PAS: A Training-Free Stabilizer for Temporal Encoding in Video LLMs

Supplementary Material

Conditions and Proof of Theorem 1 (the Phase Modulation Approximation)

Setup. Fix a paired dimension m and RoPE frequency set $\Omega = \{\omega_i\}_{i=0}^{m-1}$ with time-scale $\alpha > 0$. Let $C_i := z_i w_i^* \in \mathbb{C}$ be the complex coefficients for the i-th 2D pair. Define

$$S_\Delta := \sum_{i=0}^{m-1} C_i\, e^{\mathrm{j}\omega_i \alpha \Delta t}, \qquad S_0 := \sum_{i=0}^{m-1} C_i, \\ m_m(\Delta t) := \frac{1}{m} \sum_{i=0}^{m-1} e^{\mathrm{j}\omega_i \alpha \Delta t}. \tag{1}$$

The RoPE-rotated logit equals $\mathrm{Re}\, S_\Delta$ and the unrotated logit equals $\mathrm{Re}\, S_0$.

Assumptions. We state generic, distribution-free regularity conditions that can be verified in standard settings:

(A1) (*High dimension*) m is large and can grow; paired features are balanced so that $|C_i|$ have bounded second moments.

(A2) (*Quasi-uniform spectral energy*) The set Ω is sufficiently dense over its working band and does not exhibit large gaps; no narrow frequency dominates the sum in expectation.

(A3) (*Weak content–frequency coupling*) $\{C_i\}$ are weakly dependent on $\{\omega_i\}$ so that $\bar{C} := \frac{1}{m}\sum_i C_i$ satisfies $|\bar{C}| \geq c > 0$ and $\frac{1}{m}\sum_i |C_i - \bar{C}|^2 \leq \sigma^2$ with $\sigma < \infty$.

(A4) (*Well-scaled time*) The scale α matches the time unit so that aliasing of out-of-band frequencies into the working band is negligible for the range of Δt under study.

Key decomposition. Write

$$S_\Delta - S_0\, m_m(\Delta t) = \sum_{i=0}^{m-1} (C_i - \bar{C})\, e^{\mathrm{j}\omega_i \alpha \Delta t}, \\ \bar{C} := \frac{1}{m} \sum_{i=0}^{m-1} C_i. \tag{2}$$

Thus the approximation error is governed by how much $\{C_i\}$ deviate from their mean and how these deviations correlate with the phases $e^{\mathrm{j}\omega_i \alpha \Delta t}$.

Concentration lemma. Under (A1)–(A3), there exists a constant K (depending only on σ) such that for any $\delta \in (0, 1)$,

$$\Pr\left(\sup_{\Delta t \in \mathcal{T}} \left| S_\Delta - S_0\, m_m(\Delta t) \right| \;\leq\; K\, \sigma \sqrt{m \log \frac{|\mathcal{T}|}{\delta}} \right) \geq\; 1 - \delta, \tag{3}$$

for any finite grid $\mathcal{T}$ of lags. Using Equation (2), apply Bernstein/Hoeffding concentration to the real and imaginary parts over i with bounded second moments and union bound over $\mathcal{T}$.

Finite-m approximation theorem. Let $\mu := \mathbb{E}[\bar{C}]$ with $|\mu| \geq c > 0$, and assume (A1)–(A4). Then for any finite grid $\mathcal{T}$ and $\delta \in (0, 1)$, with probability at least $1 - \delta$,

$$\sup_{\Delta t \in \mathcal{T}} \left| \frac{S_\Delta}{S_0} - m_m(\Delta t) \right| \;\leq\; \underbrace{\frac{K\, \sigma}{|\mu|} \sqrt{\frac{\log(|\mathcal{T}|/\delta)}{m}}}_{\text{concentration}} + \underbrace{\varepsilon_{\text{spec}}}_{\text{spectral nonuniformity}}, \tag{4}$$

where $\varepsilon_{\text{spec}}$ absorbs deviations from (A2) (e.g., non-dense or heavily skewed Ω) and vanishes as the band coverage improves. Combine Equation (2) with Equation (3), then divide by $S_0 = \bar{C}\, m$; control $|S_0|$ away from zero via $|\mu| \geq c$ and standard concentration.

Consequences for logits. Taking real parts and recalling that $\langle \tilde{\mathbf{q}}(t), \tilde{\mathbf{k}}(t') \rangle = \mathrm{Re}\, S_\Delta$ and $\langle \mathbf{q}, \mathbf{k} \rangle = \mathrm{Re}\, S_0$, Equation (4) yields

$$\left\langle \tilde{\mathbf{q}}(t), \tilde{\mathbf{k}}(t') \right\rangle = \langle \mathbf{q}, \mathbf{k} \rangle \cdot \mathrm{Re}\{ m_m(\Delta t) \} \;+\; \mathcal{E}_m(\Delta t), \\ \sup_{\Delta t \in \mathcal{T}} |\mathcal{E}_m(\Delta t)| \;\lesssim\; \|\mathbf{q}\|\, \|\mathbf{k}\| \left(\frac{K'}{\sqrt{m}} + \varepsilon_{\text{spec}} \right), \tag{5}$$

for a constant K' depending on σ and $|\mu|$. Hence, as $m \to \infty$ and spectral coverage improve, the error vanishes uniformly over $\mathcal{T}$, justifying the phase modulation approximation with the IFT kernel $m_m(\Delta t)$.

Proof of Theorem 2 (Smooth IFT ⇒ Phase-Stable Attention) and an Exact Spectral Bound

Set-up. Let the RoPE frequency set be $\{\omega_i\}_{i=0}^{m-1}$ and the time-scale be $\alpha > 0$. Define complex coefficients $C_i := z_i w_i^*$, the IFT kernel

$$m(\Delta t) := \frac{1}{m} \sum_{i=0}^{m-1} e^{\mathrm{j}\omega_i \alpha \Delta t}, \tag{6}$$

and the exact RoPE-rotated logit

$$A_{\mathrm{ex}}(\Delta t) := \mathrm{Re}\left[\sum_{i=0}^{m-1} C_i\, e^{\mathrm{j}\omega_i \alpha \Delta t}\right]. \tag{7}$$

Proof under the phase modulation approximation. Assume $A(\Delta t) = \langle \mathbf{q}, \mathbf{k}\rangle\, \mathrm{Re}\{m(\Delta t)\}$. Then $A'(\tau) = \langle \mathbf{q}, \mathbf{k}\rangle\, \partial_\tau \mathrm{Re}\{m(\tau)\}$ and hence $\sup_\tau |A'(\tau)| \le |\langle \mathbf{q}, \mathbf{k}\rangle|\, L_m$ with $L_m := \sup_\tau |\partial_\tau \mathrm{Re}\{m(\tau)\}|$. By the mean value theorem, for any Δt and δt there exists ξ between them such that

$$|A(\Delta t + \delta t) - A(\Delta t)| = |A'(\xi)|\, |\delta t| \le |\langle \mathbf{q}, \mathbf{k}\rangle|\, L_m\, |\delta t|, \tag{8}$$

which is theorem 2.

Exact spectral Lipschitz bound (no approximation). Differentiating Equation (7) gives

$$A'_{\mathrm{ex}}(\tau) = \mathrm{Re}\left[\sum_{i=0}^{m-1} \mathrm{j}\,\omega_i \alpha\, C_i\, e^{\mathrm{j}\omega_i \alpha \tau}\right], \tag{9}$$

hence

$$\sup_{\tau \in \mathbb{R}} |A'_{\mathrm{ex}}(\tau)| \le \alpha \sum_{i=0}^{m-1} \omega_i\, |C_i|. \tag{10}$$

Applying the mean value theorem yields, for any $\Delta t, \delta t$,

$$\left|A_{\mathrm{ex}}(\Delta t + \delta t) - A_{\mathrm{ex}}(\Delta t)\right| \le \alpha \left(\sum_{i=0}^{m-1} \omega_i\, |C_i|\right) |\delta t|. \tag{11}$$

If one prefers a normalized sensitivity, assume $\mathrm{Re}\sum_i C_i > 0$ and divide both sides of Equation (11) by $A_{\mathrm{ex}}(0) = \mathrm{Re}\sum_i C_i$ to obtain

$$\left|\frac{A_{\mathrm{ex}}(\Delta t + \delta t) - A_{\mathrm{ex}}(\Delta t)}{A_{\mathrm{ex}}(0)}\right| \le \alpha\, \frac{\sum_{i=0}^{m-1} \omega_i\, |C_i|}{\mathrm{Re}\sum_{i=0}^{m-1} C_i}\, |\delta t|. \tag{12}$$

Thus the exact Lipschitz constant is controlled by a first spectral moment of the coefficients $\{C_i\}$; reducing high-frequency mass (or flattening its influence) tightens stability, aligning with the approximation-based bound in the main text.

Proof of Theorem 3 (Smoothing by Multi-Phase Averaging)

Set-up. Let $m : \mathbb{R} \to \mathbb{R}$ be the real IFT kernel of the RoPE spectrum and define m_{eff} by

$$m_{\mathrm{eff}}(\Delta t) = \sum_{h=1}^{H} a_h\, m(\Delta t + \delta_h), \qquad a_h \ge 0,\ \sum_h a_h = 1. \tag{13}$$

Fix $\varepsilon \in \mathbb{R}$ and consider the mean-square local variation functional.

$$\mathcal{V}_\varepsilon(f) = \limsup_{T \to \infty} \frac{1}{T} \int_0^T \big(f(\tau + \varepsilon) - f(\tau)\big)^2\, d\tau. \tag{14}$$

Part I: Time-domain inequality. For each τ, set $\Delta_h(\tau) := m(\tau + \varepsilon + \delta_h) - m(\tau + \delta_h)$. Then

$$m_{\mathrm{eff}}(\tau + \varepsilon) - m_{\mathrm{eff}}(\tau) = \sum_{h=1}^{H} a_h\, \Delta_h(\tau). \tag{15}$$

By Jensen's inequality for the convex function $x \mapsto x^2$ and $\sum_h a_h = 1$,

$$\Big(\sum_h a_h\, \Delta_h(\tau)\Big)^2 \le \sum_h a_h\, \Delta_h(\tau)^2. \tag{16}$$

Integrating over $\tau \in [0, T]$, dividing by T, and taking $\limsup_{T\to\infty}$ yields

$$\mathcal{V}_\varepsilon(m_{\mathrm{eff}}) \le \sum_h a_h\, \limsup_{T \to \infty} \frac{1}{T} \int_0^T \big(m(\tau + \varepsilon + \delta_h) - m(\tau + \delta_h)\big)^2 d\tau. \tag{17}$$

By translation invariance of the Lebesgue integral (or ergodicity for periodic/quasi-periodic m),

$$\begin{aligned} &\limsup_{T \to \infty} \frac{1}{T} \int_0^T \big(m(\tau + \varepsilon + \delta_h) - m(\tau + \delta_h)\big)^2 d\tau \\ &= \limsup_{T \to \infty} \frac{1}{T} \int_0^T \big(m(\tau + \varepsilon) - m(\tau)\big)^2 d\tau = \mathcal{V}_\varepsilon(m). \end{aligned} \tag{18}$$

Combining Equation (17) and Equation (18) gives

$$\mathcal{V}_\varepsilon(m_{\mathrm{eff}}) \le \sum_h a_h\, \mathcal{V}_\varepsilon(m) = \mathcal{V}_\varepsilon(m), \tag{19}$$

with strict inequality whenever $\{\delta_h\}$ are not all equal (modulo the period induced by ε) and at least two $a_h > 0$, unless m is affine on all segments traversed by the shifts.

Part II: Frequency-domain characterization. Let $\widehat{m}(\omega)$ be the (discrete-line) spectrum of m at angular frequencies $\{\omega_i\}$. Convolution by the discrete shift kernel $\kappa(\tau) = \sum_h a_h \delta(\tau - \delta_h)$ yields

$$\widehat{m_{\mathrm{eff}}}(\omega) = \widehat{\kappa}(\omega)\, \widehat{m}(\omega) \quad \text{with} \quad \widehat{\kappa}(\omega) = \sum_{h=1}^{H} a_h e^{\mathrm{j}\omega\alpha\delta_h}. \tag{20}$$

By the triangle inequality and $\sum_h a_h = 1$,

$$\big|\widehat{\kappa}(\omega)\big| \le \sum_h a_h\, \big|e^{\mathrm{j}\omega\alpha\delta_h}\big| = 1, \tag{21}$$

with strict inequality $|\widehat{\kappa}(\omega)| < 1$ for any $\omega \neq 0$ whenever the phases $\{\omega\alpha\delta_h\}$ are not all equal (mod 2π). Thus nontrivial multi-phase dispersion attenuates nonzero-frequency lines while preserving DC, confirming that Equation (19) reflects genuine smoothing of ripple components. This frequency view also explains why each *individual head* keeps its spectrum (no change within a head); smoothing arises only after aggregation across phase-shifted heads via $\widehat{\kappa}(\omega)$.

Proof of Theorem 4 (Spectrum Preservation under Time Shifts)

Set-up. Let $\Delta > 0$ be the sampling period and assume m is band-limited to $|\Omega| < \pi/\Delta$ in continuous time. Define $m[n] = m(n\Delta)$ and the shifted sequence $m_\delta[n] = m(n\Delta + \delta)$ for any $\delta \in \mathbb{R}$ (fractional delay). The discrete-time Fourier transform (DTFT) of m is

$$M(e^{\mathrm{j}\omega}) = \sum_{n\in\mathbb{Z}} m[n]\, e^{-\mathrm{j}\omega n}, \qquad \omega \in (-\pi, \pi]. \tag{22}$$

Fractional-delay frequency response. By the standard shift/modulation property of the DTFT (fractional delay implemented as an all-pass phase ramp),

$$\mathcal{F}\{m_\delta[n]\}(e^{\mathrm{j}\omega}) = \sum_n m(n\Delta+\delta)\, e^{-\mathrm{j}\omega n} = e^{\mathrm{j}\omega\,\delta/\Delta}\, M(e^{\mathrm{j}\omega}), \tag{23}$$

provided m is the sampled version of a band-limited continuous-time function (Nyquist) so that fractional delays are representable without changing amplitude. Hence

$$\big|M_\delta(e^{\mathrm{j}\omega})\big| = \big|M(e^{\mathrm{j}\omega})\big|, \qquad \forall\, \omega \in (-\pi, \pi]. \tag{24}$$

Finite-length DFT and circular shifts. For an N-point windowed sequence $m_W[n] = w[n]\, m[n]$ (with a fixed window w) and an integer shift $s \in \mathbb{Z}$, the k-th DFT coefficient obeys

$$\mathrm{DFT}\{m_W[n-s]\}[k] = e^{-\mathrm{j}2\pi ks/N}\, \mathrm{DFT}\{m_W[n]\}[k], \tag{25}$$

so magnitudes are preserved across bins. For fractional shifts, one may implement the all-pass phase ramp $e^{\mathrm{j}2\pi k\delta/N}$ across DFT bins, which again preserves magnitudes. Therefore, per-head temporal shifts leave the spectral support and magnitudes at RoPE's discrete lines invariant; only phases change.

Remark on aliasing. If the sampling period Δ violates Nyquist, the sampled spectrum contains aliases; then time shifts can redistribute energy among overlapped images, and magnitude invariance no longer holds in general. This regime is excluded by the theorem's bandlimit assumption and is studied empirically in the main experiments.