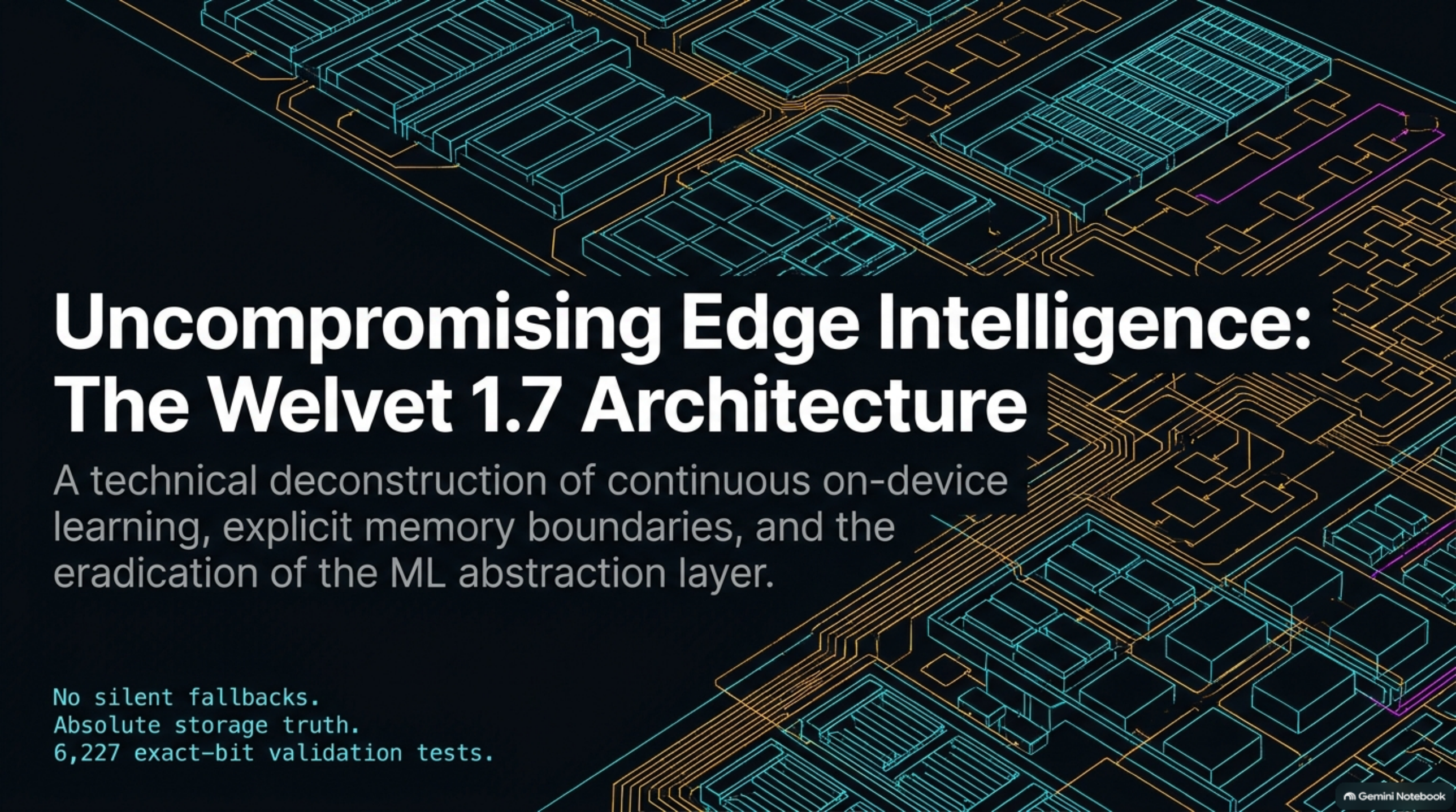


Uncompromising Edge Intelligence: The Welvet 1.7 Architecture

A technical deconstruction of continuous on-device learning, explicit memory boundaries, and the eradication of the ML abstraction layer.

```
No silent fallbacks.
Absolute storage truth.
6,227 exact-bit validation tests.
```

Gemini Notebook

Forged in the Crucible of Embodied Artificial Life





The Constraint:

Autonomous helpers required real-time, on-device observation and physical mimicry on constrained consumer hardware.

LATENCY: <5ms
COMPUTE: EDGE ONLY
INPUT: MULTIMODAL SENSOR STREAM

The Result:

The ALife imperative birthed an engine where intelligence grows natively alongside the user, rather than executing a pre-trained artifact.

ADAPTATION: ON-DEVICE
GROWTH: CONTINUOUS
ARTIFACT: NATIVE EVOLUTION

The Requirement:

Continuous, asynchronous synaptic weight updates without exhausting the VRAM of standard iPads or desktops.

VRAM FOOTPRINT: <2GB
UPDATE FREQ: ASYNC
MEMORY: DYNAMIC ALLOCATION

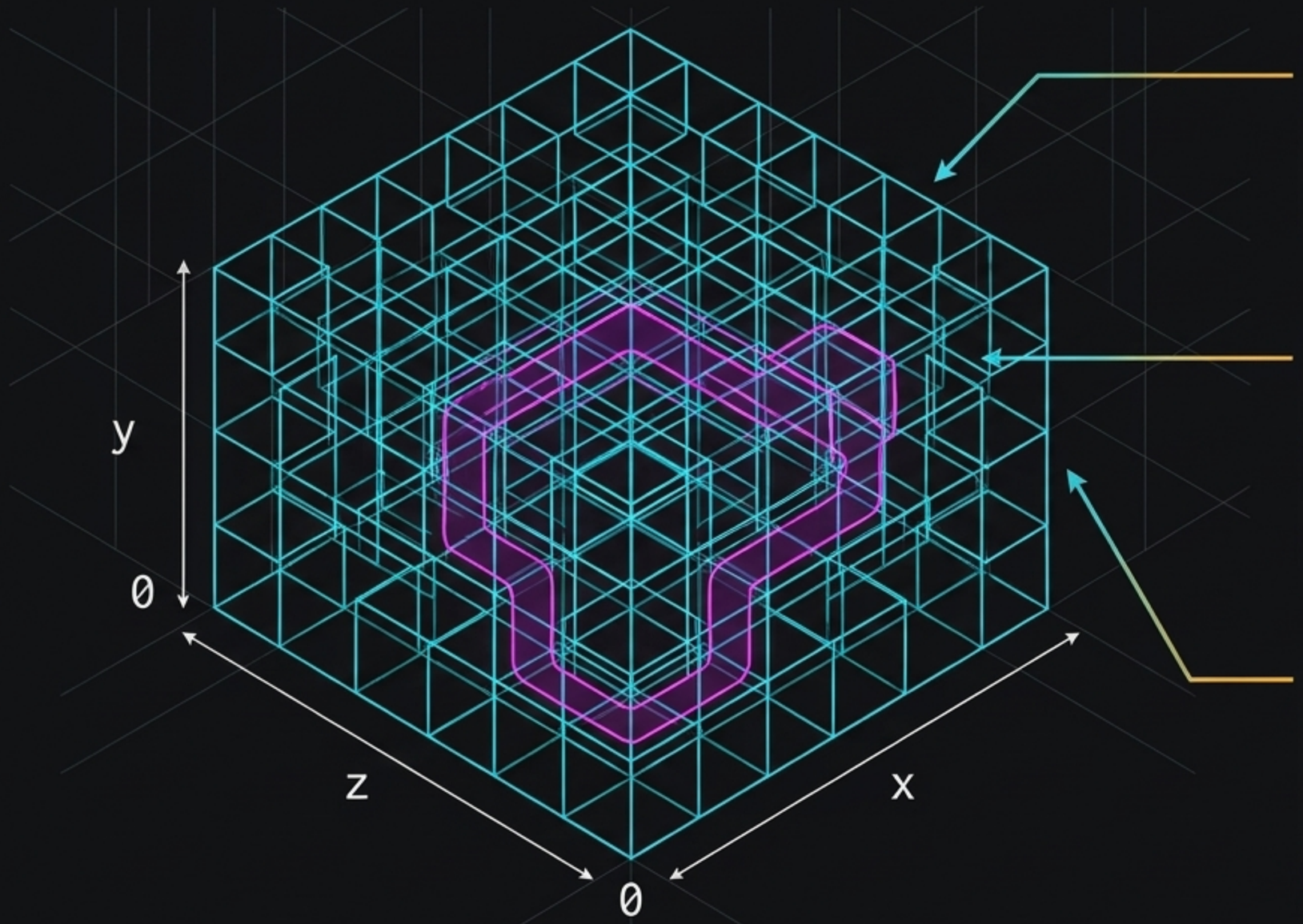
Gemini Notebook

The Rejection of Industry Standard Abstractions

	Standard ML Paradigm	Welvet's Edge-Native Continuous Learning
Topology	Sequential Linear Stacks & Directed Acyclic Graphs (DAGs)	Spatial Volumetric Grids & Remote Cyclic Linkages
Quantization	Simulated QAT (Requires FP32 master tensor)	FormatNone Absolute Storage Truth (Zero master tensor)
Hardware Dispatch	Abstracted execution with silent fallbacks to CPU	Strict explicit routing with hard terminal errors on missing paths
Validation	Simulated Jupyter Notebooks (.ipynb)	Compiled C++ Native Parity Suites (6,227 strict leaf checks)

Gemini Notebook

Pillar 1: Deconstructing the Directed Acyclic Graph



Coordinate Space

Topology is strictly defined by Depth, Rows, Columns, and LayersPerCell.

Biological Mimicry

Non-sequential routing allows a deep perceptual layer to instantly loop back to a shallow motor-control layer.

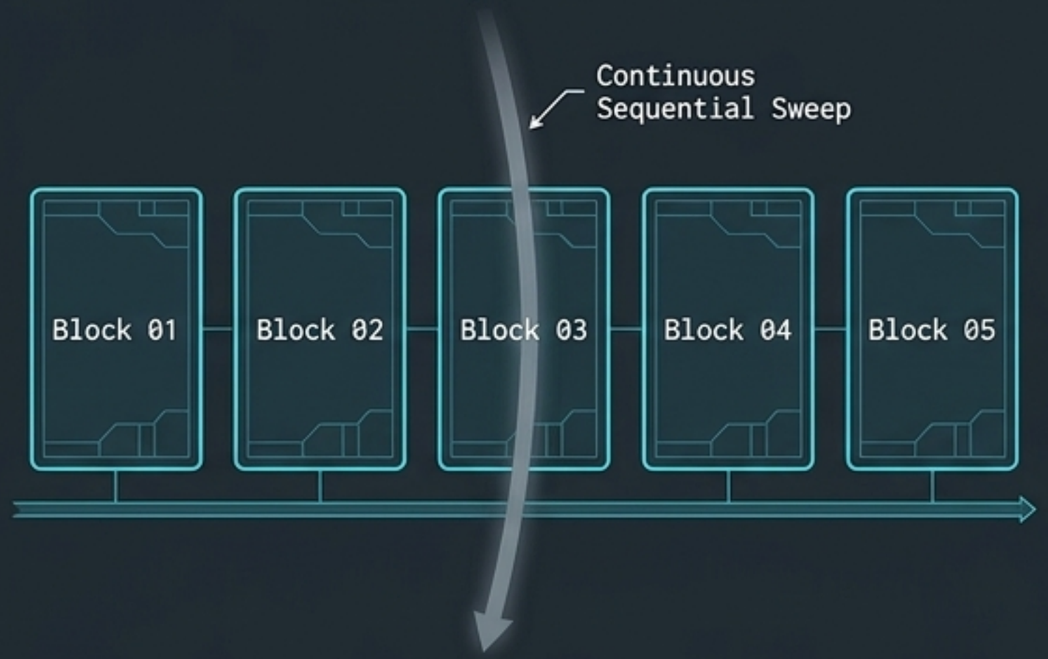
Hardware Mapping

Standard C++ configuration keys map grid sectors directly to underlying hardware memory while preserving remote spatial links.

Gemini Notebook

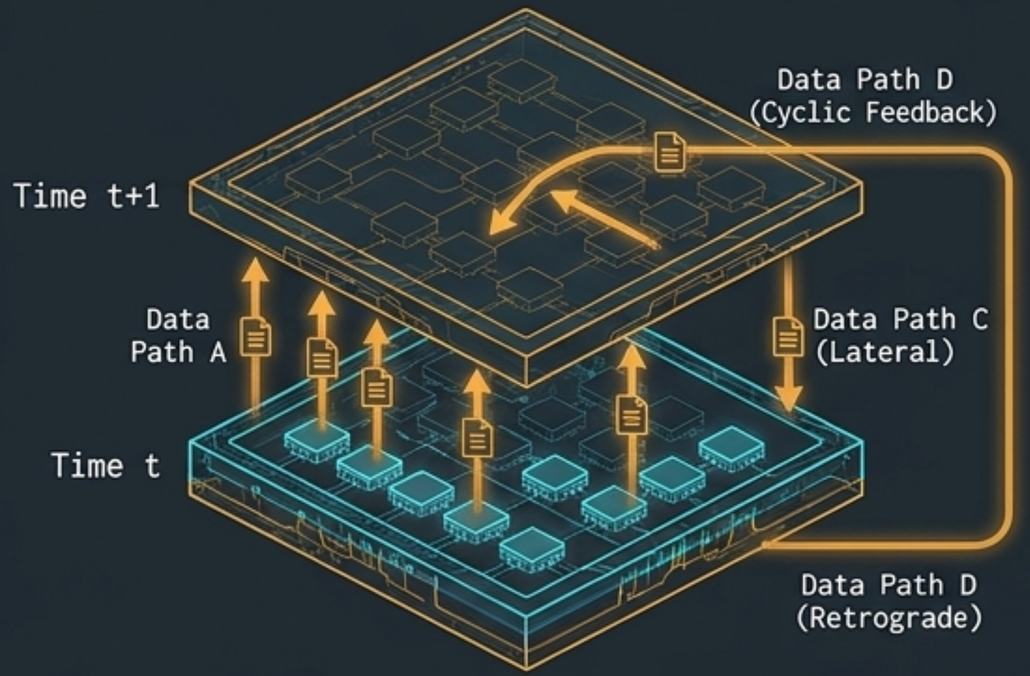
The Volumetric Step Mesh: Resolving Cyclic Instability

A Traditional Autoregressive



// Standard Linear Processing Path

B Welvet Step Mesh



// Parallel Asynchronous Updates with Double-Buffering

Double-Buffering

Explicit separation of current state (t) from calculated future state (t+1) prevents race conditions.

Asynchronous Pulsing

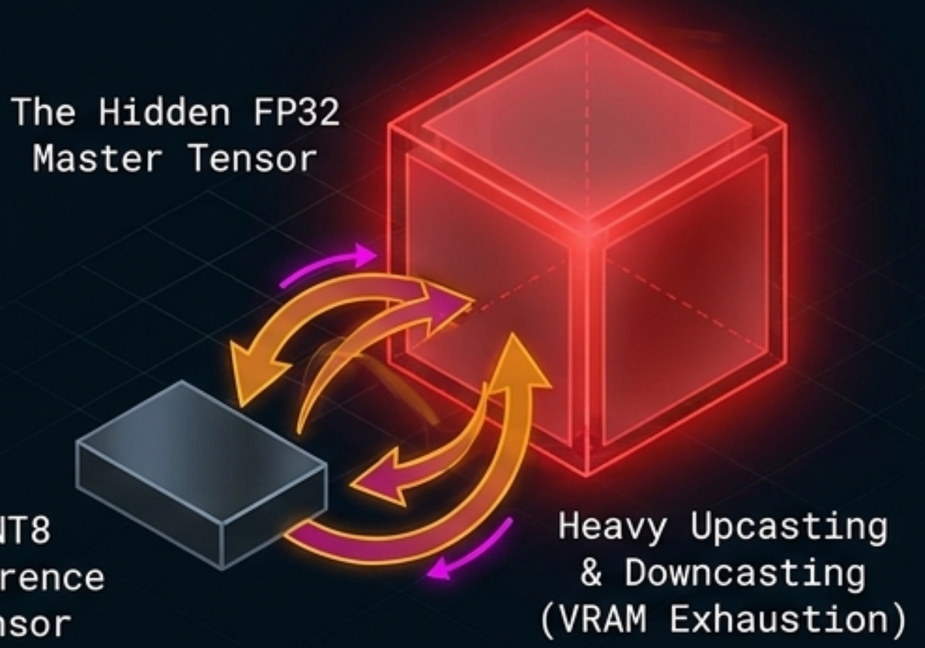
Allows cross-numeric training where sub-networks operate at different precision levels and update intervals without catastrophic numeric instability.

Gemini Notebook

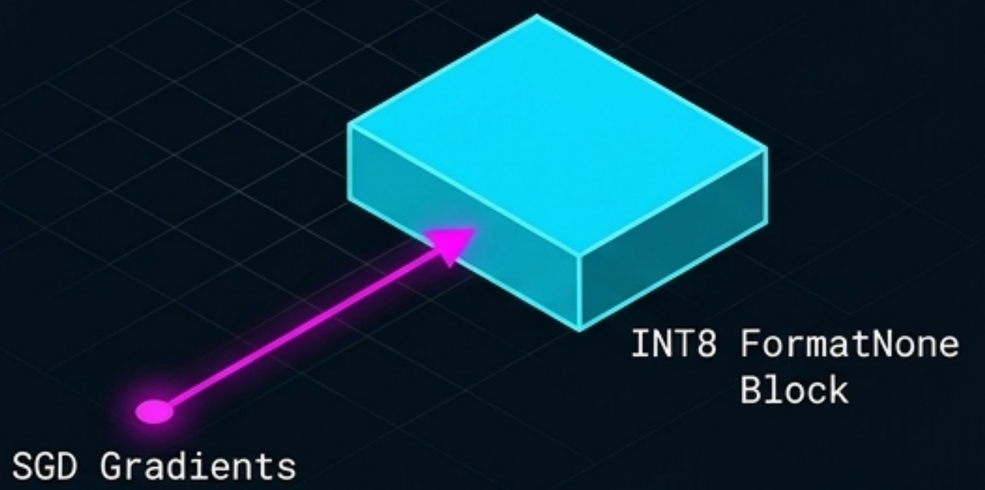
Pillar 2: The Hidden Bottleneck of QAT

The VRAM Schematic

Standard QAT



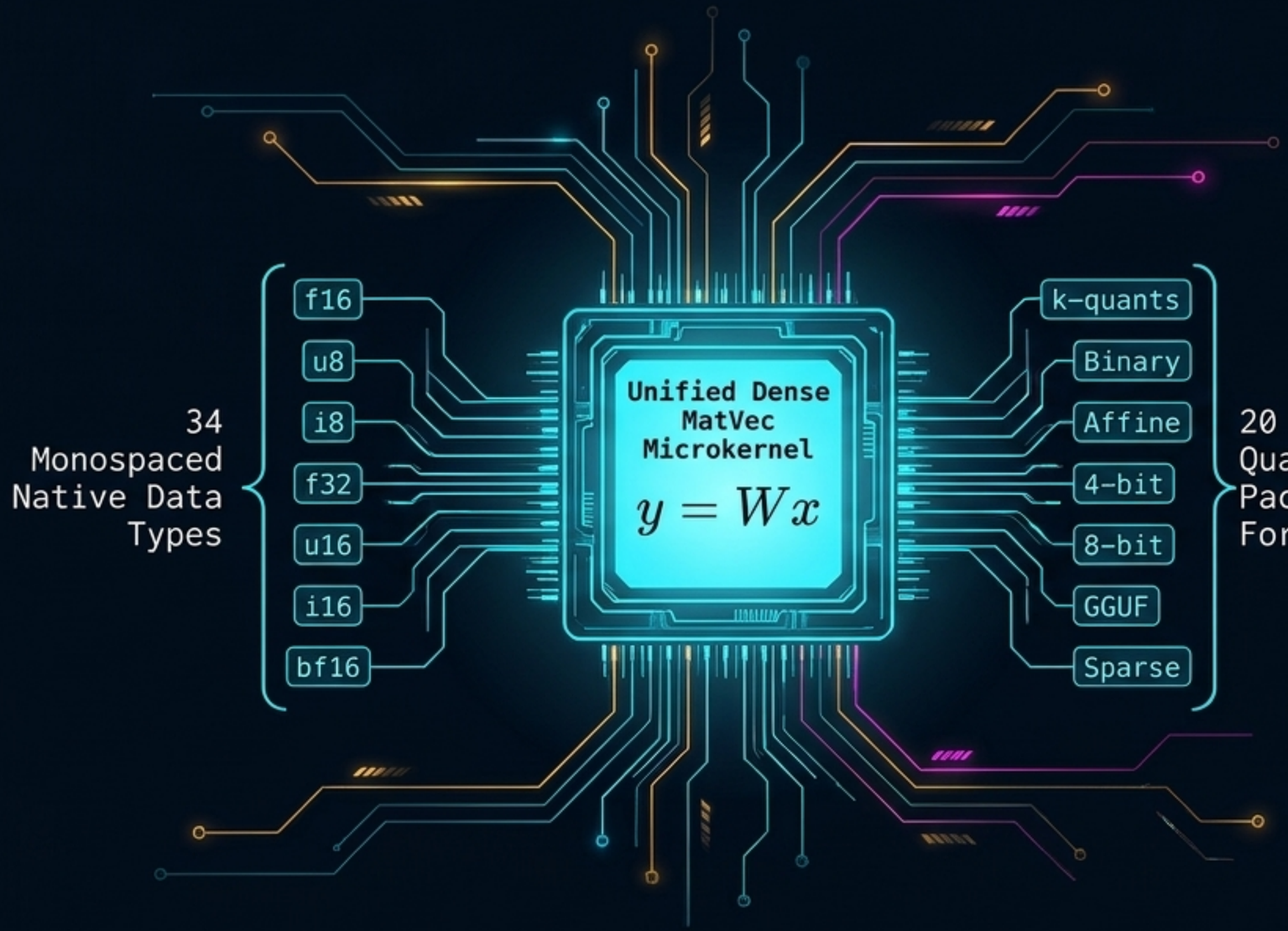
Welvet Storage Truth



Edge training is impossible when backpropagation secretly demands a full 32-bit floating-point master copy in the background.

Gemini Notebook

FormatNone: The Principle of Absolute Storage Truth



Storage Truth Rule:

The **stored dtype** is the absolute, final truth. No upcasting layer, no morphing formats during the **training loop**.

Unified Surface Area:

Centralizing all generalized **matrix-vector multiplications** eliminates subtle mathematical drift across hardware.

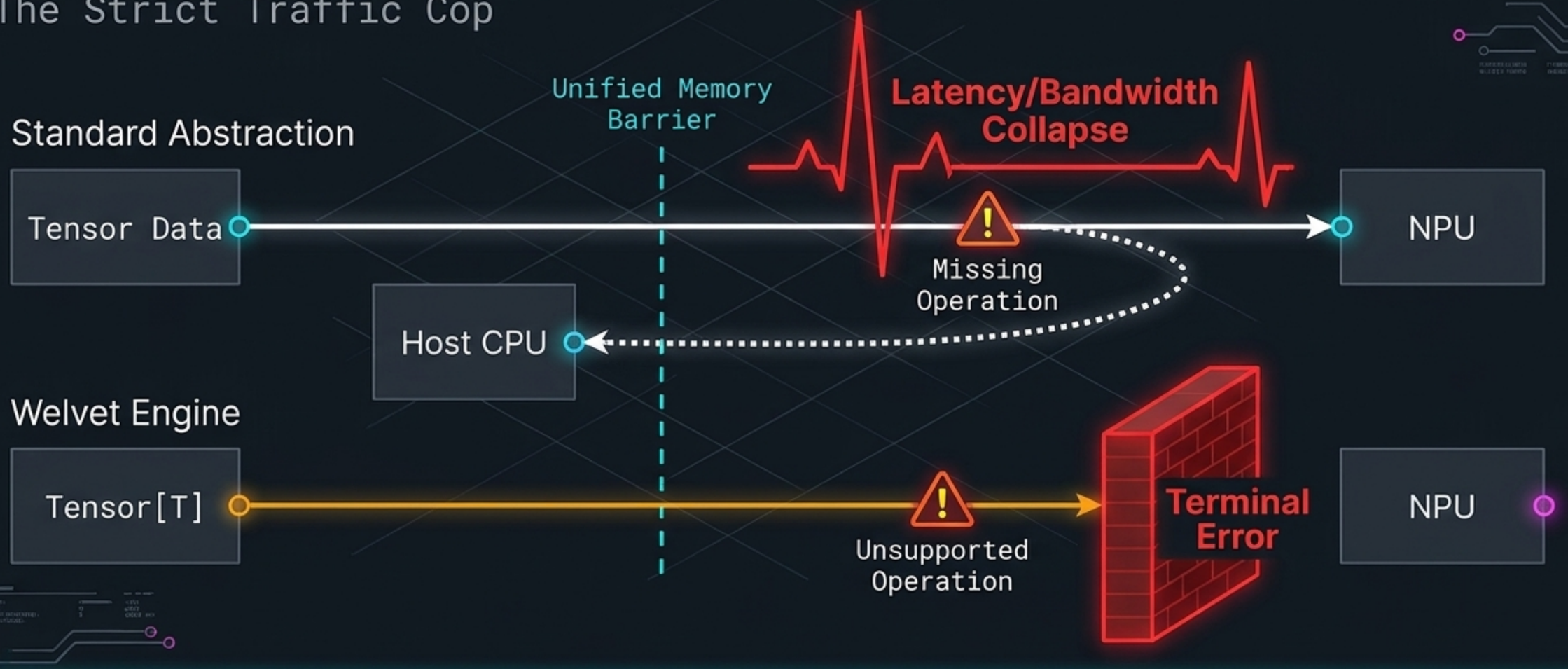
In-Dtype SGD:

Gradients are calculated and applied directly into **low-precision memory boundaries**.

Gemini Notebook

Pillar 3: The Eradication of Silent Hardware Fallbacks

The Strict Traffic Cop



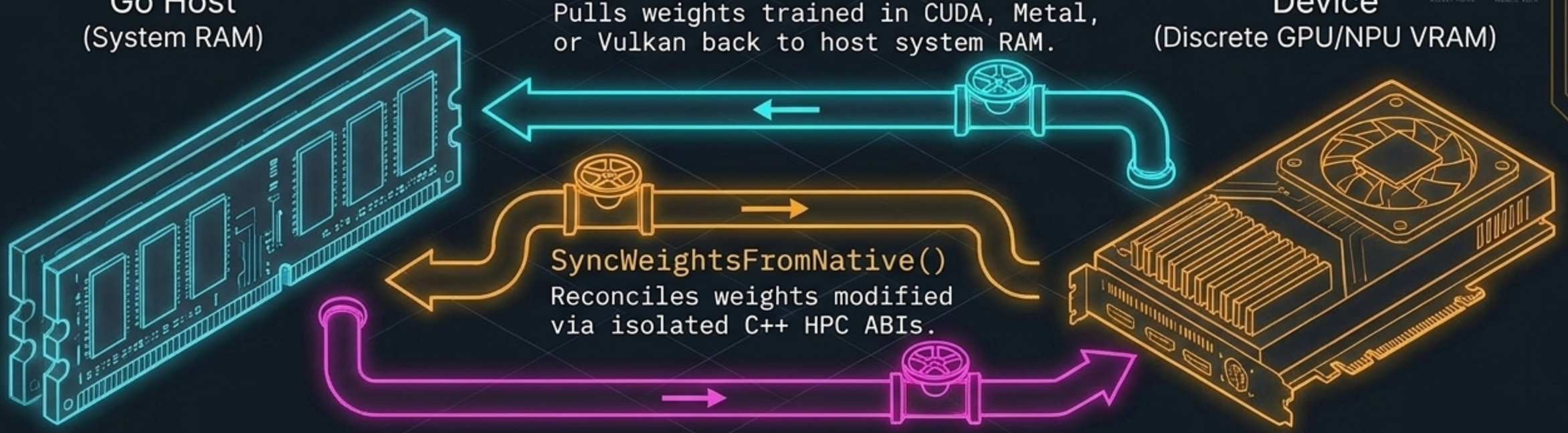
Welvet forces developers to explicitly manage graph layouts. Banning simulated software fallbacks prevents cycle times from silently bleeding back to the Go runtime.

Gemini Notebook

Enforcing Strict State Synchronization

Go Host
(System RAM)

Device
(Discrete GPU/NPU VRAM)



SyncToGo()
Pulls weights trained in CUDA, Metal, or Vulkan back to host system RAM.

SyncWeightsFromNative()
Reconciles weights modified via isolated C++ HPC ABIs.

SyncToMLX()
Pushes data into forward-only MLX execution contexts.

The Rule: Dual-state architecture completely eliminates asynchronous race conditions. Data moves only when explicitly commanded.

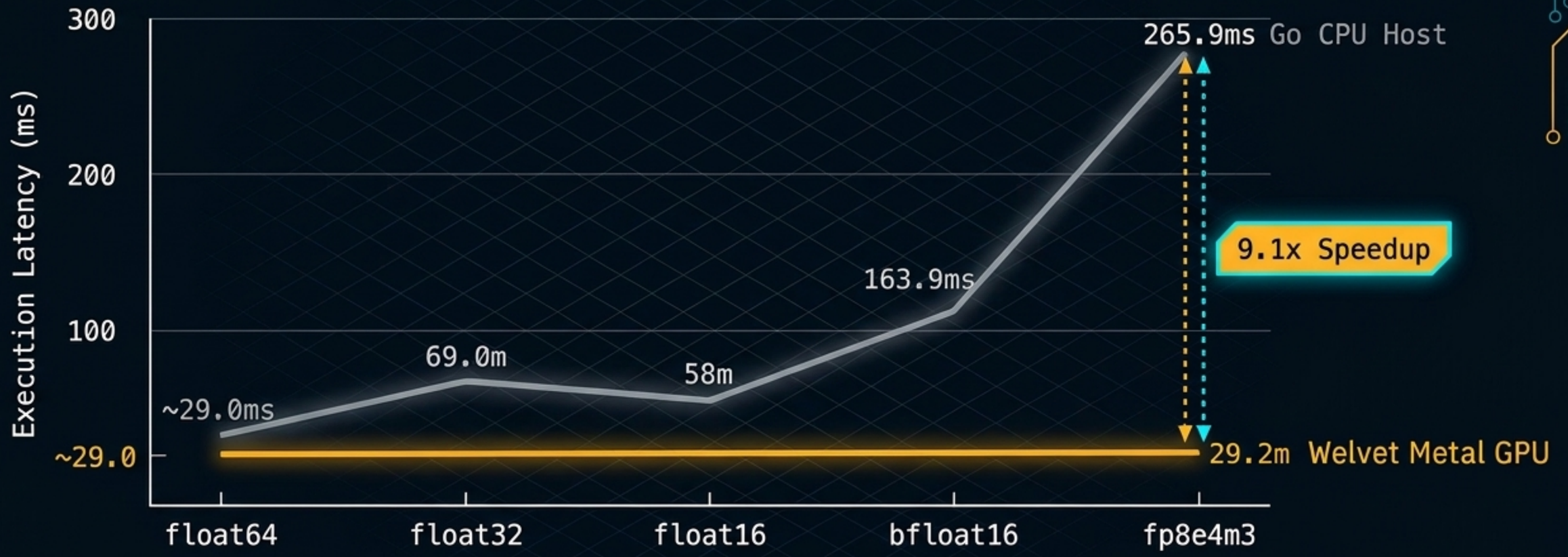
Gemini Notebook

The Explicit Application Binary Interface (ABI) Matrix

Target	ABI	Execution Insight
Snapdragon X (Hexagon)	`nativeqnn / libwelvet_qnn`	Bypasses host CPU entirely; handles highly efficient mixed-precision integer pathways.
Apple Silicon	`nativemetal / welvet.metal`	Saturates compute threshold natively; flat 29ms execution across varied datatypes.
Discrete GPUs	`nativecuda / nativerocm / nativevulkan`	Opt-in user-controlled memory mapping targeting high-end edge environments.
C++ HPC	`BackendNativeCPU / welvet.cpp`	Executes complex NormalBP and TrainMSE natively without Go overhead.

Gemini Notebook

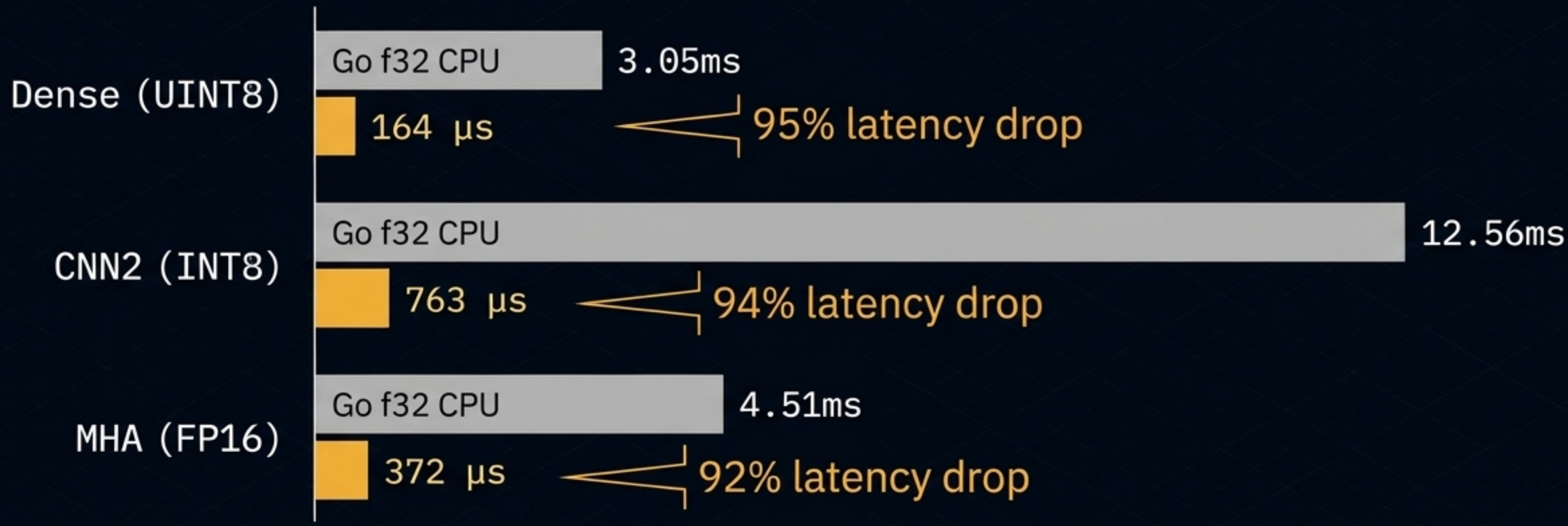
Empirical Proof: Flattening the Metal GPU Latency Curve



The static GPU execution time proves Welvet saturates the compute threshold. The GPU processes native byte types flawlessly without the latency penalty of software upcasting.

Gemini Notebook

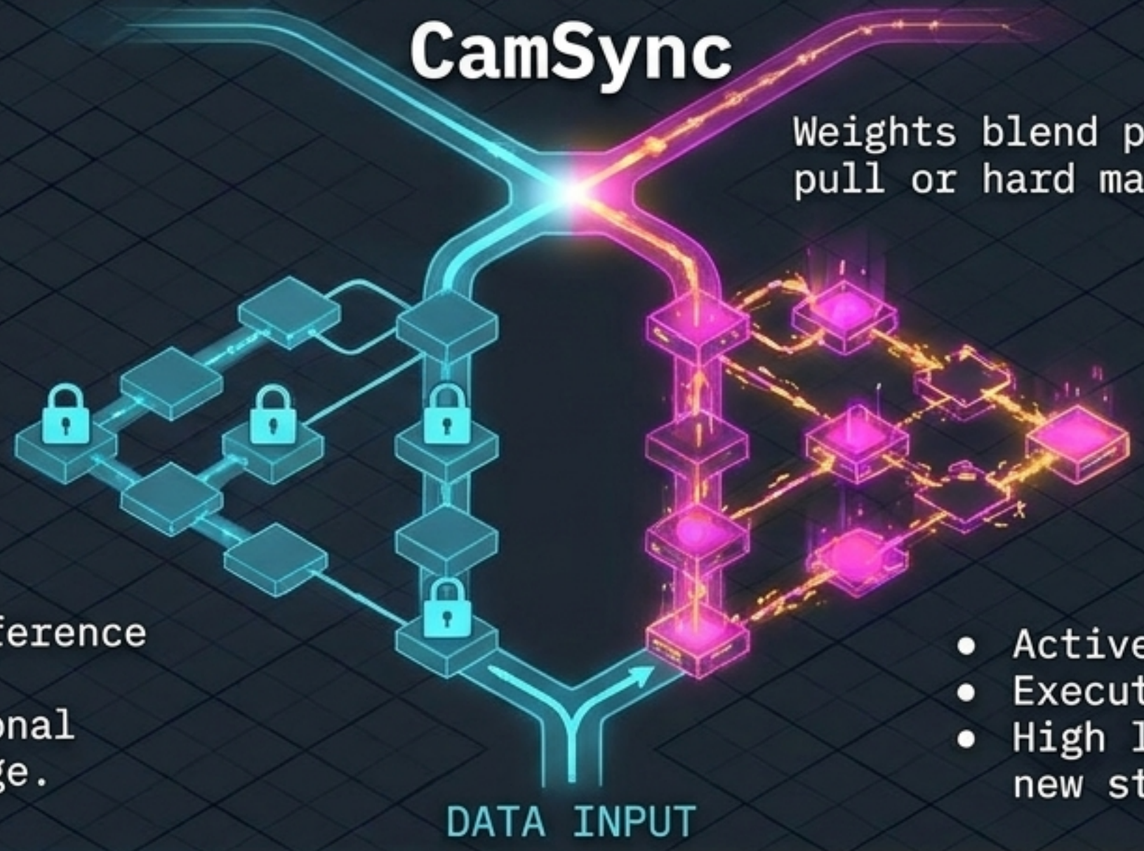
Qualcomm Hexagon: Sub-Millisecond Mixed Precision



Technical Insight: Welvet flawlessly orchestrates mixed-precision execution on the same chip. Processing quantized integers (u8/i8) for dense layers, while seamlessly transitioning to mathematically sensitive 16-bit floating points (f16) for Multi-Head Attention without unified memory bleed.

Gemini Notebook

Synthesis: Bicameral Learning Systems (CamSync)



Foundation Hemisphere

- Locked in 8-bit inference
- Executing on NPU
- Preserves foundational behavioral knowledge.

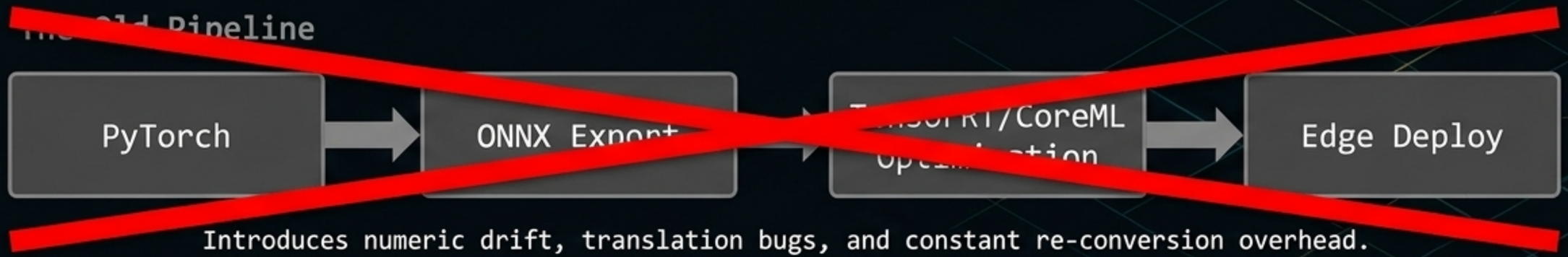
Explorer Hemisphere

- Active TrainMSE mode in FP16
- Executing on discrete GPU
- High learning rate to map new stimuli.

Core Takeaway: Welvet allows sibling hemispheres with completely different expert systems, topologies, and precisions to train dynamically and merge asynchronously, solving catastrophic forgetting on the edge.

Gemini Notebook

The Translation Bypass: Direct .safetensors Ingestion



The LOOM / Welvet Pipeline



Seamless Instantiation:

- Seamlessly instantiates models like **TinyLlama (1.1B)** and **SmolLM2 (1.7B)** via **Python**, **JS/Node**, or **C#/.NET 9.0** by simply passing the repository string.

Persistence:

- Native **.entity** file structure accurately saves topology, **FormatNone** variables, and localized states to disk after edge-training.

Gemini Notebook

Empirical Rigor: The w2a Ecosystem



6,227

Total automated exact-bit leaf test cases.

5,689

Tests dedicated strictly to native C/GPU parity.

29

Distinct spatial training modes modeled and tested.

Architectural claims are not validated by simulated Jupyter notebooks. They are validated by 6,000+ deterministic hardware executions. Welcome to close-to-the-metal intelligence.

Gemini Notebook