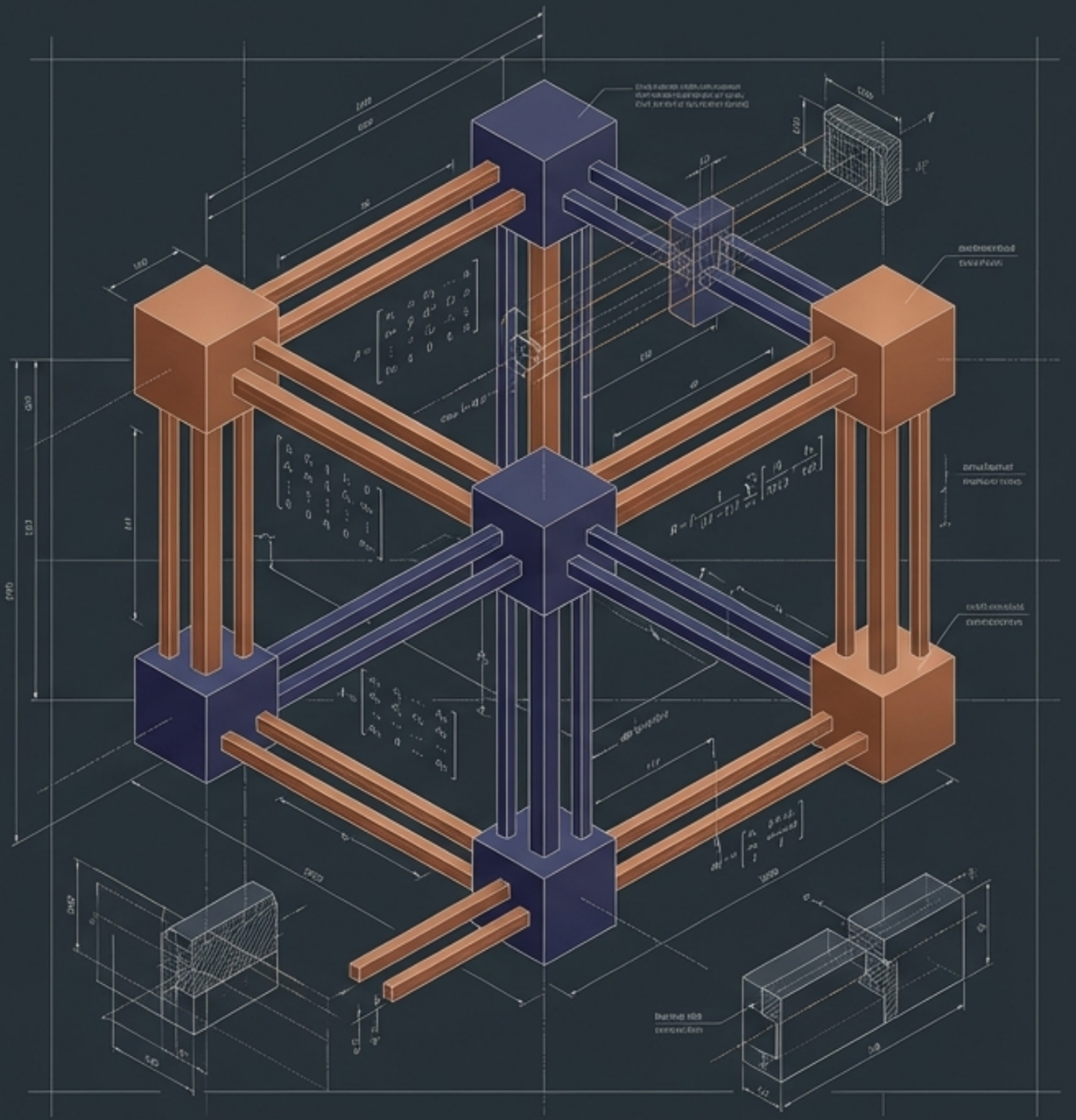
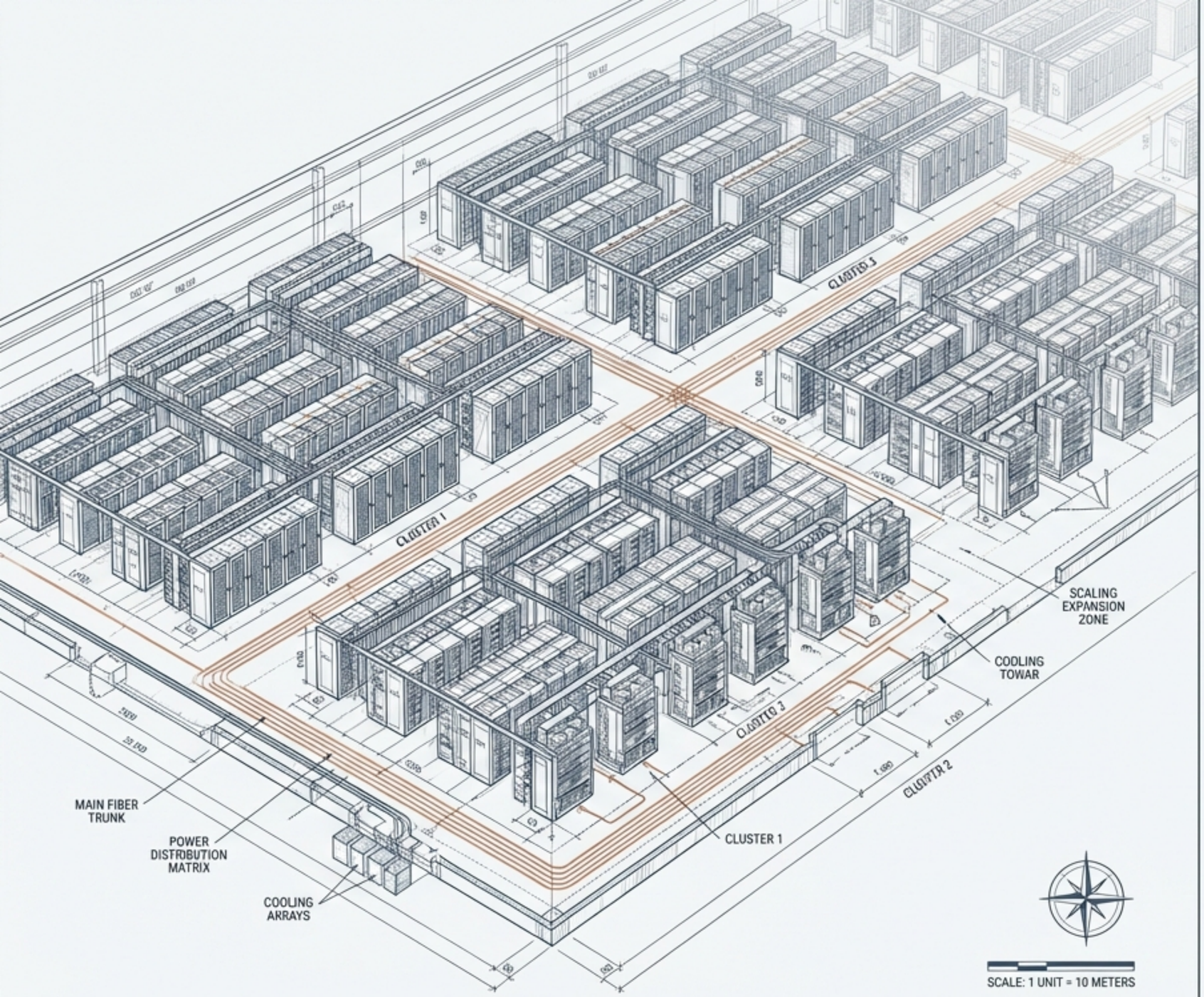
BYPASSING THE INFINIBAND PREMIUM

Deploying Predictive Tensor Control Plane (PTCP) over COTS SONiC Ethernet for 100,000-GPU AI Clusters



Operational Economics and Architectural Integration



The 100,000-GPU Scaling Frontier

As AI models advance, infrastructure demands have crossed a critical threshold. Scaling compute to 100,000-GPU clusters amplifies latent networking inefficiencies into catastrophic financial losses.

Target Scale

100,000 interconnected GPUs.

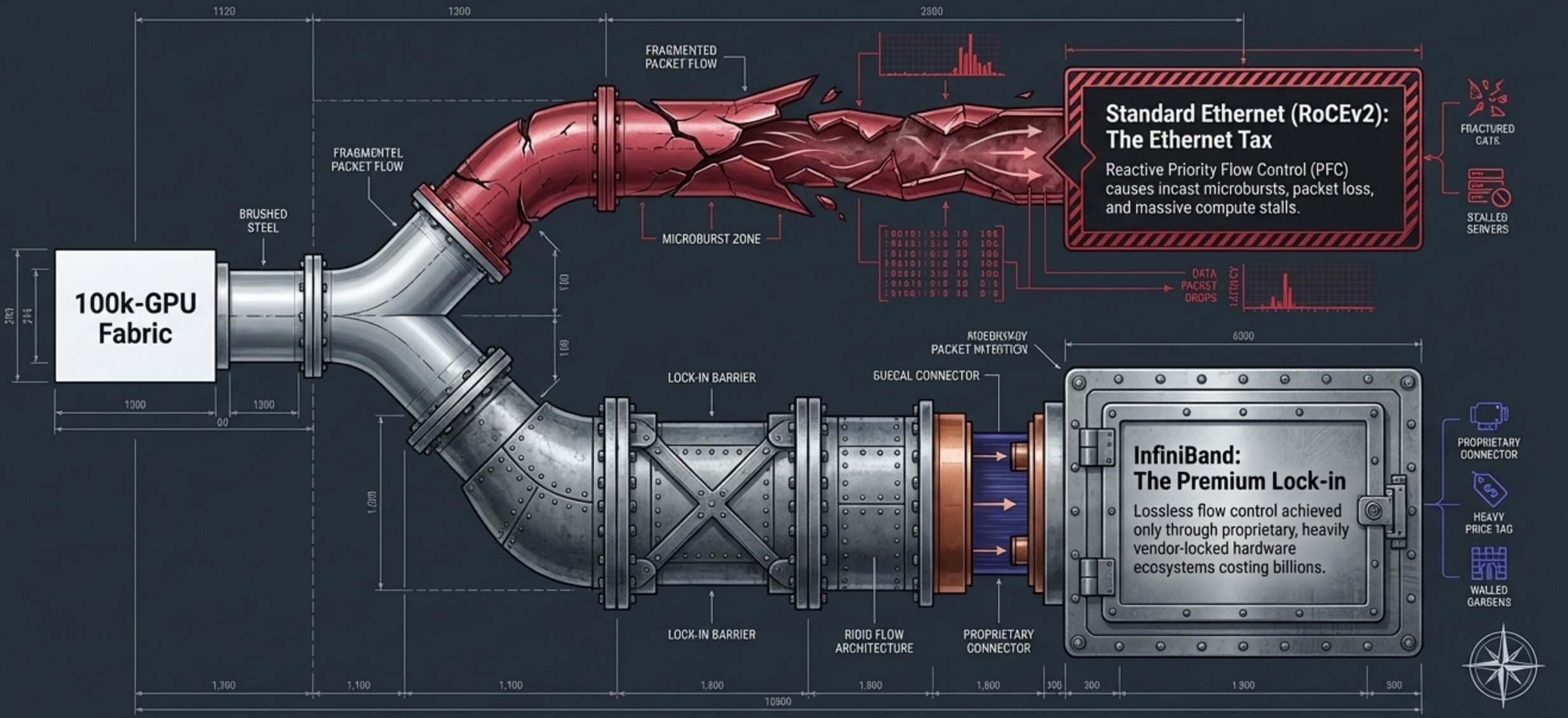
Primary Workload

Synchronized All-Reduce phases.

The Constraint

The network fabric is the absolute bottleneck determining Cluster Model Flops Utilization (MFU).

The Billion-Dollar Fabric Dilemma



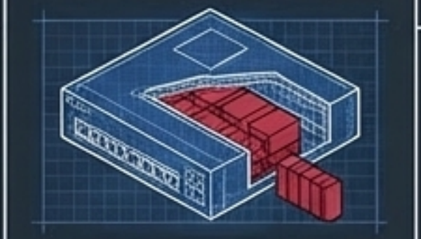
Path A: The Anatomy of an Ethernet Microburst

1. Synchronization

Thousands of GPUs synchronize during an All-Reduce phase, blasting data simultaneously.

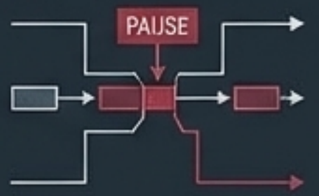
2. Buffer Overflow

Top-of-Rack (ToR) ASIC buffers reach 100% capacity instantly.



3. Reactive Pause

Priority Flow Control (PFC) sends pause frames after congestion occurs.

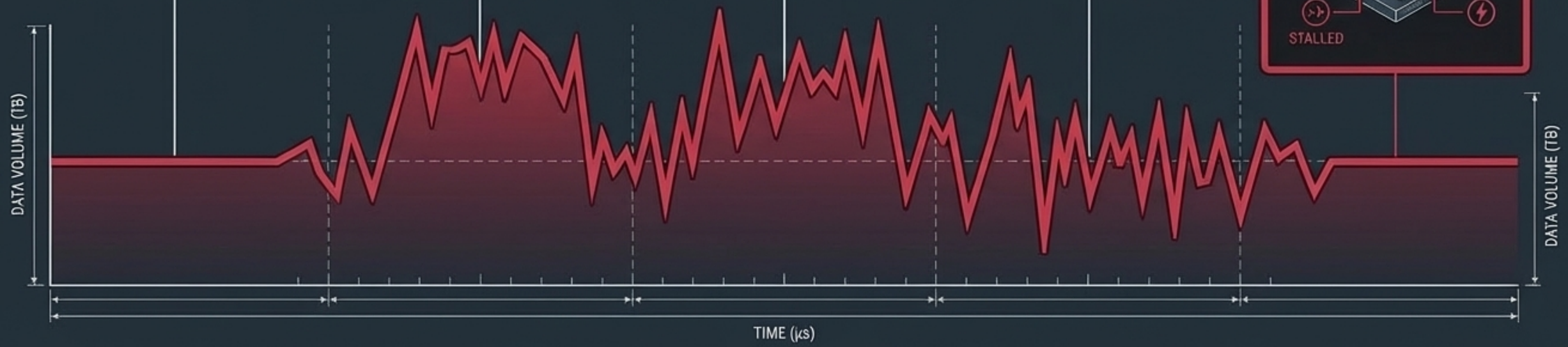


4. Packet Drop

Packets are dropped; TCP timeouts trigger.

5. Compute Stall

Expensive GPUs sit completely idle, drawing immense static power while waiting for data.



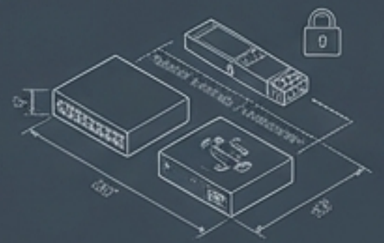
Path B: The Restrictive Vault of InfiniBand

The Credit Mechanism

Utilizes a credit-based flow control mechanism. A switch only transmits if the receiver verifies buffer space. Drops are prevented.

The Hardware Tax

Forces the acquiring entity into a completely proprietary hardware ecosystem (switches, optics, and Host Channel Adapters).

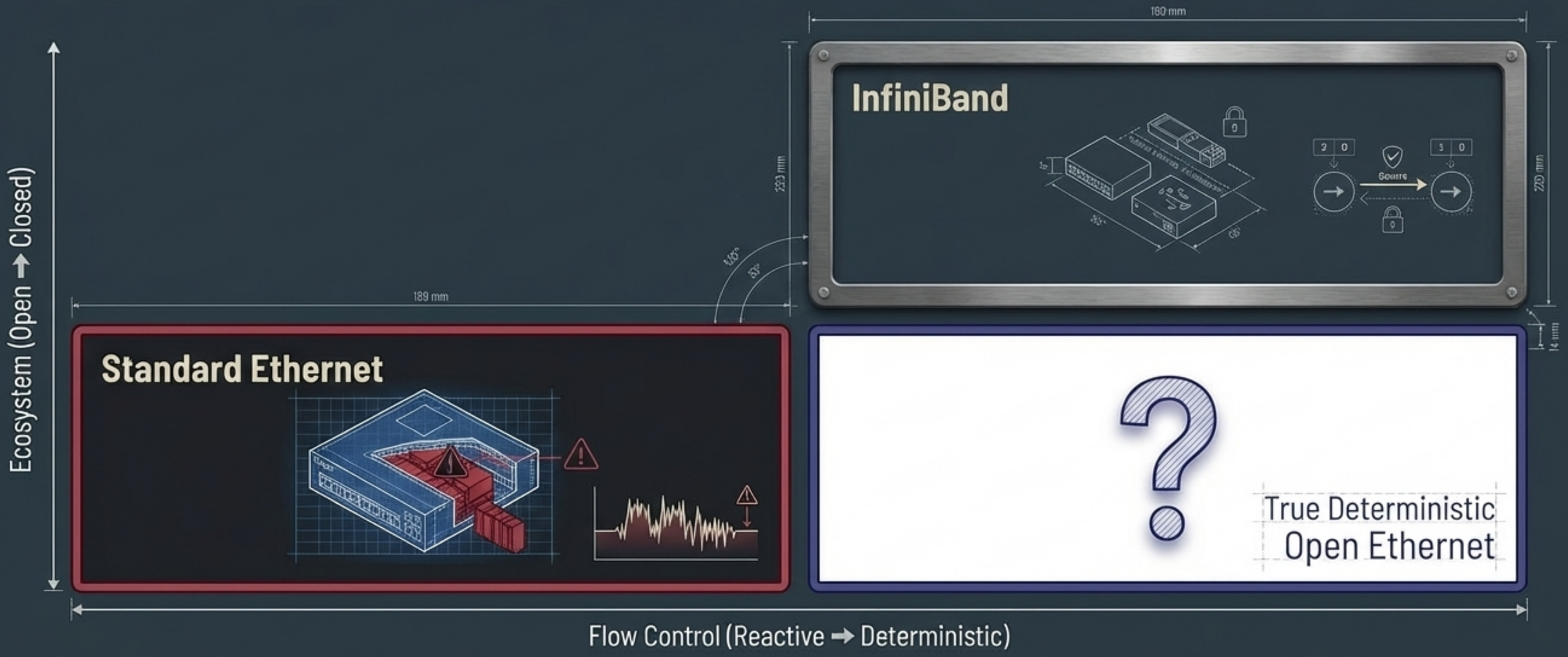


The CapEx Trap

Solves the technical problem by assuming a massive, sunk CapEx burden that scales exponentially with cluster size.

The Architectural Root Cause

Standard Ethernet is reactive; it attempts to fix congestion after the buffer overflows. InfiniBand is restrictive; it prevents overflow but closes the ecosystem. True deterministic, open Ethernet requires pre-empting the bottleneck before it occurs.



Escaping the Trap: Predictive Tensor Control Plane (PTCP)

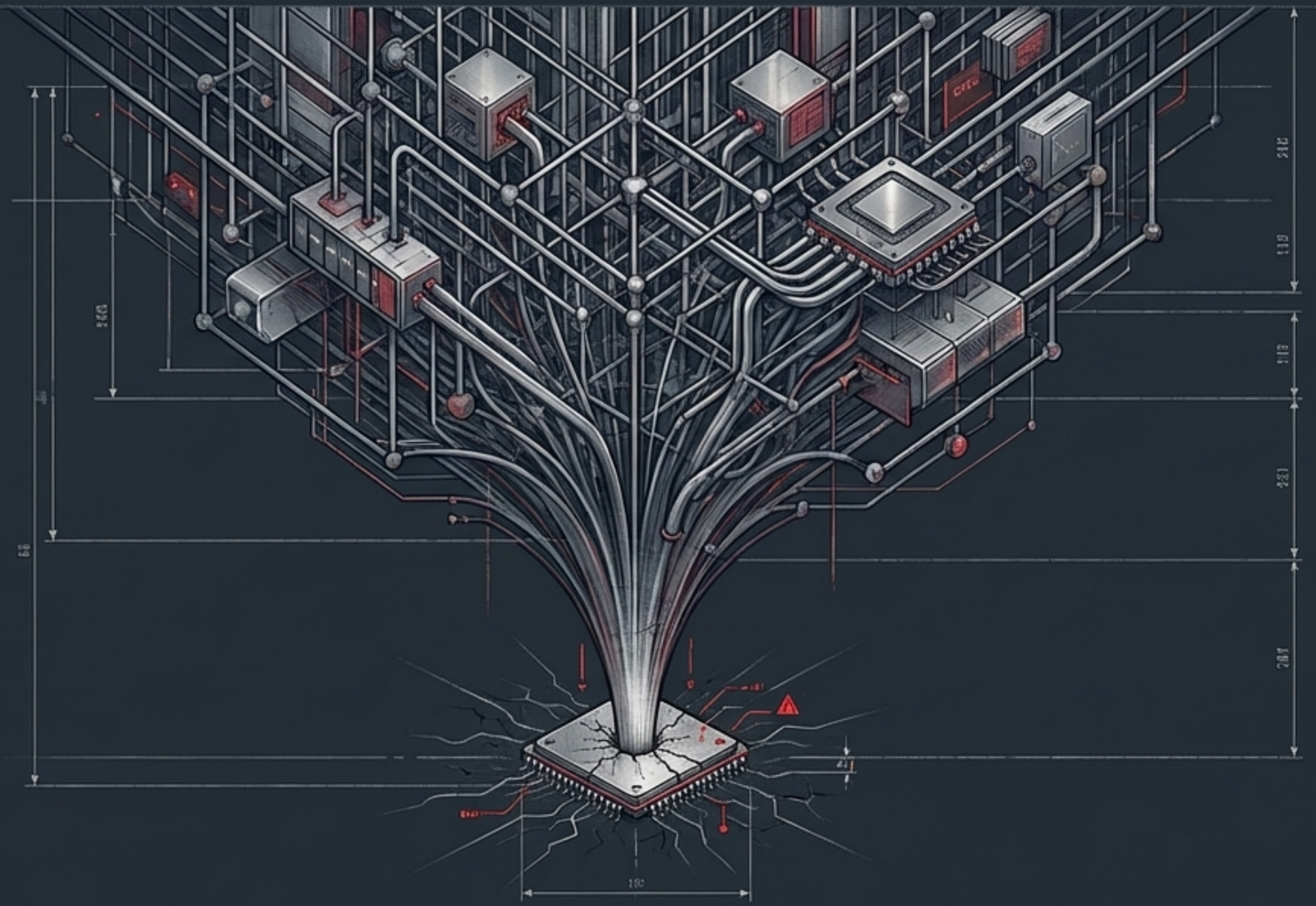
Core Value Proposition

PTCP shatters the binary choice. By deploying a containerized intelligence agent on standard white box switches, hyperscalers achieve InfiniBand-tier deterministic performance over standard commercial silicon.

The Mechanism

Predicting the exact probability of an impending incast and pacing egress queues to perfectly match uplink capacity—mathematically eliminating buffer overflows.

The Barrier: The Curse of Dimensionality



The Problem

Predicting AI network behavior across 100,000 nodes requires analyzing an immense joint probability space.

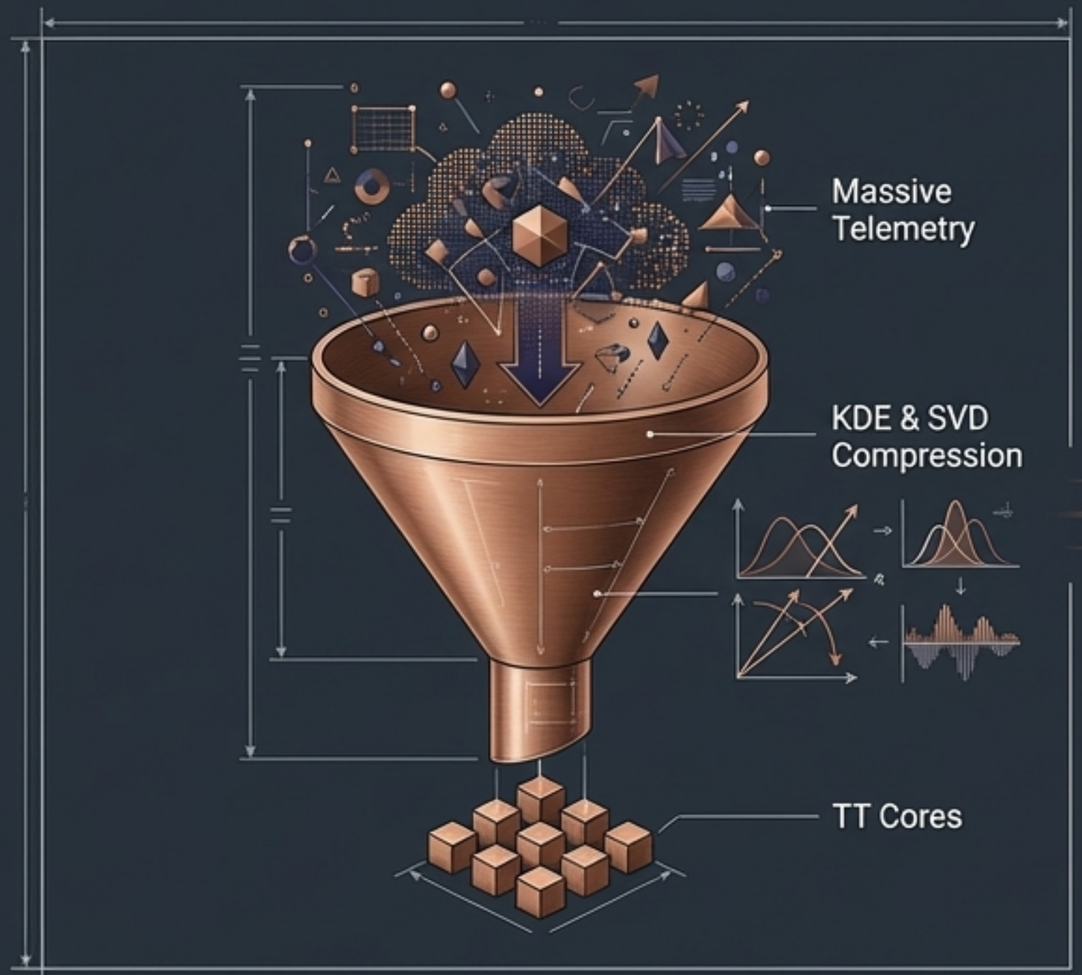
The Variables

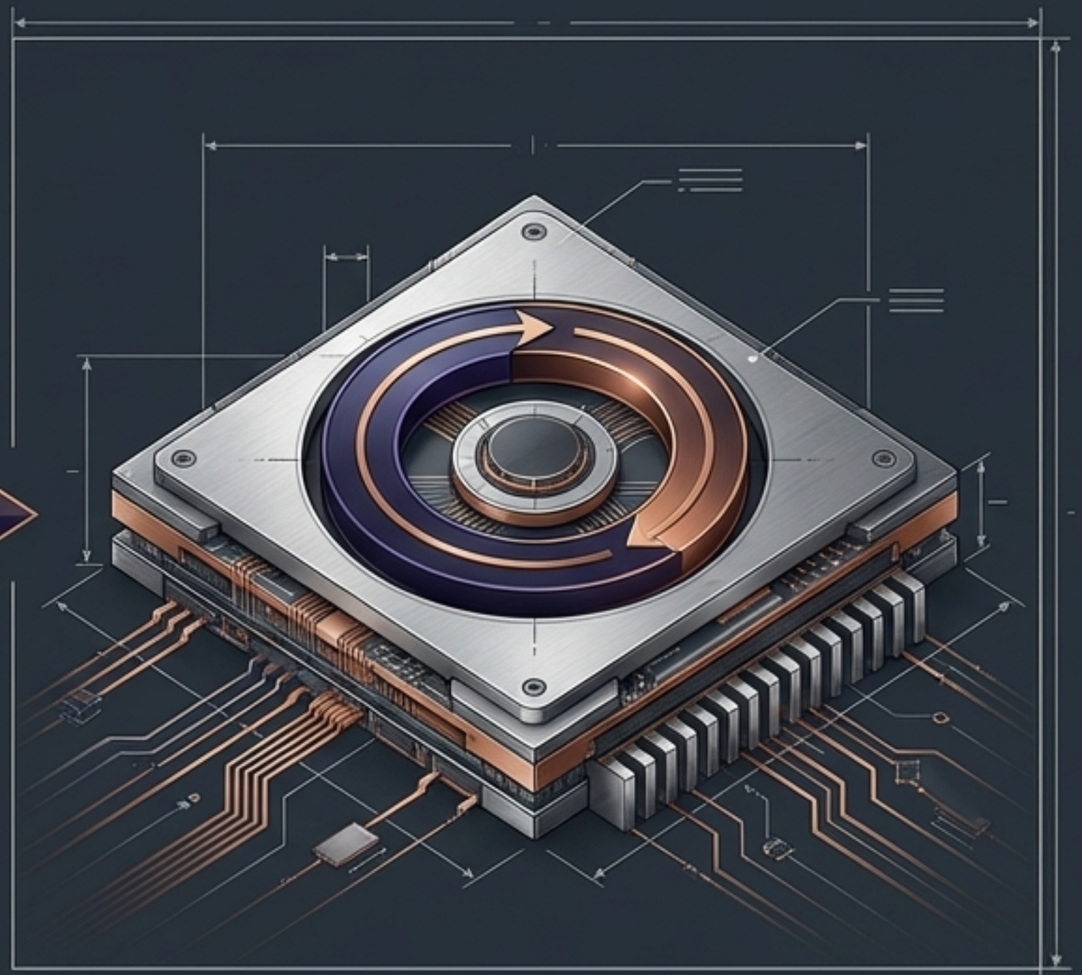
Hardware queues, dynamic workload phases, and discrete PCIe states.

The Processing Wall

Calculating this space is mathematically intractable for real-time, microsecond execution at the network edge.

The PoL-TT Mathematical Engine

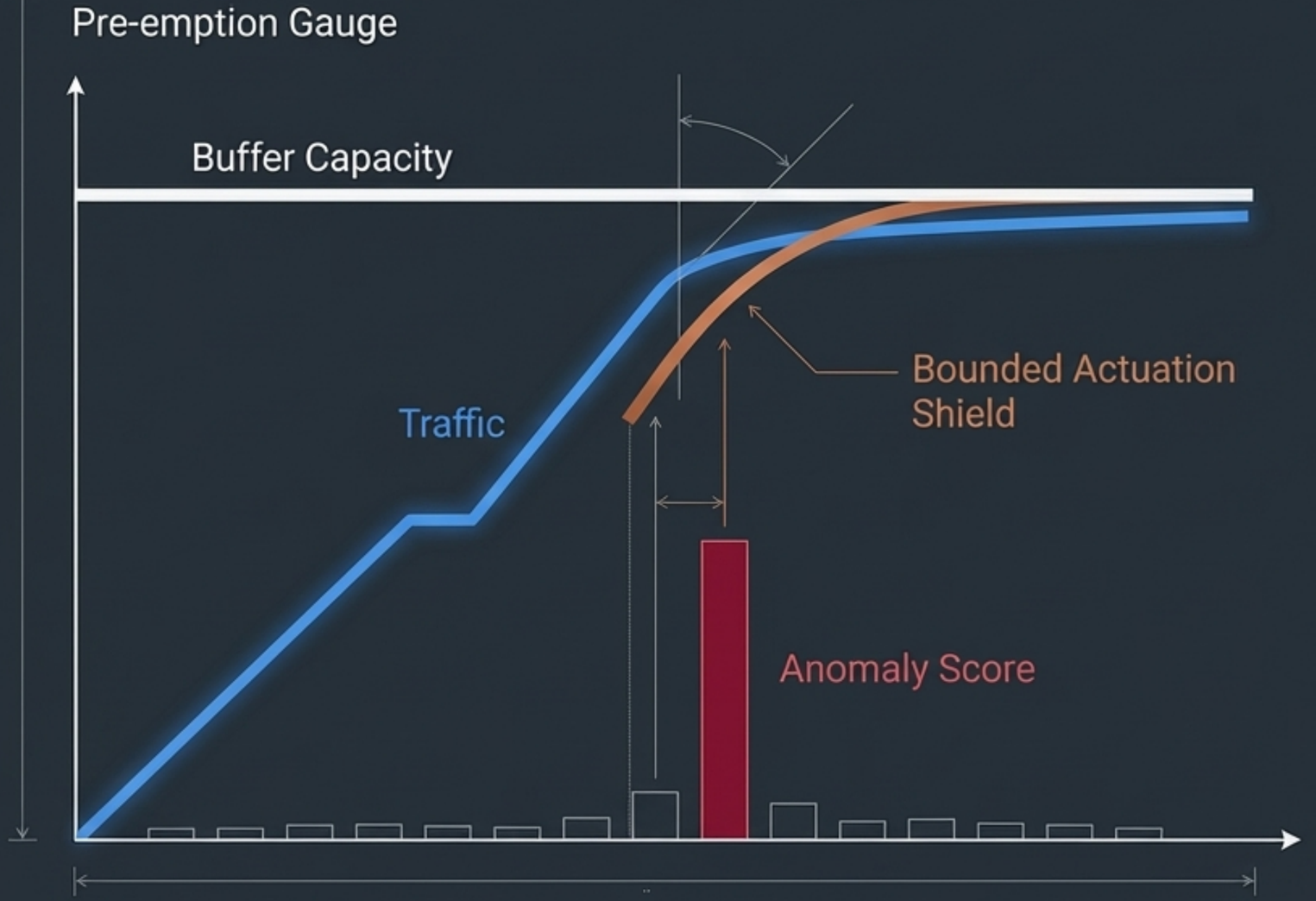




The Slow-Path (Learning): The PTCP Cluster Controller ingests massive telemetry. Kernel Density Estimation (KDE) normalizes it into a continuous probability distribution. Tensor Train (TT) decomposition and SVD rank truncation compress the model, preventing memory inflation.

The Fast-Path (Microsecond Actuation): Compressed TT cores are pushed to the Edge Agent. Execution complexity is dramatically bounded to $O(d * r^2 * 1)$.

Triggering Bounded Actuation



$$S(s) = -\log(p(s))$$

The Mechanics: For a real-time infrastructure state vector **s**, the edge agent evaluates the TT cores to output a joint probability **p(s)**.

The Result: When the anomaly score **S(s)** spikes, it mathematically proves an impending pathological incast burst. Actuation is triggered Actuation is triggered before hardware failure occurs.

Deployment Reality: Zero Host-Level Upgrades

The Commercial Advantage

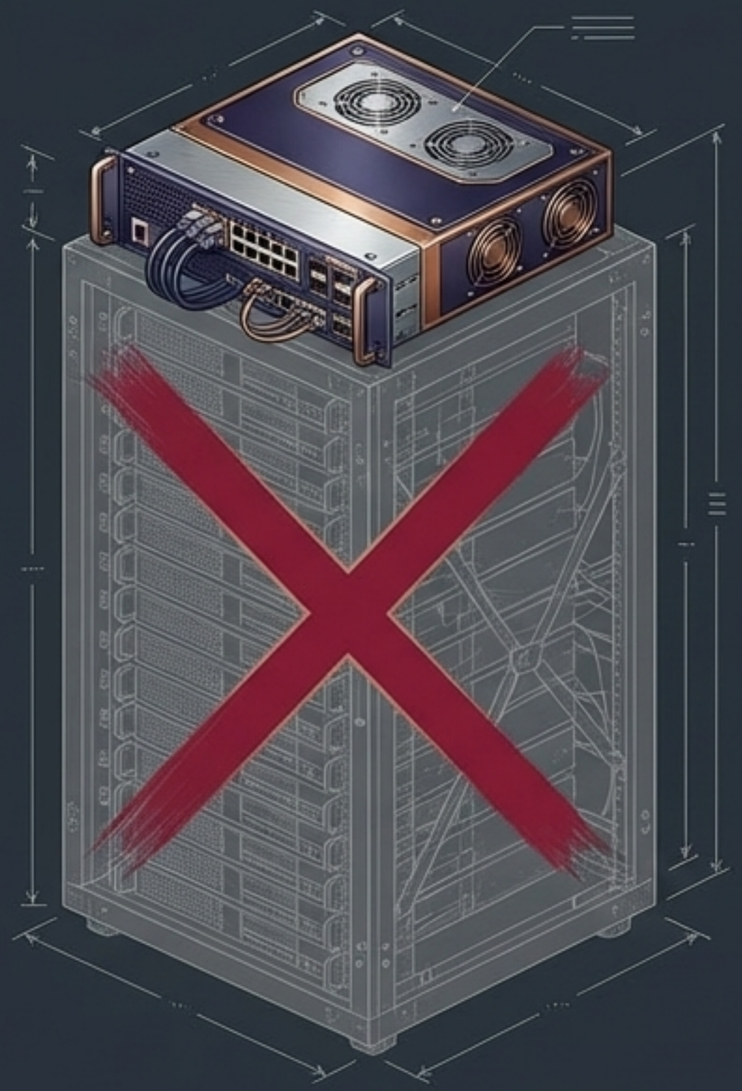
PTCP’s primary value lies in its friction-free deployment footprint.

What is NOT Required

Upgrading 100,000 host nodes with specialized SmartNICs, DPUs, or proprietary adapters.

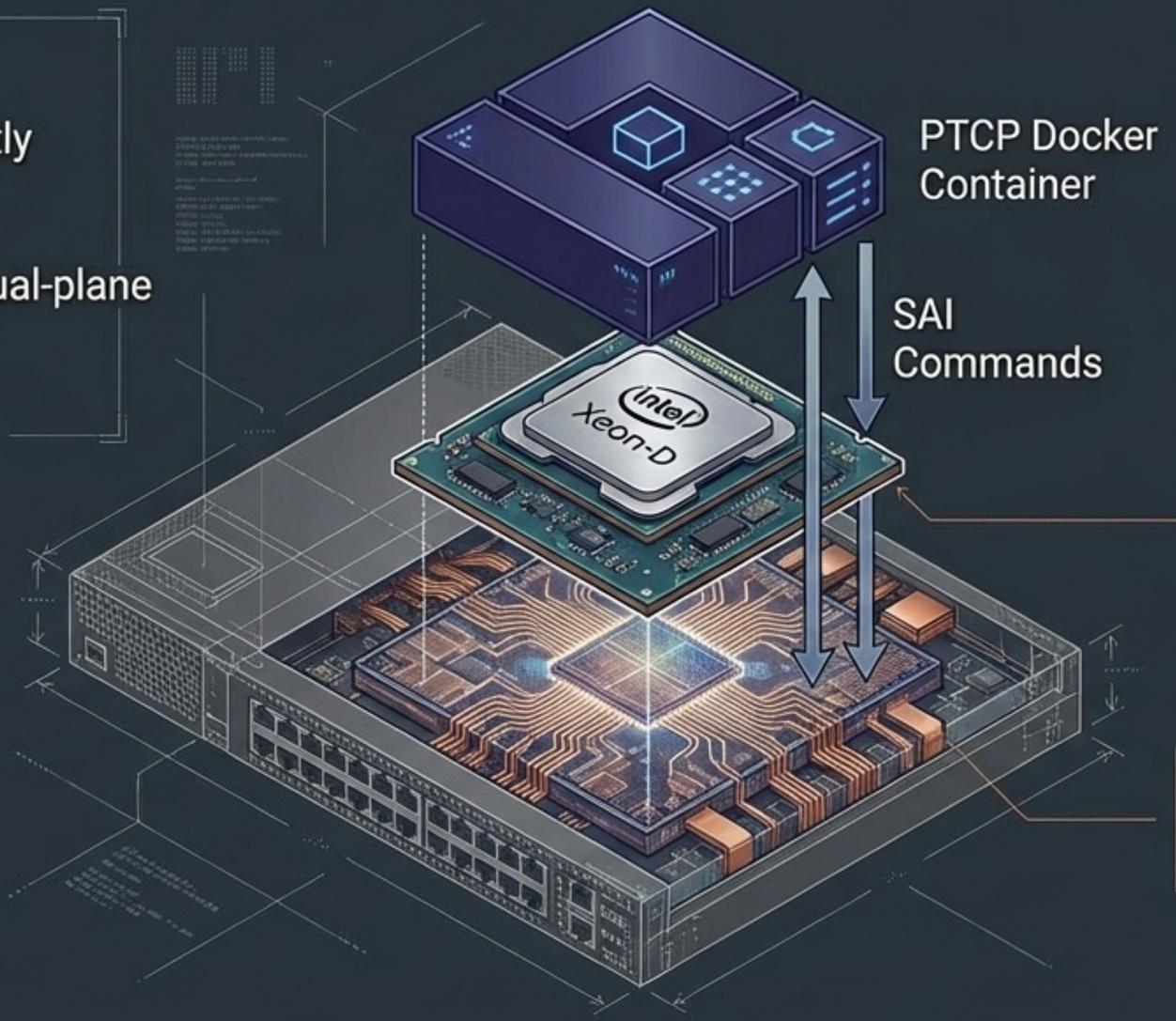
Where it Lives

Intelligence is strictly localized at the rack edge, deployed entirely within the existing Top-of-Rack white box ecosystem.



The COTS Integration Stack

Native Compatibility:
PTCP integrates directly into standard Open Network Install Environment (ONIE) dual-plane architectures without custom ASICs.

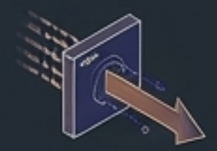


Containerized Agent:
Runs as a lightweight Docker container natively within the SONiC operating system on the local Xeon-D processor.

Control Plane:
Intel Xeon-D processor (Running SONiC)

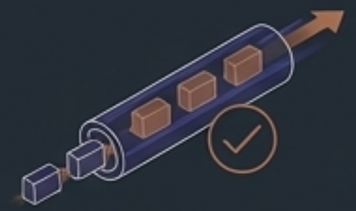
Data Plane:
Broadcom Tomahawk ASIC (Terabits of raw switching)

The Microsecond Actuation Loop



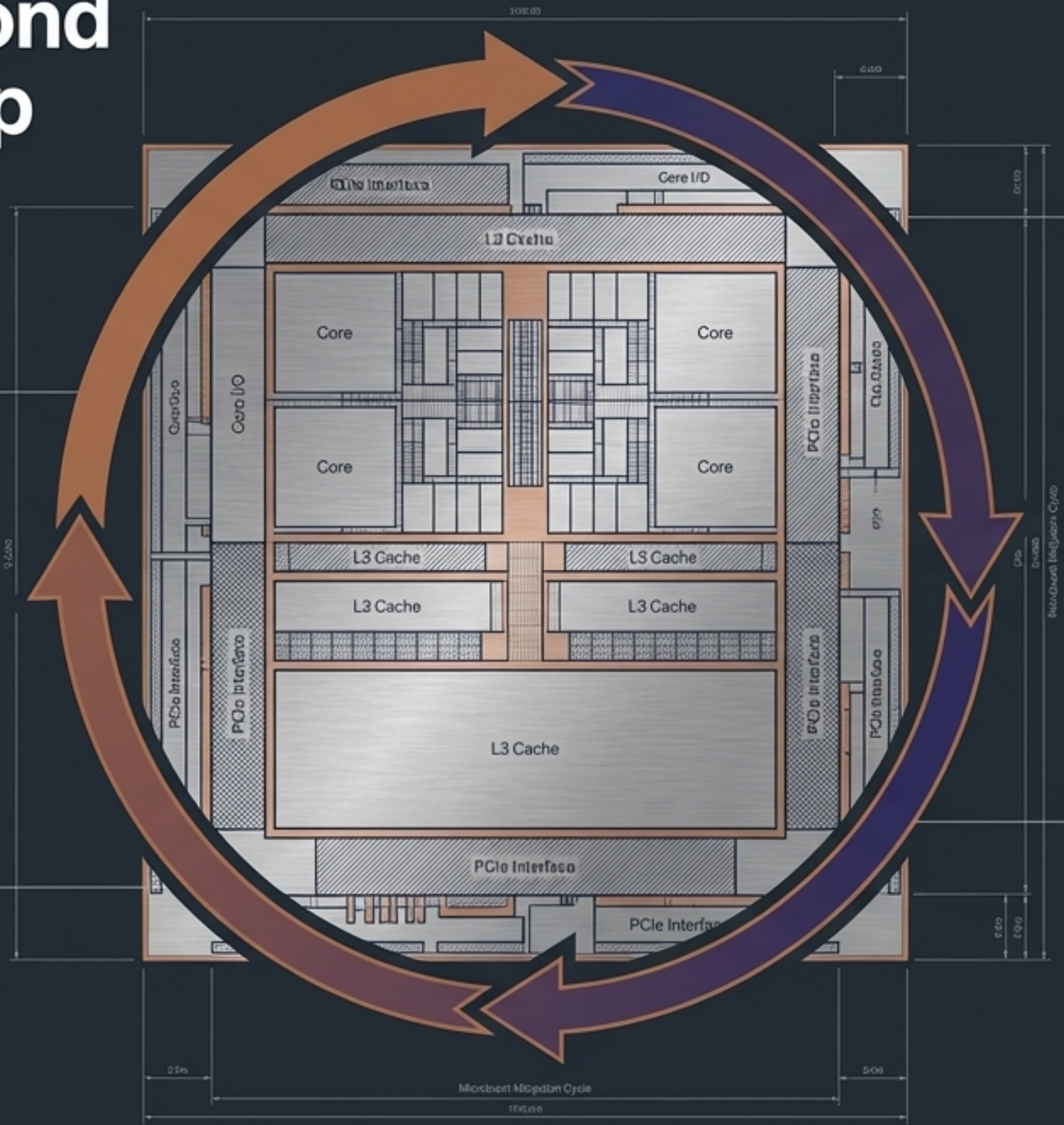
1. Telemetry Ingestion

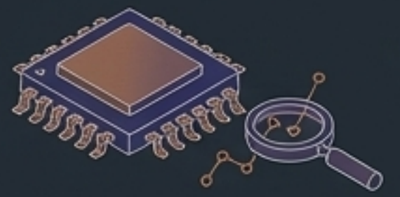
The agent continuously reads local queue depths and incoming flow rates.



4. Perfect Pacing

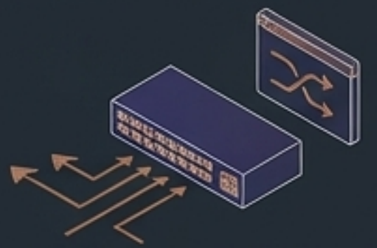
Electrical traffic is perfectly paced into the optical fabric. Buffer never hits 100%, PFC is never triggered, zero packet loss.





2. Fast-Path Evaluation

Because math is bounded to $O(d \cdot r^2 \cdot 1)$, the Intel processor calculates the anomaly score in microseconds.



3. Hardware Actuation

The PTCP container issues standard Switch Abstraction Interface (SAI) commands to dynamically throttle specific Broadcom ASIC egress queues causing the microburst.

The Diagnostic Architecture Matrix

	Standard Ethernet (RoCEv2)	InfiniBand	PTCP over COTS Ethernet
Flow Control Type	Reactive (PFC)	Deterministic (Credit)	**Predictive Tensor (PTCP)**
Hardware Requirements	COTS White box	Proprietary Fabric & HCAs	**COTS White box**
Ecosystem	Open (SONiC)	Locked	**Open (SONiC)**
Scalability Limit	Low (Incast failures)	High	**High (Zero Drop)**
Total Opportunity Loss	Massive OpEx Bleed	Zero	**Zero**

Operational Cost Delta: A $1.5 Billion Shortcut

InfiniBand
CapEx Premium

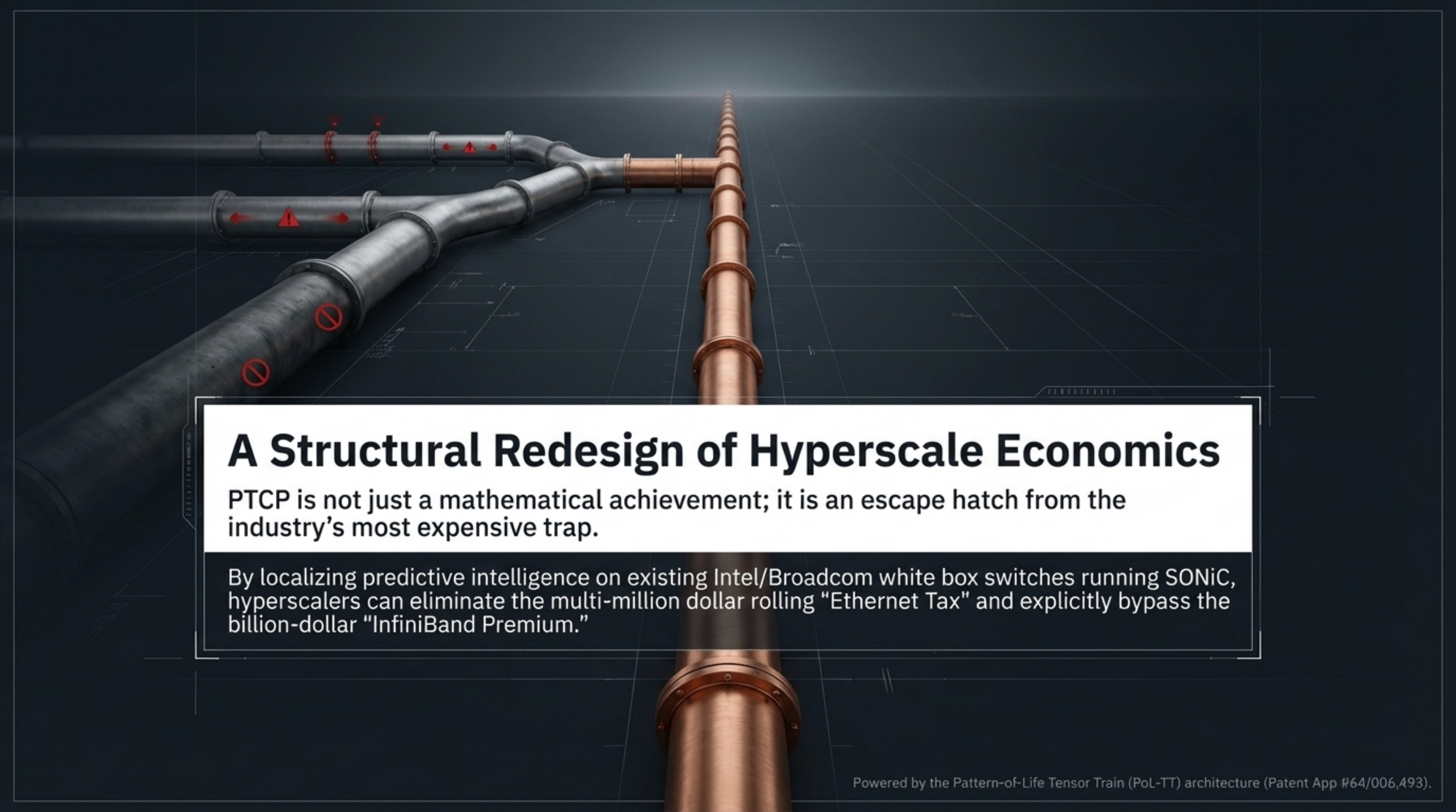
InfiniBand: ~$1.5B CapEx premium for proprietary switches and optics.

RoCEv2
OpEx Loss

Standard Ethernet: ~$650M annual OpEx loss. (Based on a mere 5% network-induced stall rate across 100k GPUs at $15/hr compute/cooling).

PTCP
Licensing

PTCP over COTS: ~$25M PTCP Software Licensing. Achieves the exact same zero-drop, deterministic flow as InfiniBand at a fraction of the cost.



A Structural Redesign of Hyperscale Economics

PTCP is not just a mathematical achievement; it is an escape hatch from the industry's most expensive trap.

By localizing predictive intelligence on existing Intel/Broadcom white box switches running SONiC, hyperscalers can eliminate the multi-million dollar rolling "Ethernet Tax" and explicitly bypass the billion-dollar "InfiniBand Premium."

Powered by the Pattern-of-Life Tensor Train (PoL-TT) architecture (Patent App #64/006,493).