

**Q: Could you please explain the following equations in detail so the user understands what is being computed and how? • For the batch  $B = \{tx_1, \dots, tx_k\}$ , compute aggregated delta:  $\Delta_B = \sum_{i=1}^k \delta(tx_i)$  For batched B:  $\mathcal{E}_\theta(S_{t+|B|}) = \mathcal{E}_\theta(S_t) + \sum_{tx \in B} \delta(tx)$  with  $|\text{error}| \leq |B| \times 10^{-9}$**

Here's a clear, detailed explanation of the two closely related equations from NERV's **Execution and Homomorphic State Update** section.

## 1. Aggregated Delta for a Batch

**Equation:**

For the batch  $B = \{tx_1, \dots, tx_k\}$ , compute aggregated delta:

$$\Delta_B = \sum_{i=1}^k \delta(tx_i)$$

**What this means in plain English**

- A **batch B** is a small group of private transactions (up to 256 in NERV) that the network processes together for efficiency.
- Each individual transaction  **$tx_i$**  (a private transfer of some amount from one hidden account to another) has its own pre-computed **delta vector  $\delta(tx_i)$** . This  $\delta(tx_i)$  is a 512-dimensional vector (512 floating-point numbers) that represents **exactly what change this single transaction would make to the neural state embedding** if applied alone.
- The network simply **adds up** (sums) all these individual delta vectors to get one single **aggregated delta  $\Delta_B$**  that represents the combined effect of the entire batch.

**Why addition ( $\Sigma$ ) works**

The core breakthrough (the “Transfer Homomorphism”) is that the trained transformer encoder  $\mathcal{E}_\theta$  has learned to represent simple balance transfers in a **linear** way in its 512-dimensional latent space.

This means:

- Subtracting an amount from sender  $\approx$  moving the embedding a certain distance in one direction.
- Adding the same amount to the receiver  $\approx$  moving the embedding the same distance in the opposite direction.
- The net effect of a transfer is just a single vector addition in this space.

Because transfers are linear, the effect of many transfers is just the sum of their individual effects — exactly like vector addition.

## How it is computed in practice

1. Each user's wallet, when preparing a transaction  $tx_i$ , runs a small ZK circuit ( $\Delta\theta$  sub-circuit) that outputs the correct  $\delta(tx_i)$  without revealing sender, receiver, or amount.
2. The transaction (including its  $\delta(tx_i)$ ) is onion-routed anonymously into the shard.
3. Inside the shard's trusted execution environment (TEE), the node collects up to 256 such transactions into a batch.
4. It performs a simple element-wise vector addition of all 512 components:  $\Delta_B[j] = \delta(tx_1)[j] + \delta(tx_2)[j] + \dots + \delta(tx_n)[j]$  for each dimension  $j = 1$  to 512.

This addition is extremely fast and cheap —  $O(1)$  cost regardless of how many accounts exist in the ledger.

## 2. Batched Homomorphic Update of the State Embedding

### Equation:

For batched B:

$$\mathcal{E}_\theta(S_{\{t+|B|\}}) = \mathcal{E}_\theta(S_t) + \sum_{tx \in B} \delta(tx) \quad \text{with} \quad |\text{error}| \leq |B| \times 10^{-9}$$

(The summation  $\sum_{tx \in B} \delta(tx)$  is exactly the same as  $\Delta_B$  above.)

### What this means in plain English

- $S_t$  is the full private ledger state at some point in time  $t$  (all hidden accounts and their balances).
- $\mathcal{E}_\theta(\cdot)$  is the fixed, public 24-layer transformer encoder that compresses the entire state  $S_t$  into a tiny 512-byte vector  $e_t = \mathcal{E}_\theta(S_t)$ . This  $e_t$  is the canonical "state root" of the shard.
- After applying the entire batch B (which contains  $|B|$  transactions), the new true state is  $S_{\{t+|B|\}}$ .
- The equation says: instead of re-running the huge transformer on the new full state  $S_{\{t+|B|\}}$  (which would be impossibly expensive), we can just **add the aggregated delta vector to the old embedding** and get almost exactly the same result:  
**new embedding  $\approx$  old embedding +  $\Delta_B$**
- The approximation is extraordinarily accurate: the total error in the 512-dimensional space is at most  $|B| \times 10^{-9}$  (for a full batch of 256 transactions, maximum error  $\leq 2.56 \times 10^{-7}$ , which is negligible for fixed-point financial calculations).

### Why this is a breakthrough

- Traditional blockchains must store and update gigabytes of state (Merkle tries, account databases, etc.) and produce large proofs for every change.
- NERV never needs to decompress or touch the full state. It just does a 512-number addition inside a hardware enclave.

- The resulting new embedding  $e_{t+1} = e_t + \Delta_B$  becomes the new canonical state root.
- A succinct zero-knowledge proof (Halo2 recursive) attests that the addition was done correctly and that the homomorphism error bound was respected.

### Visual intuition

Imagine the 512-dimensional embedding as a point in a high-dimensional space:

- Starting point = current state embedding  $e_t$
- Each transaction = a tiny arrow ( $\delta$  vector)
- Applying a batch = moving the point by the vector sum of all those arrows
- The transformer has learned the geometry of ledger updates so well that this simple vector addition lands within  $10^{-9}$  distance of where the full transformer would have placed the point if run on the updated ledger.

### Summary of the overall computation flow

1. Wallets compute individual  $\delta(tx_i)$  privately.
2. Shard batches transactions and sums the  $\delta$ 's  $\rightarrow \Delta_B$  (simple 512-dim vector addition).
3. New state embedding = old embedding +  $\Delta_B$
4. ZK proof confirms correctness and tiny error bound.
5. The network now has a new 512-byte state root representing millions of accounts and thousands of updates — with full privacy and post-quantum security.

This is what enables NERV's claimed 900× compression,  $O(1)$  state updates, and massive scalability while keeping everything private by default.

---

### Q: What's the reason for including this 'Partial BLS12-381 signature share' in the AI-Native Optimistic Consensus and how does it work?

The "**Partial BLS12-381 signature share**" is a key part of how NERV achieves **full cryptographic (economic) finality** on top of the fast, AI-driven optimistic path in its consensus mechanism. Let's break it down clearly: why it's included and exactly how it works.

### Why Include Partial BLS12-381 Signature Shares?

NERV's consensus is **hybrid optimistic**:

- **Fast path (99.99% of cases)** — Validators simply predict and vote on the next embedding hash ( $h_{pred}$ ). If  $\geq 67\%$  weighted agreement (stake  $\times$  reputation) happens quickly, the block gets **probabilistic finality** (very high confidence, sub-second).
- This is super efficient and low-latency, but it's technically "optimistic"—there's a tiny chance of disagreement or attack.

To make finality **unconditional and cryptographically irreversible** (even if the optimistic vote was wrong or attacked), NERV layers on **threshold BLS signatures**:

- BLS12-381 (Boneh-Lynn-Shacham on the BLS12-381 curve) is chosen because it supports **efficient threshold aggregation**—many validators can contribute parts of a signature, which combine into one compact signature representing the whole group.
- Once aggregated from enough honest validators, this single threshold signature **cryptographically commits** to the new embedding root. It's impossible to revert without breaking the underlying math (assuming <33% Byzantine stake).
- This provides **economic finality**: Users (and VDWs) can rely on it forever, with slashing for any misbehavior.
- Without it, the system would rely only on probabilistic/optimistic security (like some rollups), which is riskier for a base-layer chain handling value.

In short: The neural predictions give speed; the BLS shares give unbreakable finality.

## How It Works Step-by-Step

### 1. What Each Validator Broadcasts:

- In the neural vote message (128 bytes total for efficiency):
  - The predicted hash  $h_{pred}$  (32 bytes).
  - Their **partial BLS12-381 signature share** on a message that includes  $h_{pred}$  (or the full updated embedding/root).
  - Reputation score and TEE attestation.

### 2. The partial share is generated using the validator's private staking key (standard BLS signing, but only a "share" in the threshold sense).

### 3. Threshold Agreement Detection:

- Nodes collect these vote messages.
- They check if  $\geq 67\%$  weighted stake/reputation agrees on the same  $h_{pred}$ .
- Simultaneously, honest nodes **aggregate the partial BLS shares** from the agreeing validators.

### 4. Aggregation into Full Threshold Signature:

- BLS signatures are **aggregatable**: Partial shares from different validators (on the same message) can be mathematically combined into one single signature (still ~96-128 bytes).
- This is done using BLS threshold rules (e.g., Lagrange interpolation or simple addition in the group, depending on the exact scheme—NERV likely uses a standard (t-of-n) setup).
- Only needs shares from the quorum ( $\geq 67\%$ ), not all validators.
- The result: One compact **full BLS threshold signature** representing the entire agreeing majority.

### 5. Commitment and Finality:

- The aggregated threshold signature is **attached to the new embedding root** and broadcast/gossiped.

- Once visible and verified (fast BLS verify against the aggregated public key of the stake-weighted validators), the root is **cryptographically final**.
  - Light clients and users can verify this signature forever (using the public stake weights).
  - VDWs (user receipts) tie back to this signed root.
6. **In the Rare Challenge Case:**
- If disagreement triggers Monte-Carlo dispute resolution, the winning root still gets a fresh threshold BLS signature from the resolving TEEs/validators.
  - Losers get slashed.

This design keeps the common path tiny (just 128-byte votes) while upgrading to full crypto finality without extra rounds—perfect for NERV's sub-second goals.

BLS12-381 is a popular choice in real systems (e.g., Ethereum beacon chain uses BLS aggregation for validator attestations) precisely for this efficient threshold/aggregation property. It fits NERV's post-quantum goals too (though BLS itself isn't PQ, NERV layers Dilithium/ML-KEM elsewhere)

---

## Easy-to-Understand Breakdown of the LatentLedger ZK Circuit

The **LatentLedger ZK Circuit** is the heart of NERV's innovative state management system. It allows the blockchain to replace massive, traditional state databases (like Merkle trees that can be gigabytes large) with a tiny 512-byte "embedding" vector—while proving everything is correct and private using **zero-knowledge proofs (ZKPs)**.

Think of it like this: In normal blockchains, the "state" (all account balances and data) is a huge spreadsheet. Updating it requires big proofs and lots of storage. LatentLedger uses a trained neural network (transformer) to **compress** that entire spreadsheet into just 512 numbers, and the ZK circuit proves that this compression (and updates to it) was done honestly—without revealing any private details.

This is built on **ZK-ML** (zero-knowledge machine learning), a hot research area where neural networks are turned into verifiable computations. It's more efficient than general-purpose ZK systems (like zkEVMs that mimic full virtual machines) because it's custom-tailored for one job: encoding ledger states with the special "Transfer Homomorphism" property (linear updates for transfers).

The circuit uses **Halo2** (a popular ZK framework from Zcash) with recursion (Nova) for compactness, and runs inside secure hardware (TEEs like SGX/SEV) to prevent tampering. It's only ~7.9 million "constraints" (basic math checks)—much smaller than zkEVM circuits, leading to tiny proofs (~750 bytes) and fast verification (~80ms on a phone).

Now, let's go through the **core components** and **step-by-step process** in simple terms.

## Core Components (What Goes In and Out)

1. **Private Inputs (Witnesses – Secret Stuff Only the Prover Knows):**
  - **St:** The full current ledger state—a list of (blinded key, balance) pairs. Keys are hidden (e.g., hashed or encrypted with post-quantum ML-KEM) so no one sees real accounts.
  - **$\theta$  (Theta):** The transformer's weights/parameters (~1-2 million numbers, quantized to 8-16 bits for efficiency). These are public and audited.
  - **Transaction Batch:** Up to 256 transfers, used to calculate deltas ( $\delta$ ) via a small helper circuit ( $\Delta\theta$ ).
2. **Public Inputs/Outputs (Visible to Everyone):**
  - Old embedding ( $e_{t-1}$ ) from before the update (hashed to 32 bytes on-chain).
  - New embedding ( $e_t$ ) after processing.
  - Aggregated delta path: The combined change vector for the batch.
3. **Circuit Building Blocks (How Math is Handled):**
  - **Arithmetization:** Everything converted to math over a large prime field (like BLS12-381 curve) with fixed-point numbers (e.g., 16 bits integer + 16 bits fraction for decimals).
  - **Gates:** Basic/custom operations for multiplications (attention layers), additions, etc.
  - **Lookup Tables (LUTs):** Pre-made tables for tricky non-linear functions (like Softmax or GELU)—cheaper than full math, optimized with tricks like "neural teleportation" to shrink tables.
  - **Recursion (Nova):** Allows folding proofs for batches, keeping things scalable.

Constraint breakdown: ~70% attention (quadratic but padded to fixed ~1024 "accounts"), ~20% feed-forward layers, ~10% non-linearities/homomorphism checks.

## How It Works Step-by-Step (Simple Flow)

1. **Setup Phase (Done Once, Publicly):**
  - Create proving/verification keys (universal setup in Halo2—no toxic waste risk).
  - Quantize (compress) the transformer weights to fit small.
  - Pre-build LUTs for activations (e.g., approximate GELU with a small table of values).
2. *Why?* Prepares a reusable "template" for proofs—efficient and secure.
3. **Input Preparation (Prover Side, Inside Secure TEE):**
  - Load private state  $St$  as a "tokenized" sequence (like words in NLP: keys get embeddings, balances scaled).
  - Blind keys with post-quantum encryption.
  - For updates: Compute individual  $\delta(tx)$  deltas (lightweight linear math based on sender/receiver embeddings and amount).
4. *Why in TEE?* Hardware enclave ensures no cheating or leaks—even the host machine can't see secrets.
5. **Circuit Execution (Turning Neural Net into Verifiable Math):**

- Simulate the full transformer forward pass as thousands of constraint equations.
  - **Embedding Layer:** Linear projection of tokenized state → initial hidden vectors (matrix multiplies ~500K constraints).
  - **Transformer Blocks (Layers 2-23):**
    - Attention: Project to queries/keys/values, compute scaled dots, softmax (via LUT), weighted sum (~2M constraints/block, fixed heads help).
    - Feed-Forward: Linear layers + GELU (LUT) (~1M/block, sparsity skips zeros).
    - Residuals/LayerNorm: Cheap additions/sums.
  - **Output Layer:** Average/pool to final 512-byte  $e_t$  (~200K constraints).
  - **Homomorphism Check:** Verify new  $e_t \approx \text{old } e_{t-1} + \text{sum of deltas}$  (error  $< 10^{-9}$ , ~50K constraints).
6. *Why this way?* Breaks complex neural ops into simple, provable gates. Padding handles variable accounts.
  7. **Total Circuit Size:**
    - ~10K rows in the witness table, enforced with copy/permutation/lookup checks.
  8. **Proof Generation (Prover in TEE):**
    - Commit to private data (KZG polynomials).
    - Generate succinct proof  $\pi$  (Halo2 arguments + Fiat-Shamir).
    - <1s on GPU-TEE; ~750 bytes (Nova recursion shrinks it).
  9. *Why fast/small?* Tailored optimizations beat general zkEVMs.
  10. **Verification (Public, Anyone/Light Client):**
    - Check proof against public inputs (old root hash, weights hash).
    - ~80ms on mobile; accepts new  $e_t$  hash as canonical.
  11. *Why trust it?* Math guarantees correctness without seeing secrets.

## Security and Optimizations (Why It's Safe and Efficient)

- **Soundness/ZK:** Halo2 + lookups = near-zero cheating chance ( $< 2^{-128}$ ); many-to-one mapping hides state (huge entropy).
- **Post-Quantum:** Lattice sigs (Dilithium) + compatible blinding.
- **Optimizations:**
  - Quantization: Lower-bit weights → fewer multiplies.
  - Sparsity: Skip zeros (67% savings).
  - Recursion: Handle huge batches in one proof.
  - Non-Invertibility: Extra non-linear tweaks make reversing embedding impossible (new hardness assumption).

In short: LatentLedger turns a giant neural network into a verifiable, private, ultra-compact state machine—enabling NERV's privacy, speed, and scalability dreams. It's ambitious ZK-ML applied directly to the blockchain core!

---

## Easy-to-Understand Breakdown of Embedding Bisection for Sharding

**Embedding Bisection** is one of the coolest and most unique parts of NERV's "dynamic neural sharding" system. It lets the blockchain automatically split busy shards into two (like a cell dividing) or merge quiet ones — with no manual setup, no slow governance votes, and almost no downtime. The whole split takes just 3–4 seconds, even at 10,000+ transactions per second.

Traditional blockchains struggle with sharding because splitting state usually means:

- Moving huge amounts of account data around (slow and risky).
- Pre-defining shard boundaries years in advance (bad guesses lead to imbalances).
- Complex rebalancing that leaks metadata or takes days.

NERV avoids all that by working directly on the **512-byte neural embedding** ( $e_t$ ) — the compressed "DNA" of the entire shard's ledger state. Because of the Transfer Homomorphism (linear updates), the embedding naturally encodes the relationships and clusters of accounts. Bisection cleverly cuts this high-dimensional "DNA" in half mathematically, creating two balanced child shards without ever unpacking the full state.

Think of the embedding as a point floating in a 512-dimensional space where similar accounts cluster together. Bisection draws a random but deterministic "knife" (hyperplane) through that point and projects the state into two new points — one for each child shard.

This is completely novel: no other blockchain splits state by mathematically bisecting a neural latent vector.

### Core Intuition (Why This Approach?)

- **Why bisection?** The embedding captures the ledger's "shape" (account clusters, hot spots). Cutting along a smart hyperplane naturally separates busy groups of accounts into different shards, reducing future load.
- **Deterministic** Every node gets the exact same split (no forks) because it's based on a shared seed.
- **Balanced** An LSTM neural net predicts overload early and only triggers splits when needed. The cut aims for ~50/50 load split.
- **Reversible** Merging uses the reverse math — perfect for when traffic dies down.

### How It Actually Works: Step-by-Step (Simple Explanation)

1. **Trigger Detection (Every Node Watches for Overload)**
  - Each shard runs a small (1.1 MB) LSTM model that looks at the last 2 minutes of traffic stats: transactions per second (TPS), cross-shard messages, latency, etc.
  - The LSTM predicts: "What's the chance this shard gets overloaded in the next 15 seconds?"
  - If probability > 92%, any node can propose a split. It broadcasts a signed "SplitProposal" (post-quantum Dilithium signature) with the current shard ID and block height.
2. *Why proactive?* Prevents slowdowns before they happen — like predictive maintenance.



3. **Quick Approval (<1.5 seconds)**
  - Validators (weighted by stake × reputation) co-sign the proposal using BLS partial signatures.
  - Needs ≥67% agreement (same threshold as normal consensus).
  - If approved, the proposer gets rewarded; bond returned.
4. *Why fast?* Keeps the process lightweight — no full new block needed.
5. **Deterministic Bisection (The Math Magic — All Nodes Do This Identically)**
  - **Create a seed:** Hash(shard\_id + current height + a fixed chain constant) → fully deterministic, post-quantum secure (SHA3).
  - **Pick a random direction:** Use the seed to generate a random unit vector **d** in 512 dimensions (like spinning a compass in high-D space).
  - **Define the cutting plane:** A hyperplane through the current embedding **e\_t**, perpendicular to **d**. Mathematically: all points **x** where  $\mathbf{d} \cdot (\mathbf{x} - \mathbf{e}_t) = 0$ .
  - **Project to child embeddings** (without decompressing anything):
    - Compute how much **e\_t** sticks out along **d** (projection).
    - Create two new points: **e\_left** ≈ **e\_t** moved slightly left along **d** **e\_right** ≈ **e\_t** moved slightly right along **d**
    - Add tiny differential-privacy noise ( $\epsilon$ ) to prevent leakage.
    - Result: Two new 512-byte embeddings that together represent the split state.
  - **Assign accounts to children:** Future transactions route based on hash(account\_key + child\_seed) — ensures roughly 50/50 split.
  - A small Halo2 proof (~500K constraints) verifies the bisection preserved the homomorphism and balance.
6. *Why no data migration?* The state lives in the embedding — we just create two new "views" of it mathematically.
7. **Re-Execution for Safety (<2 seconds)**
  - New child shards (nodes assigned via VRF from the seed) replay the last 500 transactions on their new projected embeddings.
  - They apply the same old deltas additively inside TEEs.
  - Both children must end up with embeddings matching within  $10^{-9}$  — if not, abort and resolve via dispute.
  - New embeddings are erasure-coded (Reed-Solomon) for fault tolerance and distributed optimally (genetic algorithm for low latency).
8. *Why replay?* Guarantees no divergence — pure safety net.
9. **Finalization and Routing Update (<0.5 seconds)**
  - Update the network's DHT and anonymous mixer tables so new transactions go to the correct child shard.
  - Broadcast the new child roots globally.
  - Old shard marked as "split" in history.
  - Issue updated VDWs (receipts) for any affected transactions.

**Merging** is the reverse: when two sibling shards are both quiet for 10 minutes, automatically recombine using the original parent seed and replay recent txs.

## Underlying Mathematics (Simplified)

- **Projection onto direction d:**  $\text{proj}_d(e_t) = (d \cdot e_t / \|d\|^2) \times d$
- **Child embeddings** (approximate):  $e_{\text{left}} = e_t - (\text{proj}_d(e_t) / 2) + \text{tiny noise}$   
 $e_{\text{right}} = e_t + (\text{proj}_d(e_t) / 2) + \text{tiny noise}$
- **Why it preserves linearity:** Transfers are vectors in the same space. Intra-child transfers stay fully in one child; cross-child ones get handled by the mixer — homomorphism still holds.
- **Balance trick:** The random direction d is chosen to avoid major clusters (inspired by PCA on training data), giving even splits most of the time.

## Why This Enables "Infinite" Scalability

- No theoretical limit on shard count — busy shards keep splitting like cells.
- Quiet ones merge back → efficient resource use.
- Splits/merges in seconds → reacts instantly to real traffic (e.g., viral token launches).
- Zero state migration → no congestion or privacy leaks.

In essence, NERV treats shards like living organisms that grow and divide based on AI-predicted needs, using pure math on neural embeddings. It's elegant, fast, and unlike anything in current blockchains.

## Embedding Bisection and Shard Scaling

Each **dynamic neural shard** in NERV has exactly **one canonical 512-byte neural state embedding** ( $e_t$ ) at any height t. This single embedding compresses and represents the entire shard's state (all hidden accounts, balances, and commitments within that shard). The number of shards varies dynamically with network load:

- Simulations and projections show shard counts fluctuating, e.g., from **312** to **1,204** back to **489** in a 90-day Aurora testnet run.
- Expected range scales with adoption and TPS demand, starting at low dozens/hundreds at genesis and potentially reaching thousands or more for sustained >1M TPS (no theoretical ceiling, as splits occur automatically on overload).

Shards form a binary tree-like lattice (DAG of embeddings), but active shards at any time each maintain their own independent  $e_t$ .

## Routing New Transactions After a Split

When a shard splits:

1. The parent embedding  $e_t$  is deterministically bisected using a seed derived from  $\text{shard\_id} \parallel \text{height}$  (post-quantum secure hash).

2. This produces two child embeddings ( $e_{\text{left}}$ ,  $e_{\text{right}}$ ) and corresponding child shard\_ids (e.g., via  $\text{shard\_id} \oplus \text{hash}(\text{seed} \parallel 0/1)$ ).
3. Future transaction routing uses blinded commitments in the private tx payload:
  - The wallet includes a blinded shard selector (e.g., hash of receiver commitment  $\parallel$  sender hint, encrypted).
  - The final 5-hop relay or ingress TEE routes to the appropriate child shard using public shard root metadata without learning contents.
  - For intra-shard transfers post-split, the homomorphism ensures deltas apply to the correct child (accounts implicitly assigned via hash projections onto the bisection hyperplane).
  - Cross-shard txs are handled seamlessly via the mixer and DHT updates (gossip <800 ms global).

No "latest" single shard exists post-split; both children are equally canonical successors, inheriting subsets of the parent's logical state.

## Typical Numbers: Shards, Embeddings, Transactions/Accounts

Entity	Typical/Projected Numbers	Relationship
<b>Embeddings per shard</b>	Always <b>1</b> (512-byte $e_t$ )	Compresses entire shard state
<b>Accounts per shard</b>	Up to <b>~100 million</b> (inferred from gigabyte-scale traditional tries vs. 512-byte compression)	One embedding represents all accounts/balances in shard
<b>Transactions per batch</b>	Up to <b>256</b> (optimal for density; avg. ~2 bytes overhead per tx via homomorphic deltas)	Batches update one embedding via aggregated $\Delta_B$
<b>TPS per shard</b>	<b>~1,000</b> sustained (real traffic projections)	Drives splits when overloaded
<b>Active shards</b>	<b>Hundreds to low thousands</b> (e.g., 800–1,200 for 1M+ TPS)	Total TPS $\approx$ shards $\times$ TPS/shard (horizontal scaling)

<b>Total accounts network-wide</b>	Potentially <b>billions</b> (infinite scaling via dynamic splits/merges)	Distributed across shards; each embedding handles ~100M max before split
------------------------------------	--	--

Batching enables extreme efficiency: a 256-tx batch updates the embedding in  $O(1)$  time with ~512-byte aggregated delta.

## Genesis Initialization of $e_t$

At mainnet genesis (June 2028):

- The network starts with a small number of shards (likely 1–dozens, based on initial expected load).
- The baseline embedding  $e_0$  (genesis root) is the fixed, public output of the encoder  $\mathcal{E}_\theta$  run on the empty/initial state  $S_0$ :
  - $S_0$  includes only genesis commitments (e.g., fair emission reserves, no pre-mine).
  - $e_0 = \mathcal{E}_\theta(S_0)$ , computed once offline and burned into the genesis block (attested via Halo2 proof and TEEs).
- All shards initialize from this (or derived via early splits if needed).
- The encoder  $\mathcal{E}_\theta$  itself is public, audited, and fixed for the first epoch (30 days).

This ensures a deterministic, verifiable starting point with zero plaintext exposure. Light clients sync only the 32-byte hash chain from  $e_0$  onward (<100 KB forever).

In NERV's architecture with potentially 1000+ active dynamic neural shards, **each shard** maintains its own independent 512-byte neural state embedding ( $e_t$ ) that represents the entire private state of accounts and balances within that shard.

A single private transaction (a blind transfer) always affects **only one shard** — specifically, the shard containing both the sender's and receiver's hidden accounts (or the target shard for cross-shard routing). Therefore, there is **exactly one relevant  $e_t$**  (and one corresponding delta vector  $\delta(tx)$ ) for any given transaction.

## How a New Transaction Selects the Correct Shard and $e_t$

1. **Client-Side Preparation (Wallet):**
  - The user's wallet knows the blinded commitments or hints for the sender's and receiver's accounts (derived from private keys/seeds).
  - It deterministically computes the target **shard selector** (e.g., a blinded hash of receiver commitment || sender hint || other metadata).
  - This selector identifies the exact active shard without revealing plaintext details.

- The wallet then generates the homomorphic delta  $\delta(tx)$  using the lightweight  $\Delta\theta$  sub-circuit, tailored specifically to the transfer (amount-scaled difference in learned embeddings for sender/receiver identities).
- 2. **Blinded Routing During Ingress:**
  - The transaction payload includes the **encrypted/blinded shard selector**.
  - The 5-hop TEE mixer and final relay use the public on-chain shard registry (shard roots and lattice metadata) to route the onion-routed tx to the correct shard's encrypted mempool.
  - No plaintext is exposed; routing happens without learning sender, receiver, or amounts.
- 3. **Execution in the Target Shard:**
  - The shard loads its current canonical  $e_t$ .
  - Applies the aggregated delta  $\Delta_B$  (including this tx's  $\delta(tx)$ ) homomorphically:  $e_{t+1} = e_t + \Delta_B$ .
  - Generates Halo2 proof attesting correctness.

There are never 1000+ possible deltas for one tx — the wallet pre-computes the single correct  $\delta(tx)$  based on the known target shard's embedding space (all shards share the same public encoder  $\mathcal{E}_\theta$ , but each has its own latent vector  $e_t$ ).

Cross-shard transfers (rare, as accounts tend to cluster) are handled via private atomic swaps or future bridges, but a single NERV-side transfer remains intra-shard by design.

This ensures efficient, private execution even at extreme scale: wallets always know the precise target shard/ $e_t$  via local deterministic computation, and the network routes blindly. No global search or prediction across all shards is needed.

## Detailed Explanation of Shard Selection in NERV Wallets

In NERV's fully private-by-default design, **no public addresses or account identifiers exist on-chain**. Accounts are represented only by **private commitments** (cryptographic hiding of the account key and balance) derived from the user's wallet seed. Despite this, the wallet must route a private transfer transaction to the **exact correct dynamic neural shard** among potentially thousands of active shards — without leaking any plaintext information (sender identity, receiver identity, amounts, or even the target shard itself to observers).

This is achieved through a **deterministic, client-side shard selector** that is **blinded** (encrypted/hidden) in the transaction payload. The selector allows the network's ingress relays (operating inside attested TEEs) to route the transaction correctly using only public metadata, while preserving full privacy.

Below is a step-by-step breakdown of how the two quoted mechanisms work.

### 1. Deterministic Computation of the Target Shard Selector (Client-Side in Wallet)

The wallet performs this computation **entirely locally** before constructing the private transaction. No network queries are needed beyond syncing the light-client header chain (public embedding root hashes, <100 KB forever).

**Inputs available to the wallet (private, from local keystore):**

- Sender's full private account key (sk\_sender) or derivation path.
- Receiver's public commitment (com\_receiver): This is a post-quantum hiding commitment (e.g., ML-KEM-encapsulated or Pedersen-like on lattices) shared out-of-band (e.g., via QR code, encrypted chat). It hides the receiver's actual account identifier.
- Optional sender hint: A partial blinded value derived from sk\_sender (e.g., first 128 bits of a hash) to handle cases where sender/receiver are in different shards (rare, but possible in cross-shard setups).
- Current network state: The wallet maintains a cached **shard lattice map** from light-client sync — a compact Merkle-ized DAG of active shard\_ids and their root hashes (derived from embedding hashes).

**Computation steps:**

**a. Derive blinded identifiers locally:**

- Receiver identifier: blind\_receiver = Hash(com\_receiver || chain\_constant) (using post-quantum SHA3-256 or BLAKE3).
- Sender hint (optional): blind\_sender\_hint = Hash(sk\_sender || nonce) (short, e.g., 64-bit, refreshed per tx for freshness).

**b. Concatenate and hash to form raw selector:**

- Raw input: blind\_receiver || blind\_sender\_hint || height\_estimate || other\_metadata
  - height\_estimate: Approximate current chain height (from light-client headers) to handle shard splits/merges over time.
  - other\_metadata: Optional fields like tx type flags (all blinded).

**c. Final shard selector:**

- shard\_selector = Hash(raw\_input) → a 256-bit value.
- This is **deterministic**: Given the same inputs, every honest wallet (or replaying node) computes the identical selector.

**d. Map selector to actual shard\_id:**

- The network maintains a public **shard registry** (a compact key-value map in the global consensus state, updated on splits/merges).
- Locally, the wallet looks up: target\_shard\_id = ShardRegistry.lookup(shard\_selector)

- The registry is a hashed prefix tree or sorted list: e.g., shards sorted by their id ranges, selector mod  $2^k$  routes to the covering shard.
- Because splits/merges update the registry atomically (gossiped in <800 ms), all light clients have a consistent view.

### Why deterministic?

All participants use the same public chain constants, hash functions, and synced registry. No randomness — ensures the tx reaches the shard containing the logical accounts.

## 2. Identifying the Exact Active Shard Without Revealing Plaintext Details

The selector must be sent with the transaction, but **cannot leak information** to passive observers, active adversaries, or even honest-but-curious relays.

### Blinding and inclusion in payload:

#### a. Encrypt the selector:

- The shard\_selector is encrypted to the **public key of the ingress TEE cluster** (multi-vendor attested enclaves).
- Or more efficiently: Included as a **zero-knowledge commitment** (e.g., Pedersen or lattice-based) with a proof that it matches the routing logic.

#### b. Payload structure (simplified):

text

```
private_tx = {
    blinded_payload: Encrypt( $\delta$ (tx), Halo2_proof_validity, ...),
    blinded_shard_selector: Commit(shard_selector) OR Encrypt(shard_selector, pk_ingress),
    routing_hint: Optional short prefix of shard_selector (e.g., 32 bits) for efficient DHT pre-filtering,
    user_sig: Dilithium signature over whole blob
}
```

#### c. Onion routing and final relay handling (inside TEEs):

- The 5-hop TEE mixer forwards the onion based on outer layers (no knowledge of selector).
- The **final relay** (or a dedicated ingress enclave):
  - Decrypts/reveals the blinded\_shard\_selector inside the attested TEE.

- Queries the public shard registry (sealed inside the TEE) to resolve `shard_selector` → `target_shard_id`.
- Forwards the still-encrypted private payload to the target shard's encrypted mempool channel.
- Adds its own attestation proving correct routing.

### Privacy guarantees:

- **Passive observers:** See only fixed-size padded packets — no visible shard identifier.
- **Intermediate relays:** Never see the selector (onion layers).
- **Final relay/ingress TEE:** Sees the selector but operates inside hardware enclave with remote attestation — proven to run correct audited code, no side channels, memory wiped after.
- **Target shard nodes:** Receive the blinded transaction without knowing origin (thanks to mixer) or even full selector details.
- **Formal anonymity:** The selector is either encrypted (indistinguishable) or committed (hiding). Combined with cover traffic and jitter, ProVerif model proves no linkage even against global adversaries.

### Edge cases:

- **Shard split after tx submission:** If the target shard splits mid-flight, the ingress TEE detects via latest registry and re-routes to the correct child (using the same deterministic selector → child mapping).
- **Cross-shard (rare):** Future extensions use intent-based private matching; basic transfers are encouraged intra-shard via wallet clustering.

In summary, the **deterministic blinded hash selector** allows the wallet to privately and correctly target one shard among thousands, while the **TEE-bound routing** ensures the network can act on it without ever exposing plaintext account or shard information. This preserves NERV's core promise: infinite horizontal scaling with unconditional privacy.

### Shards are created first — independently of any specific transaction.

In NERV's dynamic neural sharding system, **shards (and their embeddings `e_t`)** exist and evolve based on **network load**, not based on individual transactions or accounts being created.

Here's the precise order and relationship:

## 1. Shard Lifecycle (Independent of Transactions)

- The network starts at genesis with a small number of initial shards (likely 1 or a handful).
- Shards **split** automatically when an existing shard's load (TPS, batch size, latency) exceeds thresholds predicted by the LSTM model (>92% overload probability).
- Shards **merge** when underutilized (e.g., <10 TPS sustained).



- These splits/merges happen proactively or reactively based on aggregate traffic patterns — **not triggered by a specific user wanting to send a transaction**.
- When a split occurs, the parent embedding is bisected deterministically, child shard\_ids are derived, and the public shard registry (part of global consensus state) is updated atomically.

→ At any moment, there is a fixed set of **active shards** with known shard\_ids and root embeddings, visible to all light clients.

## 2. Account Creation and Transaction Preparation (Dependent on Existing Shards)

- When a user first creates an account (or derives a new receiving commitment), the wallet must assign it to an existing shard.
- The wallet does this by:
  - Computing a **raw shard selector** from the account's commitment (or seed-derived key): e.g., Hash(com\_receiver || chain\_constant).
  - Looking up that selector in the **current public shard registry** (the DAG lattice map synced via light client).
  - Determining the **current active shard\_id** that covers that selector range.
- The new account is now logically "in" that shard — even though nothing is revealed on-chain (privacy by default).
- When sending a transfer:
  - The wallet uses the **receiver's commitment** (and optionally sender hint) to recompute the same raw selector.
  - It maps it to the **current shard\_id** using the latest registry.
  - It generates  $\delta(tx)$  relative to that shard's embedding space.
  - It includes the blinded selector in the tx for routing.

## 3. Key Clarification: Which Comes First?

**The shard always comes first.**

- You cannot create a transaction targeting a shard that doesn't yet exist.
- If a split occurs **after** an account was assigned but **before** a transaction is sent, the wallet will use the **updated registry** and route to the correct child shard (the deterministic selector ensures it lands in the right descendant).
- The system is designed so that account-to-shard mapping is **stable over time** unless a split intervenes — in which case the bisection preserves logical grouping (accounts that transact together tend to stay together due to latent space clustering).

## Analogy

Think of shards as **cities** that grow and divide when overcrowded:

- Cities (shards) are created and split based on population/load.
- When you get a new postal address (account commitment), it's assigned to an **existing** city based on a deterministic hash.
- When you mail a letter (transaction), you address it using the current city map — even if the city split in the meantime, the postal system (TEE routing) forwards it correctly.

**No transaction can "create" a new shard on demand.** Shards evolve autonomously with network demand. Transactions always target **existing** shards via deterministic, registry-looked-up routing.

This ensures infinite horizontal scalability while maintaining full privacy and efficient homomorphic execution.

DHT stands for **Distributed Hash Table**.

In the context of NERV (and many other peer-to-peer blockchain networks like BitTorrent, IPFS, or libp2p-based systems), a DHT is a decentralized key-value store spread across all nodes in the network. There's no central server — instead, nodes collaboratively store and retrieve data based on cryptographic hashes.

## How DHT Works in Simple Terms

- Every piece of data (e.g., a shard's metadata, a node's address, or routing tables) has a **key** (usually a hash of the data itself).
- Nodes in the network are each assigned an ID in the same key space.
- Data with a given key is stored on the nodes whose IDs are "closest" to that key (measured by XOR distance or similar).
- To find data, a node "walks" the network by asking closer and closer nodes until it reaches the ones storing the value.

## Why NERV Uses DHT

In NERV's architecture, the DHT serves several critical privacy-preserving and scalability roles:

1. **Discovery of active shards and routing tables**
  - The public shard registry (mapping shard selectors to active shard\_ids and embedding roots) is stored and gossiped via DHT.
  - Light clients and wallets query the DHT to stay updated on the current lattice of shards after splits/merges.
2. **Node discovery and peer routing**
  - Nodes find each other (for gossip, consensus messages, or mixer relays) without centralized coordinators.
3. **Anonymous ingress support**
  - Wallets query the DHT for lists of currently active, attested TEE relays to build their 5-hop onion routes.

#### 4. Fast global updates

- When a shard splits or merges, the new child shard\_ids and roots are published to the DHT with global propagation in typically <800 ms.

NERV likely uses a modern DHT implementation based on **Kademlia** (the most common design, used in Ethereum, IPFS, and libp2p), enhanced with post-quantum security and integration with the 5-hop TEE mixer for privacy.

In short: DHT is the decentralized "phone book" and bulletin board that lets NERV's thousands of nodes and dynamic shards coordinate efficiently without any single point of control or failure.

**Protobuf** (short for **Protocol Buffers**) is a language-agnostic, platform-neutral, extensible mechanism developed by Google for **serializing structured data**. It is widely used in high-performance systems for efficient data interchange, storage, and communication.

### Core Features of Protobuf

- **Compact binary format:** Much smaller and faster to encode/decode than text-based formats like JSON or XML (typically 3–10× smaller, 5–100× faster).
- **Schema-driven:** You define data structures in .proto files (human-readable schema), then generate code in many languages (Rust, Go, C++, Python, Java, etc.).
- **Backward/forward compatibility:** Fields can be added/removed without breaking old versions (via field numbers).
- **Deterministic serialization:** Critical for cryptographic applications (same input → identical bytes).
- **Efficient parsing:** No reflection needed at runtime; generated code is highly optimized.

Example .proto snippet:

proto

```
message Vdw {  
  
    bytes tx_hash = 1;  
  
    uint64 shard_id = 2;  
  
    uint64 lattice_height = 3;  
  
    bytes delta_proof = 4;  
  
    bytes final_root = 5;  
  
    bytes attestation = 6;  
  
    bytes signature = 7;  
}
```

}

## Importance in NERV's Blockchain Design

NERV is engineered for **extreme efficiency** (sub-second finality, >1M TPS, tiny proofs) while maintaining **full privacy** and **post-quantum security** in a highly distributed environment with hardware enclaves (TEEs). Protobuf plays a critical role in several areas:

### 1. Verifiable Delay Witness (VDW) Structure

- The VDW is NERV's user-facing permanent receipt (~1.4 KB average).
- Defined as a fixed-format Protobuf message for deterministic layout and minimal overhead.
- Enables precise size bounds ( $\leq 1.8$  KB max), byte-range HTTP serving, and efficient CDN/IPFS/Arweave pinning.

### 2. Network Message Efficiency

- All gossip, consensus votes, onion routing layers, and mempool messages use Protobuf.
- Critical for sub-second latencies: smaller messages → lower bandwidth → faster propagation in a network with thousands of nodes and dynamic shards.

### 3. TEE Enclave Interfaces

- Attested enclaves (SGX, SEV-SNP, etc.) seal and unseal data using strictly defined Protobuf schemas.
- Guarantees no ambiguity in what data enters/exits the enclave → stronger remote attestation and auditability.

### 4. Deterministic Cryptographic Signing

- Protobuf's canonical serialization ensures identical bytes for the same logical data.
- Essential when signing payloads with Dilithium-3 (post-quantum signatures) across the 5-hop mixer and consensus.

### 5. Cross-Language Interoperability

- NERV nodes and wallets can be implemented in multiple languages (Rust core, Go/Swift wallets, etc.).
- Protobuf-generated code ensures seamless communication without custom parsers.

### 6. Long-Term Archival and Verification

- VDWs pinned to Arweave/IPFS use Protobuf for century-scale reliability.
- Even decades later, offline wallets can parse and verify VDWs with the same schema.

In summary, **Protobuf is the glue that enables NERV's combination of extreme performance, succinctness, and cryptographic rigor**. Without a compact, deterministic, schema-driven format, achieving 1.4 KB permanent proofs, sub-second global gossip, and hardware-enforced privacy boundaries would be significantly harder or impossible at NERV's target scale. It is a foundational (if understated) component of the "infinitely scalable" promise.

## General Overview of CRYSTALS-Dilithium (ML-DSA)

**CRYSTALS-Dilithium** (standardized by NIST as **ML-DSA** in FIPS 204, August 2024) is a lattice-based digital signature scheme designed to be secure against attacks from both classical and quantum computers. It is based on the hardness of two lattice problems: **Module-Learning With Errors (Module-LWE)** and **Module-Short Integer Solution (Module-SIS)**.

Unlike classical signatures (e.g., ECDSA, which rely on elliptic curves vulnerable to Shor's quantum algorithm), Dilithium uses high-dimensional mathematical lattices — grids of points in space where finding short vectors is computationally hard, even for quantum computers.

It follows the **Fiat-Shamir with aborts** paradigm:

- A signer generates a "commitment" (masked computation).
- A challenge is derived deterministically from the message and commitment (via hash, turning interactive proof into non-interactive signature).
- The signer responds with a "response" that reveals limited information.
- Rejection sampling ("aborts") ensures the response doesn't leak the secret key — the signer loops until the response is safe to reveal.

Key advantages:

- No trusted setup.
- Strong security proofs in the random oracle model (and quantum ROM).
- Efficient, constant-time implementations (avoids Gaussian sampling for side-channel resistance).

**Dilithium-3** specifically targets **NIST Security Level 3** (equivalent to AES-192 classical security, resistant to attacks costing  $\sim 2^{143}$  operations post-quantum).

Parameters (approximate):

- Public key: ~1,952 bytes
- Private key: ~4,000 bytes
- Signature: ~3,293 bytes
- Verify speed: ~58  $\mu$ s (optimized AVX-512)

### High-Level Signing Process

#### 1. Key Generation:

- Generate secret vectors  $\mathbf{s}_1, \mathbf{s}_2$  (small coefficients).
- Generate public matrix  $\mathbf{A}$  (from seed).
- Compute  $\mathbf{t} = \mathbf{A} \cdot \mathbf{s}_1 + \mathbf{s}_2$ .
- Public key: (A seed,  $t_1$ ) where  $t$  is decomposed/compressed into high/low bits ( $t = t_1 \cdot 2^d + t_0$ ).
- Private key: ( $s_1, s_2$ , etc.).

2. **Signing** (for message  $\mu$ ):
  - Mask secret with large random  $\mathbf{y}$  (uniform high-entropy polynomial vector).
  - Compute commitment  $\mathbf{w} = \mathbf{A} \cdot \mathbf{y}$ .
  - Hash  $\mu \parallel \mathbf{w}_1$  (high bits of  $\mathbf{w}$ ) to get challenge  $\mathbf{c}$  (sparse polynomial with fixed number of  $\pm 1$  coefficients).
  - Compute response  $\mathbf{z} = \mathbf{y} + \mathbf{c} \cdot \mathbf{s}_1$ .
  - If  $\mathbf{z}$  is too large (risks leaking  $\mathbf{s}_1$ ), reject and restart (aborts ~few times on average).
  - Include "hint"  $\mathbf{h}$  to help verifier reconstruct  $\mathbf{w}_1$  from  $\mathbf{z}$ .
  - Signature:  $(\mathbf{z}, \mathbf{h}, \mathbf{c})$ .
3. **Verification**:
  - Recompute  $\mathbf{w}' = \mathbf{A} \cdot \mathbf{z} - \mathbf{c} \cdot \mathbf{t}_1 \cdot 2^d$  (using hint to adjust).
  - Check if high bits of  $\mathbf{w}'$  match the hashed challenge  $\mathbf{c}$ , and  $\|\mathbf{z}\|$  is bounded.

Security relies on: even with many signatures, recovering short  $\mathbf{s}_1/\mathbf{s}_2$  from noisy equations is hard (Module-LWE/SIS).

## Overview of Dilithium-3 as a Digital Signature Scheme

Dilithium-3 is a **non-interactive** lattice-based digital signature scheme, standardized by NIST as part of post-quantum cryptography (FIPS 204, ML-DSA). It is non-interactive, meaning there is **no back-and-forth** during the signing of a specific message. The signer computes the entire signature alone and sends it (with the message) to the verifier in one go.

However, the design is derived from an **interactive identification protocol** (a 3-move  $\Sigma$ -protocol) that is made non-interactive using the **Fiat-Shamir transform** (replacing the verifier's random challenge with a hash of the message and commitment). This is why the user's earlier high-level description feels like an "exchange" — it mirrors the underlying interactive steps.

I will explain:

1. The **underlying interactive protocol** (to show the "back-and-forth" intuition).
2. How it becomes **non-interactive** for signatures.
3. **All terms, symbols, and expressions** in detail.
4. **What is pre-computed, what each party knows, and what is exchanged.**
5. A **toy example** using simplified scalar (non-ring) parameters for clarity (real Dilithium uses polynomials and modules, but the math is identical in structure).

All operations are in the **module over the ring**  $R_q = \mathbb{Z}_q[X]/(X^{256} + 1)$ , with  $q = 8380417q = 8380417q = 8380417$ . Bold lowercase letters are vectors, uppercase are matrices. All norms are coefficient-wise infinity norms unless specified.

### Parameters for Dilithium-3 (for context)

- Dimension: module rank  $k=6$ ,  $l=5$  →  $A$  is  $6 \times 56$  matrix, vectors have 6 or 5 entries.
- $\eta=4$  (secret coefficients bounded by 4).
- $\gamma_1=219=524288$  (masking bound).
- $\gamma_2=95232$  (decomposition parameter).
- $d=13$  (rounding bits for  $t$ ).
- $\tau=49$  (challenge has exactly 49 coefficients of  $\pm 1$ ).
- $\beta=196$  (bound on  $\|c \cdot s_1\|_\infty = \tau \cdot \eta$ ).

## 1. Underlying Interactive Identification Protocol (the "back-and-forth")

This proves knowledge of short  $s_1, s_2$  such that  $t \approx A s_1 + s_2$  (approximate due to compression of  $t$ ).

**Prover (signer)** knows: private key ( $s_1, s_2$ )

**Verifier** knows: public key (seed  $p$  for  $A, t_1$ ) → can compute  $A$  and approximate  $t' = t_1 \cdot 2^d$

### Round 1 (Commitment) – Prover → Verifier

- Prover samples large random masking vector  $y$  (each coefficient uniform in  $-(\gamma_1-1)$  to  $(\gamma_1-1)$ ).
- Computes  $w = A y$  ( $k$ -vector).
- Computes  $w_1 = \text{HighBits}(w, 2\gamma_2)$  (high-order parts after decomposition).
- Sends  $w_1$  to verifier.

### Round 2 (Challenge) – Verifier → Prover

- Verifier picks a random challenge  $c$  uniformly from the set of polynomials with exactly  $\tau$  coefficients  $\pm 1$  (rest 0) and sends  $c$ .

### Round 3 (Response) – Prover → Verifier

- Prover computes  $z = y + c \cdot s_1$  (componentwise polynomial multiplication).
- If  $\|z\|_\infty \geq \gamma_1 - \beta$ , **abort and restart entire protocol** (rejection sampling — prevents information leakage about  $s_1$ ; happens with low probability in practice).
- Computes hint  $h = \text{MakeHint}(w - c \cdot s_2 + \text{low parts from decomposition}, \dots)$  — a sparse bit vector indicating where carry-over occurred in the low bits.
- Sends  $(z, h)$ .

### Verification (Verifier)

- Computes  $w' \approx A z - c \cdot t'$  (using  $h$  to adjust for possible carry in high bits).
- Checks  $\text{HighBits}(w') = w_1$  and  $\|z\|_\infty < \gamma_1 - \beta$ .
- If both hold, accept (prover knows short secret).

Security: Even with many transcripts, recovering short  $s_1/s_2$  is hard (Module-LWE and Module-SIS problems).

## 2. Non-Interactive Signature Scheme (Actual Dilithium-3)

The interactive protocol is transformed via **Fiat-Shamir**:

- The challenge  $c$  is derived deterministically as  $c \leftarrow H(\text{truncated } \mu \parallel w_1)$  (hash-to-challenge, ensuring exactly  $\tau \pm 1$ 's).
- The message  $\mu$  is bound in the hash.
- Aborts/restarts happen privately (signer loops internally,  $\sim 1-3$  times on average).

**Pre-computed (once, offline, by signer):**

- Key pair  $(pk, sk)$ .

**Key Generation (signer only):**

- Generate random seed  $p$ .
- $A \leftarrow \text{ExpandA}(p)$  (public  $k \times l$  matrix, deterministic — anyone with  $p$  can recompute  $A$ ).
- Sample short secret vectors  $s_1$  ( $l$ -vector, coeffs  $\in [-\eta, \eta][-\eta, \eta][-\eta, \eta]$ ),  $s_2$  ( $k$ -vector, same).
- Compute  $t = A s_1 + s_2$  (exact,  $k$ -vector).
- Decompose  $t = t_1 \cdot 2^d + t_0$  using Power2Round ( $t_1$  has roughly  $q/2^d$  bits per coefficient,  $t_0$  small:  $|t_0 \text{ coeffs}| \leq 2^{\{d-1\}}$ ).
- **Public key**  $pk = (p, t_1)$  (small,  $\sim 1-2$  KB).
- **Private key**  $sk = (p, t_1, s_1, s_2, \text{additional seeds/tr} = H(pk))$ .
- Signer publishes/shares  $pk$  once.

**What both parties know after key distribution:**

- Both:  $pk \rightarrow$  both can compute  $A$  and  $t' = t_1 \cdot 2^d$  (approximate  $t$ ).

**What only signer knows:**

- $s_1, s_2$  (short secrets).

**Signing a message  $\mu$  (signer only, loops until success):**

1. Sample large random  $y$  ( $l$ -vector, coeffs uniform  $-(\gamma_1-1) \dots (\gamma_1-1) \dots (\gamma_1-1)$ ).
2. Compute  $w = A y$  ( $k$ -vector).
3. Compute  $w_1 = \text{HighBits}(w, 2\gamma_2)$  (decomposes each coefficient:  $w \approx w_1 \cdot 2\gamma_2 + \text{low part}$ , with low part bounded).
4. Compute challenge  $c \leftarrow H(H(\mu) \parallel w_1)$  (hash produces bitstring expanded to sparse polynomial with exactly  $\tau \pm 1$ 's).
5. Compute  $z = y + c \cdot s_1$  (componentwise poly mult).
6. If  $\|z\|_\infty \geq \gamma_1 - \beta \rightarrow$  **reject, restart from step 1**.



7. Compute hint  $\mathbf{h} = \text{MakeHint}(\mathbf{A} \mathbf{z} - \mathbf{c} \cdot \mathbf{t}_1 \cdot 2^d - \mathbf{c} \cdot \mathbf{t}_0, \mathbf{w})$  (sparse vector indicating positions where low bits caused carry into high bits;  $\|\mathbf{h}\|_0$  bounded).
8. Output signature  $\sigma = (\mathbf{z} \text{ (compressed)}, \mathbf{h}, \mathbf{c})$ .

#### Exchange (one-way):

- Signer  $\rightarrow$  Verifier: message  $\mu$  + signature  $\sigma = (\mathbf{z}, \mathbf{h}, \mathbf{c})$ .

#### Verification (verifier, using only pk and $\mu$ ):

1. Recompute  $\mathbf{A}$  from  $\mathbf{p}$ .
2. Recompute candidate  $\mathbf{w}' \approx \mathbf{A} \mathbf{z} - \mathbf{c} \cdot \mathbf{t}_1 \cdot 2^d$ .
3. Use hint  $\mathbf{h}$  to adjust  $\mathbf{w}'$  in positions where carry likely occurred  $\rightarrow$  get  $\mathbf{w}_1' = \text{HighBits}(\mathbf{w}', 2\gamma_2)$  with hint.
4. Check  $\mathbf{H}(\mathbf{H}(\mu) \parallel \mathbf{w}_1') = \mathbf{c}$ .
5. Check  $\|\mathbf{z}\|_\infty < \gamma_1 - \beta$ .
6. If both pass, accept.

### 3. Detailed Explanation of Terms and Expressions

- $\mathbf{A}$ : Public  $k \times l$  matrix of polynomials (generated deterministically from seed  $\mathbf{p}$ ). Known to both.
- $\mathbf{s}_1, \mathbf{s}_2$ : Short secret vectors ( $l$ - and  $k$ -length). Coefficients  $\in [-\eta, \eta][-\eta, \eta][-\eta, \eta]$ . Only signer knows.
- $\mathbf{t} = \mathbf{A} \mathbf{s}_1 + \mathbf{s}_2$ : Exact public vector before compression.
- $\mathbf{t} = \mathbf{t}_1 \cdot 2^d + \mathbf{t}_0$ : Decomposition (Power2Round).  $\mathbf{t}_1$  sent in pk,  $\mathbf{t}_0$  small and secret.
- $\mathbf{y}$ : Large masking vector ( $l$ -length). Private to signer per signature. Hides  $\mathbf{s}_1$ .
- $\mathbf{w} = \mathbf{A} \mathbf{y}$ : Commitment vector ( $k$ -length).
- $\mathbf{w}_1 = \text{HighBits}(\mathbf{w}, 2\gamma_2)$ : High-order part of  $\mathbf{w}$  after decomposition (Decompose function ensures  $\mathbf{w} = \mathbf{w}_1 \cdot 2\gamma_2 + \text{low}$ ,  $|\text{low}|$  bounded).
- $\mathbf{c}$ : Sparse challenge polynomial (single element, multiplied componentwise). Exactly  $\tau$  coefficients  $\pm 1$ .
- $\mathbf{z} = \mathbf{y} + \mathbf{c} \cdot \mathbf{s}_1$ : Response vector. Large, but statistically independent of  $\mathbf{s}_1$  due to rejection.
- $\mathbf{h}$ : Hint vector ( $k$ -length, sparse). Indicates where adding 1 to high bits recovers correct  $\mathbf{w}_1$  despite noise from  $\mathbf{c} \cdot \mathbf{s}_2$  and  $\mathbf{t}_0$ .
- **Rejection sampling**: Ensures  $\mathbf{z}$ 's distribution hides  $\mathbf{s}_1$  ( $\mathbf{z}$  looks like fresh large random vector).
- **Security basis**: Module-LWE (finding short  $\mathbf{s}$  from  $\mathbf{A}, \mathbf{t}$  hard) and Module-SIS (forging without  $\mathbf{s}$  hard).

### 4. Toy Example (Simplified Scalar Version)

Real Dilithium uses polynomials and modules, but here we use plain integers mod  $q$  to show the flow numerically ( $k=l=1$ , single "A"). Parameters are tiny/toy (not secure).

Let  $q=769$ ,  $d=8$  ( $2^d=256$ ),  $\eta=2$ ,  $\gamma_1=60$ ,  $\gamma_2=100$ ,  $\tau=2$  (challenge  $\pm 1$  with 2 non-zeros, but since scalar,  $c \in \{-1,0,1\}$  with  $|c|=1$  for simplicity).

### KeyGen:

- $A = 123$  (public)
- $s_1 = 2, s_2 = -1$  (short)
- $t = 123 \cdot 2 + (-1) = 245$
- Decompose:  $t_1 = 0, t_0 = 245$  (bad example, but suppose  $t=300 \rightarrow t_1=1, t_0=300-256=44$ )
- Better:  $t=280 \rightarrow t_1=1, t_0=24$
- $pk = (A=123, t_1=1)$

### Signing $\mu$ (non-interactive):

1. Sample  $y = 55$  (large,  $|y| < 60$ )
2.  $w = 123 \cdot 55 = 6765$
3.  $w_1 = \text{HighBits}(6765, 200) \approx \text{floor}(6765 / 200) = 33$  (simplified, assume low part small)
4.  $c = H(\mu \parallel 33) \rightarrow$  suppose hash gives  $c=1$
5.  $z = 55 + 1 \cdot 2 = 57$
6. Check  $|z|=57 < 60-2=58$ ? No  $\rightarrow$  reject, restart. New  $y=50$   $w=123 \cdot 50=6150$   $w_1 \approx 30$   
 $c=H(\mu \parallel 30) \rightarrow c=-1$   $z=50 + (-1) \cdot 2=48$   $|z|=48 < 58 \rightarrow$  accept
7. Hint  $h$ : compute  $A z - c t$   $2^d = 123 \cdot 48 - (-1) \cdot 1 \cdot 256 = 5904 + 256 = 6160$   $\text{HighBits}(6160, 200)$  should equal 30 with possible  $+1$  adjustment  $\rightarrow$  suppose no carry,  $h=0$

Signature:  $(z=48, h=0, c=-1)$

### Verification:

- $A z - c t' = 123 \cdot 48 - (-1) \cdot 256 = 5904 + 256 = 6160$
- $\text{HighBits}(6160, 200) = 30$  (matches  $w_1$  that hashed to  $c$ )
- $|z| < 58 \rightarrow$  accept.

This shows masking ( $y$  hides  $s_1$ ), rejection (prevents  $z$  too large), and hint/recovery of commitment.

In real Dilithium, everything is a vector of 256-coefficient polynomials, but the structure is identical.

### What is $R_q = \mathbb{Z}_q[X]/(X^{256} + 1)$ ?

What is  $R_q = \mathbb{Z}_q[X] / (X^{256} + 1)$ ?

In Dilithium (and other lattice-based schemes like Kyber), computations happen in a special algebraic structure called a ring. This ring is denoted  $R_q = \mathbb{Z}_q[X] / (X^{256} + 1)$ . It combines ideas from polynomial arithmetic and modular arithmetic to create an efficient, structured lattice for cryptographic hardness.

I'll break this down step by step, from basic concepts to the full structure, with concrete examples.

### Step 1: Polynomial Rings – $\mathbb{Z}[X]$ and $\mathbb{Z}_q[X]$

- Start with ordinary polynomials with integer coefficients:  $\mathbb{Z}[X]$ .
  - Elements:  $f(X) = a_0 + a_1 X + a_2 X^2 + \dots + a_d X^d$  where  $a_i$  in  $\mathbb{Z}$  (any integers).
  - Addition: coefficient-wise.
  - Multiplication: standard polynomial multiplication (convolution of coefficients).
- Now reduce coefficients modulo  $q$ : this gives  $\mathbb{Z}_q[X]$  (also written  $(\mathbb{Z}/q\mathbb{Z})[X]$ ).
  - Coefficients are in  $\{0, 1, \dots, q-1\}$  (or balanced representatives like  $-\lfloor q/2 \rfloor$  to  $\lfloor (q-1)/2 \rfloor$ ).
  - For Dilithium-3:  $q = 8380417$  (a prime chosen for efficiency with NTT – Number Theoretic Transform).
  - Example (small  $q=17$  for toy):  $f(X) = 15 + 3X + 20X^2 \equiv 15 + 3X + 3X^2 \pmod{17}$  since  $20 \equiv 3 \pmod{17}$ .

### Step 2: Quotient by an Ideal – $/ (X^{256} + 1)$

- To make the ring finite and suitable for cryptography, we quotient by the ideal generated by the polynomial  $\phi(X) = X^{256} + 1$ .
  - This means we work with polynomials of degree less than 256, and we enforce the relation  $X^{256} \equiv -1 \pmod{\phi(X)}$ .
  - Higher powers reduce:  $X^{256} = -1$ ,  $X^{257} = -X$ ,  $X^{512} = 1$ , etc. (It's a cyclotomic ring, specifically suited for NTT-friendly parameters.)
- Full ring:  $R_q = \mathbb{Z}_q[X] / (X^{256} + 1)$ .
  - Elements: polynomials  $f(X) = \sum_{i=0}^{255} a_i X^i$  with each  $a_i$  in  $\mathbb{Z}_q$  (0 to  $q-1$ , or balanced).
  - There are exactly  $q^{256}$  elements (huge, but structured).
  - Addition: coefficient-wise modulo  $q$ .
  - Multiplication: standard polynomial multiplication, then reduce modulo  $X^{256} + 1$  (i.e., when degree  $\geq 256$ , subtract multiples using  $X^{256} = -1$ ).

### Example of Operations in $R_q$ (Toy Parameters)

Use tiny  $n=4$ ,  $q=17$ , ring  $R = \mathbb{Z}_{17}[X]/(X^4 + 1)$  for illustration (real Dilithium has  $n=256$ ).

- Element a:  $a(X) = 3 + 5X + 2X^2 + 15X^3$
- Element b:  $b(X) = 1 + 16X + 4X^2 + 7X^3$  (note  $16 \equiv -1 \pmod{17}$ )

Addition: coefficient-wise mod 17

$$a + b = (3+1) + (5+16)X + (2+4)X^2 + (15+7)X^3 = 4 + 21X + 6X^2 + 22X^3 \equiv 4 + 4X + 6X^2 + 5X^3 \pmod{17}$$

Multiplication: First compute raw product (degree up to 6):

- Constant:  $3 \cdot 1 = 3$
- $X$ :  $3 \cdot 16 + 5 \cdot 1 = 48 + 5 = 53 \equiv 2 \pmod{17}$
- $X^2$ :  $3 \cdot 4 + 5 \cdot 16 + 2 \cdot 1 = 12 + 80 + 2 = 94 \equiv 9 \pmod{17}$
- $X^3$ :  $3 \cdot 7 + 5 \cdot 4 + 2 \cdot 16 + 1 \cdot 15 = 21 + 20 + 32 + 15 = 88 \equiv 3 \pmod{17}$
- $X^4$ :  $5 \cdot 7 + 2 \cdot 4 + 15 \cdot 16 = 35 + 8 + 240 = 283 \equiv 13 \pmod{17}$
- $X^5$ :  $2 \cdot 7 + 1 \cdot 5 = 14 + 5 = 19 \equiv 2 \pmod{17}$
- $X^6$ :  $1 \cdot 7 = 7 \equiv 7 \pmod{17}$

Now reduce modulo  $X^4 + 1$ :

- $X^4 \equiv -1$
- $X^5 \equiv -X$
- $X^6 \equiv -X^2$

So add:

- $X^4$  term (13)  $\rightarrow 13 \cdot (-1)$  to constant
- $X^5$  term (2)  $\rightarrow 2 \cdot (-X)$  to  $X$
- $X^6$  term (7)  $\rightarrow 7 \cdot (-X^2)$  to  $X^2$

Final:

- Constant:  $3 + 13 \cdot (-1) = 3 - 13 = -10 \equiv 7 \pmod{17}$
- $X$ :  $2 + 2 \cdot (-1) = 2 - 2 = 0 \pmod{17}$
- $X^2$ :  $9 + 7 \cdot (-1) = 9 - 7 = 2 \pmod{17}$
- $X^3$ : 3

Thus  $a \cdot b \equiv 7 + 13X + 6X^2 + 3X^3 \pmod{17, X^4+1}$

In real implementations, multiplication is optimized using NTT (like FFT but over finite fields) because  $q$  is chosen NTT-friendly.

Step 3: Modules over the Ring –  $R_q^m$

- A module over  $R_q$  is like a vector space, but with scalar multiplication from the ring (not just a field).
- In Dilithium:
  - Vectors are in  $R_q^l$  or  $R_q^k$  ( $l=5, k=6$  for Dilithium-3, some parameters vary).
  - Each "entry" is a full polynomial in  $R_q$  (256 coefficients).
  - So a vector  $s$  in  $R_q^l$  is  $l$  polynomials, total  $256 \cdot l$  coefficients.
  - Matrix  $A$  in  $R_q^{k \times l}$  is  $k$  rows  $\times$   $l$  columns of polynomials.
- Operations:
  - Vector addition: component-wise (each component is a polynomial addition).

- Matrix-vector multiplication:  $A \cdot y$  means row-column dot products, where "dot" is sum of ring multiplications.
- Example: single entry of result =  $\sum_j A_{ij} \cdot y_j$  (ring multiplication and addition).

This structure gives a module lattice: the set of all  $A \cdot v$  for short  $v$  forms a lattice with good hardness properties (Module-LWE/SIS).

Why This Structure in Dilithium?

- Efficiency: Ring structure allows fast multiplication via NTT ( $O(n \log n)$  instead of  $O(n^2)$ ).
- Security: Based on hard problems on ideal lattices (Module-LWE: distinguishing  $A \cdot s + e$  from random, with  $s, e$  short).
- Compactness: Public key (matrix  $A$  + vector  $t$ ) is smaller than unstructured lattices.

In summary:

- $R_q$  elements = 256-coeff polynomials mod  $q$ , with  $X^{256} = -1$ .
- Arithmetic: poly add/mult with mod  $q$  and mod  $(X^{256} + 1)$ .
- Modules: vectors/matrices of these polynomials  $\rightarrow$  structured high-dimensional lattices.

This enables secure, practical post-quantum signatures.

### Step 3: Modules over the Ring – $R_q^m R_q^m$

- A **module** over  $R_q$  is like a vector space, but with scalar multiplication from the ring (not just a field).
- In Dilithium:
  - Vectors are in  $R_q^l$  or  $R_q^k$  ( $l=5, k=6$  for Dilithium-3, some parameters vary).
  - Each "entry" is a full polynomial in  $R_q$  (256 coefficients).
  - So a vector  $s_i \in R_q^l$  is  $l$  polynomials, total  $256 \cdot l$  coefficients.
  - Matrix  $A \in R_q^{k \times l}$  is  $k$  rows  $\times$   $l$  columns of polynomials.
- Operations:
  - Vector addition: component-wise (each component is a polynomial addition).
  - Matrix-vector multiplication:  $A \cdot y$  means row-column dot products, where "dot" is sum of ring multiplications.
  - Example: single entry of result =  $\sum_j A_{ij} \cdot y_j$  (ring multiplication and addition).

This structure gives a **module lattice**: the set of all  $A \cdot v$  for short  $v$  forms a lattice with good hardness properties (Module-LWE/SIS).

Why This Structure in Dilithium?

- **Efficiency:** Ring structure allows fast multiplication via NTT ( $O(n \log n)$  instead of  $O(n^2)$ ).
- **Security:** Based on hard problems on ideal lattices (Module-LWE: distinguishing  $A \cdot s + e$  from random, with  $s, e$  short).
- **Compactness:** Public key (matrix  $A$  + vector  $t$ ) is smaller than unstructured lattices.

In summary:

- $R_q$  elements = 256-coeff polynomials mod  $q$ , with  $X^{256} = -1$ .
- Arithmetic: poly add/mult with mod  $q$  and mod  $(X^{256} + 1)$ .
- Modules: vectors/matrices of these polynomials  $\rightarrow$  structured high-dimensional lattices.

This enables secure, practical post-quantum signatures.

### 1. Check: $H(H(\mu) \parallel w_1) = c$

What this equation means:

- This is the **Fiat-Shamir consistency check**. It verifies that the recomputed high-bits of the commitment ( $w_1$ ) — derived from the response  $z$ , hint  $h$ , and challenge  $c$  — exactly match the  $w_1$  that was originally used (by the signer) to compute the challenge  $c$  via hashing with the message.
- In other words: If the prover (signer) is honest, the verifier should recover the *same* commitment high-bits that "caused" the challenge  $c$  when hashed together with the message.
- If this holds, it proves:
  - The signature is bound to this specific message  $\mu$  (changing  $\mu$  would change the hash  $\rightarrow$  different  $c$ ).
  - The prover committed to  $w_1$  before knowing  $c$  (non-interactive simulation of the interactive protocol).
  - No forgery: An attacker without the secret couldn't produce a  $z$  and  $h$  that recover a  $w_1$  hashing to the given  $c$ .

How the variables are calculated/initialized:

- $\mu$ : The original message being signed (arbitrary byte string, provided by the signer along with the signature).
- $H$ : A cryptographic hash function. In Dilithium:
  - The outer  $H$  is typically SHAKE-256 (an extendable-output function) used in "challenge mode" to produce  $c$ .
  - There is an inner hash for message processing:
    - First, compute a **message representative**  $\mu'$  (often denoted as  $H(\mu)$  in simplified explanations, but precisely:  $\mu' = \text{SHAKE-256}(\text{domain separator} \parallel \text{tr} \parallel \mu)$ , where  $\text{tr} = H(\text{pk})$  is a public transcript hash for domain separation).

- In high-level descriptions (like the one you referenced), it's simplified as  $H(\mu)$ , but the key is that the message is hashed first to fixed length before concatenating with  $w_1$ .
- $w_1'$  (recomputed high bits):
  - Verifier computes a candidate commitment vector:
    - Start with  $Az - c \cdot (t_1 \cdot 2^d)$  (matrix-vector in the module-ring, approximating  $A(y + c s_1) - c (A s_1 + s_2 - t_0) \approx Ay - c s_2 + \text{small terms}$ ).
  - Apply the hint  $h$  (from signature) to adjust for possible carry-over in the low bits (UseHint function: adds 1 in positions indicated by  $h$  if low bits suggest a carry).
  - Then decompose:  $w_1' = \text{HighBits}(\text{adjusted value}, 2\gamma_2)$ .
  - $\gamma_2$  is a fixed parameter (for Dilithium-3:  $\gamma_2 = (q - 1)/88 \approx 95232$ ).
- $\parallel$  : Concatenation of byte strings ( $\mu$  encoded  $\parallel$  encoded  $w_1'$ ).
- $c$ : The challenge polynomial from the signature (sparse, with exactly  $\tau=49$  coefficients of  $\pm 1$  in Dilithium-3).
  - During signing:  $c$  was computed as  $c \leftarrow \text{ChallengeHash}(H(\mu \parallel w_1))$  (hash output expanded and masked to ensure exactly  $\tau$  nonzero  $\pm 1$  coeffs).
  - Verifier checks if plugging in the recomputed  $w_1'$  yields the *same*  $c$ .

#### Failure modes:

- If  $w_1' \neq w_1$  (original), the hash won't match  $\rightarrow$  reject (catches forgeries or tampering).

#### 2. Check: $\|z\|_\infty < \gamma_1 - \beta$

##### What this equation means:

- This is the **response bound check** (part of rejection sampling validation).
- It ensures that the response vector  $z$  is not too large in infinity norm (max absolute value of any coefficient across all polynomials in the vector).
- Why? During signing,  $z = y + c s_1$ :
  - $y$  is sampled uniformly from  $-(\gamma_1 - 1)$  to  $+(\gamma_1 - 1)$ .
  - $c s_1$  adds small noise (bounded by  $\beta$ ).
- If  $\|z\|_\infty \geq \gamma_1$ , the distribution of  $z$  would depend on  $s_1$  (leakage risk), so signer **rejects internally** and restarts.
- The verifier enforces a slightly stricter bound  $< \gamma_1 - \beta$  to accept only signatures where the signer *must* have accepted (no chance of leaky large  $z$  slipping through).
- This maintains **unconditional zero-knowledge** (signature reveals nothing about secret beyond public key) and **soundness** (only legitimate signer can produce passing  $z$ ).

##### How the variables are calculated/initialized:

- $z$ : The response vector from the signature ( $l$ -length vector of polynomials, compressed for transmission).
  - Decompressed by verifier to full coefficients.
  - During signing:  $z = y + c \cdot s_1$  (ring-module operations).

- $\|\cdot\|_\infty$ : Infinity norm over the entire module-vector.
  - For a vector  $\mathbf{z} \in R_q^l$ : max over all  $l$  polynomials, all 256 coefficients, of  $|\text{coeff}|$ .
  - (Not L2 or other norms — infinity for simplicity and tight bounds.)
- $\gamma_1$ : Fixed parameter bounding the masking vector  $\mathbf{y}$ .
  - $\mathbf{y}$  coefficients uniform in  $\{-(\gamma_1 - 1), \dots, +(\gamma_1 - 1)\}$ .
  - For Dilithium-3:  $\gamma_1 = 2^{19} = 524288$ .
  - Chosen so  $\mathbf{y}$  has high entropy ( $\sim 19$  bits per coeff) to mask secrets.
- $\beta$ : Fixed bound on the "noise" from challenge  $\cdot$  secret.
  - $\beta = \tau \cdot \eta$ 
    - $\tau$  = number of  $\pm 1$  in  $\mathbf{c}$  (Dilithium-3:  $\tau = 49$ ).
    - $\eta$  = secret coefficient bound (Dilithium-3:  $\eta = 4$ , secrets in  $[-\eta, \eta]$ ).
  - Thus  $\beta = 49 \times 4 = 196$ .
  - Exactly:  $\|\mathbf{c} \cdot \mathbf{s}_1\|_\infty \leq \tau \cdot \eta = \beta$  (since at most  $\tau$  nonzero  $\pm 1$  in  $\mathbf{c}$ , multiplied by  $\leq \eta$ ).

### Why $\gamma_1 - \beta$ specifically:

- Honest  $\mathbf{z}$  satisfies  $\|\mathbf{z}\|_\infty \leq (\gamma_1 - 1) + \beta$  with high probability.
- But to guarantee no leakage, signer rejects if  $\|\mathbf{z}\|_\infty \geq \gamma_1 - \beta$  (gives margin so accepted  $\mathbf{z}$  looks uniformly random in the bound).
- Verifier uses same strict bound.

### Summary of why both checks together work:

- The hash check ensures correctness and message binding.
- The norm check ensures statistical hiding (zero-knowledge).
- Passing both  $\approx$  proof of knowledge of short  $\mathbf{s}_1, \mathbf{s}_2$  satisfying the public equation (Module-LWE/SIS hardness).

These checks are efficient (mostly NTT-based multiplications) and provide strong post-quantum security.

## Explanation of the Expressions in Dilithium

These expressions appear in the context of Dilithium's (ML-DSA) signature scheme, specifically in the analysis of the **response vector**  $\mathbf{z} = \mathbf{y} + \mathbf{c} \cdot \mathbf{s}_1$  during signing and the **bound check**  $\|\mathbf{z}\|_\infty < \gamma_1 - \beta$  during verification. They deal with norms (measuring "size") of vectors in the module-ring structure.

Dilithium operates over a **module**  $R_q^l$  (or  $R_q^k$ ), where each "entry" is a polynomial in  $R_q = \mathbb{Z}_q[X]/(X^{256} + 1)$  with 256 coefficients. Norms quantify how "large" these objects are, which is crucial for security (short secrets) and correctness (rejection sampling to hide secrets).

### 1. $\|\cdot\|_\infty$ : Infinity Norm over the Entire Module-Vector

- **What it means:**



- The **infinity norm** (also called  $l_\infty$  norm or max norm) of an object measures its **largest absolute coefficient** (ignoring all others).
- For a single polynomial  $p(X) = \sum_{i=0}^{255} a_i X^i$  in  $R_q$ :
  - $\|p\|_\infty = \max_i |a_i|$  (maximum absolute value among the 256 coefficients).
- For a **vector**  $v = (v_1, v_2, \dots, v_m)$  in  $R_q^m$  (e.g.,  $m = l = 5$  or  $7$  depending on parameter set, each  $v_j$  a polynomial):
  - $\|v\|_\infty = \max\{j, i\} \mid (v_j \text{'s } i\text{-th coefficient})\}$
  - In words: the single largest absolute coefficient **anywhere** in the entire vector (across all polynomials and all their coefficients).
- This is denoted as "infinity norm over the entire module-vector" to emphasize it's global over the high-dimensional structure (total  $\sim 256 \times m$  coefficients).
- **Why infinity norm (not others like  $l_2$ )?**
  - Simple to compute and bound tightly.
  - Fits the uniform sampling of masking  $y$  (coefficients independent uniform in a range).
  - Ensures tight security reductions and efficient rejection sampling.
- **Examples in Dilithium:**
  - Secret vectors  $s_1, s_2$ :  $\|s_1\|_\infty \leq \eta$  (e.g.,  $\eta=4$ , all coeffs  $\leq 4$  in abs).
  - Masking  $y$ : coefficients uniform in  $\{-(\gamma-1), \dots, +(\gamma-1)\}$ , so  $\|y\|_\infty \leq \gamma - 1$  ( $\gamma=2^{19}$  for Dilithium-3).
  - Response  $z$ : Checked with  $\|z\|_\infty < \gamma - \beta$ .

## 2. $\|c * s_1\|_\infty \leq \tau * \eta = \beta$

- **What it means:**
  - This is a **provable upper bound** on the infinity norm of the "noise" term added to the masking vector during signing.
  - $c * s_1$ : Component-wise ring multiplication in the module:
    - $c$  is the challenge (a single polynomial in  $R_q$  with **exactly  $\tau$  nonzero coefficients**, each  $\pm 1$ , rest 0).
    - $s_1$  is the secret vector ( $R_q^l$ , each component a polynomial with coeffs in  $[-\eta, \eta]$ ).
    - Result: a vector where each entry is the ring product  $c * s_{1_j}$  (polynomial multiplication modulo  $X^{256} + 1$ ).
  - The inequality says: the largest coefficient in this noise vector is **at most  $\tau * \eta$**  in absolute value.
  - Thus,  **$\beta$  is defined as  $\beta = \tau * \eta$**  (a fixed parameter, e.g.,  $\tau=49$ ,  $\eta=4 \rightarrow \beta=196$  in Dilithium-3).
- **Why this bound holds (step-by-step derivation):**
  - Ring multiplication of two polynomials is coefficient convolution:
    - Each output coefficient is a sum of products of input coefficients.
  - For one component: product  $p = c * s$  (where  $s = s_{1_j}$ ,  $\|s\|_\infty \leq \eta$ ).
    - $c$  has exactly  $\tau$  nonzero coeffs ( $\pm 1$ ).

- In the convolution, each output coefficient of  $p$  is a sum of **at most  $\tau$**  terms (one for each nonzero in  $c$ ).
  - Each term:  $(+/-1) \times (\text{coeff of } s) \rightarrow \text{absolute value} \leq \eta$ .
  - Worst case: all  $\tau$  terms align in sign  $\rightarrow \text{sum} \leq \tau \eta$  in abs.
- Thus: every coefficient of  $p$  has  $|\text{coeff}| \leq \tau \eta \rightarrow \|p\|_\infty \leq \tau \eta$ .
- Since this holds for every component of the vector  $c * s_1$ , the global  $\|c * s_1\|_\infty \leq \tau \eta$ .
- Equality can theoretically be reached (if signs align perfectly), so the bound is tight.
- **Why this matters in the scheme:**
  - During signing:  $z = y + c * s_1$ .
    - $y$  is large ( $\|y\|_\infty \leq \gamma - 1$ ) to mask the secret.
    - Added noise  $c * s_1$  is small ( $\|\dots\|_\infty \leq \beta$ ).
    - Honest  $z$  has  $\|z\|_\infty \leq (\gamma - 1) + \beta$  with overwhelming probability.
  - Rejection sampling: Signer rejects if  $\|z\|_\infty \geq \gamma - \beta \rightarrow$  ensures accepted  $z$  is statistically close to uniform in the larger range (hides  $s_1$  completely  $\rightarrow$  zero-knowledge).
  - Verification: Enforces strict  $\|z\|_\infty < \gamma - \beta \rightarrow$  accepts only non-leaky signatures.

In summary:

- $\| * \|_\infty$  is the max absolute coefficient globally.
- The bound defines  $\beta$  as the worst-case noise from challenge  $\times$  secret, enabling secure masking and rejection.

This tight bounding is key to Dilithium's efficiency and proven security! Let me know if you'd like a toy numerical example.

## Dilithium-3 Usage in the NERV Network

NERV uses **CRYSTALS-Dilithium-3** (NIST Level 3) exclusively for all critical signatures — no legacy curves like ECDSA/EdDSA. This ensures full post-quantum security from genesis.

Sizes in NERV: pk  $\sim 1,809$  bytes, signature  $\sim 3,297$  bytes (as per whitepaper).

### Specific Steps and Uses

1. **Outer Onion Layer Signature (User-Side Spam Prevention):**
  - Wallet constructs the onion (layered post-quantum encryptions with ML-KEM-768).
  - Signs the **outermost onion blob** with the user's Dilithium-3 private key.
  - Purpose: Non-repudiation — prevents spam/flooding; nodes can reject unsigned onions.

- Verifier: First relay checks signature before processing.
- 2. **Hop-by-Hop Processing in TEE Relays:**
  - Each of the 5 relays (inside attested TEEs):
    - Decrypts its layer.
    - Generates cover traffic/jitter.
    - **Re-encrypts and forwards with a fresh remote attestation report.**
    - **Signs the outbound packet** (or attestation chain) with the enclave's Dilithium-3 key.
  - Purpose: Chain of trust — proves each hop ran correct code without side channels.
  - Cumulative: Builds a verifiable chain of 5 Dilithium-3 signatures/attestations.
- 3. **Final Relay Submission:**
  - 5th relay verifies inner Halo2 proof and user's outer Dilithium signature.
  - **Adds its own Dilithium-3 signature** and secure timestamp.
  - Forwards to shard mempool.
- 4. **VDW (Verifiable Delay Witness) Generation and Signing:**
  - After batch finalization, nodes generate VDWs inside TEEs.
  - Bundle: tx\_hash, location, delta\_proof, root, etc.
  - **Enclave signs the VDW payload** with Dilithium-3 (using attested key).
  - Purpose: Non-repudiable proof of correct computation inside genuine hardware.
  - Size contribution: ~96 bytes (sig + attestation).
- 5. **Other Consensus/Validator Messages:**
  - Neural votes, challenges, etc., include partial/full Dilithium-3 signatures where needed (though primary consensus uses BLS12-381 threshold for efficiency; Dilithium for PQ-critical parts).

In all cases:

- Signatures are over Protobuf-serialized payloads for determinism.
- Verification is fast (~58  $\mu$ s) and constant-time.
- Combined with TEE attestations, ensures no plaintext leakage or tampering, even against quantum adversaries.

NERV's heavy reliance on Dilithium-3 (for user ingress, relay chains, and proofs) makes it one of the first blockchains fully PQ-secure in authentication from day one.

**LSTM** stands for **Long Short-Term Memory**.

It is a special type of recurrent neural network (RNN) architecture designed to effectively handle and predict sequences of data — especially when patterns depend on both recent events **and** older ones.

**Why LSTM in NERV?**

In the dynamic neural sharding system:

- The shard runs a small (1.1 MB) **LSTM model** that looks at the last ~120 seconds of traffic metrics (TPS, cross-shard ratio, latency spikes, etc.).
- It predicts the probability of overload in the next 15 seconds with very high accuracy (>95% on test data).

LSTMs are perfect here because blockchain traffic is **sequential and bursty**:

- Recent spikes matter (short-term memory).
- Longer-term patterns (e.g., daily cycles, growing token popularity) also matter (long-term memory).
- Standard feed-forward nets would forget older context; LSTMs have "gates" that selectively remember or forget information over time.

So, the LSTM acts as the shard's "early warning system" — proactively spotting when a split is needed before users notice slowdowns.

---

## Detailed Explanation of the Shapley Value Formula

The **Shapley value** (invented by Lloyd Shapley in 1953) is a way to **fairly divide the total "profit" or "value" created by a group of cooperating players** (in NERV's case, nodes contributing gradients in federated learning). It ensures each player gets paid exactly in proportion to their average contribution across all possible ways the group could have formed.

The formula calculates this fair share  $\phi_i(\mathbf{v})$  for a specific node  $i$ :

$$\phi_i(\mathbf{v}) = \sum_{C \subseteq N \setminus \{i\}} \frac{|C|! \cdot (n - |C| - 1)!}{n!} [v(C \cup \{i\}) - v(C)]$$

Let's break it down **term by term**, then explain **why it works** and how it applies to NERV.

### 1. The Overall Structure

- The big  $\sum$  (**sum**) is over **every possible subset  $C$**  of players that **does NOT include  $i$**  (written  $C \subseteq N \setminus \{i\}$ ).
- For each such subset  $C$ , we compute:
  - A **weighting factor** that says how important this particular subset is.
  - The **marginal contribution** of  $i$  when added to that subset.
- We multiply them and add everything up  $\rightarrow$   $i$ 's fair share.

This averages  $i$ 's marginal contributions over **all possible orders** in which the full group  $N$  could form.

2. Term-by-Term Breakdown

Term	Meaning	Detailed Explanation
<b>N</b>	The full set of all players (nodes)	In NERV, N is the set of all nodes that submitted gradients in this federated learning round (e.g., 100–1000 nodes per shard).
<b><math>n =</math></b>	N	<b></b>
<b>C</b>	A coalition (subset) not including i	Every possible group of other nodes that could form before i joins. C can be empty ( $\varnothing$ ), size 1, 2, ..., up to size n-1.
<b></b>	C	<b></b>
<b></b>	C	<b><math>!</math></b>
<b><math>(n -</math></b>	C	<b><math>- 1)!</math></b>
<b>n!</b>	Factorial of total players	Total number of ways to order all n nodes (full permutations of N).
<b><math>\frac{</math></b>	C	<b><math>! \cdot (n -</math></b>
<b><math>v(C \cup \{i\})</math></b>	Value of the coalition after adding i	The total value (e.g., model improvement) created by the group C plus node i.

$v(C)$                       Value of the coalition                      Value created by just C alone.  
before adding i

$v(C \cup \{i\})$                       **Marginal**                      How much **extra** value i brings when joining that  
-  $v(C)$                       **contribution** of i to C                      specific pre-existing group C.

### 3. Why This Formula is "Fair"

The Shapley value satisfies four desirable fairness properties:

- **Efficiency:** The sum of all  $\phi_i$  equals the total value  $v(N)$  — everything is distributed.
- **Symmetry:** If two nodes contribute equally in every coalition, they get the same share.
- **Dummy player:** If a node adds zero marginal value everywhere, it gets zero.
- **Additivity:** Works nicely when combining games.

The weighting factor ensures we average marginal contributions **uniformly over all possible joining orders** — no bias toward early or late joiners.

### 4. Simple Numerical Example (Toy Case)

Suppose  $n = 3$  nodes: A, B, i (you want  $\phi_i$ ).

All permutations of order they join:

- |  |                                  |                                       |
|--|----------------------------------|---------------------------------------|
| 1. $C=\emptyset \rightarrow i \rightarrow (A,B)$ | Weight: $0! \cdot 2! / 3! = 1/6$ | Marginal: $v(\{i\}) - v(\emptyset)$   |
| 2. $A \rightarrow i \rightarrow B$               | Weight: $1! \cdot 1! / 3! = 1/6$ | Marginal: $v(\{A,i\}) - v(\{A\})$     |
| 3. $B \rightarrow i \rightarrow A$               | Weight: $1! \cdot 1! / 3! = 1/6$ | Marginal: $v(\{B,i\}) - v(\{B\})$     |
| 4. $A,B \rightarrow i$                           | Weight: $2! \cdot 0! / 3! = 1/6$ | Marginal: $v(\{A,B,i\}) - v(\{A,B\})$ |
| 5. $B,A \rightarrow i$                           | Weight: $2! \cdot 0! / 3! = 1/6$ | (same as above)                       |
| 6. $i$ first $\rightarrow \dots$                 | Weight: $0! \cdot 2! / 3! = 1/6$ | (same as 1)                           |

The formula automatically groups the identical ones and averages the six equally-weighted marginals.

### 5. How It Applies to NERV (The Value Function $v$ )

In NERV, the "value" of a coalition C is how much that group improves the global neural encoder:

$$v(C) = -\text{Loss}(\text{aggregated\_grads}(C))$$

or

$$v(C) = 1 / (1 + \text{homomorphism\_error}(C))$$

- Higher  $v$  = better model (lower loss or lower homomorphism error).
- Aggregated grads from  $C \rightarrow$  simulate the updated encoder  $\rightarrow$  measure improvement on a validation set.
- Marginal  $v(C \cup \{i\}) - v(C)$  = how much **extra** improvement  $i$ 's gradient adds when combined with  $C$ 's.

Nodes with consistently useful gradients (good data, honest computation) have high average marginal contributions  $\rightarrow$  high Shapley  $\rightarrow$  more token rewards.

## 6. Practical Computation in NERV

Exact Shapley is exponential ( $2^n$  terms — impossible for  $n > \sim 30$ ), so NERV uses **Monte-Carlo approximation** inside TEEs:

- Randomly sample thousands of coalitions  $C$ .
- Estimate the average weighted marginal.
- Still very accurate with 10k–100k samples, and runs fast on GPU-TEE.

This keeps rewards fair and resistant to gaming.

In summary: The formula elegantly averages "how much extra value would the team gain if  $i$  joined at every possible moment" — weighted by the likelihood of that moment occurring in a random order. It's the gold-standard fair allocation method, perfectly suited for rewarding genuine contributions in NERV's useful-work economy.

---

## Detailed Explanation of the Alternative Shapley Value Formula

This is an **equivalent formulation** of the Shapley value you saw earlier — just written in a different but mathematically identical way. It makes the "average over all possible joining orders" intuition even clearer.

The formula:

$$\phi_i = \frac{1}{n!} \sum_{\pi} [v(P_{\pi}(i) \cup \{i\}) - v(P_{\pi}(i))] \quad \phi_i = \frac{1}{n!} \sum_{\pi} [v(P_{\pi}(i) \cup \{i\}) - v(P_{\pi}(i))]$$

where the sum is over **all permutations**  $\pi$  of the  $n$  players, and  $P_{\pi}(i)$  means "the set of players who come before  $i$  in the permutation  $\pi$ " (the **predecessors** of  $i$ ).

### 1. What This Formula is Saying in Plain English

- Imagine the  $n$  nodes (players) join the coalition one by one in a **random order** (every possible order is equally likely).
- For each possible order  $\pi$ :
  - Look at the group that has already formed just **before** node  $i$  joins — that's  $P_\pi(i)$ .
  - Compute how much **extra value**  $i$  adds when it joins that group:  $v(\text{after } i \text{ joins}) - v(\text{before } i \text{ joins})$ .
- Average that marginal contribution over **all  $n!$  possible orders** (divide by  $n!$ ).
- The result is  $i$ 's fair share  $\phi_i$ .

This directly captures the idea: "On average, across every possible way the team could have assembled, how much value did  $i$  contribute when it showed up?"

## 2. Term-by-Term Breakdown

Term	Meaning	Explanation
$n!$	Total number of permutations	There are $n!$ ways to order $n$ distinct nodes. Dividing by $n!$ averages over all orders equally.
$\sum_\pi$	Sum over all permutations $\pi$ of the $n$ players	We consider every possible sequence in which the nodes could join.
$P_\pi(i)$	Predecessors of $i$ in permutation $\pi$	The set of nodes that appear <b>before</b> $i$ in the order $\pi$ . Example: If $\pi$ is (A, B, i, C), then $P_\pi(i) = \{A, B\}$ .
$P_\pi(i) \cup \{i\}$	The coalition after $i$ joins	The predecessors plus $i$ itself — the group right after $i$ arrives.
$v(P_\pi(i) \cup \{i\})$	Value of the coalition after adding $i$	How good the model (or whatever $v$ measures) is with $i$ included.



$v(P_{\pi(i)})$	Value of the coalition before adding i	How good it was without i.
$v(P_{\pi(i)} \cup \{i\}) - v(P_{\pi(i)})$	Marginal contribution of i in this specific order	The exact improvement i brings <b>when arriving at that particular moment</b> .

### 3. Why This is Exactly the Same as the Previous Formula

The earlier version grouped terms by coalition size  $|C|$ , counting how many permutations correspond to each possible predecessor set C.

This permutation version is more direct:

- For every possible predecessor set C, it appears in exactly  $|C|! \times (n - |C| - 1)!$  permutations (ways to order the predecessors, then the remaining players after i).
- So when you expand the sum over permutations and group identical C's, you recover the weighting factor  $|C|! \cdot (n - |C| - 1)! / n!$  from the first formula.

Both compute the **same weighted average of marginal contributions** — just one loops over coalitions, the other over orders.

### 4. Tiny Example to See It in Action (n=3)

Nodes: A, B, i

We want  $\phi_i$ .

All 6 permutations ( $\pi$ ) and i's marginal contribution in each:

Permutation $\pi$	$P_{\pi(i)}$ (predecessors)	$v(\text{after } i) - v(\text{before } i)$	Contribution
i, A, B	$\emptyset$	$v(\{i\}) - v(\emptyset)$	contrib1
A, i, B	$\{A\}$	$v(\{A, i\}) - v(\{A\})$	contrib2

A, B, i	{A,B}	$v(\{A,B,i\}) - v(\{A,B\})$	contrib3
B, i, A	{B}	$v(\{B,i\}) - v(\{B\})$	contrib4
B, A, i	{B,A}	$v(\{B,A,i\}) - v(\{B,A\})$	contrib5 (= contrib3)
i first (other orders)	$\emptyset$	same as first	contrib6 (= contrib1)

Then:

$$\phi_i = 1/6(\text{contrib1} + \text{contrib2} + \text{contrib3} + \text{contrib4} + \text{contrib5} + \text{contrib6})$$

Notice that the empty predecessor case appears twice (i first in two orders), the size-1 cases twice each, and size-2 once — exactly matching the weighting in the coalition version.

## 5. Why This View is Useful for NERV

Thinking in permutations makes it obvious why Shapley is **strategy-proof**: a node can't manipulate its payout by misreporting, because its marginal value is measured in every possible context. In NERV's federated learning, a node that consistently adds real improvement — no matter when its gradient is "added" in the imaginary order — gets fairly rewarded.

Both formulas are interchangeable; cryptographers and economists often prefer the permutation version for theoretical proofs, while the coalition version is more common for approximations (easier to sample subsets than full permutations).

-----

## AI Exposure to Transaction History

Validators (and the distilled predictor model used in AI-native consensus) primarily interact with the current canonical **512-byte neural state embedding** ( $e_t$ ) and the aggregated homomorphic delta ( $\Delta_B$ ) for each batch. They predict the next embedding hash based on  $e_t + \Delta_B$  without decompressing or accessing the full underlying state  $S_t$  (accounts, balances, etc.). The design explicitly avoids decompressing the embedding to preserve privacy and efficiency.

The AI models (distilled predictor, LSTM for sharding, etc.) are **not exposed to plaintext transaction history** beyond the latent effects captured in the embeddings and deltas. However:

- During useful-work federated learning (every ~15 seconds or 1000 txs), nodes train on **anonymized recent transaction batches** (blinded via the 5-hop TEE mixer) with Differential Privacy (DP-SGD,  $\sigma=0.5$ ). This exposes gradients derived from recent deltas/batches but not full history or plaintext.
- The encoder  $\mathcal{E}_\theta$  itself is trained on sequences of ledger states and updates from real-world-like datasets, but updates occur only every 30 days via federated learning on anonymized data.

No direct access to historic plaintext transactions occurs in normal operation.

## Network Storage of Historic Transactions

NERV does **not store historic transactions** on-chain in the traditional sense (no public transaction graph, no Merkle trie of history). The chain is a DAG/lattice of **512-byte embedding roots** (hashed to 32 bytes publicly), forming a tiny header chain (<100 KB forever for light clients).

- Recent batches (last ~500 txs) are temporarily retained in encrypted mempools or DHT for shard splits/merges (re-execution in TEEs, <4 seconds).
- No long-term on-chain storage of individual transactions or full history—the privacy-by-default design makes the full state irreversible and latent (reduction to Neural Network Inversion Problem).
- Embeddings are replicated via Reed-Solomon erasure coding across nodes for fault tolerance, but this is current/recent state, not infinite history.

The network focuses on the current latent state; history is proven via VDWs rather than stored transactions.

## AI Training During Testnet (Aurora Testnet)

The whitepaper mentions the **Aurora testnet** (with metrics like 28k nodes, 1.1M TPS projections), but no public testnet has launched yet (target Q2 2026).

Training follows the same principles as mainnet:

- Federated learning on **anonymized recent transaction batches** (not full historic plaintext transactions).
- Focus on improving homomorphism preservation using embedding data, deltas, and simulated/real traffic distributions.
- DP-SGD ensures privacy; no exposure to raw historic transactions beyond what's needed for gradient contributions on recent anonymized data.

## Verifiable Delay Witness (VDW) Terminology

**VDW** is a term **specific to NERV**. It is NERV's custom privacy-preserving receipt mechanism (tiny ~1.4 KB proof of inclusion via homomorphic delta paths, Halo2 proofs, and TEE attestations).

It is **not used industry-wide**. The broader crypto field uses **Verifiable Delay Functions (VDFs)** for sequential computation with fast verification (e.g., in randomness beacons like Chia or Ethereum proposals). NERV's VDW is unrelated—it's a "witness" for private inclusion, not a delay function.

## Historic VDW Storage

Yes, VDWs are designed for **permanent archival**:

- Served instantly via HTTP/3 from nodes upon issuance.
- Within **30 seconds**: Automatically pinned to **Arweave** (permanent, economic incentive) and **IPFS** (redundant).
- After **5 years**: Aggregated into daily **Merkle Mountain Range buckets** ( $\leq 1$  KB proof for century-scale retrieval).

This ensures century+ availability with succinct proofs, even for historical VDWs.

## Historic e\_t (Embedding Roots) Storage

The public chain is a sequence/DAG of embedding root hashes (32-byte BLAKE3 hashes of the 512-byte e\_t).

- Full **512-byte embeddings** are replicated across nodes via Reed-Solomon ( $k=5$ ,  $m=2$ ) for current/recent shards (40% fault tolerance).
- Long-term: Roots are archived via Arweave/IPFS pinning, similar to VDWs.
- Light clients sync only the **32-byte hash chain** (<100 KB forever).
- No explicit pruning mentioned for old embeddings, but the design implies permanent availability for verification (via archival like VDWs).

Historic roots remain verifiable forever with minimal data.

**DP-SGD** (Differentially Private Stochastic Gradient Descent) is a modified version of the standard **Stochastic Gradient Descent** (SGD) optimization algorithm used in training machine learning models, particularly deep neural networks. It incorporates **differential privacy** guarantees to protect sensitive training data from being inferred or leaked through the trained model.

Key modifications in DP-SGD (introduced in seminal work by Abadi et al., 2016):

- **Gradient clipping**: Per-sample gradients are clipped to bound their L2 norm, limiting the influence of any single data point.

- **Noise addition:** Gaussian noise is added to the aggregated (averaged) gradients before updating model parameters.
- **Privacy accounting:** Tracks the cumulative privacy loss (using metrics like  $(\epsilon, \delta)$ -differential privacy) over training iterations.

This ensures that the model's outputs are statistically indistinguishable whether trained with or without any individual data point, providing strong theoretical privacy protections. It's widely used in privacy-preserving ML (e.g., via libraries like Opacus in PyTorch or TensorFlow Privacy) and balances privacy with model utility—stronger privacy (more noise) reduces accuracy.

In the NERV context, DP-SGD (with noise scale  $\sigma=0.5$ ) is applied during federated learning to anonymize gradient contributions from recent transaction batches, preventing inversion attacks on private data.

## Is $S(t)$ Stored Anywhere?

No, the full plaintext state  $S(t)$  — the complete key-value ledger of accounts, balances, and metadata at height  $t$  — is **not stored anywhere** in the NERV network, neither temporarily nor long-term.

Core reasons from the design:

- Privacy-by-default:  $S(t)$  remains **latent and irreversible** inside the 512-byte neural state embedding  $e_t = \mathcal{E}_\theta(S_t)$ . Decompression would violate privacy (reduction to the "Neural Network Inversion Problem" hardness assumption, with  $>2^{4000}$  entropy and deliberate non-linearities).
- No plaintext ever touches untrusted memory; operations (updates, execution) occur homomorphically inside attested TEEs without revealing  $S(t)$ .
- The chain stores only the tiny embedding roots (32-byte hashes publicly, full 512-byte vectors replicated via erasure coding for current shards).

Temporary retention is limited to encrypted/anonymized recent batches (~500 txs) in mempools/DHTs for shard splits/merges, but not the full decompressed  $S(t)$ . Historical  $S(t)$  is never reconstructed or archived—proofs rely on VDWs and delta paths instead. This enables extreme compression and privacy while avoiding traditional state bloat.

---

## Detailed Explanation of the Weighted Quorum Formula

The formula is from NERV's **AI-Native Optimistic Consensus** (Section 4.1–4.2). It defines when validators reach **agreement** on the predicted next embedding hash ( $h_{pred}$ ), granting fast probabilistic finality.

**Formula:**

Let  $V$  be the set of all active validators.

Each validator  $v$  has weight  $w_v = \text{stake}_v \times \text{reputation}_v$

(Stake = tokens locked; reputation = score from useful-work/federated learning contributions.)

Quorum holds (Q) if and only if:

$$\sum_{v \in V} w_v \geq 0.67 \times \sum_{v \in V} w_v \iff \sum_{v \in V} w_v \geq 0.67 \times \sum_{v \in V} w_v$$

### What This Means:

- Validators vote by broadcasting their predicted hash  $h_v$  (plus partial BLS signature, reputation, attestation).
- Find the majority-predicted hash  $h_{\text{majority}}$  (most common one).
- Sum the weights of validators who voted for that majority hash.
- If this sum  $\geq 67\%$  of the **total weight** of all validators  $\rightarrow$  quorum achieved  $\rightarrow$  instant probabilistic finality (~600 ms median).

### Term-by-Term Breakdown

Term	Meaning	Why It Matters
$V$	Set of all active validators	Everyone participating in this round (staked and online).
$w_v = \text{stake}_v \times \text{reputation}_v$	Individual validator's voting weight	Multiplicative: High stake alone isn't enough — must also have good reputation (from honest/useful contributions). Prevents sybil attacks or lazy rich validators dominating.
$h_v$	The hash predicted by validator $v$	Their "vote": what they think the next embedding hash should be (from running the distilled predictor model).
$h_{\text{majority}}$	The most popular predicted hash	The one with the highest total weight behind it.

$\sum_{\{v: h_v = h_{\text{majority}}\}} w_v$	Total weight voting for the majority hash	How much "support" the winning prediction has.
$\sum_{\{v \in V\}} w_v$	Total weight of all validators	The full pie — 100% of voting power.
$\geq 0.67 \times \text{total}$	67% supermajority threshold	Standard BFT-like threshold: Tolerates up to ~33% Byzantine (malicious/misbehaving) weight without compromising safety.

### How It Works in Practice (Step-by-Step)

1. After a batch update, validators run the same small predictor model → get  $h_{\text{pred}}$  → broadcast vote ( $h_{\text{pred}}$  + partial BLS + reputation + TEE proof).
2. Nodes gossip votes and tally:
  - Group by  $h_{\text{pred}}$  value.
  - Compute weighted sums per group.
  - Identify  $h_{\text{majority}}$  (highest weighted group).
3. Check if that group's weight  $\geq 67\%$  of total active weight.
  - Yes → Probabilistic finality: Accept the new embedding root immediately.
  - No (<0.01% cases) → Trigger challenge/Monte-Carlo dispute for resolution.
4. Honest majority also aggregates BLS partial shares → full threshold signature for cryptographic finality (~1.8 s total).

### Why 67% Specifically?

- Classic BFT tolerance: System safe if <1/3 faulty (33% + epsilon can be tolerated).
- Weighted by stake × reputation adds extra honesty incentive (reputation slashed for bad behavior).

This hybrid (neural prediction for speed + weighted quorum for security) gives sub-second finality most of the time, with proven safety under partial synchrony (networks can delay) and bounded faults.

It mentions reputation several times as a key multiplier in voting weight ( $w_v = \text{stake}_v \times \text{reputation}_v$ ) and notes that it comes "from federated learning," but the precise mechanics are **high-level or implied** rather than fully specified.

### What the Whitepaper Does Say About Reputation

- **Section 4. AI-Native Optimistic Consensus:**
  - Validators broadcast their "current reputation score (from federated learning)".
  - Weight = stake × reputation.
  - Used in the 67% quorum threshold.
- **Section 4.3 Useful-Work Economy:**
  - Nodes are paid for gradient contributions (60% of rewards) measured via "secure Shapley-value inside TEEs".
  - Honest validation & finality: 30% of rewards (stake × reputation × uptime).
  - This strongly implies reputation is **derived (at least in part) from a node's historical Shapley-value contributions** in the federated learning process — i.e., how much they consistently help improve the network's encoders/models over time.
- Reputation acts as a **sybil-resistance and quality filter**: High-stake but low-reputation (lazy/malicious) validators have reduced influence.

## Likely How Reputation Works

Since the whitepaper ties it directly to federated learning and Shapley:

1. **Calculation:**
  - Reputation is a **running score** (e.g., 0–1 or normalized value) based on a node's **cumulative or recent Shapley-value contributions** to encoder improvements.
  - High Shapley = provided valuable, honest gradients → reputation increases.
  - Low/negative/malicious = reputation decreases (or slashing in disputes).
  - Possibly smoothed over time (e.g., exponential moving average) to reward consistent honesty.
2. **Storage:**
  - Likely **on-chain** (in the embedding/state) or in validator registry — per-validator value, updated periodically (e.g., every epoch or federated round).
  - Secured in TEEs during computation to prevent tampering.
3. **Processing/Update:**
  - During/after each useful-work cycle: Secure aggregation in TEEs computes Shapley → adjusts reputation.
  - Broadcast in votes for transparency.
  - Slashed explicitly in challenge phases (0.5–5%) for bad behavior, which would tank reputation.

This fits the "endogenous intelligence" theme: The network rewards nodes that make it smarter, boosting their future voting power.

## Why Isn't It More Detailed?

- The whitepaper is a **technical specification** but leaves some implementation details (like exact reputation formula) to the open-source code ([github.com/nerv-bit/nerv](https://github.com/nerv-bit/nerv)).



- As of now (December 2025), the project appears conceptual/pre-launch (fair launch June 2028), so full details might emerge in code, updates, or appendices.
- Common in ambitious whitepapers: Core innovations (neural embeddings, homomorphism) are specified deeply; incentive tweaks like reputation are outlined high-level.

If/when code or updates drop, reputation will likely be a **Shapley-history-based score** (possibly with decay, minimums, etc.) computed verifiably in TEEs.

## Arweave and IPFS: What They Are, How They Work, and Archiving for NERV

The whitepaper mentions pinning VDWs (and likely other data like old embeddings/headers) to **Arweave** and **IPFS** within ~30 seconds of commitment for permanent/redundant archival.

### 1. Arweave

**What it is:** A decentralized, permanent data storage blockchain (often called the "permaweb"). Pay once → data stored forever (guaranteed  $\geq 200$  years economically).

#### How it works:

- Built on a "blockweave" (graph-like structure, not linear chain): Each block links to previous + a random older block (Proof of Access consensus — miners prove they can access old data).
- Economic model: One-time upfront payment (in AR tokens) goes into an **endowment pool**. Pool pays miners over time (assumes storage costs drop  $\sim 0.5\%$ /year long-term).
- Data replicated  $\sim 20x$  across the network; incentives ensure miners store everything indefinitely.
- Truly permanent: No deletion or expiration — ideal for immutable archives.

**Examples:** Solana, Polkadot, Avalanche archive historical state/ledgers on Arweave. NFTs (e.g., via Meta/Instagram) store metadata permanently.

### 2. IPFS (InterPlanetary File System)

**What it is:** A peer-to-peer protocol/network for distributed file storage/sharing (not a blockchain). Content-addressed (files identified by cryptographic hash/CID, not location).

#### How it works:

- Files broken into blocks → hashed → stored across nodes.
- Retrieved by CID (content identifier) — network finds nearest copies via DHT (distributed hash table).
- No built-in incentives: Data persists only if nodes "pin" it (voluntarily host long-term).

- Redundant but not guaranteed permanent — unpinned data can vanish (garbage collected).
- Often paired with Filecoin (incentivized layer on top) for paid pinning.

**Examples:** Decentralized websites, NFT media, Wikipedia mirrors during censorship.

### How NERV Exports/Archives Data to Them

The whitepaper says: "Within 30 seconds of commitment, VDWs are permanently pinned to Arweave + IPFS."

**Process** (as described/implemented):

1. After embedding finality → generate VDWs (~1.4 KB each) or batch headers/old data.
2. Full nodes (or dedicated archivers) automatically:
  - Bundle data (e.g., daily/weekly batches for efficiency).
  - Upload to Arweave: Pay one-time AR fee → get permanent transaction ID.
  - Add to IPFS: Compute CID → request pinning (via services like Pinata, web3.storage, or community nodes; some pay Filecoin for reliability).
3. Store Arweave tx IDs / IPFS CIDs on-chain or in metadata for easy retrieval.
4. Users/wallets fetch from these for historical VDWs (century-scale proofs via Merkle buckets).

### Why both?

- Arweave: True permanence + economic guarantees.
- IPFS: Faster/cheaper access + broader redundancy (many gateways).

Many projects (e.g., Solana archives) use Arweave directly; some bridge IPFS → Arweave for "permapinning" (store IPFS CID on Arweave to incentivize eternal hosting).

This ensures NERV data outlives the chain itself — VDWs verifiable forever, even offline.

---

The **NERV** design incorporates several layered mechanisms to address these concerns, focusing on cryptographic verifiability, hardware-rooted trust, bounded error, and economic incentives rather than assuming perfect AI behavior.

### Preventing AI "Hallucinations" (Incorrect Predictions or Drift)

In NERV, "hallucination" primarily refers to the distilled predictor model (used in optimistic consensus) incorrectly forecasting the next embedding hash, or the encoder  $\mathcal{E}_\theta$  drifting from the transfer homomorphism property over time.

- **Optimistic fast path with challenge fallback:** In 99.99% of cases, validators agree on the predicted hash for sub-second finality. If disagreement occurs ( $<0.01\%$  projected), any validator can post a bonded challenge. This triggers 32 randomly selected TEEs to run 10,000 parallel Monte-Carlo simulations on the batch deltas, determining the majority correct embedding. The losing side is slashed (0.5–5% stake), ensuring economic penalties for bad predictions.
- **Bounded homomorphism error:** Updates require a Halo2 proof that the error remains  $\leq 10^{-9}$ . During epoch updates (every 30 days via federated learning), a new encoder  $\mathcal{E}_\theta$  must include a Halo2 proof preserving this bound. If the proof fails, the network rejects the update and stays on the old encoder ("safety first").
- **No unbounded drift:** The homomorphism is enforced cryptographically per batch/epoch, not trusted to the AI indefinitely. The distilled predictor is small (1.8 MB, fits entirely in TEE) and attested for correct execution.

These ensure predictions are verifiable and disputes resolved probabilistically/cryptographically, preventing persistent "hallucinations."

## Preventing Side-Channel Attacks

All privacy- and AI-critical operations occur exclusively inside **remotely attested hardware Trusted Execution Environments (TEEs)** (multi-vendor: Intel SGX, AMD SEV-SNP, ARM CCA, Apple Secure Enclave, NVIDIA confidential GPUs).

- **Constant-time operations:** All enclave code uses constant-time arithmetic (no secret-dependent branches or table lookups) to eliminate timing/power/EM leaks.
- **Memory obfuscation:** ORAM-lite (overhead  $<1.8\times$ ) randomizes access patterns.
- **Anomaly monitoring:** Continuous power/EM fingerprint monitoring on validator clusters; anomalies trigger automatic shutdown and slashing.
- **No plaintext in untrusted RAM:** Private data, embeddings, and model executions never leave the TEE unsealed. Remote attestations (Dilithium-signed reports) prove the exact audited binary ran without side-channel leakage.
- **Cover traffic in mixer:** The 5-hop TEE mixer adds AI-tuned chaff and exponential jitter to resist timing correlation.

This production-grade hardening (formal ProVerif-verified for the mixer) mitigates known side-channels even against advanced adversaries.

## Preventing Deliberate External Attacks to Corrupt Embeddings

Embeddings (old or new) are protected by a combination of cryptography, consensus, and incentives:

- **Homomorphic updates in TEEs:** Deltas are applied element-wise inside attested TEEs; Halo2 recursive proofs attest correct summation and no overflow/invalid ops.
- **Consensus finality:** New embeddings require  $\geq 67\%$  weighted (stake  $\times$  reputation) agreement on the hash, backed by BLS12-381 threshold signatures. Malicious proposals trigger challenges and slashing.
- **Irreversibility and replication:** Old embeddings are part of the public hash chain (immutable DAG); full vectors are Reed-Solomon encoded (survives 40% node loss). Attacks can't rewrite history without controlling  $>33\%$  weighted stake (Byzantine fault tolerance).
- **Post-quantum crypto:** All signatures (Dilithium-3), encapsulation (ML-KEM-768), and attestations are NIST PQ Level 3, resisting quantum groove or harvest-now-decrypt-later attacks.
- **Economic deterrence:** Bad actors risk slashing; honest nodes are rewarded via useful-work (Shapley-valued gradients).

Corruption requires compromising a supermajority of TEE-attested nodes simultaneously—extremely costly and detectable via attestations.

Overall, NERV treats AI components as untrusted but verifiable, rooting security in hardware attestations, succinct ZK proofs (Halo2/Nova), and economic finality rather than AI reliability alone.

---

The **30-day epoch interval** for updating the neural encoder  $\mathcal{E}_\theta$  strikes a deliberate balance between safety, stability, performance, and resource efficiency in NERV's design.

## Primary Reasons for 30-Day Epochs

The whitepaper explicitly states: "If the proof fails, the network stays on the old encoder for another epoch (**safety first**)."

- **Safety and stability** — Each update requires a succinct **Halo2 proof** that the new encoder  $\mathcal{E}_\theta'$  preserves the **transfer homomorphism** (error  $\leq 10^{-9}$ ). Frequent updates would increase the risk of deploying a flawed encoder that drifts from this critical property, potentially causing incorrect state updates, consensus failures, or subtle privacy leaks. A longer interval allows more thorough federated learning on accumulated anonymized data, yielding higher-quality encoders less prone to failure.
- **Proof overhead and network coordination** — Generating, verifying, and propagating the recursive Halo2 proof (plus new weights) across thousands of nodes/shards involves non-trivial bandwidth and validation costs. Monthly cadence minimizes disruption to ongoing transaction processing and consensus.
- **Gradual self-improvement** — The network improves continuously via frequent small gradient contributions (every  $\sim 15$ s or 1000 txs), but major encoder upgrades are batched monthly to ensure meaningful, verifiable progress without constant redeployment.

## Could Epochs Be More Frequent Without Losses?

Possibly, but with trade-offs:

- **No hard loss in transaction speed or accuracy** — Normal operations (homomorphic deltas, batch execution, AI-native consensus) use the current fixed encoder and don't depend on update frequency. TPS (>1M sustained) and finality (~600ms) would remain unaffected, as updates occur off the critical path.
- **Potential gains** — More frequent updates could adapt faster to changing traffic patterns (e.g., new transaction types), slightly improving homomorphism precision or sharding predictions over time.
- **Risks and downsides:**
  - Higher chance of proof failures → more frequent rollbacks to the old encoder, causing temporary stagnation.
  - Increased coordination overhead (proof broadcasts, node syncs) could indirectly pressure bandwidth in high-load periods.
  - Diminishing returns: Short intervals yield smaller improvements per update, as less data accumulates for federated training.

In short: Feasible technically (governance could adjust), but the 30-day default prioritizes reliability over marginal speed gains.

## Energy and Computing Power Consumption

Yes, more frequent epochs would increase consumption:

- **Higher computational load** → More Halo2 proof generations/verifications (7.9M constraints, recursive) per year—expensive on CPUs/GPUs, even in TEEs.
- **Increased network traffic** → Larger/frequent broadcasts of new weights (~1-2M params quantized) and proofs.
- **Federated learning intensity** → While gradients are lightweight and ongoing, shorter epochs might encourage denser training schedules indirectly.

Overall, the monthly schedule optimizes for **useful-work efficiency** (nodes paid for productive gradients, not wasteful frequent upgrades), aligning with NERV's low-energy philosophy versus PoW systems.

The 30-day interval reflects a conservative, security-first engineering choice common in protocols requiring verifiable upgrades (e.g., hard forks or param changes in other chains).

---

## Cryptographic Schemes in NERV and Their Usage

The NERV whitepaper specifies a targeted post-quantum (PQ) cryptography suite, focusing on NIST-standardized lattice-based primitives for critical paths involving keys and signatures, while relying on established zero-knowledge proof systems for state verification and updates. Below is a breakdown of the main schemes used at each stage of the protocol, based on explicit mentions in the document (e.g., Sections 6–7, transaction flow, consensus, and privacy infrastructure).

Stage/Component	Cryptographic Scheme(s) Used	Purpose & Reasoning
<b>Digital Signatures</b> (user tx signing, validator votes, TEE attestations, non-repudiation, onion outer layer)	CRYSTALS-Dilithium-3 (NIST Level 3)	Primary PQ signature scheme. Chosen for strong security against quantum attacks (lattice-based, ML-DSA standard), reasonable sizes (pk ~1.8 KB, sig ~3.3 KB), and fast verification (~58 $\mu$ s). No elliptic curves (e.g., ECDSA/EdDSA) in critical paths to avoid quantum vulnerability via Shor's algorithm.
<b>Key Encapsulation / Onion Routing Encryption</b> (5-hop anonymous ingress, ephemeral keys per hop)	ML-KEM-768 (formerly Kyber-768, NIST Level 3); hybrid with X25519 in some cases	PQ key exchange for encrypting onion layers. Lattice-based for quantum resistance; hybrid mode maintains compatibility/performance during transition. Ensures forward secrecy and protects against quantum groove attacks.
<b>One-Time/Genesis/Cold Keys</b>	SPHINCS+-SHA2 56-192s-robust (stateless hash-based)	Backup stateless signature for rare, high-value uses (e.g., genesis). Purely hash-based, extremely conservative PQ security (no lattices), but larger sigs (~41 KB) limit it to non-performance-critical paths.

<b>Zero-Knowledge Proofs</b> (tx validity, homomorphic delta application, recursive embedding updates, inclusion proofs in VDWs, homomorphism preservation)	Halo2 (with Nova folding for recursion)	Succinct recursive SNARKs for proving correct state updates, delta summation, and batch validity without revealing private data. Enables ~900× compression, O(1) updates, and tiny VDWs (~1.4 KB). Halo2/Nova chosen for efficiency (small proofs ~750 B, fast verification), recursion without trusted setup, and integration with circuits (e.g., LatentLedger transformer in ~7.9M constraints). Not a signature scheme—it's for verifiable computation over private state.
<b>Threshold Signatures</b> (consensus quorum, finality)	BLS12-381 (aggregated threshold shares)	Efficient aggregation for ≥67% weighted quorum in AI-native consensus. Pre-quantum (pairing-based), but used internally post-PQ-protected tx validation; economic finality relies on stake/reputation slashing rather than pure crypto for long-term PQ needs.
<b>Hashing</b> (embedding roots, tx hashes, seeds, commitments)	BLAKE3 (primary); SHA3-256 (fallback)	Quantum-resistant hashes (Grover-resistant). Fast, parallelizable BLAKE3 for performance; SHA3 for standards compliance.
<b>Overall Agility</b>	CryptoVersion enum + governance	Allows future migrations (e.g., to Dilithium-5, Falcon) via 180-day votes without breaking historic verification.

No zkEVM is mentioned—NERV avoids full VM emulation (unlike Polygon zkEVM or zkSync) in favor of custom Halo2 circuits tailored to neural embeddings and homomorphic deltas for massive efficiency gains.

## Why Dilithium for Signatures but Not for ZK Proofs (Halo2/Nova)?

Dilithium (and PQ signatures in general) and Halo2/Nova serve fundamentally different roles, so they aren't interchangeable:

- **Different Primitives:** Dilithium is a digital signature scheme—it proves authorship/authenticity of a message ("I signed this"). Halo2/Nova is a zero-knowledge succinct non-interactive argument of knowledge (zk-SNARK) system—it proves that a private computation (e.g., "I correctly applied deltas to the hidden state without revealing balances") was performed honestly. You can't "replace" one with the other; signatures authenticate intent, while ZK proofs verify private execution.
- **Performance & Size Trade-offs:** Current PQ signatures like Dilithium have larger keys/signatures (2–4 KB) compared to pre-quantum ones (~64 B). Fully PQ zk-SNARKs (e.g., lattice-based folding like LatticeFold or Neo) exist in research but are significantly less efficient today: larger proofs (often 10–100×), slower proving/verification, and higher constraints. Halo2/Nova achieves tiny proofs (~750 B) and fast times (<80 ms verification on mobile) crucial for >1M TPS and light clients. Switching to PQ ZK would degrade scalability/privacy (e.g., larger VDWs, higher latency).
- **Security Model Differences:** Halo2/Nova relies on discrete-log/pairing assumptions (vulnerable to Shor), but in NERV, ZK proofs are generated/verified inside attested TEEs (hardware enclaves) and protect private state—not public keys. The quantum threat to ZK here is lower priority than to signatures (which expose public keys). Signatures need immediate PQ replacement from genesis.
- **Maturity & Ecosystem:** Halo2/Nova is battle-tested (e.g., Zcash, scrolling rollups) with recursion optimized for NERV's folding needs. PQ ZK systems are emerging (e.g., post-2024 lattice folding) but not yet production-ready for this scale. NERV prioritizes PQ where urgent (signatures, encryption) while using best-in-class for ZK efficiency.

In summary, NERV achieves "full NIST post-quantum security from genesis" by protecting all external/exposed crypto (keys, sigs, routing) with PQ primitives, while internal verifiable computation uses optimized (pre-quantum) ZK—common in hybrid PQ designs today. Future epochs could migrate ZK if PQ variants mature without sacrificing performance.

## **Q: Could you please list all the advantages of this blockchain in replacing Merkle trees with 512-byte AI-embeddings in detail?**

Replacing traditional Merkle trees (or Merkle tries) with 512-byte neural state embeddings (produced by a transformer encoder inside a ZK circuit and attested in hardware enclaves) is the core breakthrough of NERV. This substitution delivers multiple compounded advantages across privacy, scalability, efficiency, verifiability, and long-term evolution. Below is a comprehensive list of all advantages explicitly claimed or directly implied in the whitepaper:

### **1. Massive State Compression (~900× smaller state root)**

- Traditional blockchains store the entire state as a Merkle trie, requiring gigabytes for hundreds of millions of accounts.



- NERV compresses the entire shard state (all key-value pairs) into a single fixed 512-byte floating-point vector.
- Result: State root is  $\sim 900\times$  smaller than comparable zkEVM or Merkle-based roots, dramatically reducing storage, bandwidth, and light-client sync costs.

## 2. Homomorphic Updates ( $O(1)$ state transitions without decompression)

- Balance transfers are homomorphic: a simple transfer tx induces an additive delta vector  $\delta(tx) \in \mathbb{R}^{512}$  such that  $\text{new\_embedding} = \text{old\_embedding} + \delta(tx)$  with error  $< 10^{-9}$ .
- Batched deltas for 256 transfers remain  $\sim 512$  bytes (average  $\sim 2$  bytes per tx).
- Nodes update the state root by simple vector addition — no need to load, decompress, modify, or re-hash a full trie.
- Enables constant-time state transitions regardless of state size.

## 3. Tiny Inclusion Proofs and Verifiable Delay Witnesses (VDWs)

- Traditional inclusion proofs are 300–900 bytes (Ethereum) or 300–500 bytes (zkEVMs).
- NERV inclusion proofs (part of VDWs) are  $\leq 800$  bytes for single tx, averaging  $\sim 1.6$  bytes per tx when batched.
- VDWs (permanent proof of inclusion) average 1.4 KB (never  $> 1.8$  KB) and verify in  $< 80$  ms on an iPhone 15.
- Users get permanent, offline-verifiable proof of private transactions without downloading the chain.

## 4. Irreversible Privacy (no addresses, balances, or metadata ever exposed)

- The transformer is deliberately non-invertible (many-to-one mapping with entropy  $> 2^{4000}$  and non-linearities).
- Even with unlimited quantum computation, recovering any private key, balance, or transaction detail from the embedding alone reduces to the new "Neural Network Inversion Problem" hardness assumption.
- Privacy is default and endemic — no opt-in shielding, no view keys, no metadata leaks.

## 5. Recursive Provability and ZK-Friendly Design

- The entire encoder runs inside a Halo2 circuit ( $\sim 7.9$  M constraints) with Nova folding for recursion.
- Every embedding update is accompanied by a succinct recursive proof of correct computation.
- Proof sizes are  $\sim 750$ – $800$  bytes (vs. hundreds of KB for zkEVMs), enabling efficient verification on mobile devices and light clients.

## 6. Enables Dynamic Neural Sharding (shards split/merge like cells)

- Embeddings can be deterministically bisected using a seed-derived hyperplane, splitting a shard's state into two child embeddings without data migration or decompression.
- Splits complete in <4 seconds; merges similarly fast.
- Combined with LSTM load prediction, this allows truly infinite horizontal scalability with no theoretical TPS ceiling.

## 7. Perpetual Self-Improvement via Useful-Work

- Every 30 days, federated learning produces a new encoder  $\mathcal{E}(\theta')$  that must prove (via Halo2) preserved homomorphism (error  $\leq 10^{-9}$ ).
- Nodes are paid to train these upgrades, making the embeddings progressively better at compression, homomorphism, and prediction over time.
- The network literally becomes more efficient and intelligent endogenously.

## 8. Reduced Bandwidth and Storage for All Nodes

- Validators only need to store/replicate 512-byte embeddings + recent deltas (erasure-coded).
- Historical state can be archived efficiently on Arweave/IPFS.
- Light clients sync permanently with <100 KB of data.

## 9. Faster Consensus and Finality

- AI-native optimistic consensus has validators predict the next embedding hash using a distilled model.
- Agreement on a 32-byte hash (not a full Merkle root) enables sub-second probabilistic finality (median ~600 ms).

## 10. Lower Proof Generation and Verification Costs

- Recursive proofs are dramatically smaller and faster than zkEVM equivalents.
- Enables features like instant VDWS, cheap fraud proofs, and mobile-friendly verification.

In summary, replacing Merkle trees with neural embeddings doesn't just improve one metric — it simultaneously solves the privacy-scalability-quantum trilemma while creating a living, self-improving system. The trade-off (reliance on a new cryptographic assumption and periodic model upgrades) is mitigated by formal proofs, hardware attestation, and ZK verifiability. This is why NERV claims to be the first blockchain that behaves like a "neural nervous system."

**Q: Could you please list all the aspects of this blockchain technology in detail that makes it quantum-proof?**

NERV is designed to be **fully quantum-proof from genesis block 0**, meaning no part of its critical cryptographic paths relies on algorithms vulnerable to quantum attacks (e.g., no ECDSA, EdDSA, secp256k1, or any elliptic-curve cryptography anywhere). This is a deliberate,

non-negotiable principle stated explicitly in the whitepaper (Section 7): "Zero legacy elliptic curves in any critical path – from genesis block 0." Unlike most blockchains in 2025 (including Bitcoin, Ethereum, Solana, and nearly all others, which still depend on quantum-vulnerable ECDSA/EdDSA for signatures), NERV eliminates this risk entirely at the protocol level.

Here are all the specific aspects that make NERV quantum-resistant, with detailed explanations of how and why they achieve this:

## 1. Exclusive Use of NIST-Standardized Post-Quantum Signature Algorithms

- **Primary: CRYSTALS-Dilithium-3 (NIST Level 3 security)** Used for all standard signatures (e.g., transaction signing, validator votes, attestations, VDWs).
  - **Why quantum-proof:** Dilithium is lattice-based (Module-Learning With Errors problem), which has no known efficient quantum algorithm to solve (resistant to Shor's algorithm). NIST finalized it as ML-DSA in FIPS 204 (August 2024), confirming its robustness after years of global cryptanalysis.
  - Key sizes: PK ~1.8 KB, signature ~3.3 KB; verification ~58  $\mu$ s on modern hardware.
- **Stateless backup: SPHINCS+-SHA256-192s-robust** Used for cold/genesis keys and high-caution scenarios (e.g., "quantum doomsday mode").
  - **Why quantum-proof:** Purely hash-based (no algebraic structure for quantum speedups like Grover's is only quadratic, not exponential threat). NIST finalized it as SLH-DSA in FIPS 205. Signatures are larger (~41 KB) but extremely conservative.

This ensures **no quantum-vulnerable signatures ever touch the chain**, unlike retrofitted projects that must migrate legacy keys (risking "harvest now, decrypt later" attacks).

## 2. Post-Quantum Key Encapsulation and Onion Routing

- **ML-KEM-768 (formerly Kyber-768, NIST Level 3)** Used for all key encapsulation (enclave-to-enclave exchanges, 5-hop anonymous ingress mixer).
  - **Why quantum-proof:** Also lattice-based (Module-LWE), finalized as FIPS 203 in 2024. Ciphertext ~1.1 KB, decapsulation ~42  $\mu$ s.
  - Hybrid mode with legacy curves is available temporarily but not in critical paths.
- No Diffie-Hellman or RSA equivalents—quantum computers break those via Shor's algorithm in polynomial time.

## 3. Quantum-Resistant Hashing

- **SHA3-256 + BLAKE3** Both are believed quantum-resistant (Grover's algorithm provides only quadratic speedup, requiring  $\sim 2^{128}$  operations for brute-force—still infeasible). Used for all hashing needs (tx hashes, embedding roots, etc.).

## 4. Cryptographic Agility for Future Threats

- Built-in **CryptoVersion enum** allows governance-voted migrations (180-day vote) to newer primitives (e.g., Dilithium-5, Falcon-1024, or future NIST backups like HQC, selected in March 2025).
- Historic verification remains compatible—no chain-breaking hard forks needed.

## 5. Additional Hardening via New Assumptions and Design

- **Neural Network Inversion Problem:** The privacy of neural state embeddings reduces to inverting a deliberately non-linear transformer (entropy  $>2^{4000}$ ), believed post-quantum secure.
- All critical computations (e.g., consensus, mixer) occur in hardware enclaves with remote attestation—preventing side-channel leaks exploitable by quantum or classical adversaries.

## Strong Case: Why NERV Is Truly Quantum-Proof

- **From Day One:** Most blockchains (e.g., Ethereum's planned PQC migrations, Bitcoin's lack thereof) are quantum-vulnerable today and require risky upgrades. NERV launches in June 2028 with zero legacy crypto, avoiding migration attacks where adversaries "harvest" encrypted data/signatures now for future quantum decryption.
- **Real-World Validation:** Relies exclusively on NIST's finalized 2024 standards (ML-KEM/ML-DSA/SLH-DSA), which survived intense global scrutiny—no known quantum breaks as of late 2025.
- **No Compromises:** Eliminates the "quantum trilemma" risk that plagues partial-adoption projects (e.g., Penumbra or some L2s with medium/partial resistance).
- **Forward-Looking:** Agility + backups (SPHINCS+ as hedge) ensure longevity even if lattice assumptions weaken (unlikely, per current research).

In summary, NERV isn't just "quantum-resistant"—it's designed as a post-quantum native blockchain, making it one of the safest long-term stores of value against future quantum threats. This positions it uniquely in the 2025–2030 landscape, where most chains remain exposed.