PhiTransformer: Fractal Geometry and Golden-Angle Sparse Attention for Efficient Long-Context Transformers

Jacques Venter (South Africa)

August 2025

Abstract

We propose PhiTransformer, a geometry-aware Transformer guided by phyllotaxis (golden angle) and fractal self-similarity. Tokens occupy a golden-angle spiral; attention combines annulus-banded locality with φ-scaled hierarchical "spine" links; embeddings may be tied across φ-scales. The resulting $O(n \log n)$ connectivity aims to reduce attention compute while preserving long-range information flow. We formalize the topology, provide CPU and Triton reference kernels, estimate training/inference gains, and outline an experimental plan for validation. Claims of efficiency are explicitly speculative pending experiments.

1. Introduction

Long-context modeling stresses the quadratic cost of dense attention. Windowed and dilated schemes reduce cost but often degrade cross-range interactions; kernel and low-rank methods trade exactness for approximations; SSMs offer linear scaling with different inductive biases. Our approach treats token layout as a geometric design problem: arrange positions on a golden-angle spiral (uniform packing, low aliasing), then align attention edges with local annulus bands plus logarithmically many parent links along a φ-scaled spine.

2. Related Work

Positional encodings: sinusoidal and rotary (RoPE). Efficient attention: Longformer (windows + global), BigBird (random/global), LongNet (dilations). Beyond attention: S4/Mamba (state-space), Perceiver-IO (latent array). Compression: VQ-VAE and Product Quantization. Phyllotaxis is well-studied in physics and MRI. Our novelty lies in combining golden-angle geometry, band+spine topology, and fractal embeddings within a Transformer.

3. Method

3.1 Golden-Angle Spiral Coordinates

Tokens map to polar coordinates as follows:

$$\theta_i = i\,\alpha$$

$$r_i = c\,\sqrt{i}$$

A geometric variant with explicit scales uses:

$$r_i = r_0 \varphi^{\frac{i}{k}}$$

Annuli are indexed by a(i) = ⌊log_φ (r_i / r_0)⌋, creating constant-density rings and uniform angular spacing.

3.2 Band + Spine Attention

For token i, neighbors are the union of (i) band neighbors within the same annulus inside angular window ±W, and (ii) a parent chain obtained by radial down-sampling approximately by φ^2 per step, giving O(log n) ancestors. Optional bridges connect across scales at radii φ^k r_i.

3.3 Fractal Multiscale Embeddings (Speculative)

Represent each token vector as a sum over φ-scaled basis blocks, optionally quantized per scale with parameter tying across adjacent scales.

3.4 Spiral Decoding and Memory

Coarse-to-fine decoding proceeds annulus-by-annulus; high-entropy sectors are refined in follow-up passes. KV-cache and document indices can be sharded by (annulus, sector).

4. Complexity

$$L = log_{\varphi}\, n$$

Each token attends to O(1) band neighbors and O(L)=O(log_φ n) ancestors; total edges are O(n log n).

5. Performance Estimates (Analytical)

End-to-end speedup at fixed quality:

$$speedup = \frac{1}{\left(\frac{s}{F_{deg}} + (1 - s)\right)}$$

At 64k tokens with w=128 and band±2, the degree factor is ≈9.2; for attention share s∈[0.6,0.8], this predicts ≈2.1×–3.5× step speedups.

6. Experimental Plan (Track A)

Data: long-form mix (PG-19, books, arXiv, Wikipedia, StackExchange, OSS docs). Models: ≈200M params, identical tokenizer. Baseline: RoPE + window ±128 + a few globals. Candidate: φ-spiral + band±2 + spine depth cap. Curriculum: 4k→8k→32k→64k. Metrics: PPL vs. length, Needle-in-Haystack recall@1, tokens/sec, VRAM, attention wall-time. Ablations: band width, spine depth, spiral variant, auxiliary losses.

7. Implementation Notes

Reference CPU and Triton kernels are provided for neighbor attention. Neighbors are packed in ELL [N, MAX_DEG] with a boolean mask. For backprop, implement a custom autograd wrapper around the two-pass softmax. Performance: batch multiple rows per program, tile D and neighbors, prefetch KV.

8. Novelty & Prior Art Assessment

Golden-angle positional layout: likely novel in Transformers; related polar/log mappings exist in vision. Band+spine topology: distinct from window+global (Longformer/BigBird) and dilations (LongNet). Fractal VQ/PQ embeddings with φ-tied codebooks: speculative novelty. Complexity: O(n log n) is a known target; the φ-structured route appears new. All untested claims are flagged as speculative.

9. Blockchain Layer (Speculative Infrastructure)

We do not place blockchain in the numerical inner loop. Instead, a ledger can support trust, attribution, and compliance: provenance and reproducibility (commit dataset/epoch/model hashes), verifiable inference (publish checkpoint hashes; prove outputs derive from a given model, optionally via ZK proofs), federated contribution rewards (track GPU-hours and gradient usefulness), and decentralized memory (sharing KV/annulus caches and retrieval indices across nodes with staking and eviction). A φ-inspired layered ledger—annulus sidechains committing to a spine root chain—mirrors the architecture's locality + hierarchy.

10. Broader Impact and Ethics

If validated, φ-structured attention could cut long-context energy costs and broaden access. Risks include over-claiming efficiency, misuse for mass summarization, and bias amplification. We commit to open code and datasets where licensing permits, and to reporting negative results.

Appendix A. Reproducibility Checklist

We include equations for coordinate mapping and speedup, pseudocode for band+spine attention, reference kernels and configs, an evaluation plan with ablations, and build/run instructions.

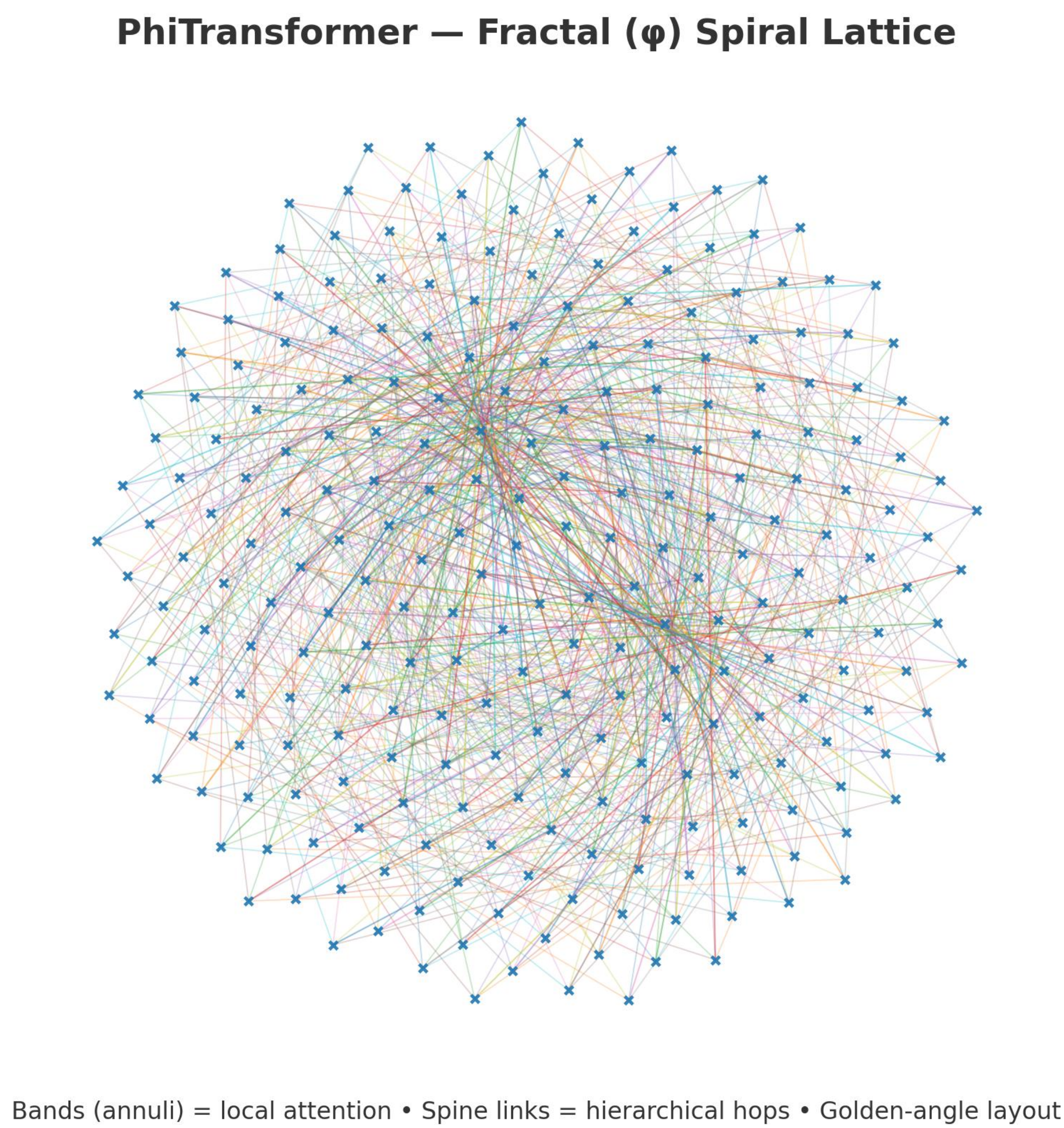

Figure: Golden-angle spiral with local bands and φ-scaled spine links (schematic).

References

[1] Su et al. (2021) RoFormer.
[2] Beltagy et al. (2020) Longformer.
[3] Zaheer et al. (2020) BigBird.
[4] Wang et al. (2023) LongNet.
[5] Gu et al. (2021) S4.
[6] Jaegle et al. (2021) Perceiver IO.
[7] van den Oord et al. (2017) VQ-VAE.
[8] Jégou et al. (2011) PQ.
[9] Esteves et al. (2017) Polar Transformers.
[10] Feng et al. (2020) Golden-angle MRI.
[11] Douady & Couder (1992) Phyllotaxis growth process.