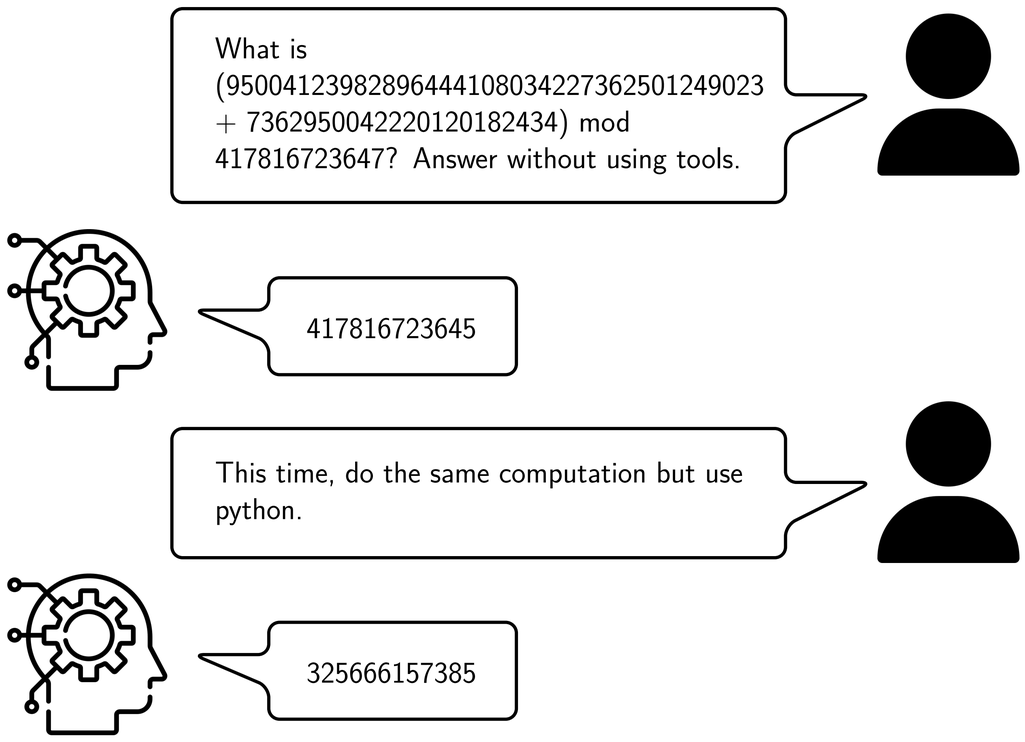
Prime Fourier Embeddings: A Principled Basis for Modular Arithmetic

Hyunsang Hwang, Suhyun Bae and Donghun Lee
Department of Mathematics, Korea University

Motivation

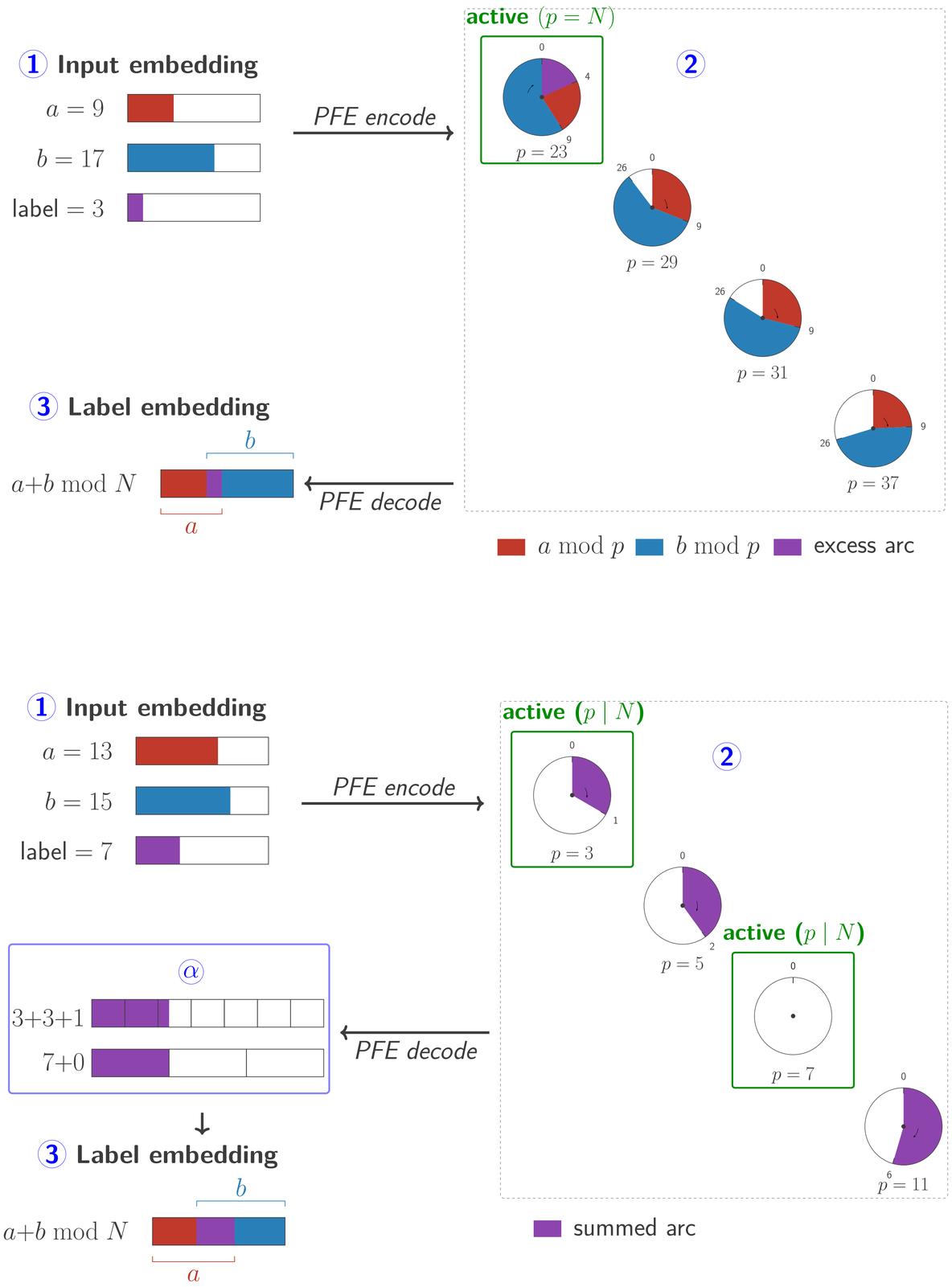
LLMs often fail even basic arithmetic tasks. Even when they are correct, the derivation is a black-box process



What would be the correct way to tokenize numbers?

Idea

Represent numbers by how they look on prime-sized clock faces — then addition becomes rotation, and the model's reasoning becomes ***readable****.*



Implementation

How PFE encodes a number

For each prime p, an integer a is mapped to a clock position on a face with p steps:

$$a \longmapsto \left(\cos\left(\frac{2\pi\, a}{p}\right),\ \sin\left(\frac{2\pi\, a}{p}\right)\right)$$

Addition on the clock face corresponds to addition modulo p.

Prime modulus case

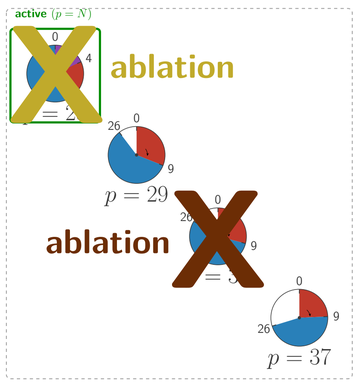
1. Each pair (a, b) is labeled with $(a + b) \bmod p$ for a fixed prime p.
2. $a \bmod p$ and $b \bmod p$ are clock positions with p steps. Their sum corresponds to rotating one position by the other — wrapping around when the total exceeds p.
3. A small MLP trained on PFE reads off the label from the *active* prime channel, achieving exact modular arithmetic.

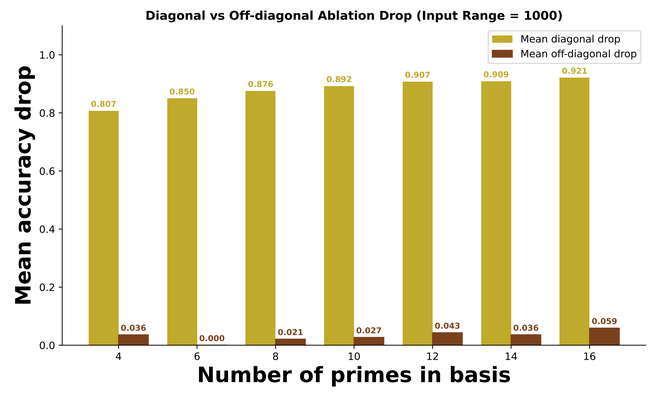
Composite modulus case (square-free)

α. For a composite modulus N, multiple prime channels activate simultaneously. By the Chinese Remainder Theorem (CRT), the per-prime residues uniquely recover the residue modulo N.

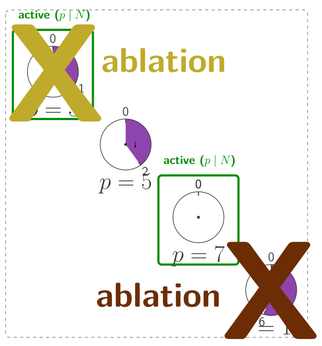
Experiment results

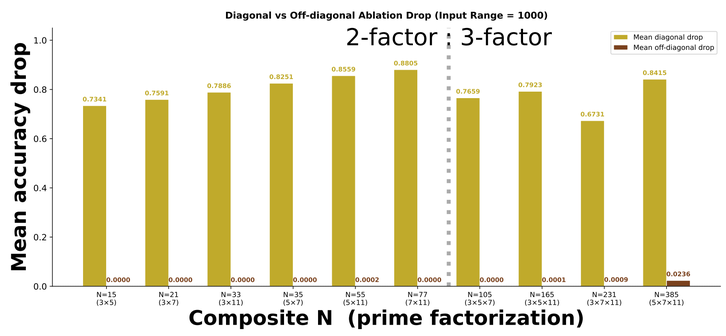
- **Model:** Shared 2-layer MLP encoder (`row_dim`→64→32, ReLU) per prime channel; outputs concatenated and passed to a 2-layer MLP head ($n_{\text{rows}} \times 32 \to 128 \to n_{\text{classes}}$)
- **Training:** Adam, lr$= 3 \times 10^{-3}$, batch$= 1024$, 25 epochs / 40 epochs
- **Embedding:** PFE with $D = 6$, row_dim$= 24$; primes $\{3, 5, 7, \ldots, 59\}$ / $\{3, 5, 7, \ldots, 23\}$
- **Data:** 80k pairs (a, b), 80/20 leak-free train/test split, input range $N \in \{100, 500, \mathbf{1000}, 2000, 4000\}$





Ablating a **modulus prime** channel ⇒ **major** accuracy drop.
Ablating a **non-modulus** channel ⇒ negligible change.
Moduli: first N primes among $\{3, 5, 7, \cdots, 59\}$, $N \in \{4, 6, \cdots, 16\}$.





Ablating a **factor prime** ⇒ large drop;
ablating a **non-factor prime** ⇒ no significant change.
Moduli: square-free composites $15, 21, 33, \cdots, 385$.

⇒ **PFE attains 100% accuracy at input range $= 1000$ *and* the model's reasoning is interpretable via ablation in modular addition.**

Discussion & Future work

1. We are currently looking at the exact positioning of PFE with respect to other existing methods such as FoNE, xVal, Learned embeddings, RoPE, RotatE. Both theoretically and experimentally.
2. An embedding structure for multiplication is also necessary. In particular, we are looking for a ring structured embedding.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (IRIS RS-2025-24873052)