HUB71+ AI · ACCESS PROGRAMME · COHORT 20

TRINITY S³AI

Synthesisable arithmetic for low-precision AI silicon

Numeric formats delivered as decoders, arithmetic cores and bit-exact conformance vectors that a chip team drops into an SoC. Precision became a hardware design variable; the standard that would settle it is not ratified. We sell the proof that a format behaves identically in silicon and in the reference model.

Dmitrii Vasilev — Founder, FPGA/RTL and hardware-AI engineer

ORCID 0009-0008-4294-6159 · t27.ai · admin@t27.ai · github.com/gHashTag

Pre-seed · bootstrapped · August 2026 · every figure in this deck is labelled by origin

PROBLEM

Low precision shipped. The conformance did not.

Four bits went mainstream in one hardware generation. What a chip team cannot buy is proof that its decoder matches the spec bit for bit.

- **Precision is now a hardware design variable, not a software detail.**
  NVIDIA reports NVFP4 holding within 1% of FP8 accuracy at roughly 25% of FP16 memory, and up to 1.59× training throughput against BF16 over trillion-token runs. [third-party: NVIDIA]

- **The format is free; the conformance is not written.**
  OCP released the MX formats in September 2023, license-free, authored by Microsoft, AMD, Arm, Intel, Meta, NVIDIA and Qualcomm. A free specification does not tell a chip team whether their decoder matches it. [third-party: OCP MX v1.0]

- **The standard that would settle it is not ratified.**
  IEEE P3109 'Arithmetic Formats for Machine Learning' has been an active PAR since February 2023 and is not approved; the working group states its drafts must not be used for conformance claims. [third-party: IEEE SA]

- **A bit-level error is no longer academic.**
  Epoch AI puts frontier training-run cost on a 2.4× per year trend since 2016, crossing a billion dollars per run before 2027. [third-party: Epoch AI]

SOLUTION AND VALUE PROPOSITION

We do not sell the format. We sell the bench behind it.

A number format cannot be sold — MX is license-free by design and IEEE formats are public. Three things are sellable, and only one of them is hard to copy.

- **The implementation — synthesisable decoders and arithmetic cores.**
  RTL that has been through an open flow on real hardware, delivered with the exact synthesis commands that produced the numbers. Yosys 0.65 + nextpnr-xilinx 1743d0f + Icarus 13.0, median of five seeds, DSP inference off, no Vivado in the loop.
- **The conformance suite — bit-exact vectors per format, plus second-oracle checks.**
  This is the part a chip team cannot download and cannot cheaply build. Copying a format takes an afternoon; reproducing a corpus a chip team will trust does not.
- **The theory that says which formats can exist at all.**
  52 theorems proved in the paper, machine-checked in Coq where stated. T25 fixes the three-symbol weight alphabet by uniqueness; T27 enumerates the multiply-free scales by companion-matrix sparsity. We can state what cannot exist — an empirical radix cannot.

Theorem 26 — a property of the algebra, not of a bit pattern

In Z[ϕ] the whole linear path carries no rounding error at any fan-in and any depth — machine-checked in Coq. The bound we state with it: this is a lattice property, not an edge unique

TRACTION

What exists before the money

Every figure below comes from an open toolchain anyone can install, run by the founder on the part named. Areas are LUT counts from that flow. Frequencies are not claimed — see the box.

---

66 LUT

GFTernary decoder, isolated, on XC7A200T; the bare-wire harness on the same part is 112 LUT

MEASURED

38×

fewer LUT — our TNF(4,8) adder is 397 against tekum8's 15,251, and TNF is the wider format

MEASURED

0

rounding error on the whole linear path in Z[ϕ], any fan-in, any depth (Theorem 26)

PROVED

28 LUT

per weight at fan-in 8, zero DSP at any fan-in; rises to 33.2 at fan-in 32

MEASURED

a tie

accuracy at equal stored width — TNF(4,8) 5.32e-3 vs takum16 5.70e-3; TNF(4,24) 1.20e-7 vs takum32 1.26e-7

MEASURED

5 / 9

ladder rungs standing in hardware; TNF64 = 7479 LUT

MEASURED

52 / 83

theorems proved in the paper (Coq where stated) / formats catalogued

PROVED

2

public preprints, arXiv 2606.05017 and 2606.09686; the theorems are in a third paper, not yet on arXiv

SPEC

Not claimed here

No silicon exists — every figure was routed on a binary FPGA (ALINX AX7203, XC7A200T) on an open flow; an ASIC will differ. No ternary fabric was available.
No timing closure. Frequencies here are nextpnr unconstrained critical-path estimates — the logs read PASS at 12.00 MHz — so no clock rate is claimed.
No energy number — energy was not measured. No accuracy lead over takum: an earlier 2.1×/2.6× figure is withdrawn, its oracle negated wrongly on negative codes and the budget was nominal rather than equal stored width. This deck was corrected on 20 August 2026, after the application was filed.

COMPETITION

Named, reproduced, and credited — not paraphrased

Every competing result on t27.ai is re-run on our bench and published with its origin. That is the same discipline we sell.

Who	What they hold	Where our line falls
MX / MXFP4 (OCP)	Released 2023 by 7 vendors, license-free — released, not ratified; hardware default below 8 bits	License-free spec with no conformance story. We supply the bench, not a rival format.
IEEE P3109	The standard-in-waiting for ML arithmetic	Active PAR since Feb 2023, not approved; drafts explicitly unusable for conformance.
takum (Hunhold)	Tapered logarithmic format, strong near unity	Accuracy at equal stored width is a three-way tie — TNF 5.32e-3 / takum16 5.70e-3 / tekum10 8.56e-3. No format wins by more than the width it gives up.
tekum (Hunhold)	Balanced-ternary tapered arithmetic, counted in trits	The one measured asymmetry, and it is hardware: our full adder 397 LUT against tekum8's 15,251 — 38×, with TNF the wider format. Decode 1 LUT vs 542.
posit (Gustafson / Calligo)	Established tapered alternative to IEEE	Compared on area on the same bench. No throughput ratio is claimed — the frequency column is unconstrained.
AetherFloat (Feb 2026)	Escapes the AMAX block-scaling penalty — our ground	Their quad-radix is proposed; ours is enumerated — T27 lists every multiply-free scale. Same outcome, different warrant.
FLoPS / ARCH HDL (2026)	Formalise a given encoding exhaustively (Lean, SMT)	They verify a bit pattern. T26 is a property of the algebra — no input enumeration applies.

MARKET

Counted from the bottom, in the market a core is actually sold into

Not the AI-chip market — agencies disagree about its 2025 base by a factor of two, and a number nobody can decompose is not evidence.

$9.8B

the semiconductor IP market, 2025 — licences, royalties and services; the top five hold 62.2%

THIRD-PARTY

~$1.2M

a proxy for deal size — Ceva's 54 agreements signed against $63.6M recognised in 2025; our quotient, not Ceva's figure

DERIVED

~5% / ~10%

royalty rates Arm itself discloses for CPU IP and for CSS

THIRD-PARTY

97.9%

Arm gross margin, Q4 FY2026 — why this model is worth building

THIRD-PARTY

- **Why the buyer is in Abu Dhabi and not only in the Valley.**
  The buyer of an arithmetic core is anyone taping out AI silicon or specifying an accelerator. Abu Dhabi is assembling exactly that stack — Core42, TII, MBZUAI, AI71 and 42 Abu Dhabi are Hub71+ AI partners, and each of them either designs or specifies compute.
- **Sovereign compute needs an auditable numeric layer.**
  A national AI programme cannot verify a vendor's low-precision datapath against a standard that is not ratified. A conformance bench is the missing instrument, and it is jurisdiction-neutral by construction.

BUSINESS MODEL

IP licensing, priced the way the IP market already prices

Three revenue layers, all recurring-capable, all sold into design teams rather than end users.

Layer	What the customer gets	Pricing	Cycle
Core licence	Synthesisable decoder / arithmetic core, RTL + synthesis commands	per engagement; market proxy ~$1.2M	6–12 months
Conformance suite	Bit-exact vectors, second-oracle checks, seeds and flow	annual subscription	recurring
Royalty	Per-unit on shipped silicon carrying the core	to be set; Arm discloses ~5% for whole-CPU IP as a comparable	life of product
Measurement service	Customer sends RTL, gets it measured on the same bench and seeds	engagement fee	30–60 days

- **Why the bench is the moat, and the patents are not.**
  Nothing is patented and nothing is filed — preprints and open RTL create priority of publication, not exclusivity, and publishing first forecloses a patent on the same disclosure. This is a deliberate trade and an investor should price it as one. The defensible asset is the corpus a chip team will trust.

FOUNDING TEAM

One engineer, one bench, and a public record of what did not work

Every number on the public site was measured by the person named here, on hardware named here, with a toolchain anyone can install.

Dmitrii Vasilev — Founder

FPGA/RTL and hardware-AI engineer · author of the GF-T ternary format

Owns the whole line himself: specification → RTL → open synthesis flow → measurement → paper. Author of two public preprints (arXiv 2606.05017, 2606.09686), a catalogue of 83 numeric formats, and a third paper carrying 52 theorems, machine-checked in Coq where stated. Built and runs the measurement bench the business is sold on — three ALINX AX7203 boards, an entirely open toolchain, five-seed medians, and a published record of the claims he has withdrawn. · t27.ai · github.com/gHashTag · linkedin.com/in/neurocoder

ORCID 0009-0008-4294-6159

- **Founder–market fit, stated as a check rather than a claim.**
  The rare part is not the maths and not the RTL — it is one person carrying a claim from a theorem to a placed-and-routed number and publishing the retractions. Two are on record: a 323 MHz figure withdrawn on t27.ai, and a 3× LUT-counting error corrected in a merged public PR. This deck itself was corrected after submission on the same rule — the withdrawn items are listed on the traction slide.
- **First hires with Hub71 capital: a verification engineer and an ASIC-track RTL engineer.**
  Both in Abu Dhabi. The bottleneck today is not ideas; it is one pair of hands against nine ladder rungs.

FUNDRAISING

Previous: none. Next: $3M, released against gates.

100% founder-held today, no prior round, no institutional investors, no revenue and no signed licensee. The plan below is what money changes; the record on the previous pages is what exists without it.

$0
raised to date — bootstrapped, 100% founder-held, no prior round

SPEC

$3M
round size sought; the asking terms are published at t27.ai/#invest and are open to discussion

PLAN

18 mo
runway planned, released in three milestone-gated tranches

PLAN

60/25/15
engineering / verification / operations split of the round

PLAN

Tranche	Gate — a check anyone can run
Months 1–6	Compute axis of the catalogue closed as far as the open FPGA flow allows, vectors published alongside, and a head-to-head against takum RTL on one flow — which needs a VHDL front end this bench does not have.
Months 7–12	Paper through peer review, and a first external design partner running the cores on their own hardware — an evaluation agreement, not a mention.
Months 13–18	Arithmetic cores taped out, so the numbers stop being FPGA numbers, and a first paid licence of a core that has been through the flow.

Hardware money is released against milestones rather than months. A gate that does not close is a reason not to release the next tranche.

PLANS FOR ABU DHABI AND HUB71

Why this company should be built here

The first three months, and what each of them is checked by.

- **Months 1–3 — relocate, incorporate in ADGM, and move the bench.**
  The founder relocates to Abu Dhabi, the company incorporates under the ADGM Tech Startup licence, and the measurement bench — boards, flow and seeds — is re-established there so every published number afterwards is measured in Abu Dhabi.
- **Months 1–3 — convert two Hub71+ AI partners into design-partner conversations.**
  Core42, TII, MBZUAI and AI71 each design or specify compute. The ask is narrow and testable: give us one low-precision datapath to check against our conformance vectors, on your hardware. That is a two-week engagement, not a procurement cycle.
- **Months 1–3 — hire the verification engineer locally.**
  Out of 42 Abu Dhabi and MBZUAI. The single-founder bottleneck is verification throughput, and the region has the pipeline for it.
- **Why the region and not only the Valley.**
  Abu Dhabi is standing up sovereign AI compute, and sovereign compute cannot verify a vendor's low-precision datapath against a standard that is not ratified. A jurisdiction-neutral conformance bench is a strategically useful thing for the UAE to own early — and it is a semiconductor-IP asset, which is exactly the layer the national stack does not yet have.

THREE THINGS TO KEEP

01 Precision became a hardware variable, and the standard that would settle it is not ratified.

The gap is not another format. It is the proof that a format behaves the same in silicon and in the model.

02 The sellable asset is the conformance bench, not the format.

A market where a comparable licence averages about $1.2M (our quotient from Ceva's disclosures) and royalties run at rates Arm publishes for CPU IP.

03 The record is checkable today, and everything it does not cover is written down.

Open toolchain, published seeds, published retractions. No silicon, no energy number, no ternary fabric — stated on the public site, not discovered in diligence.

Dmitrii Vasilev · admin@t27.ai · t27.ai

Every measured number, its toolchain version and its limits are published at t27.ai — including the claims withdrawn.