Forschungsnotiz · November 2026

Rechenleistung als Konsens

KI × Krypto × Chips 2026, ausgehend von Pearl

Pearl · Nützliche Arbeit · DePIN und Agenten · ASIC-Ökonomie · Prozessknoten · Gestufte Budgets

Tan Shuai
tanshuai.com · i@tanshuai.com



DAS FAZIT

Das Theorem hält; der Anreiz nicht

01 Pearl funktioniert als Technik, noch nicht als nützliche Arbeit

Rund 90 EH/s ganzzahliger Matrixmultiplikation sichern die Kette, aber nur etwa 1 % der Blöcke enthält echte KI-Arbeit.

02 Bezahlte Nachfrage liegt bei Inferenz, Daten und Privatsphäre

Bittensor rund 70 Mio. $ verifizierbare Einnahmen pro Jahr, Grass 17 Mio. $ im 1. Halbjahr 2026, TEEs in Produktion; Agentenzahlungen etwa 5.000–11.000 $ im Monat.

03 Ein ASIC rechnet sich nur mit großem Topf, stabilen Regeln und Zweitmarkt

Seine Gewinnobergrenze ist R × (1 − 1/k). Heute bietet kein Netz alle drei Analyse .

≈90 EH/s

Pearl am 13. November 2026: Ganzzahldurchsatz von etwa 122.000 H100

≈1 %

der Pearl-Blöcke enthalten echte KI-Arbeit (Blockworks)

4–11×

Bittensors Emissionen im Verhältnis zu seinen verifizierbaren Einnahmen

KI × Krypto in sechs Sektoren: wo bezahlte Nachfrage real ist

Ketten mit nützlicher Arbeit — 01

REAL, ABER KLEIN

≈70 Mio. $/Jahr

verifizierbare externe Einnahmen von Bittensor

Signal: Emissionen 4–11× Einnahmen; Pearl ≈1 % KI-Blöcke

Die Mechanismen funktionieren; Nutzen braucht Zahler

Dezentrale Rechenleistung — 02

REAL, ABER KLEIN

2,35 Mio. $/Jahr

Renders On-Chain-Gebühren, die größten verifizierbaren

Signal: Marktwert ≈360× Gebühren

Meist Weiterverkauf von Standard-Inferenz

Training und Inferenz — 03

GEWINN IM EIGENKAPITAL

100+ Mio. $/Jahr

Prime Intellect, ein Unternehmen mit Eigenkapital auf zentralen GPUs

Signal: Gensyns Trainings-Mainnet nicht live

Fortschritt ist real; Gewinn geht ans Eigenkapital

Verifizierbare KI — 04

FRÜH

398

zahlende Nutzer von Phalas TEE-Cloud, Ende 2025

Signal: kein offengelegtes bezahltes zkML

TEEs in Produktion; zkML früh

Agentenzahlungen — 05

VOR DER NACHFRAGE

5–11 Tsd. $/Monat

tatsächlich von Agenten ausgelöste x402-Zahlungen

Signal: ≈200 Mio. Transaktionen, meist Signalisierung

Die Schienen kamen zuerst; die Nachfrage nicht

Datennetzwerke — 06

ECHTE KÄUFER

17 Mio. $

Datenerlöse von Grass, 1. Halbjahr 2026 (Eigenangabe)

Signal: Marktwert ≈10× annualisierte Einnahmen

Einer der wenigen Sektoren mit echten Käufern

> Subvention ist Treibstoff für einen Kaltstart, kein Geschäftsmodell. Krypto leistet Vertrieb und Abwicklung; die Gewinne gehen ans Eigenkapital, zumindest 2026.

Ein Multiply-Accumulate ist ein Hash

Vom Theorem zum Mainnet

Komargodski und Weinstein (April 2025): ein Proof of Useful Work für jedes Matrixprodukt mit 1 + o(1) Overhead. Pearls Mainnet startete am 27. April 2026.

So funktioniert es

Miner multiplizieren verrauschte Matrizen, hashen den Akkumulator jeder Kachel zu einem Los und beweisen die Gewinnerkachel mit einem Plonky2-Beweis unter 60 KB.

Die Einheit

1 H ist ein angerechnetes INT7 × INT7 → INT32-Multiply-Accumulate: das Sicherheitsbudget einer Blockchain, bepreist in Tensordurchsatz.

PEARL AM 13. NOVEMBER 2026

89,77 EH/s

Netzwerk-Hashrate (unabhängiger Full Node)

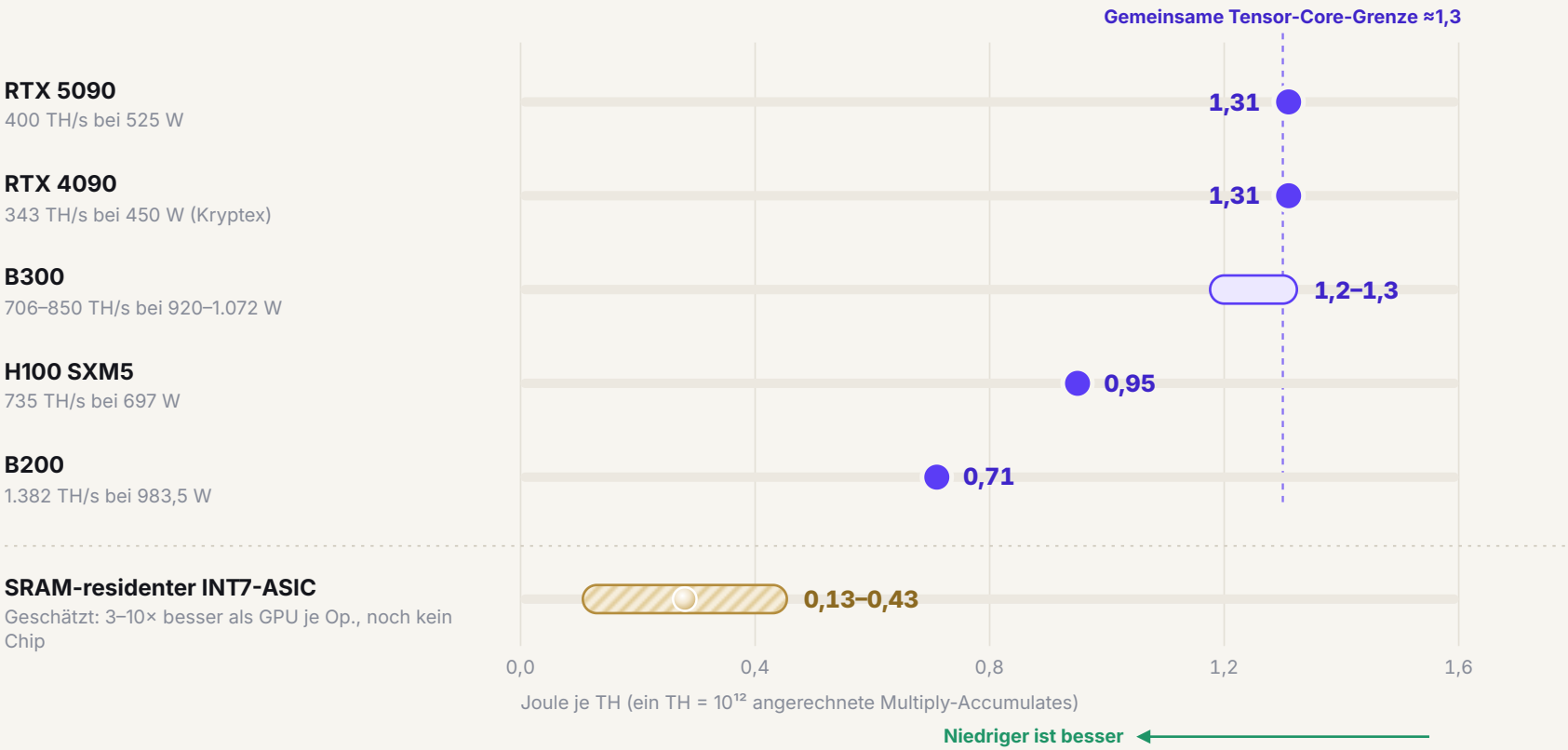
≈500 Mio. $

Marktwert; vollständig verwässert etwa 2,8 Milliarden US-Dollar

+78 %

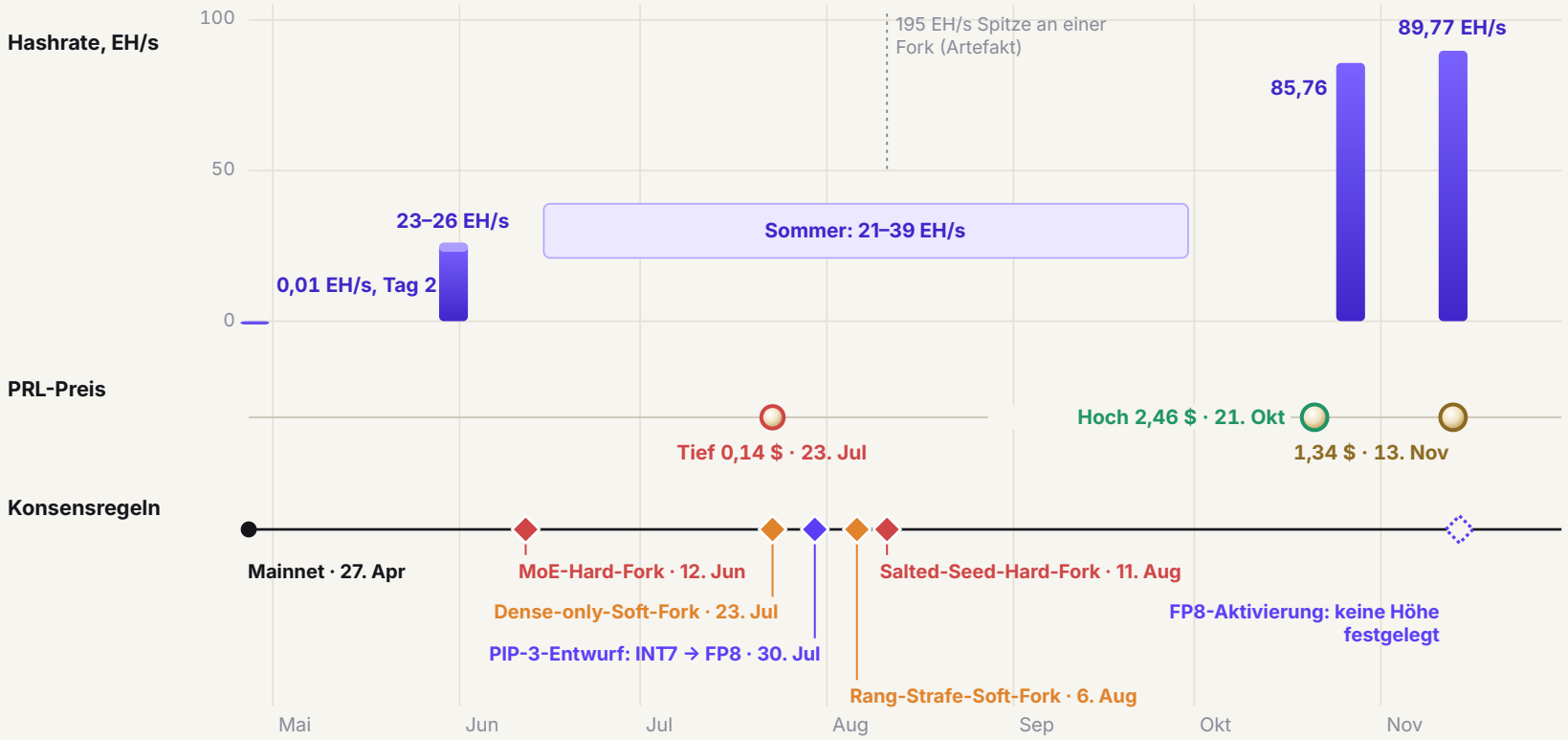
neues Angebot in den nächsten 12 Monaten

Bei Pearls Ganzzahlarbeit teilen GPUs eine Energiegrenze



> Gaming-Karten und die B300 treffen sich bei etwa 1,3 J/TH; selbst die B200 erreicht nur 0,71. Ein Chip, der nur 7-Bit-Ganzzahlen aus dem SRAM multipliziert, könnte plausibel 3- bis 10-mal besser sein **Analyse**.

Hashrate folgt dem Preis; die Regeln ändern sich ständig



Rund 120 Millionen Coins, 32 % des heutigen Angebots, wurden in den ersten sechs Tagen gemint. Eine RTX 5090 verdient jetzt etwa 5–6 $ am Tag; Karten nur zum Minen zu mieten, bringt Verlust.

Das Theorem hält; der Anreiz nicht

WAS NÜTZLICHE ARBEIT KOSTET

Inferenzseite: geringer Overhead

Pearl meldet 5,08 % mehr Zeit für Llama 70B auf vier H200; ein unabhängiger Test maß 9,2 % weniger Durchsatz auf einer Hopper-GPU.

WAS NÜTZLICHE ARBEIT EINBRINGT

Mining-Seite: zu wenig Hashrate

Eine Karte, die Inferenz bedient, mint mit etwa 27 % eines reinen Miners beim Prefill, 5,4 % beim Chat, rund 1 %, wenn Decoding dominiert.

Also multiplizieren rationale Miner Zufallsmatrizen. Nützlichkeit ist ein kommerzielles Problem außerhalb der Kette: Jemand muss für Inferenz bezahlen.

Vier Forks in zwei Monaten, und als Nächstes FP8

12. Juni
Mixture-of-Experts-Hard-Fork (Block 71.935)

23. Juli
Dense-only-Soft-Fork (Block 91.630)

6. August
Rang-Strafe-Soft-Fork: manche Setups verlieren 50–87,5 % effektiver Hashrate

11. August
Salted-Seed-Hard-Fork: alte Miner liefern nur noch ungültige Shares

Offen
PIP-3: INT7 zu FP8, nur Hopper und Blackwell zugelassen

2 von 3

nützlich für KI, bitgenau prüfbar, hardwareneutral: ein Zahlenformat schafft höchstens zwei

13 Bit

behält Hoppers FP8-Akkumulation, mit Abschneiden: nicht über Generationen reproduzierbar

≈82 %

der Hashrate auf Consumer-Karten, die PIP-3 nicht zulassen würde (Modell geringer Sicherheit)

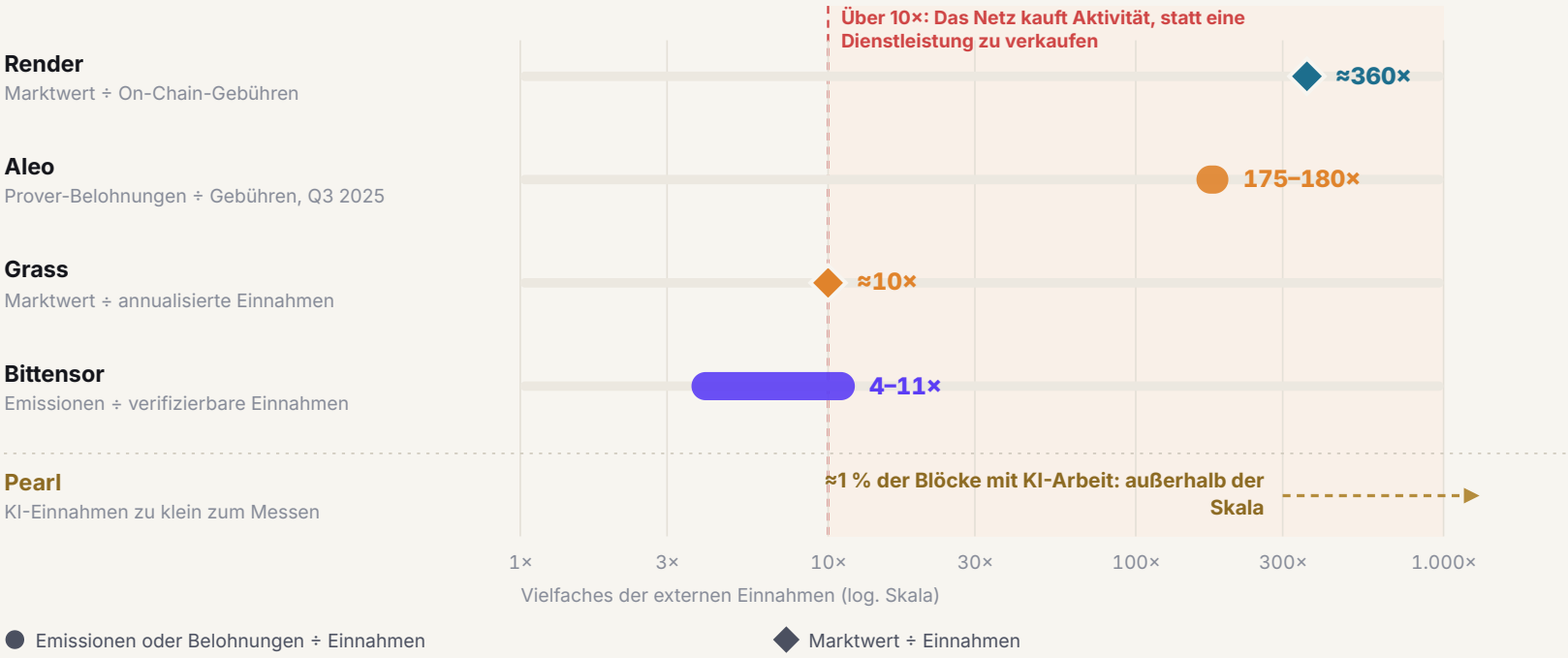
Wer die Arbeitsfunktion kontrolliert, bestimmt das Schicksal der Hardware-Investoren Analyse.

Weder Matrizen noch Zero Knowledge stoppen ASICs

Netz	Was geschah	Was es lehrt
Kaspa	Ein 64 × 64-Matrixprodukt zwischen Hashes; ASICs nach 16 Monaten, etwa 1.000× GPU-Effizienz; der beste verdient jetzt 0,86 $ am Tag	Eine Matrix ist weder nützlich noch ASIC-fest
Aleo	Ein synthetisches SNARK-Rätsel; Spezialmaschinen nach fünf Monaten; Prover-Belohnungen 175–180× Gebühren	Synthetische Arbeit läuft auf Subvention
Qubic	Behauptet KI-Training; seine Rechenleistung minte Monero (51 % behauptet) und jetzt Dogecoin	Allzweck-PoW lässt sich von außen kapern
Bittensor	Stake-gewichtete Bewertung; etwa 70 Mio. $ verifizierbare Einnahmen pro Jahr gegenüber mehr als 300 Mio. $ Emissionen	Das einzige mit externen Einnahmen in Größenordnung
Monero, Sia, Ethereum	Forks machten ASICs zu Elektroschrott; nach dem Merge fanden unter 20 % der Miner neue Arbeit	Hardware mit Zweitmarkt überlebt

SUBVENTION

Das Subventionsvielfache ist der ehrlichste Gesundheitscheck



Ab etwa dem 10-Fachen der externen Einnahmen kauft ein Netz Aktivität, statt eine Dienstleistung zu verkaufen. Pearls KI-Einnahmen sind zu klein für die Skala.

Reale, aber kleine Nachfrage, und Wert wandert ins Eigenkapital

Rechenmärkte

Renders On-Chain-Gebühren etwa 2,35 Mio. $ pro Jahr; Akashs H100 für 2,73 $ pro Stunde schlägt AWS und Google, nicht aber Vast.ai oder RunPod.

Training

8–40 Milliarden Parameter über das Internet trainiert; INTELLECT-3 kehrte zu 512 H200 mit InfiniBand zurück.

Verifizierbare KI

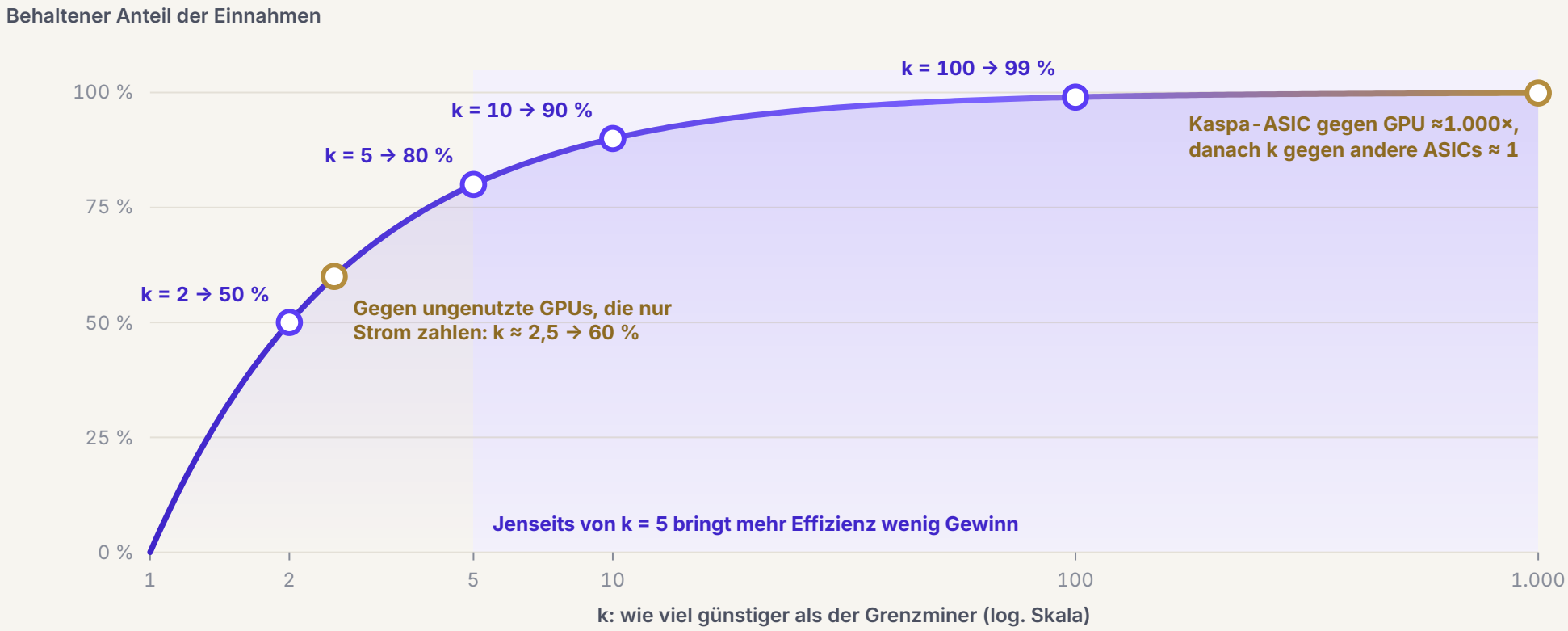
TEEs kosten unter 7 % und laufen in Produktion; TEE.fail brach die Attestierung mit einem 1.000-$-Interposer; zkML beweist GPT-2 mit 174 Token pro Minute.

Agenten und Daten

Von Agenten ausgelöste x402-Flüsse etwa 5.000–11.000 $ im Monat; Grass 17 Mio. $ Datenerlöse im 1. Halbjahr 2026.

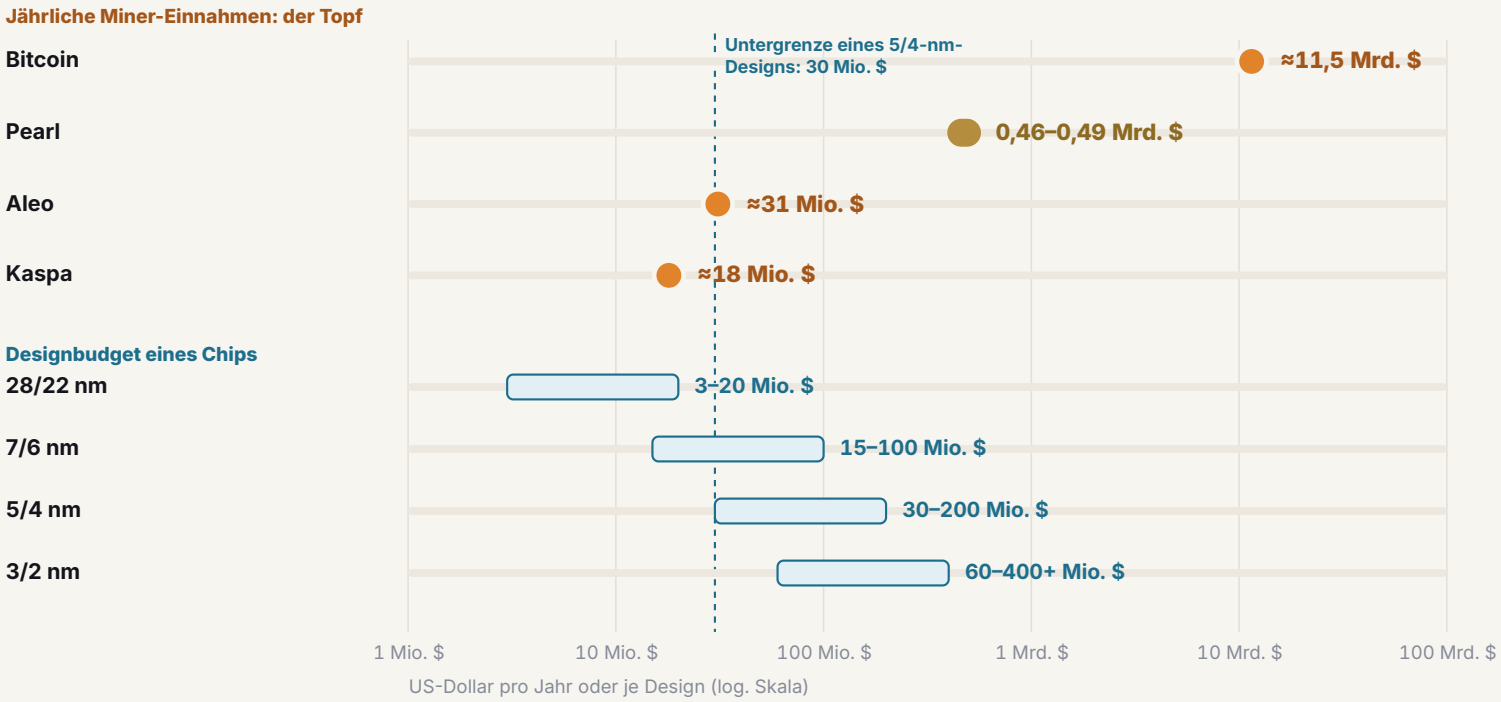
Vier Unternehmen mit Eigenkapital auf zentraler Infrastruktur sind etwa 8,5 Milliarden US-Dollar wert, ungefähr so viel wie alle KI-Token der Protokoll- und Anwendungsebene.

Warum Effizienz sich ab k = 5 nicht mehr auszahlt



Gewinnobergrenze ≈ R × (1 − 1/k). Sobald ASICs die Grenzminer sind, setzt sich k gegenüber anderen ASICs zurück; gegenüber ungenutzten GPUs, die nur Strom zahlen, liegt k bei etwa 2,5 **Schätzung**.

Die meisten Töpfe sind kleiner als das Designbudget eines Chips



Amortisations-Hashrate h* = NRE ÷ [R × (1 − 1/k) × T × p]. Für Pearl bei 20 Mio. $ NRE und k = 5: etwa 4,9 EH/s für ein Jahr, wenn die Regeln halten, 9,8 EH/s, wenn sie mit gleicher Wahrscheinlichkeit halten **Schätzung**.

Fünf Chiptypen, fünf Antworten



PoUW-Mining-Chip à la Pearl

NICHT JETZT

k: ohne HBM und Allzweckeinheiten: 3–10× je Op. (geschätzt)

Zweitmarkt: KI-Inferenz, bei allgemeinem INT8/FP8-GEMM

Auf einen Inferenz-Chip setzen, der minen kann, nicht auf ein reines INT7-Teil



Bitcoin-SHA-256-Mining-Chip

FERNBLEIBEN

k: ≤1 gegenüber den Etablierten

Zweitmarkt: fast keiner

Nur mit eigenem Strom, Nischenform oder als IP



KI-Trainingschip

IP STATT CHIPS

k: 1,1–1,8× bei gleicher Präzision

Zweitmarkt: —

Packaging und HBM entscheiden; IP, Chiplets, Dienste bauen



KI-Inferenzchip

BAUEN

k: Allgemeinheit aufgeben: 10–20× behauptet

Zweitmarkt: er ist der Hauptmarkt

Erst den Kunden gewinnen, dann die Festlegung wählen



ZK-Proving-Beschleuniger

ABWARTEN, ABSICHERN

k: 10–100× GPUs behauptet

Zweitmarkt: begrenzt außerhalb von ZK

Programmierbare MSM- und NTT-Engines; zuerst FPGA



FHE-Beschleuniger

FRÜH POSITIONIEREN

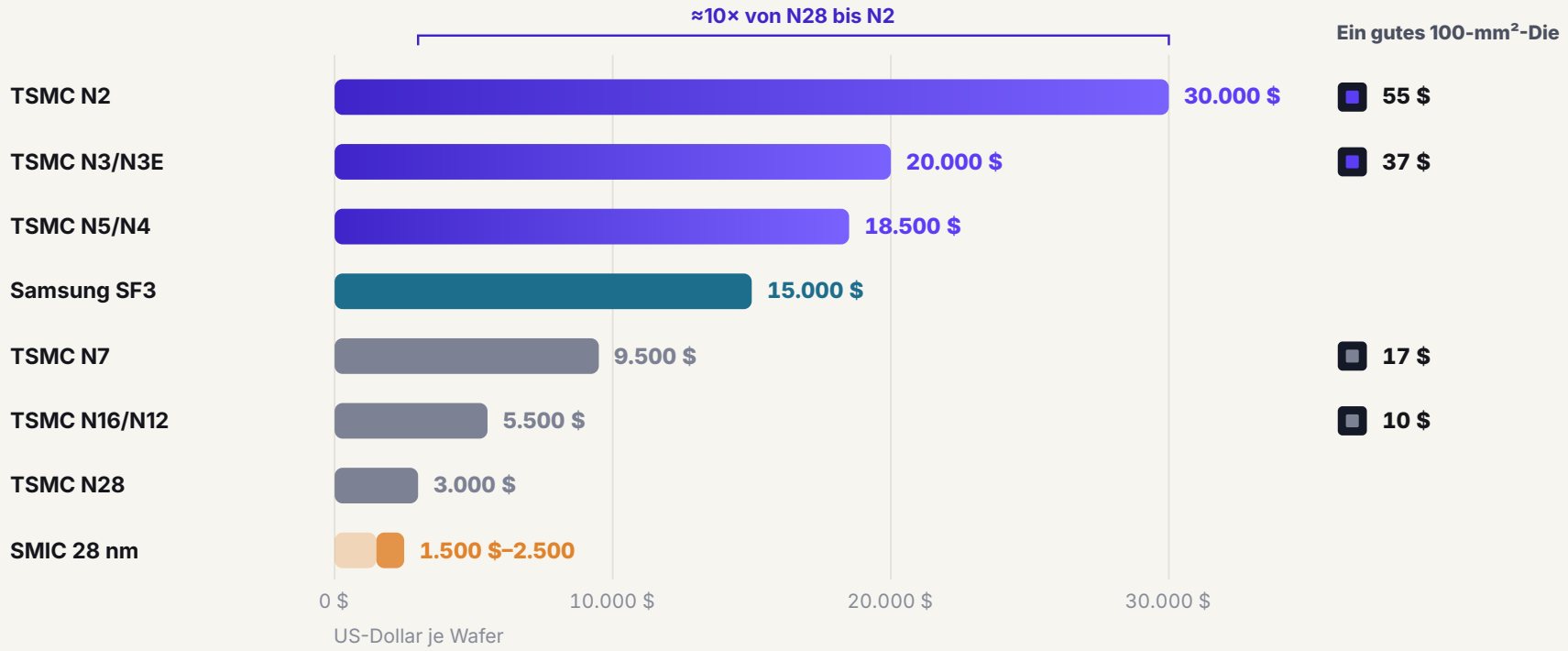
k: 1.074–5.547× ein Xeon (Heracles)

Zweitmarkt: verschlüsseltes Rechnen, private KI

Ein Fenster 2027–2029; mit wenig Kapital validieren

> Kein Netz bietet heute großen Topf, stabile Regeln und k über 1 zugleich: Bitcoins k gehört den Etablierten, Pearls Regeln sind am wenigsten stabil, Kaspas und Aleos Töpfe sind zu klein.

Ein Wafer am Leading Edge kostet zehnmal so viel wie ein ausgereifter



Die größten Lebenszeitkosten sollen entscheiden: Energie bei Dauerlast-Rechenchips, Packaging und HBM bei Trainingschips (63 % der Kosten einer H100), die Wahl SRAM oder HBM bei Inferenzchips.

Ein Knoten für jeden Chip und jedes Team

Chip	Kleines Team	Etabliertes Unternehmen	Weg in Festlandchina
PoUW-Mining-Chip	28 oder 16/12 nm MPW; erste Serie N7/N6	N3/N3E oder Samsung SF3	SMIC N+1/N+2 funktioniert, etwa 40–50 % mehr je Die
SHA-256-Mining-Chip	Nicht der Spitze nachjagen	3-nm-Klasse	Effizienz führender Knoten schwer erreichbar
KI-Trainingschip	IP, Chiplets, Designdienste	N3P, dann N2/A16 mit CoWoS und HBM	SMIC N+2/N+3 mit heimischem HBM
KI-Inferenzchip	SRAM-zentriert auf N6/N7 oder N4/N5/SF4	N3/N2 mit HBM oder SRAM-Chiplets	SMIC N+2 mit LPDDR oder GDDR
ZK- und FHE-Beschleuniger	Zuerst FPGA oder 16/12 nm MPW	N5/N3; FHE mit HBM	ZK machbar; FHE durch HBM begrenzt

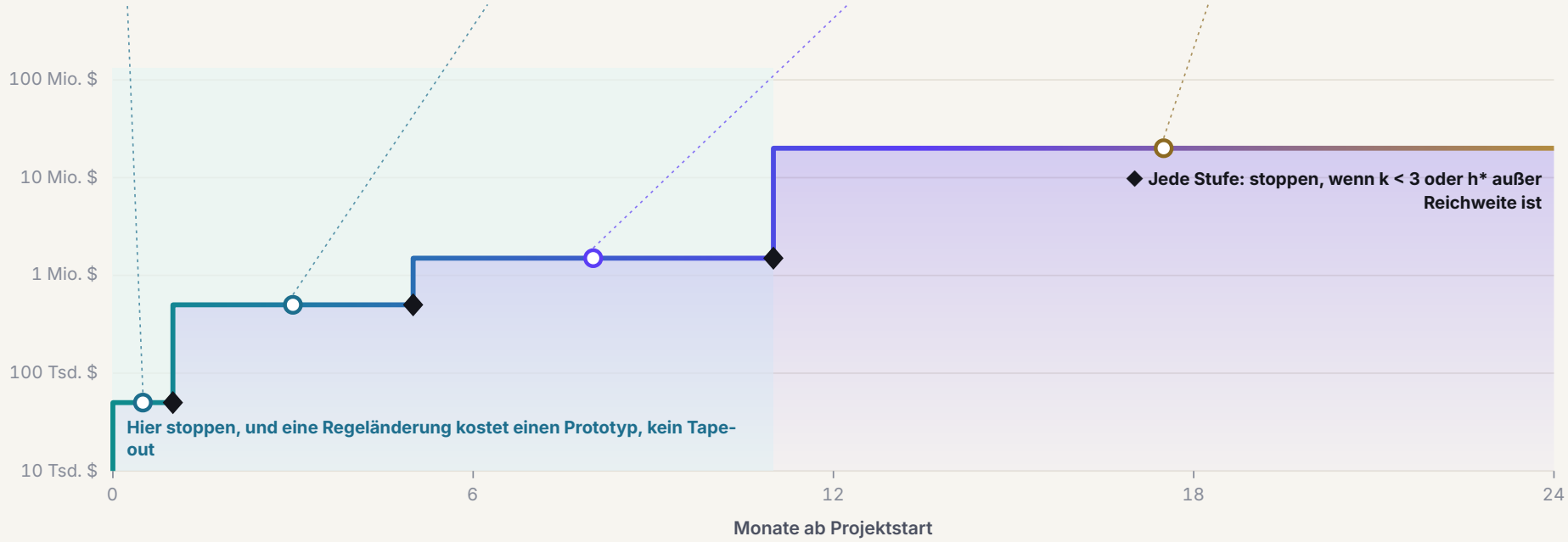
Belege vor Masken kaufen

Stufe 0 · Modell
≈50 Tsd. $ · Wochen
J/Op. und $/Op. aus Arithmetik, SRAM und I/O

Stufe 1 · FPGA
≈0,5 Mio. $ · Monate
Die echte Last durchgängig: Shares oder Token

Stufe 2 · MPW
≈1,5 Mio. $ · Monat 11
Gemessenes Silizium ersetzt das Modell

Stufe 3 · Masken und erstes Los
≈20 Mio. $ · Monat 24
N7-Masken ≈3,5 Mio. $ in einem Design für 15–100 Mio. $



Ein Modell, ein FPGA und ein MPW führen für etwa 1,5 Mio. $ in elf Monaten zu gemessenem Silizium; Maskensatz und erstes Los kosten mehr als zehnmal so viel. Vor den Masken stoppen, bis Käufer oder Protokollversion feststehen.

Krypto als erster Kunde, KI als Zweitmarkt

Kundenspezifische Chips

- Die Nachfrage ist überwiegend KI: Broadcom etwa 58 Mrd. $ KI-Halbleiter im Geschäftsjahr 2026
- Mining-Chip-Know-how ist übertragbar: Niederspannungs-Arrays, Kilowatt-Stromversorgung, schnelle Tape-outs
- Mit eigenem Marktzugang gewinnen, nicht über Mining-Kanäle

KI-Entwicklung

- Stunden sind Massenware: 25–49 $ pro Stunde auf Clutch
- Der Wert liegt an den Nahtstellen: Kernels für eigene Chips, Mining im Inferenzbetrieb
- Verifizierbare Inferenz: TEEs plus ZK-Stichproben

KI × Krypto-Systeme

- Gestufte Verifikation: TEE oder Stichproben, dann Streitverfahren oder ZK-Stichproben
- Vertrieb und Liquidität sind die Hürden, nicht die Kosten
- Erst den Compute-Käufer gewinnen, dann den Token

Drei Wege, jeder mit einem kleinen ersten Schritt

Weg	Erster Schritt
A. Ein inferenzorientierter Beschleuniger, der auch minen kann	Eine vierwöchige Studie zu J/Op. und $/Op. eines SRAM-residenten INT8/FP8-Arrays gegenüber RTX 5090 und H100, dann FPGA oder MPW auf ausgereiftem Knoten
B. Custom-Silicon-Dienste: zuerst Krypto, dann KI	Das NRE- und Break-even-Modell dieses Berichts in bezahlte Machbarkeitsstudien mit Meilensteinen und fester Protokollversion verwandeln
C. Ein KI-natives Netz mit Käufer vor dem Token	Eine Designnotiz zu gestufter Verifikation und herstellerneutraler kanonischer Arithmetik, geschrieben mit einem namentlich bekannten Inferenzkäufer

Höchstens drei Wege, jeder geprüft, bevor Geld gebunden wird Analyse.

Worauf zu achten ist, und Fragen, die man stellen sollte

WEGMARKEN

Jederzeit	Pearls FP8-Aktivierungshöhe und ob sich die Kette spaltet
Monatlich	Pearls Anteil an Blöcken mit KI-Arbeit
Quartalsweise	Bittensors verifizierbare Einnahmen gegenüber Emissionen; von Agenten ausgelöste x402-Flüsse
Januar 2027	Berichtete Preiserhöhung von TSMC um 3–6 % für 2 nm und 3 nm
Q1 2027	FP4 in Pearl, laut Berichten aus dem AMA der Gründer
April 2028	Bitcoins nächstes Halving

FRAGEN, DIE MAN STELLEN SOLLTE

01 Wer kontrolliert die Arbeitsfunktion?

02 Wie hoch ist das Subventionsvielfache?

03 Was kostet die nächste Stufe, und was stoppt Sie?

04 Wie hoch ist die Amortisations-Hashrate, samt p?

05 Was ist der Zweitmarkt des Chips?

06 Wer ist der erste zahlende Kunde, namentlich?

Das sind Felder, die wir gemeinsam erkunden können, im fachlichen Austausch und in technischer Zusammenarbeit: kundenspezifische Chips, KI-Entwicklung und KI × Krypto-Systeme.

i@tanshuai.com · tanshuai.com