollar | — |
| **TSMC N7** | Omkring 9 500 dollar (8 100–10 900 dollar) | 3,5 miljoner dollar / 5–10 miljoner dollar | 7/6 nm: 15–100 miljoner dollar | Kommersiellt 100 000–500 000 dollar |
| **TSMC N16/N12** | Omkring 5 500 dollar | 1,8 miljoner dollar | 16/12 nm: 6–40 miljoner dollar | Akademiskt omkring 35 000 dollar per mm² |
| **TSMC N28** | Omkring 3 000 dollar | 1 miljon dollar / 1–3 miljoner dollar | 28/22 nm: 3–20 miljoner dollar | Kommersiellt 50 000–150 000 dollar; akademiskt omkring 12 000 dollar per mm² |
| **SMIC 28 nm** | 1 500–2 500 dollar | — | — | — |
| **SMIC 7 nm-klass (N+2/N+3)** | 40–50 % över TSMC | — | — | — |

**FIG. 08** **En wafer i spetsnod kostar tio gånger en mogen**

Uppskattat pris per 300 mm-wafer per nod, och kostnaden för en fungerande 100 mm²-die vid 85 % utbyte




**Källa:** Silicon Analysts waferpriser (uppskattningar, N5 låg säkerhet), Wccftech (SMIC); die-kostnaderna är författarens beräkning med antaget utbyte.

Sedan fysiken. Horowitz klassiska mätningar visar att en programmerbar instruktion kostar omkring 70 pJ medan själva aritmetiken kostar några pikojoule: **aritmetik är billig och overhead är dyr**, vilket är skälet till att systoliska matriser och dedikerade chip finns. Att halvera strukturstorleken halverar ungefär energin per operation, men matningsspänningen har stannat kring 0,9–1,1 V. En N2-wafer kostar 50 % till 67 % mer än N3, så om inte densiteten ökar mer än omkring 1,5 gånger går kostnaden per transistor upp i stället för ned uppskattning: de främsta noderna måste betala sig genom effektivitet och prestanda. För en die på 100 mm² vid 85 % utbyte är kostnaden för

varje fungerande die omkring 55 dollar på N2, 37 dollar på N3, 29–34 dollar på N5, 17 dollar på N7 och 10 dollar på N16 [uppskattning; utbytet är ett antagande].

## Att välja nod efter chiptyp och team

Min regel är: **låt den största posten i chipets livstidskostnad välja nod.** För beräkningschip som går fullt belastade dygnet runt (mining, ZK) dominerar elen livstidskostnaden, så när algoritmen är stabil bör de flytta till N5/N4 och N3/SF3; så länge algoritmen är oklar eller pengarna knappa bör man validera på MPW i mogen nod först. För träningschip är flaskhalsen kapsling och minne (HBM och CoWoS utgör omkring 63 % av materialkostnaden för en H100), så noden är sekundär. För inferenschip är det första beslutet minnesarkitekturen: SRAM-centrerade konstruktioner kan stanna på N6/N7 till N4, medan HBM-konstruktioner ärver träningschipens kostnadsstruktur. FHE är bundet av minnesbandbredden och kan inte undvika HBM.

| Chiptyp | Startup eller litet team | Skalat företag | Väg i Fastlandskina | Nyckelpunkter |
|---|---|---|---|---|
| **PoUW-mining-chip för matris-multiplikation (SRAM-baserat, utan HBM)** | Prototyp på 28 nm eller 16/12 nm MPW; första volym på N7/N6; N5/N4 när algoritm och intäkter är stabila; dies på 50–150 mm² i standardkapslar med flip-chip | N3/N3E eller Samsung SF3 | SMIC N+1/N+2 är genomförbart (varken HBM eller CoWoS behövs), till omkring 40–50 % mer per fungerande die | Energin dominerar; protokollreglerna är den största risken |
| **SHA-256-miningchip** | Jaga inte spetsen; nischformer kan använda mogna noder | 3 nm-klass; de etablerade går från TSMC 5 nm till 2 nm och från Samsung SF3 till 2 nm (sekundära källor) | Svårt att matcha spetsnodernas effektivitet | Elpriset och halveringscykeln avgör marginalerna |
| **AI-träningschip** | Inget fristående chip: IP, chiplets, designtjänster | N3P, sedan N2/A16; 2–8 chiplets; säkra CoWoS och HBM tidigt | SMIC N+2/N+3 med inhemsk HBM (Huaweis modell); FP8 per chip omkring 0,2–0,5 gånger NVIDIAs, kompenserat av systemskala | Kapsling och HBM är flaskhalsen |
| **AI-inferenschip** | SRAM-centrerat eller hårdkodat, N6/N7 eller N4/N5/SF4; bara INT8, FP8 och FP4 | N3/N2 med HBM (lång kontext, mixture of experts), eller SRAM i flera chiplets | SMIC N+2 med LPDDR eller GDDR; minne i HBM-klass är strypningspunkten | Graden av arkitekturinlåsning är det centrala beslutet |
| **ZK-accelerator** | Börja på FPGA eller 16/12 nm MPW; sedan N7/N6 och N5 | N5/N3 | SMIC N+1/N+2 är genomförbart | Heltalsbaserat och deterministiskt; lätt att verifiera |
| **FHE-accelerator** | FPGA (mot referensen med öppen källkod) eller 16/12 nm MPW; volym kräver HBM | N5/N3 med HBM | Begränsat av tillgången till HBM | Minnesbandbredden, inte aritmetiken, sätter hastigheten |

## Avancerat eller moget, fyra slags foundries, och chiplets

| Alternativ | Främsta styrkor | Främsta svagheter | Bäst för |
|---|---|---|---|
| **Spetsnoder (N3/N2, SF3/SF2, Intel 18A)** | Varje hel nod halverar ungefär energin per operation; tät SRAM | Wafers 3–10 gånger, masker 5–15 gånger och designbudgetar 10–20 gånger de mogna nodernas; ledtider över 50 veckor; kostnaden per transistor kan stiga på N2 | Beräkningschip med stabila algoritmer som går dygnet runt; träningschip |

| Alternativ | Främsta styrkor | Främsta svagheter | Bäst för |
|---|---|---|---|
| **Mogna noder och mellannoder (N7/N6, N16/12, N28)** | Masker i miljondollarklassen, billig MPW, snabb iteration, många foundries (TSMC, UMC, GlobalFoundries, SMIC) | Effektivitet som inte konkurrerar vid full belastning dygnet runt | Prototyper, nischmarknader, inferens i kanten |
| **TSMC** | Bäst utbyte och ekosystem | Högst priser; knapp kapacitet i N2/N3; långa köer för CoWoS | Team som jagar effektivitetsgränsen med volymåtaganden |
| **Samsung Foundry** | Wafers omkring 20–25 % billigare (SF3 omkring 15 000 dollar mot omkring 20 000 dollar för N3); används av bland andra Groqs andra generation | Foundryverksamheten går med förlust, med sikte på nollresultat i slutet av 2027 | Kostnadskänsliga beräkningschip |
| **Intel Foundry** | Komplexa chip som Heracles bevisade på Intel 3 | Inga offentliga externa priser eller kunddata för 18A; foundryverksamheten går med förlust | Utvärdera ekosystem och stöd från fall till fall |
| **Foundries i Fastlandskina (SMIC med flera)** | Nära en stor inhemsk marknad; 28 nm-wafers för 1 500–2 500 dollar | Kapacitet i 7 nm-klassen omkring 45 000–60 000 wafers i månaden, till stor del upptagen av Huawei; utbyte omkring en tredjedel av TSMC:s; 40–50 % högre kostnader; bara DUV | Inferens-, mining- och ZK-chip för den inhemska marknaden |
| **Chiplets** | Överskrider gränserna för reticle och utbyte; standard för HBM-system (MI355X, Ironwood, Trainium3) | Högre kapslingskostnad och beroende av CoWoS; testning av kända fungerande dies (KGD) ökar kostnaden | Tränings- och stora inferenschip; behövs inte för små mining- eller ZK-chip |

KAPITEL 10

10

EN BUDGET I STEG

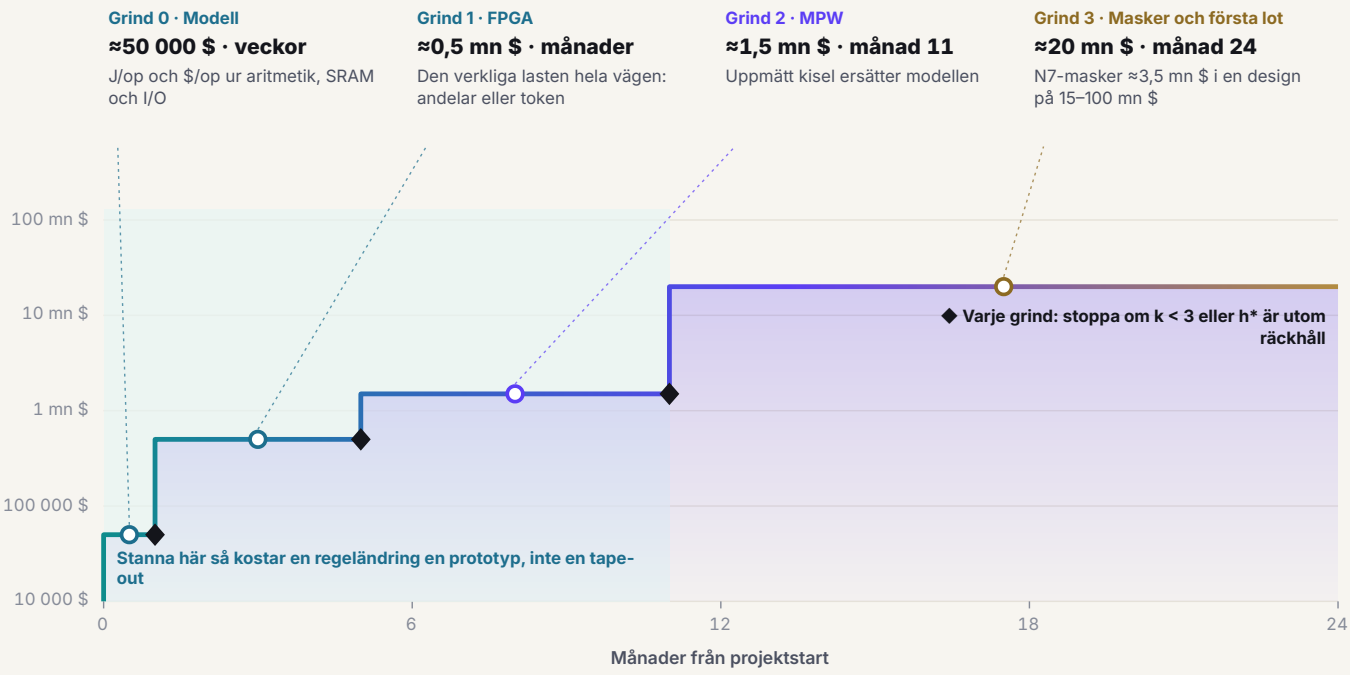
# Från modell till kisel: en budget i steg för ett första chip

## Köp belägg före masker

Ett helt kundanpassat chip är en serie satsningar, var och en större än den förra, och disciplinen består i att köpa belägg innan man köper masker. Jag använder fyra grindar analys. Grind 0 är en modell: en arkitekturstudie som uppskattar J/op och $/op utifrån aritmetiken, SRAM-minnet och I/O, mot referenshårdvaran i kapitel 7; den kostar veckor och tiotusentals dollar. Grind 1 är en FPGA-prototyp som kör den verkliga lasten hela vägen, miningandelar eller inferenstoken, och bevisar datavägen och protokollintegrationen; den kostar några månader och några hundra tusen dollar. Grind 2 är en multiprojektwafer: en kommersiell plats kostar omkring 50 000–150 000 dollar på N28 och 100 000–500 000 dollar på N7, och uppmätt kisel ersätter modellens energiuppskattningar. Grind 3 är det fullständiga masksetet och den första produktionsloten: masker för omkring 1 miljon dollar på N28, 3,5 miljoner dollar på N7 och 15 miljoner dollar på N3, inom en designbudget på 15–100 miljoner dollar för 7/6 nm. Varje grind behöver en siffra som stoppar projektet, till exempel k under 3 mot den marginella hårdvaran, eller en hashkraft för återbetalning h* som fönstret inte kan nå.

**FIG. 09** **Köp belägg före masker: en budget i steg för ett första chip**

Ackumulerade utgifter i US-dollar för ett N7/N6-projekt genom fyra grindar, logaritmisk skala; varje grind har en siffra som stoppar projektet




**Källa:** Priser för wafers, MPW och masker som i kapitel 9 (Silicon Analysts, AnySilicon, modellvärden); grindarnas tidsplan och utgifter är författarens illustrativa antaganden.

Trappan är brant. I ett N7/N6-projekt kostar de tre första grindarna tillsammans omkring 1,5 miljoner dollar och omkring elva månader; den fjärde kostar mer än tio gånger så mycket uppskattning. Det är därför ordningen betyder mer än noden: om protokollet ändrar sin aritmetik, eller köparen går ifrån, förlorar ett team som stannade vid grind 2 priset för en prototyp, inte för en tape-out.

## Vad varje chip kostar, och vad som måste stämma

| Projekt | Grind 2 (prototyp) | Grind 3 (nod och designbudget) | En fungerande die på 100 mm² | Vad som måste stämma för återbetalning |
|---|---|---|---|---|
| **Inferensaccelerator som kan utvinna** | MPW på 16/12 nm eller N7, 0,1–0,5 miljoner dollar | N6/N7, omkring 15–30 miljoner dollar uppskattning | Omkring 17 dollar på N7, före kapsling och minne | En tecknad inferenskund bär större delen av NRE; miningintäkter är en uppsida, inte grundscenariot |
| **Dedikerad INT7-miner för Pearl (rekommenderas inte nu)** | MPW på 28 nm, 0,05–0,15 miljoner dollar | N7/N6, omkring 20 miljoner dollar | Omkring 17 dollar på N7 | Omkring 4,9 EH/s under ett helt år med p = 1, eller 9,8 EH/s med p = 0,5 (kapitel 7) |
| **ZK-accelerator** | FPGA, sedan MPW på 16/12 nm | N7/N6, sedan N5 | Omkring 10 dollar på N16, 17 dollar på N7 | Bevissystemen stannar vid MSM- och NTT-kärnor som en fast konstruktion kan betjäna |
| **FHE-accelerator** | FPGA mot referensen med öppen källkod | N5/N3 med HBM, 30 miljoner dollar eller mer | Domineras av HBM och kapsling | En köpare av krypterad AI i volym, som ännu inte finns |

Två kommersiella slutsatser följer. För det första är den billigaste effektiviteten arkitektonisk: att hålla operanderna i SRAM och utelämna HBM och CoWoS tar bort större delen av en GPU:s kostnad (omkring 63 % av materialkostnaden för en H100) innan någon nod har valts. För det andra är tid en kostnad: ett nytt team behöver omkring 18 till 24 månader från start till volym och ledtiderna i spetsnoderna överstiger 50 veckor, så ett protokoll som ändrar sin aritmetik med några månaders mellanrum kan göra ett chip föråldrat innan grind 3 har betalat sig.

### Tre regler för ett litet team

Sammantaget ger siffrorna tre regler analys. För det första: **stanna vid grind 1 eller 2 tills det finns en namngiven köpare eller en fryst protokollversion**; varken tokenpriset eller hashkraften ersätter någon av dem. För det andra: **köp k med arkitekturen först och med noden i andra hand**: en SRAM-baserad matris på N7 kan slå en generell krets på N3 för en enskild last, till en bråkdel av maskkostnaden. För det tredje: **skriv ned hashkraften eller volymen för återbetalning, med p, före tape-out**, och prissätt betalningar per milstolpe mot en fast protokollversion, så att kunden, inte designhuset, bär regelrisken.

KAPITEL 11

# 11 MÖJLIGHETER
# Var möjligheterna finns, och vad jag skulle göra

### Kundanpassade chip: efterfrågan kommer överväldigande från AI

År 2026 kommer efterfrågan på kundanpassade chip överväldigande från AI. Broadcoms intäkter från AI-halvledare under tredje kvartalet av räkenskapsåret 2026 var 16,7 miljarder dollar, en ökning med 221 % jämfört med året innan, omkring 58 miljarder dollar för året, varav 73 % från XPU:er för sex kunder (enligt en automatisk transkription av resultatsamtalet). Global Unichips intäkter under andra kvartalet steg 128 % jämfört med året innan, med intäkter från nyckelfärdig produktion upp 188 % och en bruttomarginal på 21,5 %. Alchips intäkter under första kvartalet fördelade sig per nod med 21 % på 3/2 nm, 63 % på 7/5 nm och 16 % på 12 nm och större. VeriSilicons nya order för 2026 nådde 8,24 miljarder RMB (omkring 1,16 miljarder dollar) i slutet av april, varav 91,37 % för AI-beräkning. Marknaden har tre nivåer: överst flerårig samutveckling värd miljarder; i mitten projekt från 7 nm till 2 nm med NRE från några miljoner till över hundra miljoner dollar (5–30 miljoner dollar i mellannoder, 30–100 miljoner dollar eller mer i spetsen); och en lång svans av ASIC:er i mogna noder med NRE på 0,5–5 miljoner dollar. Kunder inom krypto och mining är en liten, cyklisk nisch, men de är den snabbaste och mest tillståndsfria ”första kunden” som betalar för beräkningskraft i dag.

De förmågor hos ett miningchipteam som går bäst att överföra, drift nära tröskelspänningen, massivt replikerade beräkningsmatriser, kraftmatning och kylning i kilowattklassen och snabba tape-outs i spetsnoder, är precis vad inferenschip och ZK- och FHE-acceleratorer behöver. Kapitalmarknaderna prissätter detta: Velaura AI (tidigare Auradine) tog in 110 miljoner dollar i augusti till en värdering över 1 miljard dollar. Motexemplet är Canaans enhet för AI-chip i kanten, som hade intäkter på 0,9 miljoner dollar räkenskapsåret 2024 medan den förbrukade 15 % av rörelsekostnaderna, och som stängdes i juni 2025. Skillnaden ligger i en egen väg till marknaden, inte i att återanvända miningkanaler.

### AI-utveckling: timmar är en handelsvara; värdet sitter i skarvarna

I den lägre änden är AI-utvecklingstjänster en handelsvara: typiska timpriser på Clutch är 25–49 dollar i timmen, och 50–99 dollar i USA, Kanada och liknande marknader; ett utvecklingsteam för blockkedjor kostar omkring

15 000 till mer än 90 000 dollar i månaden. Att sälja timmar har ingen vallgrav. Värdet sitter i skarvarna mellan fälten: kärnor och kompilatorer för kundanpassade chip, mining inbäddad i inferenstjänster (Pearls vLLM-insticksprogram är ett exempel), verifierbar inferens som kombinerar TEE med stickprov med ZK, samt betalningar och identitet för agenter.

## System för AI × krypto: teckna köparen av beräkningskraft före token

För den som konstruerar en AI-nativ kedja eller ett beräkningsnätverk har verktygslådan för verifiering 2026 tagit form. I ordning efter kostnad per arbetsenhet, från lägst till högst: statistiskt stickprov (Hyperbolics proof of sampling), delegering med skiljedomare (Gensyns Verde, som bara behöver en ärlig verifierare men kräver bitexakt exekvering), TEE (nära inbyggd hastighet, men förtroende för chipleverantören), full redundans och generell zkML. Den praktiska konstruktionen är skiktad: TEE eller stickprov för låg latens, tvistspel eller stickprov med ZK för slutgiltighet; en artikel från december 2025 simulerade konfidentiell beräkning på H100 med bedrägeribevis och slumpmässiga ZK-stickprov vid 99 % av centraliserad genomströmning och en marginalkostnad på 0,07 dollar per fråga. Hårdvarufilosofin avgör säkerhet och distribution: GPU-nativ PoUW skyddas till en början av hyrbara GPU:er i molnet, så kostnaden för en attack är knuten till hyrespriserna, och Qubic visar att proof of work på generell hårdvara kan kapas i grossistskala av externa incitament.

Kostnad och tid är inte hindret. Pearl gick från artikel (april 2025) till huvudnät på omkring 12 månader, forkat från btcd, med fullnod, plånbok, lätt klient och miner vid lanseringen. De flesta DeFi-granskningar kostar 25 000–100 000 dollar, granskningar för företag eller flera kedjor 150 000–250 000 dollar eller mer, ZK-kretsar 80 % till 120 % mer, och projekt med verkligt värde på spel budgeterar 150 000–500 000 dollar om året för säkerhet. De verkliga hindren är distribution och likviditet. Grundare uppger att Binance har begärt 0 % till 3,5 %, och enligt uppgift 8 %, av tokenutbudet för en notering (Binance förnekar att bolaget tjänar på noteringar); alla sex projekt som Cointelegraph granskade och som använde avtal om marknadsgaranti av typen ”tokenlån plus köpoption” föll, och Movements marknadsgarant sålde tokens för omkring 38 miljoner dollar på noteringsdagen. Alternativen är likviditet i kedjan, som med Hyperliquid, som passerade 10 miljarder dollar i marknadsvärde utan notering på Binance och vars eviga terminsmarknader enligt HIP-3 passerade 3,2 miljarder dollar i öppna positioner i juni, och offentliga försäljningar via stora börser, som när Monad tog in 269 miljoner dollar via Coinbase. Ett högt fullt utspätt värde med låg flytande andel var den tydligaste prediktorn för nedgång efter lanseringen 2025–2026.

Tokendesign är en del av systemkonstruktionen. Emissionen bör följa betald efterfrågan i stället för att gå före den: de nätverk i rapporten som har sundast subventionsmultiplar knyter belöningarna till arbete som en extern köpare betalar för, och de som har sämst betalar för hashkraft eller insats utan något bakom. En praktisk regel är att före lanseringen offentliggöra vilken subventionsmultipel nätverket siktar på under första, andra och tredje året, och att sänka emissionen när målet missas analys.

| Inriktning | Vem som köper, och betalningsvilja | Fönster | Nyckelförmågor | Huvudrisk |
|---|---|---|---|---|
| **Inferensaccelerator som kan utvinna** | Neoclouds, privata driftsättningar, PoUW-nätverk | 2027–2028 (beroende på forkarna till FP8 och FP4) | GEMM för INT8, FP8 och FP4; SRAM-arkitektur; protokollintegration | Protokollets godkännande, arkitekturinlåsning |
| **SRAM-baserat eller hårdkodat inferenschip** | Neoclouds, datacenter i Fastlandskina, kanten | Nu | Kompilatorer och modellanpassning; design på N6–N4 | Modellerna ändras för snabbt |
| **Miningchip (nisch eller vertikalt integrerat)** | Miners med egen el; miners för hem och uppvärmning | Nu, begränsat av halveringen och hashpriset | Lågspänningsmatriser, effektivitet, leveranskedja för volym | Priskrig från de etablerade, elpriser |

| Inriktning | Vem som köper, och betalningsvilja | Fönster | Nyckelförmågor | Huvudrisk |
|---|---|---|---|---|
| **Acceleration av ZK-bevisning** | Bevisarnätverk, rollups | Vänta tills beviss-ystemen konvergerar | Programmerbar MSM och NTT | Efterfrågan beror på tokenbelöningar |
| **Acceleration av FHE** | Krypterad AI för företag och myndigheter; integritetskedjor | 2027–2029 | NTT, HBM-system, kryptogra-fisk ingenjörskonst | Efterfrågan ännu inte formad; långa säljcykler |
| **Härdning av TEE och verifierbar inferens** | Nätverk för konfidentiell inferens, DePIN | Nu | Minneskryptering med integri-tet, oberoende attesteringsan-kare, ZK-stickprov | Beroende av leve-rantörernas ekosystem |
| **AI-nativ kedja eller beräkningsnätverk** | Inferensköpare, miners, utvecklare | Nu | Verifieringsdesign, emissions-design, distribution | Nedgång efter lan-seringen, tokenlikviditet |

## Tre vägar jag skulle ta, var och en med ett första steg

Om jag måste välja skulle jag ta högst tre vägar, var och en med ett litet första steg som prövar den innan pengar binds analys.

- **Väg A: en inferensaccelerator som också kan utvinna.** Första steg: en fyra veckor lång arkitekturstudie som modellerar J/op och $/op för en SRAM-baserad INT8/FP8-GEMM-matris mot RTX 5090 och H100, både på Pearls arbete och på transformer-inferens, följd av validering på FPGA eller MPW i mogen nod med en NRE i storleksordningen en miljon dollar.
- **Väg B: tjänster för kundanpassat kisel med kryptot som första kund och AI som andra marknad.** Första steg: gör rapportens modell för NRE och break-even till en betald förstudie för projekt med beräkningschip, med betalning per milstolpe och godkännande mot en fast protokollversion.
- **Väg C: ett AI-nativt nätverk som tecknar sin köpare av beräkningskraft före sin token.** Första steg: ett de-signnotat om skiktad verifiering (TEE plus stickprov plus ZK-stickprov) och om en leverantörsneutral kanonisk aritmetik, skrivet tillsammans med en namngiven inferensköpare före all tokendesign.

KAPITEL 12



12

SIGNALER ATT FÖLJA

# Vad man ska följa, november 2026 till 2027

| När | Vad man ska följa | Varför det är viktigt |
|---|---|---|
| **När som helst** | Pearls aktiveringshöjd för FP8 | Avgör om heltalshårdvara och kort som inte är från NVIDIA förblir godkända, och om kedjan delas |
| **Månadsvis** | Pearls andel block som bär AI-arbete | Det enskilt bästa testet av om ”nyttigt” håller på att bli sant |
| **Månadsvis** | Om inferensleverantörer för vidare miningintäkter till kunderna | Betald efterfrågan möter emission: den modell Pearl behöver |
| **Kvartalsvis** | Bittensors verifierbara intäkter mot dess emissioner | Om det största nätverket för nyttigt arbete håller på att sluta sitt subventionsglapp |
| **Kvartalsvis** | Andelen x402-flöden som faktiskt initieras av agenter | Om agentbetalningar går från signalering till handel |
| **Januari 2027** | TSMC:s rapporterade prishöjning på 3 % till 6 %, mest för 2 nm och 3 nm | Kostnaden för tape-outs i spetsnoder |
| **Första kvartalet 2027** | FP4 i Pearl, enligt uppgifter från grundarnas AMA | En ytterligare förskjutning av arbetet mot NVIDIAs nyaste format |
| **Slutet av 2027** | Samsung Foundrys mål att nå nollresultat | Om en billigare andra källa förblir livskraftig |
| **April 2028** | Bitcoins nästa halvering | Halverar potten för SHA-256-miners |

KAPITEL 13

# 13 FRÅGOR ATT STÄLLA
# Frågor du inte ställde men borde ha ställt

Frågorna nedan är de jag skulle ställa innan jag satsade pengar eller kisel på något projekt inom AI × krypto.

**Vem styr arbetsfunktionen?** Om ett litet team kan ändra aritmetiken vid en hårdkodad höjd är din hårdvaras livslängd deras beslut. Be om ett skriftligt åtagande om stabilitet, eller konstruera hårdvara med en andra marknad.

**Hur stor är subventionsmultipeln?** Dela emissionerna (eller marknadsvärdet) med intäkter som betalas utifrån nätverket. En multipel över 10 betyder att nätverket köper aktivitet, inte säljer en tjänst.

**Vad kostar nästa grind, och vilken siffra stoppar dig?** En multiprojektwafer kostar tiotusentals eller hundratusentals dollar; ett maskset kostar miljoner. Bestäm stoppsiffran innan du spenderar.

**Hur stor är hashkraften för återbetalning, med p?** Dela NRE med $R \times (1 - 1/k) \times T \times p$. Om svaret är en stor andel av nätverket är chipet en satsning på reglerna, inte på kislet.

**Vilken är chipets andra marknad?** Om protokollet forkas eller token faller, vad mer kan kislet säljas till? GPU:er överlevde Ethereums Merge eftersom de hade en; ASIC:er för en enda algoritm gjorde det inte.

**Hamnar värdet i token eller i aktierna?** År 2026 är de lönsamma företagen inom AI × krypto mest aktiebolag med kryptodistribution; bestäm vilket av de två du bygger innan du tar in pengar.

**Vem är den första betalande kunden, med namn?** Varje hållbart fall i denna rapport, från Grass till Bittensors subnät för beräkning, började med en namngiven köpare utanför nätverket.

Det här är områden jag gärna utforskar tillsammans med andra, genom tekniskt utbyte och tekniskt samarbete: kundanpassade chip för mining, AI-träning och inferens samt acceleratorer; AI-utveckling; och system för AI × krypto från kedja till nod till chip. Du når mig på i@tanshuai.com.

## Begränsningar i denna rapport

Det här är skrivbordsforskning baserad på offentliga dokument fram till 13 november 2026. Jag gjorde inga primärintervjuer och inga oberoende mätningar för denna utgåva. Flera nyckeltal är uppgifter från bolag eller leverantörer, som Aethirs intäkter, Grass intäkter, Etcheds och Taalas prestanda och Cysics effektivitet; de är märkta som sådana. Kedjemått för Pearl kommer från en oberoende fullnod och från pooler och ändras timme för timme; jag använde avläsningarna omkring kl. 07:13 UTC den 13 november och anger intervall för samma dag. Kostnader för wafers, masker och design är uppskattningar och modellvärden från tredje part som inte stämmer överens med varandra; jag visar intervall. Min break-even-kalkyl för ASIC:er och budgeten i steg använder illustrativa antaganden (k, T, p, NRE och kostnader per grind) som läsaren bör ersätta med sina egna.

# Om författaren

Tan Shuai är teknolog och grundare med mer än tjugo års arbete inom chip, hårdvara och systemprogramvara, och 20 beviljade patent. Han har många års forskning, utveckling och ledning inom fabless-chip bakom sig, bland annat helt kundanpassad design av ASIC-datavägar för miningchip för kryptovalutor och multimediachip för kameror, vilket är den lins rapporten ser genom: energi per operation, kostnad per wafer, vad ett protokoll låter dig gjuta i kisel, och vem som betalar. Hans arbete har tagit honom genom Silicon Valley, New York, Chicago, Taiwan, Greater Bay Area (Shenzhen och Hongkong), Vietnam och Malaysia. I dag arbetar han som AI-forskare och AI-utvecklare.

**Teknisk diskussion: i@tanshuai.com · tanshuai.com**

### Anmärkningar och ansvarsfriskrivning

Denna rapport är oberoende skrivbordsforskning baserad på offentliga dokument som fanns tillgängliga fram till 13 november 2026. Den tillhandahålls enbart för information och lärande och är inte investerings- eller teknisk rådgivning. Den behandlar enbart teknik och affärer. Priser, kedjemått och protokollregler ändras; kontrollera allt du förlitar dig på mot primärkällan. Siffror märkta leverantörsuppgift, uppskattning eller analys ska läsas som sådana. Priserna på kryptotillgångar är mycket volatila.

© 2026 Tan Shuai. Du får dela rapporten oförändrad med angivande av källa.

# Källor

Alla lästa mellan 10 och 13 november 2026. Grupperade efter kapitel.

### Pearl: protokoll, kod och kedjedata


- Komargodski och Weinstein, ”Proofs of Useful Work from Arbitrary Matrix Multiplication”, arXiv 2504.09971: https://arxiv.org/abs/2504.09971
- Pearl Research Labs, vitbok om INT: https://pearlresearch.ai/research/int-whitepaper
- Pearls källkod: genesisblock: https://github.com/pearl-research-labs/pearl/blob/main/node/chaincfg/genesis.go ; konsensusparametrar och fork-höjder: https://github.com/pearl-research-labs/pearl/blob/main/node/chaincfg/params.go ; redovisning av hashkraft: https://github.com/pearl-research-labs/pearl/blob/main/zk-pow/src/api/sanity_checks.rs ; uppgraderingsguide för forken med saltat frö: https://github.com/pearl-research-labs/pearl/blob/main/docs/salted-seed-fork-upgrade-guide.md ; parametrar i grenen fp8-scheme: 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, mining med flyttal: https://github.com/pearl-research-labs/pips/blob/main/PIPS/pip-0003.md
- prlstats (oberoende fullnod): nätverk: https://prlstats.com/api/network ; serier: https://prlstats.com/api/series ; startsida och modell för minerintäkter: https://prlstats.com/api/home ; utvunnet utbud: https://prlstats.com/api/mined ; pooler: https://prlstats.com/api/pools ; innehavare: https://prlstats.com/api/holders ; antaganden om hårdvara och merkostnad: https://prlstats.com/api/assumptions
- Kryptex Pool: Pearls nätverk: https://pool.kryptex.com/prl/about-coin ; hashkraft: 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 ; prishistorik: https://api.coingecko.com/api/v3/coins/pearl-2/market_chart?vs_currency=usd&days=365&interval=daily ; sökning på tickern ”Perle”: 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”, 2 juni 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, partnerskap med Pearl Research Labs: https://together.ai/blog/together-ai-partners-with-pearl-research-labs ; modellsida: https://together.ai/models/gemma-4-31b-it-pearl
- Basu, inventering av Pearls mininginstanser, arXiv 2606.04819: https://arxiv.org/html/2606.04819v2
- Blockworks, nyhetsbrevet 0xResearch, 28 oktober 2026: https://blockworks.com/newsletter/0xresearch/issue/post_daf0d74a-2bf4-41e4-86d3-2a1275475a24
- Datawallet, ”Pearl Explained”: https://www.datawallet.com/crypto/pearl-explained
- Coingabbar om TheMarkers rapport om finansieringen av Pearl Research Labs: https://www.coingabbar.com/en/pearl-research-labs-funding-130m-stealth-round-themarker


### Kedjor för nyttigt arbete och historik


- Grothe m.fl., säkerhetsanalys av Gridcoin, USENIX WOOT 2017: https://www.usenix.org/system/files/conference/woot17/woot17-paper-grothe.pdf
- Förfalskning av proof of learning, arXiv 2208.03567: https://arxiv.org/abs/2208.03567
- hoge.gg, verifiering och förtroende i decentraliserad inferens: https://hoge.gg/decentralized-inference-2026-verification-trust/ ; Bittensors intäkter och kassa: https://hoge.gg/bittensor-quantum-wallet-revenue-tao-treasury-2026/
- Kaspa: matriskoden i kHeavyHash: https://github.com/kaspanet/rusty-kaspa/blob/master/consensus/pow/src/matrix.rs ; lansering: https://coinranking.com/coin/V8GxkwWow+kaspa-kas/about ; IceRivers tillkännagivande av KS1/KS2: https://www.iceriver.io/2023/03/14/1677644580/ ; historiken från GPU till ASIC: https://captainmining.com/blogs/crypto-news/the-evolution-of-kaspa-mining-from-gpus-to-asic-powerhouses ; hashkraft för GPU:er: https://coincalculators.io/algorithms/kheavyhash ; KS7: https://www.asicminervalue.com/miners/bitmain/antminer-ks7-40th ; nätverk: https://minerstat.com/coin/kas ; 1 EH/s 2024: https://ourcryptotalk.com/news/kaspa-reaches-1-ehs-in-hashrate ; Crescendo: https://ourcryptotalk.com/news/kaspa-crescendo-live
- Aleo: frågor och svar om huvudnätet: https://aleo.org/post/aleo-mainnet-faq/ ; decentraliserad bevisning: https://aleo.org/post/decentralized-proving-advantages/ ; ASIC:er: 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 om omorganisationen i Monero: 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 ; mining av Dogecoin: https://hackernoon.com/qubic-is-now-mining-dogecoin-on-its-ai-compute-infrastructure ; CoinGecko: https://www.coingecko.com/en/coins/qubic

- Bittensors dokumentation om konsensus: 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 och forkar för ASIC-resistens: https://lbry.com/news/forking-for-asic-resistance ; Sias fork: https://hotfix.coindesk.com/markets/2018/11/01/sia-completes-hard-fork-as-bricked-miners-form-rival-blockchain ; Obelisks priser: https://www.cryptoglobe.com/latest/2018/11/obelisk-hardware-skyrockets-in-selling-price-after-sia-hard-fork/ ; miners efter the Merge: https://hotfix.coindesk.com/business/2022/11/11/ex-ethereum-miners-token-hop-to-stay-alive-after-the-merge


### Sektorer inom AI × krypto


- DefiLlama: Render: https://defillama.com/protocol/render-network-bme ; Livepeer: https://defillama.com/protocol/livepeer
- Cointelegraph om Messaris uppgifter om DePIN-intäkter: https://cointelegraph.com/news/depin-revenues-grow-despite-token-losses
- Aethir, sammanfattning av 2025: https://aethir.com/blog-posts/aethirs-2025-wrap-up-decentralized-gpu-cloud-milestones
- CoinGecko, AI-kategorin: https://www.coingecko.com/en/categories/artificial-intelligence
- Akash, hyrespriser för H100: https://akash.network/the-bid/h100-rental-price-2026-cost-per-hour/ ; rapport för kv1 2026: https://akash.network/blog/akash-network-q1-2026-report/
- io.net, affär på 8 miljoner dollar och dagliga token: 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 Computes kontrakt: 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 om INTELLECT-3: https://www.implicator.ai/prime-intellects-intellect-3-open-source-ambition-meets-centralized-reality/
- SiliconANGLE, Prime Intellects finansiering: 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 finansiering: https://runtimewire.com/article/nous-research-75-million-round-1-5-billion-valuation
- ownyourmind, Gensyn: https://ownyourmind.ai/projects/gensyn/ ; Phemex, Ritual (endast rubrik): https://phemex.com/news/article/ritual-launches-ai-layer-1-mainnet-and-ritual-token-price-dives-72-from-high-101259
- Merkostnad för konfidentiell beräkning, arXiv 2409.03992: https://arxiv.org/abs/2409.03992v2 ; Phalas rapport för 2025: 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, mätning av AI-agenters betalningar: https://www.trmlabs.com/trm-tech-blog/whos-actually-paying-measuring-ai-agent-payments-onchain ; Forkast om 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, Tempos huvudnät: https://www.theblock.co/amp/post/394131/tempo-mainnet-goes-live-with-machine-payments-protocol-for-agents ; BlockEden om Virtuals aGDP: https://blockeden.xyz/blog/2026/04/21/agdp-agent-gdp-virtuals-protocol-ai-blockchain-valuation-primitive-tvl-displacement/ ; The Block om ElizaOS: https://theblock.co/post/410774/eliza-labs-native-token-dead
- Dealroom, Grass intäkter: https://dealroom.co/news/157794-grass-backed-by-multicoin-projects-75m-in-2026-data-revenue/ ; CryptoSlate, Fetch.ai och Ocean: https://cryptoslate.com/cryptos-flagship-ai-pact-fracture-fetch-sues-ocean-over-263m-fet-community-sales/ ; Fortune om World: https://fortune.com/2026/06/01/internet-according-to-sam-altman-seeing-orb/
- Galaxy, riskkapital till krypto: https://www.galaxy.com/insights/research/crypto-blockchain-venture-capital-q3 ; Crypto Briefing (endast rubrik): https://cryptobriefing.com/crypto-vc-investment-q3-2026-galaxy/ ; Blockhead om Tiger Research: https://www.blockhead.co/2026/10/14/bitcoin-rose-43-in-q3-crypto-venture-funding-did-not-follow/ ; ForkLog om tokenlanseringar 2025: https://forklog.com/en/85-of-tokens-launched-in-2025-fell-below-initial-prices/


### ASIC-ekonomi och chiptyper


- Hashrate Index, sammanfattning 9 november 2026: https://hashrateindex.com/blog/hashrate-index-roundup-november-9-2026/
- AnySilicon, kostnad för ASIC-design per nod 2026: https://anysilicon.com/news/how-much-does-asic-design-cost-in-2026-real-price-ranges-by-process-node/ ; Silicon Analysts, kostnadsguide för tape-out i fabless-bolag: https://siliconanalysts.com/analysis/fabless-startup-tapeout-cost-guide ; marknadsdata: https://siliconanalysts.com/market-data
- MillionMiner, Bitmain S23 XP Hyd: https://millionminer.com/news/bitmain-wdms-2026-s23-xp-hyd-launch
- Khattak och Mikaitis, FP8-ackumulering i tensorkärnor, arXiv 2512.07004: https://arxiv.org/html/2512.07004v2
- Canaan, resultat för andra kvartalet 2026: https://finviz.com/news/389337/canaan-inc-reports-unaudited-second-quarter-2026-financial-results ; Bitdeer, resultat för tredje kvartalet 2026: 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 och Block: https://theenergymag.com/news/2026-07-29/core-scientific-exit-block-bitcoin-deal-ai ; ekonomin i AI-leasing: https://theenergymag.com/news/2026-08-06/miner-weekly-ai-pumpability-out-fuel
- Blockworks, Cipher och 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, resultat för kv1 räkenskapsåret 2027: https://globenewswire.com/news-release/2026/11/05/3397020/0/en/iren-reports-q1-fy27-results.html
- Spark, marknadskoncentration bland ASIC:er för bitcoin: https://www.spark.money/research/bitcoin-asic-manufacturer-market-concentration

- Google, dokumentation för TPU7x: 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 ; artikeln om TPU v1, 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, prissättning av CoWoS: 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, jämförelse av acceleratorer för AI-inferens: https://themenonlab.blog/blog/ai-inference-accelerators-compared ; Introl, AI-accelerators bortom GPU:er: https://introl.com/blog/ai-accelerators-beyond-gpus-tpu-trainium-gaudi-cerebras-2025
- StratNews (Reuters), inhemska AI-chip i statligt finansierade datacenter: https://stratnewsglobal.com/team-sng/china-orders-state-funded-data-centres-to-use-domestic-ai-chips/ ; TNW, import av H200: https://thenextweb.com/news/nvidia-h200-chips-hong-kong-china-import-approval
- Succinct, bevisning i realtid på 16 GPU:er: https://blog.succinct.xyz/real-time-proving-16-gpus/ ; Cysics ZK-ASIC-produkter: 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 och SemiFive: https://evertiq.com/design/2026-02-20-semifive-secures-design-win-with-niobium-for-fhe-accelerator ; The Quantum Insider, Niobiums finansiering: 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 på FPGA: https://www.zama.org/post/announcing-hpu-on-fpga-the-first-open-source-hardware-accelerator-for-fhe


### Processnoder


- Silicon Analysts, waferpriser: https://siliconanalysts.com/data/wafer-pricing
- Tom's Hardware, rykten om TSMC:s priser för A16: 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:s priser: https://wccftech.com/smic-5nm-chip-prices-up-to-50-percent-more-expensive/amp/
- EUROPRACTICE, TSMC:s MPW-schema 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, färdplanen för Huawei Ascend: 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, Kinas leveranskedja för AI-chip: 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


### Möjligheter och lanseringsekonomi


- Webull (Benzinga), Broadcoms resultatsamtal: 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, Canaans AI-enhet: https://www.theblock.co/post/359250/crypto-miner-canaan-to-shutter-non-core-ai-semiconductor-business-unit-amid-strategic-realignment
- Clutch, priser för AI-utveckling: https://clutch.co/developers/artificial-intelligence/pricing ; PixelPlex, kostnad för blockkedjeutveckling: 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 ; skiktad verifiering, arXiv 2512.20176: https://arxiv.org/abs/2512.20176v1
- Sherlock, prisreferens för granskningar 2026: https://sherlock.xyz/post/smart-contract-audit-pricing-a-market-reference-for-2026
- Blockworks, Binances noteringskrav: https://blockworks.com/news/binance-listing-standards-scrutiny ; Cointelegraph, avtal med marknadsgaranter: 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

BEVIS PÅ NYTTIGT ARBETE · KRYPTO × KISEL

# Beräkningskraft som konsensus

AI × krypto × chip 2026, med Pearls bevis på nyttigt arbete som utgångspunkt

Varje nyckeltal i rapporten anges med källa och datum. Där författaren själv härleder en siffra märks den som uppskattning; där bara säljaren uppger den märks den som leverantörsuppgift. Rapporten behandlar enbart teknik och affärer.



**Tan Shuai**

Teknolog och grundare · AI-forskare och AI-utvecklare

tanshuai.com · i@tanshuai.com

OFFENTLIG UTGÅVA · Data per 13 november 2026