Experimental Test Plan: Thermodynamic Truth vs. PBFT & HoneyBadger BFT

Objective: To rigorously quantify the performance, scalability, and resilience of the Thermodynamic Truth Protocol relative to industry-standard Byzantine Fault Tolerance (BFT) mechanisms.

1. Experimental Design

The comparative analysis will focus on three primary dimensions: **Scalability** (Latency/Throughput vs. Node Count), **Resilience** (Performance under Byzantine faults), and **Resource Efficiency** (Energy/Bandwidth usage).

1.1 Protocols Under Test

Protocol	Type	Complexity (Theoretical)	Key Characteristic
PBFT (Castro & Liskov)	Deterministic	$O(n^2)$	Low latency in small clusters; high message overhead.
HoneyBadger BFT	Asynchronous	$O(n^2 \log n)$	Robust to network timing; high bandwidth usage.
Thermodynamic Truth	Probabilistic	$O(n)$	Physics-based convergence; localized interactions.

2. Hardware & Environment Configuration

To ensure reproducibility and isolate variables, all tests will be conducted in a controlled environment simulating a Wide Area Network (WAN).

2.1 Infrastructure Specification

- **Cluster Size**: 4 to 100 nodes (incrementally: 4, 16, 36, 64, 100).
- **Node Specification**:
  - **CPU**: 4 vCPU (AWS c5.xlarge equivalent)
  - **RAM**: 8 GB
  - **Network**: 10 Gbps interface
- **Network Simulation**: `tc` (Linux Traffic Control) will be used to inject artificial latency and jitter.
  - **Latency**: 100ms $\pm$ 20ms (simulating global distribution).
  - **Packet Loss**: 0.1% (simulating real-world imperfections).

3. Test Scenarios & Metrics

Scenario A: Scalability (The "Happy Path")

Goal: Measure baseline performance as the network grows.

- **Conditions**: 0 Byzantine nodes, stable network.
- **Metrics**:
  - **Latency to Finality**: Time from transaction submission to irreversible consensus.
  - **Throughput**: Transactions Per Second (TPS) at saturation.
  - **Message Complexity**: Total bytes transmitted per finalized block.

Scenario B: Byzantine Resilience

Goal: Quantify degradation under active attack.

- **Conditions**: $f = \lfloor (n-1)/3 \rfloor$ Byzantine nodes.
- **Attack Vectors**:
  - **Equivocation**: Sending conflicting values to different peers.
  - **Silence**: Refusing to propose or vote.

    - **Sybil/Energy Spam**: (Specific to ThermoTruth) Attempting to overwhelm with low-energy proofs.
- **Metrics**:
    - **Throughput Drop**: % decrease in TPS compared to Scenario A.
    - **Recovery Time**: Time to restore consensus after a partition heals.

Scenario C: Asynchronous Stress Test

Goal: Evaluate performance under extreme network conditions.

- **Conditions**: Random message delays up to 5 seconds; 5% packet loss.
- **Metrics**:
    - **Liveness**: Does the protocol stall? (Yes/No)
    - **Latency Variance**: Standard deviation of block times.

4. Data Collection & Analysis

All metrics will be captured using Prometheus exporters attached to the gRPC endpoints of each node.

4.1 Key Performance Indicators (KPIs)

Metric	Unit	Definition
Finality Latency	Milliseconds (ms)	$T_{final} - T_{submit}$
Throughput	Tx/sec	Total committed transactions / Test duration
Bandwidth Usage	MB/sec	Average ingress + egress per node
CPU Load	% Utilization	Average CPU usage across all honest nodes
Energy Efficiency	Joules/Tx	(Est. CPU power * Time) / Tx count

5. Success Criteria

The Thermodynamic Truth Protocol will be considered superior for the target use case if:

1. **Scaling**: It maintains $< 1s$ latency at 100 nodes (where PBFT typically exceeds 5s).
2. **Throughput**: It achieves > 1000 TPS on standard hardware.
3. **Resilience**: It successfully converges (Error $< 0.05°C$) even with 33% Byzantine nodes.

6. Execution Schedule

1. **Day 1**: Infrastructure provisioning (Terraform/Ansible) and baseline calibration.
2. **Day 2**: Execution of Scenario A (Scalability) for all three protocols.
3. **Day 3**: Execution of Scenario B (Byzantine) and C (Asynchrony).
4. **Day 4**: Data aggregation, visualization generation, and report writing.