

# Novelty Assessment Report

**Paper:** Probing Rotary Position Embeddings through Frequency Entropy

**PDF URL:** <https://openreview.net/pdf?id=1JZuEDq62N>

**Venue:** ICLR 2026 Conference Submission

**Year:** 2026

**Report Generated:** 2026-01-05

## Abstract

Rotary Position Embeddings (RoPE) are widely used in Transformers to encode positional information in token representations, yet the internal frequency structure of RoPE remains poorly understood. Previous studies have reported conflicting findings on the roles of high- and low-frequency dimensions, offering empirical observations but no unifying explanation. In this paper, we present a systematic framework that bridges these disparate results. We introduce Frequency Entropy (FE), a metric that quantifies the effective utilization of each RoPE frequency dimension, and we provide an analysis of how RoPE's sinusoidal components contribute to model representations on a per-dimension basis. Based on an analysis of the Llama-4 model, which incorporates both RoPE and NoPE layers, we find that the periodicity captured by FE appears in RoPE layers but not in NoPE layers. Furthermore, FE identifies dimensions in which energy concentrates under RoPE. These characteristics are observed across the spectrum rather than being confined to specific dimensions. Moreover, attenuating extreme-entropy dimensions at inference yields downstream accuracy that is statistically indistinguishable from the baseline, with modest perplexity improvements on average, suggesting that such dimensions are often redundant. Overall, FE provides a simple, general diagnostic for RoPE with implications for analysis and design.

### Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

## Core Task Landscape

This paper addresses: **Understanding Internal Frequency Structure of Rotary Position Embeddings**

A total of **35 papers** were analyzed and organized into a taxonomy with **15 categories**.

### Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Frequency Analysis and Theoretical Foundations**
- **Length Extrapolation and Context Extension**
- **Architectural Variants and Generalizations**
- **Domain-Specific Extensions and Applications**
- **Efficiency and Compression Methods**
- **Mechanistic Interpretability and Model Behavior**

### Complete Taxonomy Tree

- Understanding Internal Frequency Structure of Rotary Position Embeddings Survey Taxonomy
- Frequency Analysis and Theoretical Foundations
  - Frequency Dimension Analysis ★ (4 papers)
    - [0] Probing Rotary Position Embeddings through Frequency Entropy (Anon et al., 2026) [View paper](#)
    - [12] Rotary outliers and rotary offset features in large language models (Jonasson, 2025) [View paper](#)
    - [14] On the token distance modeling ability of higher RoPE attention dimension (Jiang Che, 2024) [View paper](#)
    - Spectral Theory and Matrix Properties (2 papers)
    - [16] Fourier Position Embedding: Enhancing Attention's Periodic Extension for Length Generalization (Erkai Hua, 2024) [View paper](#)
    - [19] Unpacking Positional Encoding in Transformers: A Spectral Analysis of Content-Position Coupling (Gu Zihan, 2025) [View paper](#)
    - Emergent Properties and Wavelet Behavior (2 papers)
    - [7] Round and round we go! what makes rotary positional encodings useful? (Barbero, 2024) [View paper](#)
    - [15] Beyond Position: the emergence of wavelet-like properties in Transformers (Nanni Umberto, 2025) [View paper](#)
- Length Extrapolation and Context Extension
  - Training-Free Extrapolation Methods (3 papers)
    - [8] DoPE: Denoising Rotary Position Embedding (Jing Xiong, 2025) [View paper](#)
    - [25] LaMPE: Length-aware Multi-grained Positional Encoding for Adaptive Long-context Scaling Without Training (Zhang Si-kui, 2025) [View paper](#)
    - [33] HoPE: A Novel Positional Encoding Without Long-Term Decay for Enhanced Context Awareness and Extrapolation (Chen, 2024) [View paper](#)
  - Adaptive Training and Fine-Tuning Methods (2 papers)
    - [20] Gpt rotational position embedding for length extrapolation (Zhijie Qu, 2023) [View paper](#)
    - [31] Efficient Long-Context Fine-Tuning of LLaMA 2 7B with QLoRA, Dynamic RoPE, and FlashAttention (Burak Aktas, 2025) [View paper](#)
- Architectural Variants and Generalizations
  - Learnable and Trainable RoPE Extensions (2 papers)
    - [10] ComRoPE: Scalable and Robust Rotary Position Embedding Parameterized by Trainable Commuting Angle Matrices (Yu Hao, 2025) [View paper](#)
    - [24] Head-wise Adaptive Rotary Positional Encoding for Fine-Grained Image Generation (Li Jiaye, 2025) [View paper](#)

- Selective and Input-Dependent Variants (2 papers)
- [2] Context-aware Rotary Position Embedding (Veisi Ali, 2025) [View paper](#)
- [6] Selective Rotary Position Embedding (Sajad Movahedi, 2025) [View paper](#)
- Additive and Alternative Formulations (1 papers)
- [32] Additive Rotary Embedding - A Competitive Alternative to RoPE (Chang, 2025) [View paper](#)
- Domain-Specific Extensions and Applications
  - Vision and Image Applications (2 papers)
  - [5] Rotary position embedding for vision transformer (Byeongho Heo, 2024) [View paper](#)
  - [17] SpatialSpectral Transformer With Patch-Local Mixed-Axis 2-D Rotary Position Embedding for Hyperspectral Image Classification (Zirak Khan, 2026) [View paper](#)
  - Video and Temporal-Spatial Modeling (3 papers)
  - [1] VideoRoPE: What Makes for Good Video Rotary Position Embedding? (Wei, 2025) [View paper](#)
  - [9] RoPECraft: Training-Free Motion Transfer with Trajectory-Guided RoPE Optimization on Diffusion Transformers (Bilecen, 2025) [View paper](#)
  - [30] PAS: A Training-Free Stabilizer for Temporal Encoding in Video LLMs (Bowen Sun, 2025) [View paper](#)
  - Multimodal and Vision-Language Models (2 papers)
  - [11] Revisiting Multimodal Positional Encoding in Vision-Language Models (Huang, 2025) [View paper](#)
  - [35] HoPE: Hybrid of Position Embedding for Long Context Vision-Language Models (Li, n.d.) [View paper](#)
  - Specialized Signal and Time-Series Domains (6 papers)
  - [4] Lightweight Spatio-Temporal Attention Network with Graph Embedding and Rotational Position Encoding for Traffic Forecasting (Xiao, 2025) [View paper](#)
  - [18] SPECTRE: Spectral Pre-training Embeddings with Cylindrical Temporal Rotary Position Encoding for Fine-Grained sEMG-Based Movement Decoding (Zihan Weng, 2025) [View paper](#)
  - [22] PriorSinger: Singing Voice Synthesis Model with Prior Condition Cross Attention (Zehua Zhang, 2025) [View paper](#)
  - [27] Edge-Deployed Band-Split Rotary Position Encoding Transformer for Ultra-Low-Signal-to-Noise-Ratio Unmanned Aerial Vehicle Speech Enhancement (Feifan Liu, 2025) [View paper](#)
  - [28] BiMa-Former: A Dual-Token Hybrid Model with Bidirectional Mamba and Transformer for Temporal-Multivariate Decoupled Forecasting (Jin Li, 2025) [View paper](#)
  - [34] MIRA: Medical Time Series Foundation Model for Real-World Health Data (Li Hao, 2025) [View paper](#)
- Efficiency and Compression Methods
  - KV Cache Compression (2 papers)
  - [3] KV-Latent: Dimensional-level KV Cache Reduction with Frequency-aware Rotary Positional Embedding (Shi Luohe, 2025) [View paper](#)
  - [23] EliteKV: Scalable KV Cache Compression via RoPE Frequency Selection and Joint Low-Rank Projection (Zhou, 2025) [View paper](#)
  - Parameter-Efficient Fine-Tuning (1 papers)
  - [21] RoSA: Enhancing Parameter-Efficient Fine-Tuning via RoPE-aware Selective Adaptation in Large Language Models (Dayan Pan, 2025) [View paper](#)
- Mechanistic Interpretability and Model Behavior (2 papers)
  - [13] How large language models encode theory-of-mind: a study on sparse parameter patterns (Yuheng Wu, 2025) [View paper](#)
  - [29] Sensitivity Meets Sparsity: The Impact of Extremely Sparse Parameter Patterns on Theory-of-Mind of Large Language Models (Wu, 2025) [View paper](#)

## Narrative

Core task: Understanding the internal frequency structure of Rotary Position Embeddings (RoPE). The field has organized itself around six main branches that reflect both theoretical inquiry and practical adaptation. Frequency Analysis and Theoretical Foundations examines the mathematical underpinnings of how RoPE encodes positional information through sinusoidal frequencies, including spectral properties and dimension-wise behavior. Length Extrapolation and Context Extension addresses the challenge of generalizing trained models to longer sequences, often by adjusting frequency bases or interpolation schemes. Architectural Variants and Generalizations explores modifications such as multi-dimensional rotations for vision or video (VideoRoPE[1], Rotary Vision Transformer[5]) and hybrid designs that blend RoPE with other encoding strategies. Domain-Specific Extensions tailor RoPE to specialized modalities like audio or multimodal settings (Multimodal Positional Encoding[11]), while Efficiency and Compression Methods seek to reduce computational overhead through selective application (Selective Rotary[6]) or pruning techniques. Finally, Mechanistic Interpretability investigates how RoPE influences attention patterns and model reasoning, bridging theory with observed behavior.

Several active lines reveal key trade-offs: works in frequency analysis (Rotary Outliers[12], Token Distance RoPE[14]) probe how individual frequency bands contribute to representation quality, whereas length-extension studies balance interpolation fidelity against training cost. The original paper, Frequency Entropy RoPE[0], sits squarely within Frequency Dimension Analysis, emphasizing entropy-based metrics to characterize how information is distributed across RoPE's frequency spectrum. This contrasts with neighboring efforts like Rotary Offset Features[26], which modifies the phase structure directly, or Token Distance RoPE[14], which reinterprets frequencies in terms of token separation. By quantifying frequency utilization through entropy, Frequency Entropy RoPE[0] offers a diagnostic lens that complements both the spectral-theoretic perspective (Fourier Position Embedding[16], SPECTRE[18]) and the practical tuning strategies seen in context-extension work (Context-aware RoPE[2], RoPECraft[9]).

## Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

### 1. Rotary outliers and rotary offset features in large language models

**Authors:** A Jonasson | **Year/Venue:** 2025 | **URL:** [View paper](#)

#### Abstract

∅ The high-norm features we identify in rotary embeddings could affect the quantization errors ∅; the outliers in rotary features are concentrated in rotary frequencies which never complete a ∅;

#### Relationship Analysis

Both papers belong to the Frequency Dimension Analysis category, examining individual frequency dimensions and their roles in RoPE. They overlap in analyzing how different RoPE frequency components contribute to model behavior, with both identifying patterns in high- and low-frequency dimensions. However, the original paper introduces Frequency Entropy (FE) as a unified metric to quantify frequency utilization and reconcile conflicting findings, while the candidate paper focuses on 'rotary offset features'—specific frequency

pairs with large magnitudes that create attention patterns—and derives mathematical bounds to predict which frequencies exhibit these features based on context length.

---

## 2. On the token distance modeling ability of higher RoPE attention dimension

**Authors:** Jiang Che, Meng, Fandong, Qi, Bqing, et al. (10 authors total) | **Year/Venue:** 2024 • Conference on Empirical Methods in Natural Language Processing | **URL:** [View paper](#)

### Abstract

Length extrapolation algorithms based on Rotary position embedding (RoPE) have shown promising results in extending the context length of language models. However, understanding how position embedding can capture longer-range contextual information remains elusive. Based on the intuition that different dimensions correspond to different frequency of changes in RoPE encoding, we conducted a dimension-level analysis to investigate the correlation between a hidden dimension of an attention head and...

### Relationship Analysis

Both papers belong to the Frequency Dimension Analysis category, examining how individual frequency dimensions in RoPE contribute to model behavior. They overlap in analyzing dimension-level utilization patterns and identifying that certain frequency dimensions play distinct roles in capturing positional information. However, the original paper introduces Frequency Entropy as a quantitative metric to measure spectral complexity and periodicity across all dimensions, while the candidate paper focuses specifically on identifying 'Positional Heads' that utilize higher-frequency dimensions for long-range dependencies through correlation analysis.

---

## 3. Rotary Offset Features in Large Language Models

**Authors:** Jonasson, Andr  , Andr  e Jonasson | **Year/Venue:** 2025 | **URL:** [View paper](#)

### Abstract

Transformer-based Large Language Models (LLMs) rely on positional encodings to provide sequence position information to their attention mechanism. Rotary Positional Encodings (RoPE), which encode relative position by rotating queries and keys, have become widely used in modern LLMs. We study the features and patterns that emerge in queries and keys when using rotary embeddings and introduce the concept of rotary offset features. Our analysis reveals that these features, which frequently exhibit ...

### Relationship Analysis

Both papers belong to the Frequency Dimension Analysis category, examining individual frequency dimensions and their roles in RoPE. They overlap in analyzing how different RoPE frequency components contribute to model behavior, with both identifying patterns in query/key representations across frequency dimensions. However, the original paper introduces Frequency Entropy (FE) as a unified metric to quantify utilization and periodicity across the spectrum, while the candidate paper focuses on 'rotary offset features'—specific frequency pairs with large magnitudes that never complete full rotational periods—and derives geometric bounds predicting their emergence based on context length and rotation angles.

---

## Contributions Analysis

**Overall novelty summary.** The paper introduces Frequency Entropy (FE), a metric quantifying how effectively each RoPE frequency dimension is utilized, and proposes a systematic framework to reconcile conflicting empirical findings on high- versus low-frequency roles. It resides in the Frequency Dimension Analysis leaf, which contains four papers examining individual frequency dimensions and their utilization patterns. This leaf sits within the broader Frequency Analysis and Theoretical Foundations branch, indicating a moderately populated research direction focused on understanding RoPE's internal mechanisms rather than adapting them for specific tasks.

The taxonomy reveals that neighboring leaves—Spectral Theory and Matrix Properties (two papers) and Emergent Properties and Wavelet Behavior (two papers)—pursue complementary angles: spectral analysis of Toeplitz matrices and wavelet-like multi-resolution processing. The original paper's dimension-level entropy approach bridges these perspectives by providing a per-dimension diagnostic tool, whereas spectral methods examine global matrix properties and emergent-behavior studies focus on training dynamics. The broader Frequency Analysis branch thus encompasses theoretical, dimension-wise, and emergent viewpoints, with the original work contributing a quantitative lens for dimension utilization.

Among 21 candidates examined, the Frequency Entropy metric itself (Contribution A: 10 candidates, zero refutations) appears novel within this limited search scope. The systematic framework bridging disparate findings (Contribution B: 10 candidates, one refutation) shows overlap with at least one prior effort to unify RoPE observations, suggesting incremental consolidation rather than a wholly new synthesis. The weighted RoPE intervention method (Contribution C: one candidate, zero refutations) was minimally tested but shows no immediate prior work in the examined set. These statistics reflect a top-K semantic search, not an exhaustive survey.

Overall, the paper occupies a moderately explored niche within RoPE frequency analysis. The FE metric and intervention method appear relatively fresh given the limited candidate pool, while the unifying framework builds on existing attempts to reconcile empirical discrepancies. The analysis covers approximately 21 semantically related papers, leaving open the possibility of additional relevant work outside this scope.

---

This paper presents **3 main contributions**, each analyzed against relevant prior work:

### Contribution 1: Frequency Entropy (FE) metric for RoPE analysis

**Description:** The authors propose Frequency Entropy as a quantitative framework comprising two complementary metrics: Spectrum Frequency Entropy and Sequence Frequency Entropy. These metrics measure the spectral behavior of RoPE on a per-dimension basis, providing a model-agnostic, scale-free diagnostic tool that quantifies how each rotary pair is utilized in transformer models.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. LoFormer: Local Frequency Transformer for Image Deblurring

**URL:** [View paper](#)

#### Brief Assessment

LoFormer[39] focuses on frequency-domain operations for image deblurring using local channel-wise self-attention, not on analyzing or quantifying rotary position embedding dimensions in transformers.

---

### 2. Mel-RoFormer for vocal separation and vocal melody transcription

**URL:** [View paper](#)

#### Brief Assessment

Mel-RoFormer[41] applies RoPE transformers to music audio separation and transcription tasks, but does not propose or analyze frequency entropy metrics for quantifying RoPE dimension utilization. The candidate focuses on architectural design for music processing, not on diagnostic tools for understanding positional embeddings.

---

### 3. Lightweight Spatio-Temporal Attention Network with Graph Embedding and Rotational Position Encoding for Traffic Forecasting

URL: [View paper](#)

#### Brief Assessment

Lightweight Spatio-Temporal[4] applies RoPE to traffic forecasting with grid search for optimal frequencies, but does not propose quantitative metrics to measure frequency dimension utilization or spectral behavior of RoPE itself.

---

### 4. Base of rope bounds context length

URL: [View paper](#)

#### Brief Assessment

Base of RoPE[43] focuses on deriving theoretical bounds for RoPE's base parameter to achieve specific context lengths, not on quantifying frequency dimension utilization through entropy metrics. The candidate addresses a different research question about the relationship between base values and context length capabilities.

---

### 5. Edge-Deployed Band-Split Rotary Position Encoding Transformer for Ultra-Low-Signal-to-Noise-Ratio Unmanned Aerial Vehicle Speech Enhancement

URL: [View paper](#)

#### Brief Assessment

Band-Split RoPE[27] applies RoPE to speech enhancement in UAV scenarios but does not propose metrics for quantifying RoPE frequency dimension utilization. The candidate focuses on engineering a lightweight transformer architecture rather than analyzing RoPE's internal frequency structure.

---

### 6. Optimizing the learnable rope theta parameter in transformers

URL: [View paper](#)

#### Brief Assessment

Learnable RoPE Theta[38] focuses on optimizing the theta parameter through learning strategies, not on quantifying frequency dimension utilization through entropy-based metrics. The candidate does not propose any diagnostic framework for measuring spectral behavior or dimension-wise usage patterns in RoPE.

---

### 7. KV-Latent: Dimensional-level KV Cache Reduction with Frequency-aware Rotary Positional Embedding

URL: [View paper](#)

#### Brief Assessment

KV-Latent[3] focuses on modifying RoPE's frequency sampling mechanism for dimensional reduction in KV cache compression, not on developing metrics to quantify RoPE frequency dimension utilization or measuring spectral behavior.

---

### 8. Breaking the stage barrier: A novel single-stage approach to long context extension for large language models

URL: [View paper](#)

#### Brief Assessment

Single-Stage Context Extension[40] focuses on head-adaptive RoPE base frequencies for long-context extension during continual pretraining, not on quantifying frequency dimension utilization through entropy metrics. The candidate does not propose methods to measure spectral behavior of RoPE dimensions.

---

### 9. Extending context window in large language models with segmented base adjustment for rotary position embeddings

URL: [View paper](#)

#### Brief Assessment

Segmented Base Adjustment[42] focuses on extending context windows through base adjustment of RoPE across dimensions, not on quantifying frequency dimension utilization through entropy metrics. The candidate does not propose any diagnostic framework for measuring spectral behavior or per-dimension usage of RoPE.

---

### 10. Round and round we go! what makes rotary positional encodings useful?

URL: [View paper](#)

#### Brief Assessment

Round and Round[7] analyzes RoPE frequency usage through 2-norm measurements and visual inspection of frequency bands, but does not propose a quantitative entropy-based framework like Spectrum FE and Sequence FE for measuring spectral behavior.

---

## Contribution 2: Systematic framework bridging disparate RoPE findings

**Description:** The authors develop a unified analytical framework that reconciles previously conflicting empirical observations about the roles of high- and low-frequency dimensions in RoPE. This framework moves beyond coarse frequency classifications to provide spectrum-aware analysis that explains mixed prior findings through per-dimension entropy measurements.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

### 1. On the token distance modeling ability of higher RoPE attention dimension

URL: [View paper](#)

#### Brief Assessment

Token Distance RoPE[14] focuses on identifying 'positional heads' that capture long-range dependencies through dimension-level correlation analysis, rather than developing a unified framework to reconcile conflicting findings about high/low-frequency roles in RoPE.

---

### 2. Hierarchical spatio-temporal state-space modeling for fmri analysis

URL: [View paper](#)

#### Brief Assessment

Hierarchical Spatio-Temporal[36] focuses on fMRI analysis using state-space models for brain network connectivity, not on analyzing rotary position embeddings in transformers. The technical domains are entirely different.

---

### 3. Edge-Deployed Band-Split Rotary Position Encoding Transformer for Ultra-Low-Signal-to-Noise-Ratio Unmanned Aerial Vehicle Speech Enhancement

URL: [View paper](#)

#### Brief Assessment

Band-Split RoPE[27] does not analyze high-frequency versus low-frequency roles in RoPE or attempt to reconcile conflicting empirical observations about RoPE dimensions. It uses RoPE as a component in a speech enhancement system without investigating its frequency characteristics.

---

### 4. VideoRoPE: What Makes for Good Video Rotary Position Embedding?

URL: [View paper](#)

#### Brief Assessment

VideoRoPE[1] focuses on extending RoPE to video data with spatio-temporal structure, not on reconciling conflicting findings about high/low-frequency roles in general RoPE analysis. The papers address different problems: the original develops a unified analytical framework using frequency entropy for general RoPE understanding, while VideoRoPE[1] proposes design principles for video-specific positional encoding.

---

### 5. KV-Latent: Dimensional-level KV Cache Reduction with Frequency-aware Rotary Positional Embedding

URL: [View paper](#)

#### Brief Assessment

KV-Latent[3] addresses RoPE stability in lower-dimensional spaces for efficiency purposes, rather than providing a unified analytical framework to reconcile conflicting empirical observations about high- versus low-frequency roles in RoPE.

---

### 6. Rotary outliers and rotary offset features in large language models

URL: [View paper](#)

#### Brief Assessment

Rotary Outliers[12] focuses on identifying large-magnitude features (rotary offset features) at specific frequency pairs and deriving bounds for their emergence, rather than reconciling conflicting empirical observations about high- vs. low-frequency roles through entropy-based spectrum-aware analysis.

---

### 7. PET-NeuS: Positional Encoding Tri-Planes for Neural Surfaces

URL: [View paper](#)

#### Brief Assessment

PET-NeuS[37] focuses on neural surface reconstruction using tri-plane representations and positional encoding for 3D geometry, not on analyzing rotary position embeddings (RoPE) in transformers or reconciling findings about frequency dimensions in language models.

---

### 8. Round and round we go! what makes rotary positional encodings useful?

URL: [View paper](#)

#### Prior Art Analysis

Round and Round[7] explicitly addresses the same problem of reconciling conflicting empirical observations about high- and low-frequency roles in RoPE. The paper states that previous analyses have reported conflicting findings and provides a unified framework to explain these mixed results through per-dimension analysis. Both papers identify the same gap in the literature regarding contradictory findings about frequency dimension roles and propose systematic frameworks to resolve these inconsistencies.

#### Evidence

Evidence 1 - **Rationale:** Both papers propose per-dimension analysis methods to quantify frequency utilization in RoPE. Round and Round[7] uses 2-norm measurements as a proxy for frequency usage, providing a quantitative per-dimension framework. - **Original:** we introduce frequency entropy (fe), a metric that quantifies the effective utilization of each rope frequency dimension, and we provide an analysis of how rope's sinusoidal components contribute to model representations on a per-dimension basis. - **Candidate:** to measure the usage of frequencies, we start by noting that by cauchy-schwarz, the effect of the k-th frequency component on the activation  $a_{i,j}$  is upper bounded by the 2-norm of the query and key components... it is therefore natural to look at the mean 2-norm for each  $k$  in  $\text{gemma } 7b$  over long seq...

Evidence 2 - **Rationale:** Both papers explicitly state their goal of providing a unified framework to reconcile conflicting prior observations about RoPE, with implications for future positional encoding design. - **Original:** fe provides a spectrum-aware, model-agnostic lens that reconciles prior mixed observations and informs pruning, reweighting, and the design of future positional schemes. - **Candidate:** we overall see our work as providing a more nuanced understanding of rope and hope that it can lead to an improved understanding of how llms use positional encodings, ultimately leading to performance gains - especially over long contexts.

---

### 9. HoPE: A Novel Positional Encoding Without Long-Term Decay for Enhanced Context Awareness and Extrapolation

URL: [View paper](#)

#### Brief Assessment

HoPE[33] focuses on removing long-term decay from RoPE and identifying 'activated' components that cause u-shape attention patterns, rather than providing a unified analytical framework that reconciles conflicting findings about high- and low-frequency roles through spectrum-aware entropy measurements.

---

### 10. RoPECraft: Training-Free Motion Transfer with Trajectory-Guided RoPE Optimization on Diffusion Transformers

URL: [View paper](#)

#### Brief Assessment

RoPECraft[9] applies RoPE to video motion transfer in diffusion models, not to analyzing frequency structure or reconciling conflicting findings about high/low-frequency roles in transformers.

---

## Contribution 3: Weighted RoPE intervention method

**Description:** The authors introduce Weighted RoPE, a targeted attenuation method that reduces the contribution of specific rotation pairs during inference based on their Frequency Entropy values. This intervention approach enables probing the functional relevance of different RoPE dimensions without fine-tuning, revealing which components are redundant versus essential for model performance.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

## 1. Selective Rotary Position Embedding

URL: [View paper](#)

### Brief Assessment

Selective Rotary[6] introduces input-dependent rotations with learnable angles for positional encoding, whereas the original paper's Weighted RoPE is an inference-time intervention that attenuates specific rotation pairs based on pre-computed Frequency Entropy values without modifying the underlying RoPE mechanism or requiring learning.

---

## Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

## References

---

- [0] Probing Rotary Position Embeddings through Frequency Entropy [View paper](#)
- [1] VideoRoPE: What Makes for Good Video Rotary Position Embedding? [View paper](#)
- [2] Context-aware Rotary Position Embedding [View paper](#)
- [3] KV-Latent: Dimensional-level KV Cache Reduction with Frequency-aware Rotary Positional Embedding [View paper](#)
- [4] Lightweight Spatio-Temporal Attention Network with Graph Embedding and Rotational Position Encoding for Traffic Forecasting [View paper](#)
- [5] Rotary position embedding for vision transformer [View paper](#)
- [6] Selective Rotary Position Embedding [View paper](#)
- [7] Round and round we go! what makes rotary positional encodings useful? [View paper](#)
- [8] DoPE: Denoising Rotary Position Embedding [View paper](#)
- [9] RoPECraft: Training-Free Motion Transfer with Trajectory-Guided RoPE Optimization on Diffusion Transformers [View paper](#)
- [10] ComRoPE: Scalable and Robust Rotary Position Embedding Parameterized by Trainable Commuting Angle Matrices [View paper](#)
- [11] Revisiting Multimodal Positional Encoding in Vision-Language Models [View paper](#)
- [12] Rotary outliers and rotary offset features in large language models [View paper](#)
- [13] How large language models encode theory-of-mind: a study on sparse parameter patterns [View paper](#)
- [14] On the token distance modeling ability of higher RoPE attention dimension [View paper](#)
- [15] Beyond Position: the emergence of wavelet-like properties in Transformers [View paper](#)
- [16] Fourier Position Embedding: Enhancing Attention's Periodic Extension for Length Generalization [View paper](#)
- [17] SpatialâSpectral Transformer With Patch-Local Mixed-Axis 2-D Rotary Position Embedding for Hyperspectral Image Classification [View paper](#)
- [18] SPECTRE: Spectral Pre-training Embeddings with Cylindrical Temporal Rotary Position Encoding for Fine-Grained sEMG-Based Movement Decoding [View paper](#)
- [19] Unpacking Positional Encoding in Transformers: A Spectral Analysis of Content-Position Coupling [View paper](#)
- [20] Gpt rotational position embedding for length extrapolation [View paper](#)
- [21] RoSA: Enhancing Parameter-Efficient Fine-Tuning via RoPE-aware Selective Adaptation in Large Language Models [View paper](#)
- [22] PriorSinger: Singing Voice Synthesis Model with Prior Condition Cross Attention [View paper](#)
- [23] EliteKV: Scalable KV Cache Compression via RoPE Frequency Selection and Joint Low-Rank Projection [View paper](#)
- [24] Head-wise Adaptive Rotary Positional Encoding for Fine-Grained Image Generation [View paper](#)
- [25] LaMPE: Length-aware Multi-grained Positional Encoding for Adaptive Long-context Scaling Without Training [View paper](#)
- [26] Rotary Offset Features in Large Language Models [View paper](#)
- [27] Edge-Deployed Band-Split Rotary Position Encoding Transformer for Ultra-Low-Signal-to-Noise-Ratio Unmanned Aerial Vehicle Speech Enhancement [View paper](#)
- [28] BiMa-Former: A Dual-Token Hybrid Model with Bidirectional Mamba and Transformer for Temporal-Multivariate Decoupled Forecasting [View paper](#)
- [29] Sensitivity Meets Sparsity: The Impact of Extremely Sparse Parameter Patterns on Theory-of-Mind of Large Language Models [View paper](#)
- [30] PAS: A Training-Free Stabilizer for Temporal Encoding in Video LLMs [View paper](#)
- [31] Efficient Long-Context Fine-Tuning of LLaMA 2 7B with QLoRA, Dynamic RoPE, and FlashAttention [View paper](#)
- [32] Additive Rotary Embedding - A Competitive Alternative to RoPE [View paper](#)
- [33] HoPE: A Novel Positional Encoding Without Long-Term Decay for Enhanced Context Awareness and Extrapolation [View paper](#)
- [34] MIRA: Medical Time Series Foundation Model for Real-World Health Data [View paper](#)
- [35] HoPE: Hybrid of Position Embedding for Long Context Vision-Language Models [View paper](#)
- [36] Hierarchical spatio-temporal state-space modeling for fmri analysis [View paper](#)
- [37] PET-NeuS: Positional Encoding Tri-Planes for Neural Surfaces [View paper](#)
- [38] Optimizing the learnable rope theta parameter in transformers [View paper](#)
- [39] LoFormer: Local Frequency Transformer for Image Deblurring [View paper](#)
- [40] Breaking the stage barrier: A novel single-stage approach to long context extension for large language models [View paper](#)
- [41] Mel-RoFormer for vocal separation and vocal melody transcription [View paper](#)
- [42] Extending context window in large language models with segmented base adjustment for rotary position embeddings [View paper](#)
- [43] Base of rope bounds context length [View paper](#)