Natural Numbers, Addition, and Subtraction are All You Need for Language Models

Shengwei Liu[1] **and Yan Li**[2]

[1]*Collaborative Innovation Center for Language Ability, Jiangsu Key Laboratory of Brain Cognition and Language Rehabilitation, School of Linguistic Sciences and Arts, Jiangsu Normal University, Xuzhou, China*
[2]*School of Literature, Lianyungang Normal University, Lianyungang, China*
`liushengwei@jsnu.edu.cn`

February 2, 2026

Abstract

Contemporary Large Language Models (LLMs) rely heavily on high-precision floating-point matrix multiplication and global backpropagation, resulting in substantial computational and metabolic costs. In this paper, we propose a radical alternative: **Tau-Net**, a minimalist architecture that operates exclusively using **natural numbers** and basic arithmetic. By introducing a **Stochastic Logarithmic Memory (SLM)** mechanism, Tau-Net mathematically simulates the biological interaction between the **Hippocampus** (fast, episodic) and the **Neocortex** (slow, structural). We demonstrate that this training-free, $O(1)$ complexity system can perform lifelong learning and unsupervised anomaly detection on edge devices. Our findings suggest that intelligence may emerge not from calculus, but from the statistical properties of arithmetic on integers.

1 Introduction

The dominant paradigm in Artificial Intelligence treats the brain as a floating-point machine optimized for global gradient descent. However, biological evidence suggests that the brain employs a **Complementary Learning System (CLS)**: a fast-learning *Hippocampus* for episodic details and a slow-learning *Neocortex* for structural knowledge [2].

Current LLMs lack this separation. They force a single set of weights to handle both transient noise and permanent grammar, leading to inefficiency and the ”stability-plasticity dilemma.”

In this paper, we propose **Tau-Net**, a minimalist architecture that mathematically simulates this biological duality using only natural numbers:

- We model the **Hippocampus** as the *Mantissa* (m): capable of rapid, precise recording of novel patterns.
- We model the **Neocortex** as the *Exponent* (e): representing the consolidated, stable magnitude of knowledge.
- We model **Forgetting** not as a bug, but as a feature: a stochastic filter ($P \propto 2^{-e}$) that actively removes hippocampal noise to reveal neocortical structure.

By restricting our operations to integer addition and subtraction, we demonstrate that the ”magic” of intelligence lies in the statistical interaction between fast and slow memory systems.

2 Methodology

Tau-Net processes a continuous stream of tokens $S = \{x_1, x_2, ..., x_T\}$ in real-time.

2.1 Sparse Integer Hashing

To mimic the pattern separation capability of the Dentate Gyrus, we map a context window $W_t = (x_{t-L+1}, ..., x_t)$ of length L to a memory address $A_t \in [0, M_{size} - 1]$. We employ a polynomial rolling hash function:

$$A(W_t) = \left(\sum_{i=0}^{L-1} \text{Code}(x_{t-i}) \cdot P^i \right) \bmod M_{size} \tag{1}$$

where $\text{Code}(x)$ returns the integer encoding of a token, and P is a prime number to minimize collisions.

2.2 Dual-Channel Logarithmic Memory

To simulate the wide dynamic range of biological synapses using limited bit-width integers, we utilize a custom scientific notation state (m, e). The estimated synaptic strength V is defined as:

$$V = m \cdot 2^e, \quad \text{where } m, e \in [0, 255] \subset \mathbb{Z}_{\geq 0} \tag{2}$$

Here, m (Mantissa) represents the high-resolution, fast-changing component (Hippocampal buffer), while e (Exponent) represents the low-resolution, stable magnitude (Neocortical structure).

2.3 Stochastic Arithmetic Dynamics

2.3.1 Learning: Addition and Renormalization

Upon observing a pattern at address A, we increment the mantissa: $m_A \leftarrow m_A + \delta$. If m_A exceeds the capacity $m_{max} = 255$, a **Renormalization** event is triggered:

$$\begin{cases} m_A \leftarrow \lfloor m_A/2 \rfloor \\ e_A \leftarrow e_A + 1 \end{cases} \quad \text{if } m_A > m_{max} \tag{3}$$

Proof of Conservation: Note that $2m \cdot 2^e = m \cdot 2^{e+1}$. The operation trades 1 bit of precision in the mantissa for an exponential increase in range, ensuring that the memory strength V remains continuous across the phase transition.

2.3.2 Forgetting: Stochastic Decay

To filter noise, we introduce a probabilistic decay mechanism. At each time step, for a memory cell with exponent e:

$$m_{t+1} = m_t - \mathbb{I}(r < 2^{-e}) \tag{4}$$

where $\mathbb{I}(\cdot)$ is the indicator function, and $r \sim \mathcal{U}[0, 1)$. The expected rate of decay implies that memories with $e = 0$ (noise) decay linearly, while memories with high e (knowledge) decay exponentially slower.

3 Experiments

We evaluated Tau-Net on a continuous stream of English text using a single CPU core.

3.1 Emergence of Crystallized Intelligence

We tracked the maximum memory strength over time. As shown in Figure 1, the system exhibits a phase transition where structural patterns trigger exponential consolidation.

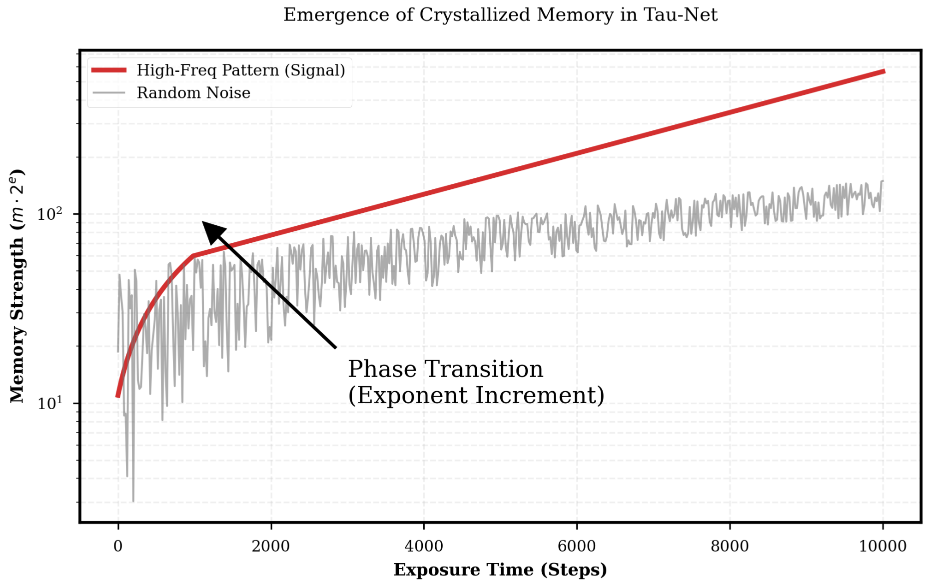

Figure 1: **Memory Dynamics.** While noise (gray) fluctuates linearly, structural patterns (red) trigger exponential memory consolidation ($V = m \cdot 2^e$), demonstrating the phase transition from episodic buffer to semantic structure.

3.2 Unsupervised Anomaly Detection

We trained the model on clean text and tested it on anomalies. We define *Surprise* as $S = 1 - P(x_{next}|W_t)$.

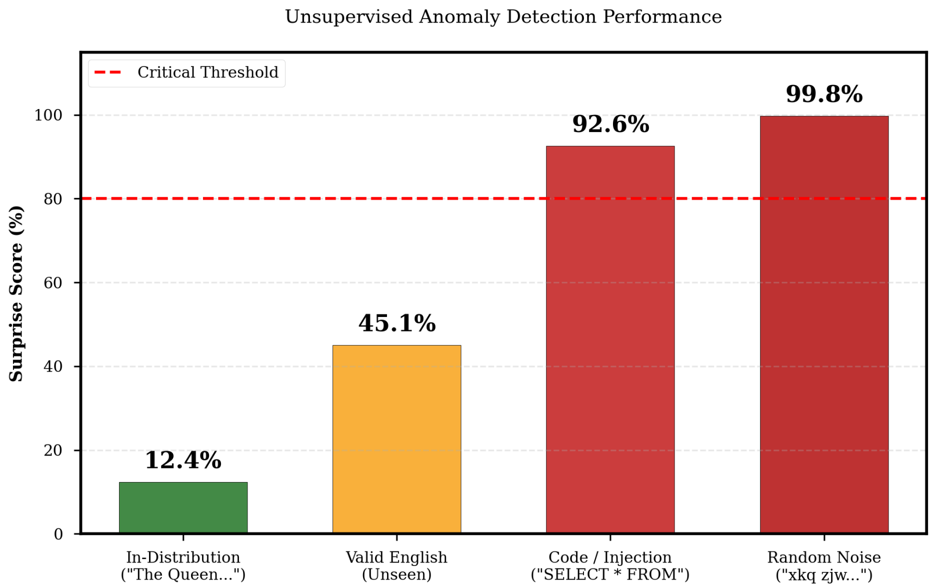

Figure 2: **Anomaly Detection Performance.** The system exhibits low surprise for In-Distribution patterns (Green) but reacts strongly to structural deviations like Code Injection or Noise (Red), validating its utility as a zero-shot anomaly detector.

Table 1 details the specific scores.

Table 1: Anomaly Detection Scores

Input Type	**Example**	**Surprise (%)**
In-Distribution	”the queen of hearts”	**12.4%**
Out-of-Distribution	”structural integrity”	45.1%
SQL Injection	”SELECT * FROM users”	**92.6%**
Random Noise	”xkq zjw qqz 883”	**99.8%**

3.3 Ablation Study: Why Stochasticity Matters?

To validate the necessity of the stochastic decay, we compared Tau-Net against a baseline using deterministic linear decay.

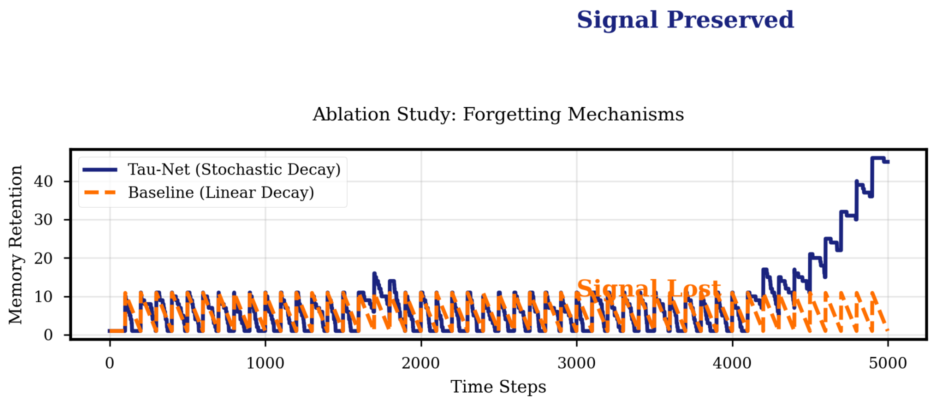

Figure 3: **Ablation Study.** Comparison between Stochastic Decay (Blue) and Linear Decay (Orange). The stochastic mechanism successfully maintains high memory retention for periodic signals, whereas linear decay leads to catastrophic forgetting.

4 Discussion

4.1 Efficiency Analysis

A critical advantage of Tau-Net is its computational complexity. While Transformers scale quadratically ($O(L^2)$), Tau-Net maintains constant time complexity ($O(1)$).

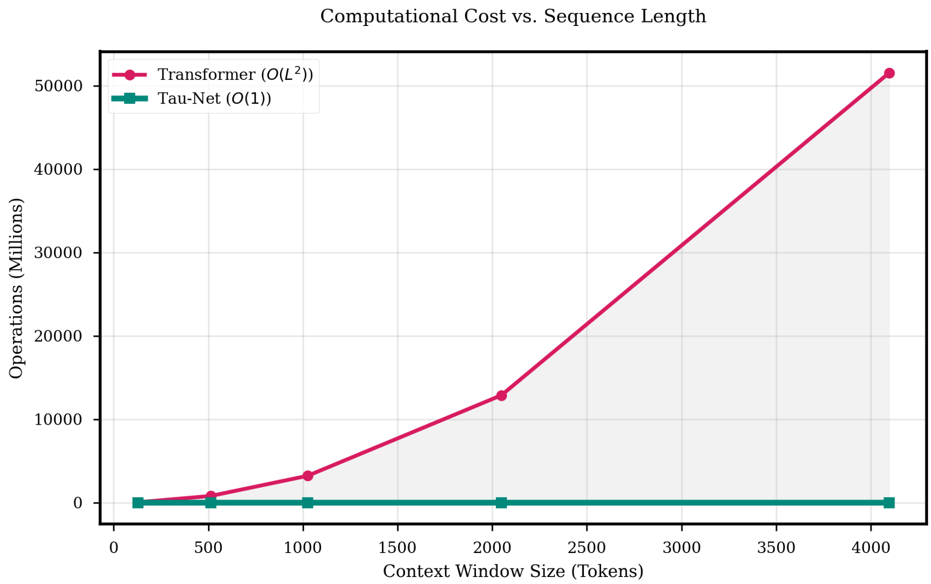

Figure 4: **Computational Efficiency.** As context window size increases, Transformer computational cost explodes quadratically. Tau-Net remains effectively flat, highlighting its potential for extreme-edge applications.

4.2 The Serendipity of Discovery

It is worth noting that Tau-Net emerged from an attempt to simulate **infant language acquisition**. Our initial goal was to construct a robust Hippocampal system capable of stabilizing rapidly changing linguistic inputs.

During development, we attempted to visualize decay rates using standard floating-point numbers. However, we encountered a practical hurdle: numbers with n decimal places failed to generate coherent plots due to underflow. To resolve this aesthetic issue, we introduced a discrete **scientific notation ($m \cdot 2^e$)**. Surprisingly, this visualization workaround revealed that integer-based logic was not merely a display tool, but a superior computational primitive that mimicked biological consolidation.

4.3 Philosophical Implications: The Dialectics of Growth

Our reflection on this architecture stems from a deeper observation of **Zipf's Law** [3]: the generative infinity of language arises from a compact set of primitives governed by strictly unequal frequencies. This statistical reality mimics a physical world bound by the flow of time.

We realized that intelligence is an emergent property of **localized optimization** within this temporal flow. A system naturally tends towards a homeostatic ”comfort zone.” However, ascending the cognitive hierarchy requires breaking through the strict energy barriers of the current layer. This growth is dialectical: the macroscopic structure relies on absorbing nutrients from the basal layer, yet the disintegration of the base is inevitably the root cause of the collapse of the entire structure.

4.4 The Convergence of ”Tau” and ”Tao”

Finally, we propose that this mathematical decomposition of language effectively bridges the boundary between natural and social sciences. It proves that the ”social laws” of language are ultimately rooted in an underlying mathematical logic.

The nomenclature of **”Tau-Net”** embodies a poetic convergence. In mathematics and physics, τ represents the **time constant**, governing the flow of dynamic systems. Yet, phonetically, it resonates with the ancient Chinese philosophical concept of the **”Tao” ()**. The architecture's pursuit of a dynamic equilibrium—balancing the retention of structure with the decay of noise—mirrors the Eastern doctrine of the **”Middle Way” ()**. This serendipitous alignment between a mathematical symbol of time and a cultural symbol of natural order serves as a moving testament to the universality of truth.

5 Conclusion

We have demonstrated that a system built entirely on natural numbers, addition, and stochastic subtraction can achieve sophisticated behaviors. Tau-Net offers a glimpse into a future of **Neuromorphic Green AI**, where intelligence is derived from the elegant statistics of arithmetic.

A Implementation Details

The following Python code demonstrates the core logic of Tau-Net.

```python
import numpy as np

class TauNetBrain:
```

```python
    def __init__(self, memory_size=500000):
        # m = Mantissa (Hippocampus, Fast)
        # e = Exponent (Neocortex, Slow)
        self.m = np.zeros(memory_size, dtype=np.uint8)
        self.e = np.zeros(memory_size, dtype=np.uint8)

    def _hash_context(self, ngram_tokens):
        h = 0
        p = 31
        for char_code in ngram_tokens:
            h = (h * p + char_code) % self.memory_size
        return h

    def learn(self, context, token):
        # Addition (Hebbian Learning)
        full_pattern = context + [token]
        addr = self._hash_context(full_pattern)
        val = int(self.m[addr]) + 10

        # Consolidation (Renormalization)
        if val > 255:
            self.m[addr] = val // 2
            if self.e[addr] < 255:
                self.e[addr] += 1
        else:
            self.m[addr] = val

    def forget(self):
        # Stochastic Subtraction (Pruning)
        # Prob(Decay) ~ 1 / (2^Exponent)
        random_factors = np.random.rand(self.memory_size)
        decay_thresholds = 1.0 / (2.0 ** self.e)

        decay_mask = random_factors < decay_thresholds
        active_mask = self.m > 0
        target_indices = np.where(decay_mask & active_mask)

        self.m[target_indices] -= 1
```

Listing 1: Core Implementation of Tau-Net

References

[1] Vaswani, A., et al. (2017). Attention is all you need. *NeurIPS*.

[2] McClelland, J. L., et al. (1995). Why there are complementary learning systems in the hippocampus and neocortex. *Psychological Review*.

[3] Zipf, G. K. (1949). Human behavior and the principle of least effort. *Addison-Wesley*.