THE PRICE OF RELIABLE ROTARY RETRIEVAL: TASK GEOMETRY, SHARED SPECTRA, AND MINIMAL ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Distinct rotary encodings need not permit reliable retrieval at a reasonable query–key norm. We study this gap in a supplied-address, position-only soft-attention channel, measuring reliability by total off-target mass. Our central result determines the exact price of reusing a representation: relative to an m-block spectrum redesigned for each task, a spectrum frozen across all lengths and addresses needs exactly $m + 1$ two-dimensional rotation blocks for constant-factor amplitude competitiveness. The obstruction is geometric. Task-specific boundary retrieval can exploit a linear slope, but a frozen spectrum efficient on interior tasks eventually surrounds the target in phase space and cancels that advantage. Stability imposes a separate constraint. For maximum displacement L, the maximal critical frequency-error radius is $1/(2L)$, with feasibility strictly below the threshold. No finite frozen spectrum retains this tolerance order at every length; one length-adaptive frequency restores it while preserving the optimal amplitude order. A finite convex calibration with a dimension-independent approximation factor turns these guarantees into checkable budget certificates. Standard-frequency comparisons and reported checkpoint interventions distinguish representation cost from phase and allocation failures. Together, the results characterize reliable rotary addressing under an explicit resource contract.

1 INTRODUCTION

RoPE encodes relative position through rotations in query–key dot products (Su et al., 2021). Position-selective constructions, nonuniform use of frequency bands, and learned or rescaled frequencies are well established (Barbero et al., 2024; Oka et al., 2026; Karypis et al., 2026). These findings motivate a complementary resource question: *how much score amplitude does a task require, and can one reusable representation retain the efficiency of a task-specific design?*

A target can uniquely maximize its logit and still receive little attention when many competitors retain mass. Reliable retrieval therefore requires more than the right score ordering: it requires enough amplitude to suppress the *sum* of off-target contributions. We isolate this cost in a supplied-address, position-only channel. Values are arbitrary, content cannot remove competitors, and every temperature or query–key gain counts toward the budget.

The price of reuse. Allowing an m-block frequency table to be redesigned for each retrieval task gives a strong benchmark. We prove that one spectrum, frozen before the length or address is known, competes with every such task using exactly $m + 1$ rotation blocks. One common key suffices; the query changes with the task. With only m blocks, even arbitrary signed sine and cosine coefficients cannot retain constant-factor competitiveness. The constants are independent of length, address, and reliability, while amplitude and evaluation precision may grow.

The reason is more subtle than eventual collision. A boundary task can exploit a one-sided slope, giving a better length exponent than an interior task. A frozen spectrum efficient on every interior task, however, has forward orbit points that surround the target *in phase space*. Their convex combination cancels the slope even though they all lie on one temporal side. One additional plane compensates for the resulting exponent loss (Figure 1).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

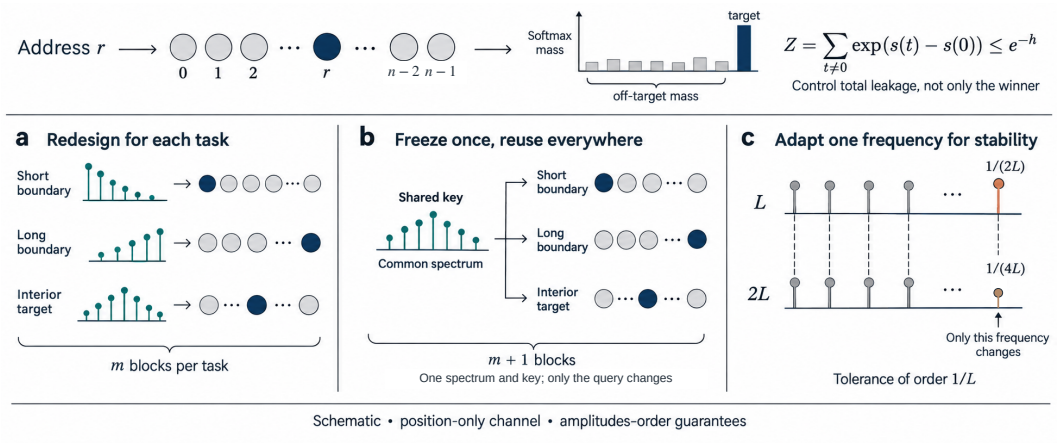


Figure 1: One retrieval contract, different design permissions (schematic; stems do not count blocks). Reliable readout controls total off-target mass. (a) Each task gets a redesigned m-block spectrum. (b) One jointly designed spectrum and common key serve all lengths and addresses; exactly $m + 1$ blocks attain constant-factor competitiveness. (c) One frequency changes from $1/(2L)$ to $1/(4L)$ as length doubles, while fast frequencies stay fixed, restoring order-$1/L$ coherent frequency-error tolerance at optimal amplitude order. This is not a minimum total dimension under adaptation.

Stable reuse with one adaptive frequency. Frequency uncertainty introduces a second obstruction: a common rational return can make retrieval impossible regardless of amplitude. Every finite frozen spectrum loses order-$1/L$ error tolerance along some lengths. We show that a fixed fast spectrum with one length-adaptive clock restores this tolerance *at the optimal amplitude order*. Single-band interpolation has already been evaluated in LeRoPE (Karypis et al., 2026, §7); here it enters a necessity-and-sufficiency resource guarantee.

From asymptotic laws to finite budgets. Convex calibration measures the required query–key norm for a specified spectrum, key, and task. At length 4096, allocation optimization reduces the standard uniform-key budget from 82.331 to 44.117. Reported checkpoint interventions address the complementary question: can failed learned scores be repaired without changing their spectrum, key, or budget? These comparisons separate *representation cost* from *the parameters training selected*. Sections 3–4 establish the resource laws; Section 5 tests their finite implications within the channel defined next.

2 The Supplied-Address Task and Its Resource Contract

The retrieval interface. There are n keys at positions $0, \dots, n-1$, with values $v_0, \dots, v_{n-1}$ and a requested address r. The query is outside this candidate set. A shared phase rule receives r and the query position, but not v_r, a target-specific key marker, or a mask deleting other keys. Retrieval uses the rotary score and softmax over *all* candidates. This models the positional routing step after an upstream component has selected a location, such as a prescribed look-back or a controller-selected index. The address is supplied; learning the controller is outside the task. Our question is the norm cost of carrying out that read through this constrained channel.

Notation and cost. Throughout the main text, k is the number of two-dimensional rotation blocks, $d_{\text{head}} \geq 2k$ is the ambient head width, and $\mathcal{C}$ denotes a competitor set. The separate symbol m is used only for the task-specific benchmark when comparing dimensions. Frequencies θ_j are measured in cycles per position. Translating the target to zero gives

$$\mathcal{C}_{n,r} = \{-r, \dots, -1, 1, \dots, n-1-r\}. \tag{1}$$

A position-only score has the form

$$s(t) = \sum_{j=1}^{k} \left[a_j \cos(2\pi\theta_j t) + b_j \sin(2\pi\theta_j t)\right], \qquad A_{\text{coeff}}(s) = \sum_j \sqrt{a_j^2 + b_j^2}. \tag{2}$$

After attention scaling, $A_{\mathrm{coeff}} \le A_{\mathrm{QK}} := \|Q\|\,\|K\|/\sqrt{d_{\mathrm{head}}}$. A task-specific query and key can attain equality by balancing their block norms. If a *common* key has positive block-energy fractions w_j, however, the exact realization cost is

$$N_w(a,b) = \left(\sum_j \frac{a_j^2 + b_j^2}{w_j}\right)^{1/2} = A_{\mathrm{QK}}, \qquad \sum_j w_j = 1. \tag{3}$$

Zero-energy blocks must have zero coefficients. Our lower bounds allow the relaxed coefficient cost; the principal constructions use equal amplitudes and a common key uniform on the active blocks, attaining both costs simultaneously. Fixed-key experiments optimize (3), not the coefficient norm evaluated after optimizing a different objective.

Reliable readout, not just the largest score. Define total leakage and target mass by

$$Z_{\mathcal{C}}(s) = \sum_{t\in\mathcal{C}} e^{s(t)-s(0)}, \qquad \alpha_0 = (1+Z_{\mathcal{C}})^{-1}. \tag{4}$$

For arbitrary values of norm at most one, the worst readout error is exactly $2(1-\alpha_0)$: choose the target value as a unit vector and all others as its negative. We require $Z_{\mathcal{C}} \le e^{-h}$, with log-reliability $h \ge \log 3$. Logit errors bounded in magnitude by ν multiply worst-case leakage by $e^{2\nu}$, so target mass $1-\varepsilon$ is guaranteed with $h = \log((1-\varepsilon)/\varepsilon) + 2\nu$. We use exact softmax or separately controlled arithmetic, not underflow-induced hard selection.

Write $A_\theta(\mathcal{C}, h)$ for the optimal coefficient amplitude at a fixed spectrum, and

$$A_{\mathrm{or},m}(n,r,h) = \inf_{\theta\in\mathbb{R}^m} A_\theta(\mathcal{C}_{n,r}, h)$$

for a benchmark that may redesign its entire spectrum and coefficient allocation for each task. Neither a common key nor a frozen spectrum has those permissions. Complete notation and realization details are in Appendix A.

3 The exact cost of a spectrum frozen across tasks

The central comparison fixes the representation *before* the length, address, and reliability are specified. The query can change; its resource cost is measured for each task.

Theorem 1 (Exactly one extra plane for frozen reuse)**.** *For each fixed oracle budget $m \ge 1$, the minimum number k of rotation blocks for which there exist a fixed spectrum, a common unrotated key, and a finite C_m satisfying*

$$\forall n \ge 2,\ \forall r,\ \forall h \ge \log 3, \quad \exists Q_{n,r,h}: \quad Z_{\mathcal{C}_{n,r}} \le e^{-h}, \qquad A_{\mathrm{QK}} \le C_m A_{\mathrm{or},m}(n,r,h) \tag{5}$$

is exactly $k = m+1$.

The frequency table and key are shared across *all* lengths. The query's amplitude and address phase, and the precision used to evaluate the same symbolic frequencies, may grow. The constant depends on m, not on n, r, h. All frequencies may be redesigned once; appending a block to an arbitrary existing table is not guaranteed to work.

3.1 The task-specific advantage that reuse must preserve

Let $E_L = \{-L, \dots, -1\}$ and $S_L = \{-L, \dots, -1, 1, \dots, L\}$. The following benchmark supplies the two different costs used in the central proof.

Proposition 2 (Boundary and interior resource orders)**.** *For each fixed $k \ge 1$, optimizing over all k-block spectra and signed scores gives*

$$A_E^*(L,k,h) = \Theta_k(L^{2/(k+1)}h), \qquad A_S^*(L,k,h) = \Theta_k(L^{2/k}h), \tag{6}$$

for $L \ge 2$ and $h \ge \log 3$. At a fixed design length the two orders can share one spectrum and one key, independent of h.

Appendix B proves the stronger expressions uniform in k, L, h. Matching candidate counts changes only constants. At logarithmic block count both costs become $\Theta(\log L + h)$; adding dimension does not remove softmax's total-mass requirement (Veličković et al., 2024).

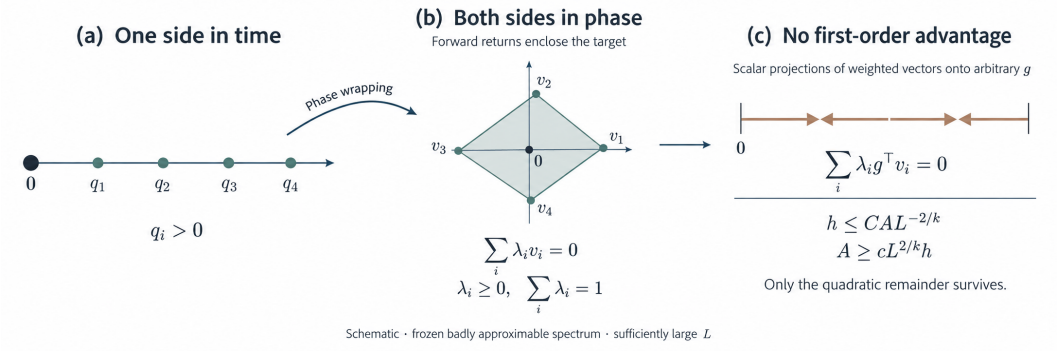


Figure 2: Why frozen reuse loses the boundary advantage (schematic). For a badly approximable spectrum and sufficiently large L, (a) positive-time competitors provide (b) local phase vectors enclosing zero. (c) Their convex weights cancel the projection onto every score gradient g, including arbitrary signed scores. Only a quadratically bounded remainder remains, forcing $A \geq cL^{2/k}h$. The polygon and opposing arrows illustrate the argument, not measured returns or individual score values.

Why the exponents differ. Place the $L+1$ orbit points $i\theta \bmod 1$ into phase cubes of side $1/Q$. For a one-sided task, nearby phases can still differ by a *linear* sine contribution. Subdivide each cube along the score's gradient direction at width $1/Q^2$. There are $O(Q^{k+1})$ cells, and a same-cell difference has score gap $O(A_{\text{coeff}}/Q^2)$. Counting all such pairs, while dividing by the number that can produce the same integer displacement, gives the one-sided lower bound. For a two-sided task, pairing t and $-t$ removes the need for gradient subdivision; $O(Q^k)$ cells suffice. The different cell counts produce the different exponents; Appendix B gives the full counting argument.

The matching construction uses one slow sine for E_L or slow cosine for S_L, together with the same fast blocks. A sine provides a linear local gap, whereas a cosine peak provides a quadratic one. Finite-grid conditional expectation selects the fast frequencies. The shared construction, even-score penalty, and its deterministic implementation follow from this benchmark.

3.2 Why a frozen spectrum loses the boundary advantage

For a fixed spectrum, define its closest joint return through length L as

$$\Delta_\theta(L) = \min_{1 \leq q \leq L} \max_j \|q\theta_j\|_{\mathbb{R}/\mathbb{Z}}.$$

A return with small Δ yields a hard pair of competitors. Conversely, pairwise separation of the orbit limits how many competitors lie in each phase-distance shell. Together these facts give

$$A_\theta(S_L, h) = \Theta_k\big(h/\Delta_\theta(L)^2\big)\,. \tag{7}$$

The full finite-constant inequalities are in Lemma 5. It follows that a fixed spectrum stays within order $L^{2/k}h$ on every interior task if and only if it is *simultaneously badly approximable*: there is a $\kappa > 0$ such that

$$\max_j \|q\theta_j\|_{\mathbb{R}/\mathbb{Z}} \geq \kappa q^{-1/k} \quad \text{for every integer } q \geq 1. \tag{8}$$

This is a specific Diophantine condition, not a property of every irrational frequency table.

Such an orbit also covers the torus at radius $O_{\theta,k}(L^{-1/k})$ (Dick et al., 2025). Thus, at sufficiently large L, positive-time competitors provide small phase vectors v_i whose convex hull contains zero. There are $\lambda_i \geq 0$ with $\sum_i \lambda_i = 1$ and $\sum_i \lambda_i v_i = 0$. Taylor expansion of the phase score F then yields

$$h \leq \sum_i \lambda_i [F(0) - F(v_i)] \leq O_{\theta,k}(A_{\text{coeff}} L^{-2/k}). \tag{9}$$

The first-order terms cancel *exactly*, leaving a curvature cost even though all selected competitors occur on one temporal side (Figure 2). Appendix C supplies the forward-covering and convex-enclosure lemmas.

For necessity in Theorem 1, suppose a frozen m-block spectrum were competitive on all tasks. Its interior costs force (8). Equation (9) then forces boundary cost $\Omega_{\theta,m}(n^{2/m}h)$, whereas Proposition 2 gives the boundary oracle $O_m(n^{2/(m+1)}h)$. The ratio diverges as $n^{2/[m(m+1)]}$. The failure length may depend on the spectrum: this is not a uniform finite-length impossibility result.

3.3 An explicit reusable representation

For sufficiency, take $k = m + 1$, $\alpha = 2^{1/(k+1)}$, and the classical algebraic vector $\theta_j = \alpha^j$, $j = 1, \ldots, k$ (Dick et al., 2025). A self-contained norm argument gives Euclidean phase separation at least $(10\sqrt{2})^{-1}L^{-1/k}$. Equal-amplitude, target-aligned scores suffice:

$$s_r(j) = \frac{A}{k}\sum_{b=1}^{k}\cos(2\pi\theta_b(j-r)), \qquad A = 25k(n-1)^{2/k}(h + k\log 5 + \log 3). \tag{10}$$

Every task-specific m-block oracle costs at least $n^{2/(m+1)}h/4096$, because every address has a long side of competitors. Comparing this lower bound with (10) proves the result; a conservative C_m is $2^{20}(m+1)^2$. One uniform key realizes $A_{\mathrm{QK}} = A_{\mathrm{coeff}} = A$.

The lower bound is pointwise in the frozen spectrum, including a spectrum selected by training (Appendix C.6). Its all-length quantifier is essential: selecting a different spectrum for each finite design horizon is a distinct comparison.

4 Stable reuse and finite-cost calibration

We now change one permission: installed frequencies may differ from their nominal values by $\widehat{\theta} = \theta + e$, $\|e\|_\infty \le \sigma$. The same vector e is used at every position, and the query's address phases use that installed spectrum. This is a coherent parameter box, not arbitrary tokenwise errors or an unknown key spectrum paired with a blindly frozen physical query.

Let $u_q(\theta) = \max_j \|q\theta_j\|_{\mathbb{R}/\mathbb{Z}}$ and define

$$R_\theta(L,\sigma) = \min_{1\le q\le L}(u_q - q\sigma)_+, \qquad \rho_\theta(L) = \min_{1\le q\le L}\frac{u_q}{q}. \tag{11}$$

The first quantity is the smallest joint separation left in the error box; the second is its critical radius. Let $\mathcal{A}_\theta$ be the worst, over installed spectra in the box, of the optimal coefficient amplitude on S_L.

Theorem 3 (Exact threshold and minimal length adaptation)**.** *For every finite k-block spectrum, $\mathcal{A}_\theta < \infty$ if and only if $\sigma < \rho_\theta(L)$. For $R = R_\theta(L,\sigma) > 0$,*

$$\frac{h + \log 2}{2\pi^2 R^2} \le \mathcal{A}_\theta \le \frac{k}{8R^2}(h + k\log 5 + \log 3). \tag{12}$$

Furthermore,

$$\sup_\theta \rho_\theta(L) = \frac{1}{2L}, \qquad \textit{every finite frozen } \theta \textit{ satisfies } \liminf_{L\to\infty} L\rho_\theta(L) = 0. \tag{13}$$

For $k \ge 2$, $k-1$ fixed badly approximable fast frequencies and one clock $1/(2L)$ attain the maximal radius. At each fixed $0 \le \lambda < 1$, an equal-amplitude score with a common key succeeds throughout $\|e\|_\infty \le \lambda/(2L)$ with $O_{k,\lambda}(L^{2/k}h)$ amplitude.

At the threshold, the box contains a common rational spectrum p/q, $q \le L$. The target and $\pm q$ have identical positional scores for *every* coefficient choice. No temperature can repair that collision. Below threshold, the shell bound applies uniformly to every installed spectrum. The construction yields one relative score that is feasible for every error in the box.

How the clock closes the resource gap. Fast badly approximable blocks protect short displacements by a $q^{-1/(k-1)}$ separation. The clock protects longer displacements by a term proportional to q/L, even after perturbation. Balancing the two gives $R = \Omega_{k,\lambda}(L^{-1/k})$, and hence the claimed amplitude. With $k = m + 1$ and $L = n - 1$, the construction competes with every ideal m-block task-specific

oracle, even though the implemented spectrum is uncertain. Zero adjustable coordinates cannot give the required $1/L$ tolerance at every length, whereas one suffices. This is a minimality result for the *number of adjustable frequencies*; the minimum total dimension in the adaptive model is a different, unresolved question.

The exact clock tolerates a range of coefficients. Appendix D.5 proves that replacing $1/(2L)$ by c_L/L preserves both orders whenever c_L stays in a fixed compact subinterval of $(0, 1)$. The guaranteed radius becomes at least $\min\{c_L, 1 - c_L\}/L$; $c_L = 1/2$ attains the extremal value.

Finite guarantees without the worst-case constants. Maximal error radius and minimal operating cost are different objectives: at a fixed subcritical physical error, the cheapest clock coefficient need not be $c = 1/2$. Fix S_L, a common key w, and $\sigma < \rho_\theta(L)$. Let B_w be the minimum N_w of *one relative coefficient vector* that succeeds for every error in the box. Define the coordinatewise worst gaps

$$G_{qj} = 1 - \cos\big(2\pi(\|q\theta_j\|_{\mathbb{R}/\mathbb{Z}} - q\sigma)_+\big), \qquad 1 \le q \le L.$$

The finite convex program

$$\widehat{\mathsf{B}}_w = \min_{a \ge 0} \left\{ N_w(a, 0) : 2\sum_{q=1}^{L} e^{-(Ga)_q} \le e^{-h} \right\} \tag{14}$$

produces a feasible query. Proposition 11 proves

$$\mathsf{B}_w \le \widehat{\mathsf{B}}_w \le \left(1 + \frac{\log L}{h + \log 2}\right) \mathsf{B}_w. \tag{15}$$

The factor is independent of dimension, Diophantine constants, and distance to the threshold. Every successful common query must suppress each position pair under that pair's worst coherent error. Scaling these necessary gap constraints adds only $\log L$ to log-reliability, producing a sufficient envelope without equating it to the exact coherent maximum. A rational allocation certifies an upper bound; a probability-weighted gap certifies a lower bound, often much tighter than (15). Computing the $L \times k$ gap matrix, solving the convex program, and verifying these witnesses replaces conservative global constants with finite budgets. The comparison fixes the spectrum and requires one relative query for the entire box.

5 Finite tests: spectral cost, learned scores, and stability

Three finite questions connect the theory to implementation: a fixed spectrum's cost, the causes of learned-score failure at a given budget, and a clock's robust operating cost. Budget computations and certificates are reproducible from the included code and witnesses; checkpoint results are reported aggregates without raw training artifacts. Appendix F.5 maps their evidence.

5.1 Measuring spectral cost under a common key

For a symmetric competitor set, evenization and removal of negative cosine coefficients reduce the exact signed-score problem to a convex nonnegative allocation problem. The fixed-key objective is N_w, rather than $\sum_j a_j$. A feasible allocation supplies an upper bound, and an entropy-based dual supplies a lower bound. Appendix E derives both the symmetric and all-address programs.

Table 1 compares all-address retrieval with identical candidate sets, one uniform key, and equal allocation permissions. Standard RoPE uses angular frequencies $\omega_j = 10000^{-j/k}$, $j = 0, \dots, k-1$. The algebraic schedules use the constructions above; each clock variant replaces one of the eight blocks by $1/[2(n-1)]$ in cycle units. Full schedules are in Appendix F.

Two effects are visible. First, allocation matters within a spectrum: optimizing standard RoPE lowers its budget from 82.33138 to 44.11702. Second, a finite spectral difference remains after this optimization: the algebraic and algebraic-plus-clock budgets are lower by 39.14% and 33.17%. Their equal-allocation costs are also lower. The hardest equal-allocation displacement is $q = 1$ for standard RoPE but $q = 2974$ for the algebraic schedule. Indeed, adjacent positions account for 99.994568% of standard RoPE's calibrated equal-allocation leakage, so that baseline's failure at budget 64 is local, not the onset of a distant alias at length 4096.

Table 1: Finite resource costs at $n = 4096$, $k = 8$, $h = \log 100$, with a common uniform key and equal optimization permissions. These are numerical optima, not interval-certified endpoints. The last column identifies the equal-allocation profile's hardest displacement.

Spectrum	Equal-allocation A_{QK}	Optimized A_{QK}	Hardest q
Standard	82.33138	44.11702	1
Algebraic	27.54879	26.85136	2974
Algebraic + clock	31.48528	29.48389	343
Standard + clock	82.33134	43.78801	1

Interval certificates give a separate, rigorous finite comparison at $k = 64$, $L = 8192$ competitors on each side, and $h = \log 10^6$: the standard uniform-key optimum lies in $[163.8370, 163.8373]$. The smaller coefficient-budget interval $[53.2618, 53.2621]$ permits a different key allocation and is not a uniform-key baseline. Appendix F.3 gives these certificates and the corresponding equal-weight costs.

5.2 Repairing phases and allocation in the same checkpoints

The reported experiment uses five runs per schedule, training lengths $32, 64, 128, 256, 1024$ AdamW updates, $d_{\mathrm{head}} = 16$, and enforced $A_{\mathrm{QK}} = 64$. Payloads enter values only. At $n = 4096$, all twenty original profiles have a reported strictly negative off-target gap; increasing their amplitude cannot make them reliable (Lemma 4). We compare two interventions: align the query phases while preserving every block norm, then optimize query allocation while retaining the *original key and frequencies*.

Table 2: Reported same-checkpoint interventions at $n = 4096$, $A_{\mathrm{QK}} = 64$ (five seeds per schedule). Success requires $Z \le 10^{-2}$ at every address. Both interventions retain the original spectrum and key; fixed-key optimization uses the evaluation task. Aligned training is a separate control (Table 3).

Spectrum	Original	Phase repair	Fixed-key optimum
Standard	0/5	1/5	**5/5**
Algebraic	0/5	**5/5**	**5/5**
Algebraic + clock	0/5	**5/5**	**5/5**
Standard + clock	0/5	2/5	**5/5**

Phase repair restores finite-amplitude feasibility in all twenty profiles and success at budget 64 in thirteen. Fixed-key query reallocation makes all twenty succeed at that budget (Table 2). Thus the reported failures do not require an explanation based on insufficient capacity of those spectra and keys: phase choice suffices to explain thirteen cases, with allocation also needed in seven. For a mirror-closed address family, phase alignment cannot increase worst-address leakage (Appendix E.2); the guarantee does not apply to an isolated boundary task.

The optimized queries use the evaluation task, so this comparison diagnoses capacity rather than trained extrapolation. It also does not establish that the original runs failed to optimize their short-length loss. A separate aligned-initialization training control reaches $17/20$ at length 4096; its paired-length counts, calibrated costs, and different parameterization are kept in Appendix F.2.

5.3 Certifying a clock's finite operating cost

The maximal-radius clock need not minimize amplitude at a fixed subcritical error. In a 450-case grid, $c = 0.60$ lowers all 30 matched equal-allocation envelope budgets relative to $c = 0.50$ at common physical error $\sigma = 0.1/L$. This ranks sufficient budgets; the common-query optimum requires lower bounds as well.

At $k = 8$, $L = 65536$, $h = \log 10^6$, and the same $\sigma = 0.1/L$, interval certificates give robust common-query cost brackets $[168.928, 189.526]$ for $c = 0.50$ and $[149.036, 168.075]$ for $c = 0.60$,

using one uniform key and optimized allocation. The second upper bound is below the first lower bound, establishing an operating-cost improvement even after allocation is optimized. This is compatible with the half-clock's maximal critical radius: tolerance and subcritical cost are different objectives.

Appendix F.6 provides all six robust-cost certificates and the independently reconstructed grid; Appendix F.4 records the additional finite-return and integer-box checks. Arithmetic-path diagnostics are treated separately in Appendix G, since fixed installed-frequency errors and subsequent phase-evaluation errors are different resources.

6 Related work and scope of comparison

Expressivity and task scale. Jerad et al. (2026) characterize fully uniform finite-precision Transformers and certify fixed-offset look-back over bounded lengths. Our frozen-spectrum quantifier is different: the spectrum and key are fixed, but amplitude and evaluation precision may grow. Wu et al. (2026) maximize average single-frequency contrast subject to $\omega W \leq \pi$, obtaining $\omega^* = \pi/W$. We instead optimize the total leakage of a joint signed score, allow phase wrapping, and distinguish boundary from interior competitors.

Observed and learned frequency bands. Oka et al. (2026) study high-norm bands, their dependence on base and training length, and interpolation–extrapolation trade-offs in language models. LeRoPE learns per-frequency scales and reports language-model and retrieval improvements from 52M to 2.5B parameters (Karypis et al., 2026). Its analysis also tests dominant-band-only interpolation, making it particularly close to our one-coordinate construction. The distinction is the object being guaranteed: necessary amplitude and all-length reuse within a specified positional channel, plus an exact coherent-error threshold. Neither the optimal allocation in Table 1 nor our clock theorem predicts the frequency band learned by those models.

Constructions and inherited tools. Position-selective RoPE scores (Barbero et al., 2024), directional Fourier priors (Litman & Guo, 2026), and context-extension rescaling (Peng et al., 2023; Wang et al., 2024b) motivate our implementation interfaces. Our proofs use classical group-code geometry, convex duality, and badly approximable Kronecker sequences (Torezzan et al., 2012; Boyd & Vandenberghe, 2004; Dick et al., 2025). In particular, the algebraic frequency vector is inherited. The contribution is the retrieval-resource comparison and the obstruction explaining why one extra plane is necessary, together with the distinct minimal-adaptation result.

7 Discussion and limitations

Task-specific boundary geometry permits a first-order advantage; all-length freezing can erase it; exactly one additional rotation plane compensates. Frequency uncertainty creates a separate obstruction, for which one length-adaptive coordinate restores order-$1/L$ tolerance without sacrificing the amplitude order. The first result minimizes total blocks under freezing; the second minimizes adjustable frequencies, not total dimension under adaptation.

These guarantees concern supplied-address, position-only soft attention. They do not establish language-model performance or context limits for content-dependent, multihead, multilayer networks. Amplitude and evaluation precision may grow. Competitive constants depend on fixed dimension, the obstruction's onset can depend on the frozen spectrum, and appending a block to an arbitrary existing RoPE table need not achieve the theorem. Standard geometric spectra are not assumed to be fully badly approximable.

Within this scope, finite calibration makes the resource distinction testable. Holding the key and candidate set fixed separates spectral cost from allocation, while same-checkpoint interventions distinguish capacity from learned phase choice. The training evidence is limited to reported aggregates, with available and missing artifacts documented in Appendix F.5. The main guarantees instead rest on the proofs and finite certificate checks. Same-dimensional finite-horizon reuse, minimum total dimension under adaptation, and the optimal joint critical-error limit remain open within this framework.

AI USE STATEMENT

Generative AI was used substantively in the research and writing. ChatGPT assisted with the resource formulation, proposed and revised mathematical claims, supplied proof ideas and proof text, participated in correctness and literature audits, designed experiments, implemented reference and certificate code, interpreted experimental reports, and drafted and edited the manuscript. Codex was used for the remote experimental implementation described in those reports. Image generation assisted the conceptual figure design; exact labels and tables were composed and checked separately. These uses extend beyond language polishing. Verification consists of explicit derivations, numerical checks, and separately executable interval or integer certificate verifiers, not proof-assistant formalization or independent human certification of AI-generated work. The human authors remain responsible for the final claims, references, artifacts, and disclosure.

REPRODUCIBILITY STATEMENT

Appendices A–D give the model, proofs, and constructions; Appendix E derives the finite certificate duals. The companion archive includes reference code, exact witnesses and verifiers for 93 certificate cases, the 450-configuration floating clock reconstruction, and four uniform-key budget recomputations. It also includes transcribed training and arithmetic aggregates. Original report documents, checkpoints, per-seed training records, and remote executors are not included, so those aggregate results are reported rather than independently replicated. Appendix F.5 and `README.md` identify the available evidence, missing artifacts, and executable checks.

ETHICS STATEMENT

The experiments use synthetic positional retrieval, with no human-subject recruitment or personal datasets. The results concern the specified positional channel rather than general language-model performance or deployment safety. We distinguish mathematical proofs, numerical certificates, and reported aggregates, retain failed cases, and disclose substantive AI involvement. No deployment-safety or broad social benefit is established by these experiments.

REFERENCES

Federico Barbero, Alex Vitvitskyi, Christos Perivolaropoulos, Razvan Pascanu, and Petar Veličković. Round and Round We Go! What Makes Rotary Positional Encodings Useful? *arXiv preprint arXiv:2410.06205*, 2024. URL https://arxiv.org/abs/2410.06205.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. URL https://web.stanford.edu/~boyd/cvxbook/.

Josef Dick, Takashi Goda, Gerhard Larcher, Friedrich Pillichshammer, and Kosuke Suzuki. On the Quasi-uniformity Properties of Quasi-Monte Carlo Point Sets and Sequences – Part I: Lattices and Kronecker Sequences. *arXiv preprint arXiv:2502.06202; version 2, revised 2026*, 2025. URL https://arxiv.org/abs/2502.06202v2.

Bram L. Gorissen and Dick den Hertog. Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research*, 227(1):30–34, 2013. doi: 10.1016/j.ejor.2012.10.007. URL https://arxiv.org/abs/1501.02656.

K.-L. Hsiung, S.-J. Kim, and Stephen Boyd. Tractable Approximate Robust Geometric Programming. *Optimization and Engineering*, 9(2):95–118, 2008. URL https://web.stanford.edu/~boyd/papers/rgp.html.

Selim Jerad, Anej Svete, Jiaoda Li, and Ryan Cotterell. Disentangling the Expressivity of RoPE. *arXiv preprint arXiv:2608.11909*, 2026. URL https://arxiv.org/abs/2608.11909.

Petros Karypis, Sean O'Brien, Shreyas Kadekodi, Rui Zhu, and Julian McAuley. LeRoPE: Learnable RoPE Frequencies Improve Language Modeling. *arXiv preprint arXiv:2607.10134v1*, 2026. URL https://arxiv.org/abs/2607.10134.

Elon Litman and Gabe Guo. You Need Better Attention Priors. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2601.15380v2.

Yui Oka, Itsumi Saito, Kyosuke Nishida, and Kuniko Saito. Frequency Bands in RoPE: Base Frequency and Context Length Shape the Interpolation–Extrapolation Trade-off. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/993dbaec0418a6449090f1debbcb8844-Abstract-Conference.html.

Bowen Peng, Jeffrey Quesnelle, Honglu Fan, and Enrico Shippole. YaRN: Efficient Context Window Extension of Large Language Models. *arXiv preprint arXiv:2309.00071*, 2023. URL https://arxiv.org/abs/2309.00071.

Jianlin Su, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. RoFormer: Enhanced Transformer with Rotary Position Embedding. *arXiv preprint arXiv:2104.09864*, 2021. URL https://arxiv.org/abs/2104.09864.

Matthew Tointon. New Bounds in the Discrete Analogue of Minkowski's Second Theorem. *Discrete Analysis, 2024:7*, 2024. URL https://doi.org/10.19086/da.122972.

Cristiano Torezzan, Jo ao E. Strapasson, Sueli I. R. Costa, and Rogério M. Siqueira. Optimum Commutative Group Codes. *arXiv preprint arXiv:1205.4067*, 2012. URL https://arxiv.org/abs/1205.4067.

Petar Veličković, Christos Perivolaropoulos, Federico Barbero, and Razvan Pascanu. Softmax Is Not Enough (for Sharp Size Generalisation). *arXiv preprint arXiv:2410.01104*, 2024. URL https://arxiv.org/abs/2410.01104.

Haonan Wang, Qian Liu, Chao Du, Tongyao Zhu, Cunxiao Du, Kenji Kawaguchi, and Tianyu Pang. When Precision Meets Position: BFloat16 Breaks Down RoPE in Long-Context Training. *arXiv preprint arXiv:2411.13476*, 2024a. URL https://arxiv.org/abs/2411.13476.

Suyuchen Wang, Ivan Kobyzev, Peng Lu, Mehdi Rezagholizadeh, and Bang Liu. Resonance RoPE: Improving Context Length Generalization of Large Language Models. In *Findings of the Association for Computational Linguistics: ACL 2024*, 2024b. URL https://aclanthology.org/2024.findings-acl.32.

Xinyi Wu, Siyuan Liu, and Ali Jadbabaie. How Data Shapes RoPE Frequency Usage: From Positional Scale Matching to Length Generalization. *arXiv preprint arXiv:2607.07678*, 2026. URL https://arxiv.org/abs/2607.07678.

A NOTATION, RESOURCE ACCOUNTING, AND ELEMENTARY REDUCTIONS

All positions and target indices in this manuscript are zero-based. A head of width d_{head} provides k two-dimensional rotation blocks; unused blocks and extra coordinates are zero. The number m denotes the task-specific oracle's block budget and need not equal k. Frequencies θ are in cycles per position, so the angular frequency is $2\pi\theta$. The quantities n, L, and the candidate count are not interchangeable: $\mathcal{C}_{n,r}$ has $n-1$ competitors; S_L has $2L$ competitors and $2L+1$ candidates including its target.

Symbol	Meaning throughout this manuscript
k	Number of rotary two-dimensional blocks
m	Oracle block budget, only when comparing representations
d_{head}	Ambient attention-head width; $d_{\text{head}} \geq 2k$
$\mathcal{C}_{n,r}$	Competitor displacements for address r among n keys
L	Maximum displacement in E_L or S_L
h	Log-reliability: success means $Z \leq e^{-h}$
$A_{\text{coeff}}, A_{\text{QK}}$	Coefficient amplitude and actual scaled query–key norm
θ, σ	Frequencies and coherent error radius, in cycles/position

The archived code and CSVs retain their original field names `d` and `D`, corresponding to k and d_{head} here; the data are unchanged.

A.1 EXACT QUERY–KEY INTERPRETATION

For the rotation $R_\phi = \left(\begin{smallmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{smallmatrix}\right)$, a block satisfies

$$Q_j^T R_\phi K_j = a'_j \cos\phi + b'_j \sin\phi, \qquad (a'_j)^2 + (b'_j)^2 = \|Q_j\|^2 \|K_j\|^2 .$$

After dividing by $\sqrt{d}_{\text{head}}$, its coefficient amplitude is $\rho_j = \|Q_j\| \|K_j\| / \sqrt{d}_{\text{head}}$. Summing and applying Cauchy–Schwarz proves $A_{\text{coeff}} \leq A_{\text{QK}}$. If each nonzero block is balanced with $\|Q_j\| = \|K_j\| = d_{\text{head}}^{1/4} \sqrt{\rho_j}$, the desired signs can be implemented by its relative phase and equality holds.

Now fix a key with $\|K_j\|^2 = w_j \|K\|^2$, $w_j > 0$. Realizing amplitude ρ_j requires

$$\|Q_j\| = \frac{\sqrt{d}_{\text{head}}\, \rho_j}{\|K\| \sqrt{w_j}}, \qquad \frac{\|Q\|^2 \|K\|^2}{d_{\text{head}}} = \sum_j \frac{\rho_j^2}{w_j}.$$

Each relative phase is freely realizable in its plane, proving (3). When $w_j = 0$, that block is omitted or its coefficient is constrained to zero. Reoptimizing w separately per query is not a common-key implementation. Our equal-amplitude constructions use $K_j = (1, 0)$ and query amplitudes $A\sqrt{d}_{\text{head}}/k$, so $\|K\| = \sqrt{k}$, $\|Q\| = A\sqrt{d_{\text{head}}/k}$, and $A_{\text{QK}} = A_{\text{coeff}} = A$.

A.2 VALUES, LOGIT ERROR, AND FINITE-AMPLITUDE FEASIBILITY

Let the target be indexed by zero and include it in the softmax denominator. Then $\alpha_0 = (1+Z)^{-1}$. For $\|v_t\| \leq 1$,

$$\left\| \sum_t \alpha_t v_t - v_0 \right\| \leq (1-\alpha_0) \|v_0\| + \sum_{t \neq 0} \alpha_t \|v_t\| \leq 2(1-\alpha_0).$$

Choose any unit vector u, with $v_0 = u$ and $v_t = -u$ for every competitor, to attain equality. This is an arbitrary bounded-value model, not an assertion about which pairs a specific vocabulary contains.

If every logit error satisfies $|\xi_t| \leq \nu$, then $\xi_t - \xi_0 \leq 2\nu$. Taking $\xi_0 = -\nu$ and all competitor errors $+\nu$ attains the simultaneous worst case $\sup_\xi Z(s+\xi) = e^{2\nu} Z(s)$. This justifies the definition of h. Arithmetic error, frequency uncertainty, and value representation error are separate budgets.

Lemma 4 (Scaling a fixed score)**.** *For $\lambda \geq 1$, $Z_{\mathcal{C}}(\lambda s) \leq Z_{\mathcal{C}}(s)^\lambda$. For a nonempty finite $\mathcal{C}$ and $h > 0$, there is some finite $\lambda \geq 0$ with $Z_{\mathcal{C}}(\lambda s) \leq e^{-h}$ if and only if $\min_{t \in \mathcal{C}} [s(0) - s(t)] > 0$.*

Proof. Write $x_t = e^{s(t)-s(0)} \geq 0$. Then $\sum x_t^\lambda \leq (\sum x_t)^\lambda$ for $\lambda \geq 1$. If a gap is nonpositive, its summand after any nonnegative scaling is at least one. Otherwise, with minimum gap $\mu > 0$, $Z(\lambda s) \leq |\mathcal{C}|e^{-\lambda\mu}$ and a sufficiently large finite λ succeeds. A finite search ceiling is not a proof of infeasibility when $\mu > 0$. □

B Complete proof of the task-specific resource laws

The fixed-dimension statement in Proposition 2 follows from the following stronger, uniform-parameter version. Let $N = L + 1$. For all $L \geq 2$, $k \geq 1$, and $h \geq \log 3$,

$$A_E^*(L, k, h) \asymp k\big(N^{2/(k+1)} - 1\big) + N^{2/(k+1)}h, \tag{16}$$

$$A_S^*(L, k, h) \asymp k\big(N^{2/k} - 1\big) + N^{2/k}h. \tag{17}$$

The comparison constants do not depend on L, k, h. At each fixed design length one common spectrum and one common key attain both orders, with frequencies independent of h. We prove this complete version below. The auxiliary partitions used for lower bounds may depend on the particular score; they are existence certificates for every score, not a preprocessing scheme independent of the query.

B.1 A collision-multiplicity identity

Put $N = L + 1$ orbit points $x_i = i\theta \bmod 1$ into at most B cells. If a cell contains n_c points, the number of same-cell unordered pairs is

$$P = \frac{1}{2}\left(\sum_c n_c^2 - N\right) \geq \frac{1}{2}\left(\frac{N^2}{B} - N\right). \tag{18}$$

A pair $i < j$ gives a displacement $t = j - i \in \{1, \ldots, L\}$. The same displacement occurs at most $N - t \leq N$ times. Thus a same-cell score-gap upper bound $s(0) - s(t) \leq \delta$ gives

$$NZ_{\{1,\ldots,L\}} \geq Pe^{-\delta}. \tag{19}$$

If instead each pair supplies $e^{s(t)-s(0)} + e^{s(-t)-s(0)} \geq 2e^{-\delta}$, then

$$NZ_{S_L} \geq 2Pe^{-\delta}. \tag{20}$$

Counting P as distinct displacements would incorrectly strengthen these bounds by a length factor.

B.2 One-sided lower bound

Suppose first $A_{\text{coeff}} = A > 0$. Partition $[0, 1)^k$ into Q^k half-open cubes of side $1/Q$, for an integer $Q \geq 1$. Inside each cube, the linear functional $b \cdot x/A$ has range of length at most $1/Q$, since $\sum_j |b_j| \leq A$. Subdivide this range into half-open intervals of width $1/Q^2$. At most $Q + 1 \leq 2Q$ intervals suffice, including boundary conventions. There are at most $2Q^{k+1}$ cells in total.

For two points in one cell, $v = x_j - x_i$ obeys

$$\|v\|_\infty \leq Q^{-1}, \qquad |b \cdot v| \leq AQ^{-2}.$$

The phase function $F(x) = \sum_j [a_j \cos(2\pi x_j) + b_j \sin(2\pi x_j)]$ has gradient $2\pi b$ at zero. Its diagonal second derivatives have absolute values at most $4\pi^2\rho_j$. Consequently,

$$|F(v) - F(0) - 2\pi b \cdot v| \leq 2\pi^2 \sum_j \rho_j v_j^2 \leq 2\pi^2 A/Q^2.$$

Periodicity gives $F(v) = s(j - i)$ exactly, because $v - (j - i)\theta$ is integral. No assumption about a phase representative avoiding a global torus cut is needed for this identity. With $K_E = 2\pi + 2\pi^2$,

$$|s(j - i) - s(0)| \leq K_E A/Q^2.$$

Using (18)–(19), and reflecting b for the negative one-sided set, proves

$$Z_{E_L} \geq \left(\frac{N}{4Q^{k+1}} - \frac{1}{2}\right)_+ e^{-K_E A/Q^2}. \tag{21}$$

For $A = 0$ the score is constant and $Z_{E_L} = L$; the displayed inequality holds directly. Rational, repeated, zero, or resonant frequencies have not been excluded.

Set $R = N^{1/(k+1)}$. If $R \geq 8$, take $Q = \lfloor R/4 \rfloor$. Then $R/8 \leq Q \leq R/4$ and the prefactor in (21) is at least $4^k - 1/2 \geq 2^k$. Success therefore requires

$$A \geq \frac{R^2}{64K_E}(h + k \log 2). \tag{22}$$

Since $R^2(h + k \log 2) \geq (\log 2)[k(R^2 - 1) + R^2 h]$, this bounds the claimed expression from below by a uniform constant.

If $R < 8$, the score range gives $Z_{E_L} \geq Le^{-2A}$, hence $A \geq (\log L + h)/2$. Put $x = \log R^2 \in (0, \log 64)$. Since $(e^x - 1)/x$ is increasing and $kx \leq 2 \log N$,

$$k(R^2 - 1) \leq \frac{126}{\log 64} \log N, \qquad R^2 h \leq 64h.$$

For $L \geq 2$, $\log N \leq (\log 3/\log 2) \log L$. Thus the entire right side of (16) is at most $64(\log L + h) \leq 128A$. Combining the cases permits, for example, lower constant 10^{-4}.

A simpler bound is useful for all-address comparison, including $L = 1$:

$$A^*_E(L, k, h) \geq \frac{1}{2048}(L + 1)^{2/(k+1)} h. \tag{23}$$

For $R \geq 8$, use (22) and $64K_E < 2048$. For $R < 8$, a single competitor already requires $2A \geq h$, which is stronger. Applying this bound with the oracle budget m, every address has one side of length at least $(n - 1)/2$; because $2/(m + 1) \leq 1$, it follows that

$$A_{\mathrm{or},m}(n, r, h) \geq \frac{1}{4096} n^{2/(m+1)} h. \tag{24}$$

B.3 Two-sided lower bound

Use only the Q^k phase cubes. For a same-cube displacement with centred difference v of maximum norm at most $1/Q$, exact pairing gives

$$\begin{aligned} e^{s(t)-s(0)} + e^{s(-t)-s(0)} &= 2e^{\sum_j a_j(\cos(2\pi v_j)-1)} \cosh\left(\sum_j b_j \sin(2\pi v_j)\right) \\ &\geq 2e^{-2\pi^2 A/Q^2}. \end{aligned}$$

The inequality uses $1 - \cos z \leq z^2/2$, $\cosh \geq 1$, and $\sum |a_j| \leq A$. It does not remove a sine term in an individual summand. Equations (18) and (20) imply

$$Z_{S_L} \geq \left(\frac{N}{Q^k} - 1\right)_+ e^{-2\pi^2 A/Q^2}. \tag{25}$$

Let $R = N^{1/k}$. If $R \geq 4$, take $Q = \lfloor R/2 \rfloor \geq R/4$. The prefactor is at least $2^k - 1$, and $h \geq \log 3$ implies $h + \log(2^k - 1) \geq (h + k \log 2)/2$. Hence

$$A \geq \frac{R^2}{64\pi^2}(h + k \log 2),$$

which is a uniform constant multiple of $k(R^2 - 1) + R^2 h$. If $R < 4$, use $Z_{S_L} \geq 2Le^{-2A}$ together with

$$k(R^2 - 1) \leq \frac{30}{\log 16} \log N, \qquad R^2 h \leq 16h.$$

As $N \leq 2L$, this also yields the claimed uniform lower bound. Again 10^{-4} is a conservative common constant.

B.4 A DETERMINISTIC SPECTRUM SHARED BY BOTH TASKS

Let $\ell = \min\{k, \lceil \log(L+1) \rceil\}$. Use a slow frequency $\theta_0 = 1/(4L)$ and $\ell - 1$ fast frequencies u_j/p, where p is a prime with $128L < p < 256L$ (Bertrand's postulate) and $u_j \in \{1, \dots, p-1\}$. Initially choose the u_j independently and uniformly. Define integers

$$\tau_E = \left\lceil 1024 L^{2/(\ell+1)} \right\rceil, \qquad \tau_S = \left\lceil 1024 L^{2/\ell} \right\rceil.$$

The base scores are

$$s_E^0(t) = \tau_E \left[\sum_{j=1}^{\ell-1} \cos(2\pi u_j t/p) + \sin(\pi t/(2L)) \right], \tag{26}$$

$$s_S^0(t) = \tau_S \left[\sum_{j=1}^{\ell-1} \cos(2\pi u_j t/p) + \cos(\pi t/(2L)) \right]. \tag{27}$$

When $\ell = 1$ the fast sum is empty. Integer powers and integer root comparisons compute the ceilings without an exact real-comparison oracle.

For either τ, the average contribution of a fast block is

$$H_p(\tau) = \frac{1}{p-1} \sum_{a=1}^{p-1} e^{-\tau(1-\cos(2\pi a/p))} \le \frac{1}{p} + \sqrt{\frac{\pi}{8\tau}} \le \tau^{-1/2}.$$

To prove the first inequality, include the zero residue (whose value one is maximal), use $1 - \cos(2\pi x) \ge 8\,\mathrm{dist}(x, \mathbb{Z})^2$, and bound the nonzero Gaussian sum by its integral. For the second, $\tau \le 1025L^2$, $p > 128L$, and $\sqrt{\tau}/p + \sqrt{\pi/8} < 33/128 + 2/3 < 1$. Multiplication by each $1 \le |t| \le L < p$ permutes the nonzero residues, so this average bound holds at every competitor displacement.

For $1 \le q \le L$, concavity of sine and the elementary cosine bound give

$$\sin \frac{\pi q}{2L} \ge \frac{q}{L}, \qquad 1 - \cos \frac{\pi q}{2L} \ge \frac{q^2}{2L^2}.$$

The slow block therefore satisfies

$$\sum_{q=1}^{L} e^{-\tau \sin(\pi q/(2L))} \le \frac{1}{e^{\tau/L} - 1} \le \frac{L}{\tau}, \quad 2\sum_{q=1}^{L} e^{-\tau(1-\cos(\pi q/(2L)))} \le \frac{\sqrt{2\pi} L}{\sqrt{\tau}}.$$

The first expression has the correct sign for $t = -q$ in (26). Independence of fast blocks yields

$$\mathbb{E} Z_E(s_E^0) \le L \tau_E^{-(\ell+1)/2} \le 32^{-(\ell+1)}, \qquad \mathbb{E} Z_S(s_S^0) \le \sqrt{2\pi} L \tau_S^{-\ell/2} \le \sqrt{2\pi}\, 32^{-\ell}.$$

Their *sum* is at most $e^{-2\ell}$ for $\ell \ge 1$. Minimizing this joint potential by conditional expectation chooses a single spectrum good for both tasks. Separate existence arguments would not establish shared frequencies.

Finite computation and derandomization. Choose each u_j by enumerating $p - 1$ candidates and minimizing the conditional expectation of the sum of the two leakages. Evaluate each candidate potential within absolute error $\delta = e^{-2\ell}/(4\ell)$. The selected candidate increases the true potential by at most 2δ; after at most $\ell - 1$ choices it is less than $\frac{3}{2} e^{-2\ell} \le e^{-\ell}$. Each individual leakage is thus at most $e^{-\ell}$.

This procedure does not rely on exact comparisons of transcendental numbers. All exponential arguments are nonpositive; terms below a controlled threshold can be discarded within the error budget. There are $O(L)$ terms and $\ell = O(\log L)$ factors. The required precision, including the factor $\tau = O(L^2)$ in angular errors, is $O(\log L)$ bits. Standard convergent series with remainder bounds therefore give a deterministic implementation polynomial in L. Direct enumeration uses $O(\ell p L) = O(\ell L^2)$ term evaluations. We do not claim a running time polynomial in $\log L$.

Choose $\lambda = \max\{1, \lceil h/\ell \rceil\}$. By Lemma 4, $Z(\lambda s^0) \le e^{-\ell\lambda} \le e^{-h}$, while $\ell\lambda \le \ell + h$. Hence

$$A_E \le 1025 L^{2/(\ell+1)}(\ell + h), \qquad A_S \le 1025 L^{2/\ell}(\ell + h). \tag{28}$$

To compare uniformly, put $z = \log N > 1$. If $\ell = k \le \lceil z \rceil$, then $2z/(k+1) \ge 2/3$ and $2z/k \ge 1$, so kN^β is a fixed constant multiple of $k(N^\beta - 1)$ for the respective exponents. If $\ell = \lceil z \rceil < k$, both $L^{2/\ell}$ and $L^{2/(\ell+1)}$ are at most e^2 and $\ell \le 2z$, while

$$k(N^{2/(k+1)} - 1) + N^{2/(k+1)} h \ge \log N + h,$$

$$k(N^{2/k} - 1) + N^{2/k} h \ge 2 \log N + h.$$

Thus (28) matches both lower bounds with a universal upper constant, for instance 2×10^4. The frequency choices depend on L, k but not h. Equal amplitudes on the ℓ active blocks and a common key uniform on those blocks prove the shared-key statement of Proposition 2. If $\ell < k$, both query and key are zero in the unused blocks; the common key may depend on the fixed design length, as allowed in that proposition.

B.5 The even-score penalty

Evenization $s_{\mathrm{ev}}(t) = [s(t) + s(-t)]/2$ does not increase coefficient cost. By convexity of the exponential it does not increase Z_{S_L} either. The optimum on S_L can therefore be approached by even scores. For such a score $Z_{S_L} = 2Z_{E_L}$, giving $A^*_{E,\mathrm{even}} \le A^*_S$. Conversely, if an even score has $Z_{E_L}(s) \le e^{-h}$, then

$$Z_{S_L}(2s) = 2Z_{E_L}(2s) \le 2Z_{E_L}(s)^2 \le 2e^{-2h} \le e^{-h}.$$

Thus $A^*_S \le 2A^*_{E,\mathrm{even}}$, with infima handled by arbitrarily close feasible scores. Combining with Proposition 2 gives the stated fixed-k penalty $\Theta_k(L^{2/[k(k+1)]})$. Evenness is not equivalent to the weaker condition $s'(0) = 0$.

C Frozen spectra: return separation, coverage, and minimum dimension

C.1 Finite-scale resource characterization

Lemma 5 (Resources at a fixed spectrum)**.** *Let* $\Delta = \min_{1 \le q \le L} \max_j \|q\theta_j\|_{\mathbb{R}/\mathbb{Z}}$. *If* $\Delta = 0$, S_L *retrieval with* $h > 0$ *is impossible. Otherwise,*

$$\frac{h + \log 2}{2\pi^2 \Delta^2} \le A_\theta(S_L, h) \le \frac{k}{8\Delta^2}(h + k \log 5 + \log 3).$$

Proof. Take a minimizing q and its centred phase representative v, $\|v\|_\infty = \Delta$. Pairing q and $-q$ as before yields $Z \ge 2e^{-2\pi^2 A \Delta^2}$. This gives the lower bound and impossibility when $\Delta = 0$.

For the upper bound, choose $s(t) = (A/k)\sum_j \cos(2\pi\theta_j t)$. The centred representatives x_t, $t = 0, \dots, L$, are pairwise at least Δ apart in Euclidean distance: their index difference is at most L, its torus maximum distance is at least Δ, and representative distance cannot be smaller. Their nonzero distances from zero are also at least Δ. In the shell $\ell\Delta \le \|x_t\|_2 < (\ell+1)\Delta$, $\ell \ge 1$, there are at most $(2\ell+3)^k$ points. Indeed, disjoint radius-$\Delta/2$ Euclidean balls around them lie in a radius-$(\ell + 3/2)\Delta$ ball about zero.

Since $1 - \cos(2\pi x) \ge 8x^2$ on $[-1/2, 1/2]$, the score gap is at least $8A\|x_t\|_2^2/k$. Count the positive and negative halves separately; no return bound at length $2L$ is needed. With $B = 8A\Delta^2/k$ and $B > k \log 5$,

$$Z_{S_L} \le 2\sum_{\ell \ge 1}(2\ell+3)^k e^{-B\ell^2} \le 2\sum_{\ell \ge 1} e^{-\ell(B - k\log 5)} = \frac{2}{e^{B - k\log 5} - 1}.$$

Here $2\ell + 3 \le 5^\ell$ and $\ell^2 \ge \ell$. Setting $B = h + k\log 5 + \log 3$ makes the bound at most e^{-h}, since $h \ge \log 3$. The same upper-bound proof accepts any valid Euclidean separation lower bound in place of Δ. □

A vector is simultaneously badly approximable if some $\kappa > 0$ satisfies

$$u_q(\theta) := \max_j \|q\theta_j\|_{\mathbb{R}/\mathbb{Z}} \geq \kappa q^{-1/k} \quad \text{for every } q \geq 1. \tag{29}$$

By Lemma 5, this is equivalent to $A_\theta(S_L, h_0) \leq CL^{2/k}$ at every L for any one fixed $h_0 \geq \log 3$. The forward implication uses $\Delta(L) \geq \kappa L^{-1/k}$. In the reverse direction apply the lower bound at $L = q$ and use $u_q \geq \Delta(q)$. It is also equivalent to a uniform upper bound $O_{\theta,k}(n^{2/k}h)$ for all addresses and all $h \geq \log 3$, by embedding $\mathcal{C}_{n,r}$ in S_{n-1}.

C.2 From bad approximability to a forward nonzero covering net

The following mechanism is classical; it is also contained in the quasi-uniformity theorem of Dick et al. (2025). We supply the exact direction needed here. We invoke Minkowski's second theorem for a symmetric convex body: for successive minima λ_i, volume V, and determinant $\det \Lambda$, $\prod_i \lambda_i V \leq 2^\ell \det \Lambda$ in dimension ℓ (see the classical statement discussed by Tointon (2024)).

Lemma 6 (Forward covering)**.** *If* (29) *holds, there are $C_{\theta,k} < \infty$ and L_0 such that, for every $L \geq L_0$, $\{q\theta \bmod 1 : 1 \leq q \leq L\}$ is a torus maximum-norm net of radius $C_{\theta,k}L^{-1/k}$.*

Proof. For $T \geq 1$ form the determinant-one lattice

$$\Lambda_T = \{(q/T,\ T^{1/k}(q\theta - p)) : q \in \mathbb{Z},\ p \in \mathbb{Z}^k\} \subset \mathbb{R}^{k+1}.$$

The determinant follows from the block triangular map with diagonal scaling T^{-1} and $T^{1/k}I_k$. Let $c = \min\{1, \kappa^{k/(k+1)}\}$. A nonzero lattice vector with $q = 0$ has maximum norm at least $T^{1/k} \geq 1$. If $q \neq 0$, put $x = |q|/T$. Then its norm is at least $\max\{x, \kappa x^{-1/k}\} \geq c$. Thus the first successive minimum for the unit cube is at least c.

For the unit cube, Minkowski's product upper bound is one, so the largest successive minimum is at most c^{-k}. Choose $k + 1$ linearly independent lattice vectors of maximum norm at most c^{-k}. They need not form an integer basis of the full lattice. Their full-rank sublattice suffices: express any point in this real linear basis and round each coefficient to the nearest integer. The resulting lattice point is within maximum distance

$$M = \frac{k+1}{2}c^{-k}.$$

For $L \geq 2M + 2$, set $T = L/(2M + 2)$. Given a torus point represented by $y \in \mathbb{R}^k$, approximate $(M + 1, T^{1/k}y)$ by the preceding sublattice. The corresponding integers satisfy

$$T \leq q \leq (2M + 1)T < L, \qquad \|q\theta - p - y\|_\infty \leq MT^{-1/k}.$$

The integer q is positive and nonzero, not a spurious target or negative-time point. We may take $C_{\theta,k} = M(2M + 2)^{1/k}$. The finite lower-length threshold is allowed to depend on the fixed spectrum. □

C.3 Forward competitors can eliminate the first-order advantage

Lemma 7 (Convex phase enclosure)**.** *If the positive-time orbit $1, \ldots, L$ covers the torus with radius $\delta < 1/6$, then every successful score on either one-sided task has*

$$A \geq \frac{h}{18\pi^2\delta^2}.$$

Proof. For every sign vector $\sigma \in \{-1, 1\}^k$, cover the centre $2\delta\sigma \bmod 1$ by a positive-time orbit point. Its centred representative v_σ obeys

$$\|v_\sigma - 2\delta\sigma\|_\infty \leq \delta, \qquad \|v_\sigma\|_\infty \leq 3\delta < 1/2.$$

The neighbourhoods stay away from the centred coordinate cut, so these Euclidean representatives are valid simultaneously. For any nonzero linear functional u, take $\sigma = -\operatorname{sign}(u)$, choosing arbitrary signs where $u_j = 0$. Then $u \cdot v_\sigma \leq -\delta \|u\|_1 < 0$. No supporting hyperplane separates zero from

their convex hull. In fact zero is interior to that hull, so there are nonnegative λ_σ with $\sum \lambda_\sigma = 1$ and $\sum \lambda_\sigma v_\sigma = 0$.

For the phase score F, Taylor's bound is $|F(v) - F(0) - 2\pi b \cdot v| \le 2\pi^2 A \|v\|_\infty^2$. Each selected competitor must individually satisfy $F(0) - F(v_\sigma) \ge h$, because its leakage term cannot exceed the total e^{-h}. After taking their convex combination, the linear term cancels *exactly*, yielding

$$h \le \sum_\sigma \lambda_\sigma [F(0) - F(v_\sigma)] \le 18\pi^2 A \delta^2.$$

Reflect the frequencies or sine coefficients for the negative-time task. All competitors in this proof lie on one temporal side; it is their phase geometry, not temporal symmetry, that cancels the gradient. □

Corollary 8 (Fixed-spectrum dichotomy)**.** *For any fixed k-block spectrum, exactly one of the following holds. If it is not badly approximable, $A_\theta(S_L, h_0)/L^{2/k}$ has no finite upper bound over L for fixed $h_0 \ge \log 3$. If it is badly approximable, then for all sufficiently large n, every r, and every $h \ge \log 3$,*

$$A_\theta(\mathcal{C}_{n,r}, h) = \Theta_{\theta,k}(n^{2/k} h).$$

Proof. The first statement follows from the equivalence following Lemma 5; an exact common period gives outright impossibility. For the second, the upper bound comes from that lemma and the lower bound applies Lemmas 6–7 to a side of length at least $(n-1)/2$. Constants and the onset of the lower bound may depend on the spectrum. □

C.4 The explicit algebraic separation constant

The spectrum below is not a new algebraic construction; it is explicitly discussed by Dick et al. (2025). This proof provides the constant required for the resource comparison.

Lemma 9 (Algebraic Euclidean separation)**.** *Let $k \ge 1$, $\alpha = 2^{1/(k+1)}$, and $\theta = (\alpha, \dots, \alpha^k)$. For every $q \ge 1$,*

$$\min_{p \in \mathbb{Z}^k} \|q\theta - p\|_2 \ge \frac{1}{10\sqrt{2}} q^{-1/k}.$$

Proof. Let $\ell = k+1$, let p_j be a nearest integer to $q\alpha^j$, and write $e_j = p_j - q\alpha^j$, $|e_j| \le 1/2$. The polynomial $x^\ell - 2$ is irreducible by Eisenstein's criterion. The algebraic integer

$$\beta = 2q + \sum_{j=1}^k p_j \alpha^{\ell - j}$$

is nonzero: its defining integer polynomial has degree below ℓ and nonzero constant term. Its field norm is therefore a nonzero integer. Let $\zeta = e^{2\pi i/\ell}$ and index embeddings by $\alpha \mapsto \alpha\zeta^u$. For the principal embedding, substituting $p_j = q\alpha^j + e_j$ gives $|\beta_0| \le 2\ell q + k \le (3k+2)q$. For $u = 1, \dots, k$, the q terms cancel by the sum of roots of unity, leaving

$$\beta_u = 2 \sum_{j=1}^k e_j \alpha^{-j} \zeta^{-uj}.$$

Parseval's identity for this finite Fourier transform gives $\sum_{u=1}^k |\beta_u|^2 \le 4(k+1)\|e\|_2^2$. Applying the arithmetic–geometric mean inequality to their squared magnitudes,

$$1 \le |\operatorname{Norm}(\beta)| \le (3k+2)q \left(2\sqrt{\frac{k+1}{k}} \|e\|_2 \right)^k.$$

Hence

$$\|e\|_2 \ge \frac{1}{2}\sqrt{\frac{k}{k+1}} (3k+2)^{-1/k} q^{-1/k} \ge \frac{1}{10\sqrt{2}} q^{-1/k},$$

using $3k+2 \le 5^k$. The maximum-norm constant is smaller by a possible $\sqrt{k}$; we do not conflate the two norms. □

C.5 Proof of the minimum-dimension theorem

For sufficiency in Theorem 1, use $k = m + 1$ and Lemma 9. On the orbit $0, \dots, n-1$, the Euclidean separation is at least $(10\sqrt{2})^{-1}(n-1)^{-1/k}$. The shell proof in Lemma 5 gives (10). Comparing with (24),

$$\frac{A_U(n,h)}{A_{\mathrm{or},m}(n,r,h)} \le 102400k\left(2 + \frac{k\log 5}{\log 3}\right) \le 2^{20}k^2.$$

Every $\mathcal{C}_{n,r}$ is a subset of S_{n-1}, so the same amplitude works at every address. For a query at position i, set

$$K_j = (1,0), \qquad Q_{i,j} = \frac{A_U\sqrt{d}_{\mathrm{head}}}{k} R_{2\pi\theta_j(r-i)}(1,0).$$

RoPE at positions i and ℓ produces the desired score:

$$\frac{(R_iQ_i)^T R_\ell K}{\sqrt{d_{\mathrm{head}}}} = \frac{A_U}{k}\sum_j \cos(2\pi\theta_j(\ell - r)).$$

This uses one fixed key and attains $A_{\mathrm{QK}} = A_U$ without an address-specific parameter table.

For necessity, suppose a fixed m-block θ satisfies (5) with some finite constant. Fix $h_0 \ge \log 3$. The two-sided task-specific upper bound in Proposition 2 forces $A_\theta(S_L, h_0) = O_m(L^{2/m})$ for all L, so θ is badly approximable. Corollary 8 then gives endpoint cost $\Omega_{\theta,m}(n^{2/m}h_0)$. But the endpoint oracle costs only $O_m(n^{2/(m+1)}h_0)$, and their ratio diverges as $n^{2/[m(m+1)]}$. This contradicts a length-independent competitive constant. Fewer blocks can be embedded in the m-block coefficient model by adding zero coordinates and are also excluded. Necessity is proved in the more permissive per-task coefficient/key allocation model, so it applies to a common key as well.

The infimum need not be attained: a feasible score within a fixed factor of any finite infimum suffices for the upper-bound implication, while all lower bounds hold for every feasible score. The precise negative quantifier is $\forall\theta\,\forall C < \infty\,\exists n, r$ at a fixed h_0. There is no exchange to $\exists n\,\forall\theta$, nor to the distinct problem $\forall n\,\exists\theta_n\,\forall r$.

C.6 Spectra chosen by a training algorithm

The quantifier in Theorem 1 is pointwise in the frozen spectrum. It therefore also covers a spectrum selected by any deterministic or randomized training procedure, once that procedure returns a finite vector that is held fixed across all future lengths. For a returned spectrum with at most m blocks and any finite C, there is a length and address violating the C-competitive guarantee against the m-block oracle. This is an immediate consequence of the theorem, not a new learning lower bound. The counterexample may depend on the returned spectrum; nothing here gives a single finite test length or a training-failure probability at that length. Content-dependent keys, multiple heads, and learned representations in a full LeRoPE model remain outside this positional-channel statement.

D Frequency-box robustness and one-coordinate adaptation

D.1 The exact robust return profile

For a fixed q and coordinate j, torus distance is Lipschitz and movement toward a nearest integer achieves

$$\min_{|e_j|\le\sigma} \|q(\theta_j + e_j)\|_{\mathbb{R}/\mathbb{Z}} = \left(\|q\theta_j\|_{\mathbb{R}/\mathbb{Z}} - q\sigma\right)_+.$$

Because the box is a Cartesian product across coordinates,

$$\min_{\|e\|_\infty\le\sigma} \max_j \|q(\theta_j + e_j)\|_{\mathbb{R}/\mathbb{Z}} = (u_q(\theta) - q\sigma)_+.$$

The frequency box is compact, there are finitely many q, and the objectives are continuous. Thus

$$\min_{\|e\|_\infty\le\sigma} \min_{1\le q\le L} u_q(\theta + e) = \min_{1\le q\le L} \min_{\|e\|_\infty\le\sigma} u_q(\theta + e) = R_\theta(L,\sigma). \tag{30}$$

This exchanges two minimizations, not an adversarial maximum and a minimum. The minimizing pair (e_*, q_*) uses *one* frequency error vector at every position. It need not minimize every displacement simultaneously.

For the lower bound in (12), choose such (e_*, q_*) and apply the paired-return lower bound to every choice of coefficients. It has the quantifiers $\exists e_* \, \exists q_* \, \forall(a, b)$ and even permits the query to know e_*. For the upper bound, every installed spectrum in the box has finite-scale separation at least R, so the same relative coefficients $a_j = A/k, b_j = 0$ work through the shell bound. In particular,

$$\sup_e \inf_c A_e(c) \le \inf_c \{A(c) : \forall e,\ Z_e(c) \le e^{-h}\} \le \frac{k}{8R^2}(h + k \log 5 + \log 3).$$

No general equality of the first two optimization problems is used. Their lower and upper bounds nevertheless match up to fixed-k constants.

The coefficients c here are in target-relative coordinates. The physical query includes the rotation $2\pi(\theta_j + e_j)(r - i)$ and therefore uses the installed spectrum. This model does not protect a query that is blind to a different unknown key spectrum.

D.2 Strict feasibility and a concrete failure witness

By definition, $R > 0$ if and only if $\sigma < \rho_\theta(L) = \min_{q \le L} u_q/q$. Below this threshold, the preceding upper bound is finite. At or above it, choose a minimizing q and nearest integers p_j to $q\theta_j$. Then

$$\widehat{\theta}_j = p_j/q, \qquad \left\| \widehat{\theta} - \theta \right\|_\infty = u_q/q \le \sigma.$$

The same installed frequency vector has period q in every block. The scores at $0, q, -q$ coincide for arbitrary (a, b), so $Z_{S_L} \ge 2$ and the target mass is at most $1/3$. This proves both the strict boundary and the impossibility of repairing it with amplitude.

For standard angular frequencies $(1, 0.1, 0.01, 0.001)$, $L = 8192$, the concrete choice $q = 6283$, $p = (1000, 100, 10, 1)$ gives a cycle-frequency perturbation below 4.694025×10^{-6}. It is an admissible coherent-box counterexample, not a claim that round-to-nearest hardware selects this particular perturbation.

One may safely upper-bound worst-case total leakage by the sum of separately worst terms, but the inequality can be strict:

$$\sup_e \sum_q f_q(e) \le \sum_q \sup_e f_q(e).$$

For example, in one block at $\theta = 1/3$, $\sigma = 1/12$, the cosine leakage terms at displacements one and two are maximized at different box endpoints. The certificates used here label the sumwise quantity as an upper bound, not an exact worst-case leakage formula.

D.3 The optimal critical radius and the obstruction to freezing

For any spectrum, $u_L \le 1/2$, so $\rho(L) \le 1/(2L)$. If a coordinate equals $1/(2L)$, then for $1 \le q \le L$ its distance is exactly $q/(2L)$. Thus $u_q/q \ge 1/(2L)$ and equality is attained. This is an elementary exact identity; its significance is whether it can coexist with low amplitude and all-length reuse.

For a frozen spectrum with an exact common period, $\rho(L) = 0$ for all sufficiently large L. Otherwise the simultaneous Dirichlet pigeonhole argument produces arbitrarily large q_ℓ with $u_{q_\ell} \to 0$; bounded q_ℓ along this construction would imply an exact return. Taking $L = q_\ell$ yields $L\rho(L) \le u_{q_\ell} \to 0$. Therefore no fixed finite spectrum, regardless of amplitude, can tolerate all boxes of radius c/L with a fixed $c > 0$ at every length. The error witnessing failure is allowed to differ between tasks; this is not a claim that one error stays inside every shrinking box.

For a badly approximable k-block spectrum, the lower bound $\rho(L) \ge \kappa L^{-1-1/k}$ is immediate. Conversely, set $Q = \lfloor L^{1/k} \rfloor$. Among $Q^k + 1$ phase points, the cube argument finds $1 \le q \le Q^k \le L$ with $u_q \le 1/Q$. Bad approximability forces $q \ge \kappa^k Q^k$, whence

$$\rho(L) \le u_q/q \le \kappa^{-k} Q^{-k-1} \le 2^{k+1} \kappa^{-k} L^{-1-1/k}.$$

The same coarse constant covers $L^{1/k} < 2$. This proves the claimed $\Theta_{\theta,k}$ scale. If $\sigma \le (\kappa/2)L^{-1-1/k}$, then $R \ge (\kappa/2)L^{-1/k}$, preserving optimal fixed-spectrum amplitude order.

D.4 A FIXED FAST SPECTRUM AND ONE MOVING CLOCK

Take $k \geq 2$, $\ell = k - 1$, and a fixed ℓ-dimensional spectrum η with $\max_j \|q\eta_j\|_{\mathbb{R}/\mathbb{Z}} \geq \kappa q^{-1/\ell}$. Let

$$\theta(L) = (\eta_1, \ldots, \eta_\ell, 1/(2L)), \qquad \sigma = \lambda/(2L), \quad 0 \leq \lambda < 1.$$

For a displacement q, write $x = q/L$, $a = (1-\lambda)/2 > 0$, $c = \lambda/2$. The worst possible joint phase distance M_q obeys

$$M_q \geq ax, \qquad M_q \geq \kappa q^{-1/\ell} - cx.$$

The first bound implies $cx \leq (c/a)M_q$; substitute into the second to obtain $M_q \geq 2a\kappa q^{-1/\ell}$. Thus

$$M_q \geq a \max\{q/L, 2\kappa q^{-1/\ell}\} \geq a(2\kappa)^{\ell/(\ell+1)} L^{-1/(\ell+1)}.$$

The final inequality balances an increasing and a decreasing function and holds for every positive real q, not merely its optimizer. Set $G = (2\kappa)^{\ell/(\ell+1)}$. Taking the minimum over integer q gives $R \geq (1-\lambda)GL^{-1/k}/2$. Equation (12) therefore supplies the explicit sufficient amplitude

$$A = \frac{k}{2(1-\lambda)^2 G^2} L^{2/k}(h + k\log 5 + \log 3). \tag{31}$$

It succeeds uniformly over the full box, and the clock coordinate yields critical radius exactly $1/(2L)$. A zero-error k-block two-sided task already needs $\Omega_k(L^{2/k}h)$, so the L, h dependence matches for fixed k, λ. The factor $(1-\lambda)^{-2}$ is a sufficient bound, not a claimed optimum when λ itself tends to one.

For an explicit fast spectrum use $\eta_j = \alpha^j$, $\alpha = 2^{1/(\ell+1)}$, $j = 1, \ldots, \ell$. Lemma 9's sharper intermediate Euclidean constant and $\|x\|_\infty \geq \|x\|_2 / \sqrt{\ell}$ give

$$\kappa = \frac{(3\ell+2)^{-1/\ell}}{2\sqrt{\ell+1}}.$$

For $k = 1$, the clock alone gives $R = (1-\lambda)/(2L)$ and $A = O_\lambda(L^2 h)$ directly. Finally, taking $k = m + 1$ and $L = n - 1$ in (31), then using (24), gives constant-factor competitiveness with the ideal m-block task-specific oracle for every address. A uniform key again attains the coefficient budget.

This proves all parts of Theorem 3. The minimality claim is only about the number of length-adaptive coordinates: none is insufficient in every finite dimension, one suffices with the stated dimension. Theorem 1's total-dimension lower bound has different quantifiers and is not transferred to the adaptive model.

D.5 A NON-EXACT CLOCK PRESERVES THE RESOURCE ORDERS

The extremal clock $1/(2L)$ is not a knife-edge requirement for order-optimal stability. The next corollary permits a range of clock coefficients and, in particular, a coefficient that varies with the design horizon.

Corollary 10 (Clock-coefficient tolerance)**.** *Let $k \geq 2$ and suppose the fixed $(k-1)$-block vector η obeys $\max_j \|q\eta_j\|_{\mathbb{R}/\mathbb{Z}} \geq \kappa q^{-1/(k-1)}$. Choose any $c_L \in (0,1)$ and set*

$$\theta(L) = (\eta, c_L/L), \qquad \gamma_L = \min\{c_L, 1 - c_L\}, \qquad \sigma = \lambda\gamma_L/L, \quad 0 \leq \lambda < 1.$$

Then $\rho_{\theta(L)}(L) \geq \gamma_L/L$ and

$$R_{\theta(L)}(L, \sigma) \geq (1-\lambda)\gamma_L^{1/k} \kappa^{(k-1)/k} L^{-1/k}. \tag{32}$$

Consequently a common-key, equal-amplitude query succeeds throughout that box with

$$A = \frac{kL^{2/k}}{8(1-\lambda)^2 \gamma_L^{2/k} \kappa^{2(k-1)/k}} (h + k\log 5 + \log 3). \tag{33}$$

In particular, for any fixed $0 < \gamma_0 \leq 1/2$, every choice $c_L \in [\gamma_0, 1 - \gamma_0]$ gives the same $L^{2/k}h$ amplitude order and $1/L$ uncertainty order, uniformly over L, at fixed k, λ.

Proof. For $1 \le q \le L$, put $x = q/L$. Since $0 < c_L x < 1$,

$$\|c_L x\|_{\mathbb{R}/\mathbb{Z}} = \min\{c_L x, 1 - c_L x\} \ge \gamma_L x.$$

Here the second branch obeys $1 - c_L x \ge (1 - c_L)x$ because $x \le 1$. Thus the clock alone implies $u_q/q \ge \gamma_L/L$. After a full frequency-box perturbation of radius σ, the joint worst phase separation $M_q = (u_q - q\sigma)_+$ satisfies

$$M_q \ge (1-\lambda)\gamma_L x, \qquad M_q \ge \kappa q^{-1/(k-1)} - \lambda\gamma_L x.$$

The first inequality gives $\lambda\gamma_L x \le \lambda M_q/(1-\lambda)$; substituting in the second yields $M_q \ge (1-\lambda)\kappa q^{-1/(k-1)}$. Hence

$$M_q \ge (1-\lambda)\max\{\gamma_L q/L, \kappa q^{-1/(k-1)}\} \ge (1-\lambda)\gamma_L^{1/k}\kappa^{(k-1)/k}L^{-1/k}.$$

The last inequality balances an increasing and a decreasing function of a positive real q and therefore holds for every allowed integer q. Taking the minimum gives (32); the shell bound gives (33). For $c_L = 1/2$ the constants reduce exactly to (31), and the critical radius is the extremal $1/(2L)$. For other c_L the displayed lower bound need not be the exact full-spectrum critical radius. The argument uses a single coherent error vector at all positions, as in Theorem 3. □

This is a tolerance corollary of the existing robust-return bound, not a claim that training produces a particular clock or that learned fast bands satisfy bad approximability. It also does not identify an optimal joint limit as $\lambda \uparrow 1$ or $\gamma_0 \downarrow 0$.

D.6 Finite calibration without worst-case packing constants

The clock-coefficient experiment compares sufficient-envelope budgets, whereas the closed-form bound uses a uniform Diophantine constant and shell counting. Neither quantity was previously a tight certificate for the true finite-window robust optimum. We now provide such a bridge with an explicit, dimension-independent approximation factor. The maximization over a coherent error is not exchanged with summation. Conservative reformulations of robust sums are a classical issue (Gorissen & den Hertog, 2013; Hsiung et al., 2008); the elementary reductions below exploit the specific nonnegative rotary gap to quantify their cost here.

One fixed relative query, not an adaptive oracle. Fix S_L, a spectrum θ, $h > 0$, and a common key with $w_j > 0$, $\sum_j w_j = 1$. Let $\mathcal{E} = \{e : \|e\|_\infty \le \sigma\}$, and require one relative coefficient vector to work for every $e \in \mathcal{E}$; physical address phases still use the installed spectrum. Define

$$\mathsf{B}_w = \inf_{a,b}\{N_w(a,b) :\ \forall e \in \mathcal{E},\ Z_{S_L,e}(a,b) \le e^{-h}\}.$$

This is not the adaptive quantity $\sup_e \inf_{a,b} N_w(a,b)$. Evenization and removal of negative cosine coefficients are valid for every e simultaneously. Therefore $b = 0, a \ge 0$ attain the same infimum for B_w. Write

$$\ell_{qj} = (\|q\theta_j\|_{\mathbb{R}/\mathbb{Z}} - q\sigma)_+, \qquad G_{qj} = 1 - \cos(2\pi\ell_{qj}), \qquad H_0 = h + \log 2, \tag{34}$$

$$\widehat{\mathsf{B}}_w = \min_{a \ge 0}\{N_w(a,0) :\ 2\sum_{q=1}^{L} e^{-(Ga)_q} \le e^{-h}\}, \tag{35}$$

$$\mathsf{M}_w = \min_{a \ge 0}\{N_w(a,0) :\ Ga \ge \mathbf{1}\}. \tag{36}$$

The envelope program is finite and convex. If some row of G vanishes, a common return makes all three costs infinite. We assume $R_\theta(L,\sigma) > 0$ below.

Proposition 11 (A certified finite-window bridge)**.** *For every such spectrum, key and uncertainty box,*

$$H_0\mathsf{M}_w \ \le\ \mathsf{B}_w \ \le\ \widehat{\mathsf{B}}_w \ \le\ (H_0 + \log L)\mathsf{M}_w. \tag{37}$$

In particular the envelope allocation is a feasible common-query robust design within factor $1 + \log L/(h + \log 2)$ *of the optimal common-query design, independently of the number of blocks, their arithmetic properties, and the distance to the feasibility threshold. The factor concerns this shared-coefficient model and not an installed-frequency-dependent query oracle.*

Proof. For each fixed q, the coordinates of e can be moved independently toward nearest integers. For nonnegative a, this attains

$$\min_{e\in\mathcal{E}} \sum_j a_j[1-\cos(2\pi q(\theta_j+e_j))] = (Ga)_q.$$

A successful common a must suppress the pair $\pm q$ at this particular error, so $(Ga)_q \geq H_0$ for every q. The minimizing errors may differ between q; these are necessary constraints on the *same* query because its guarantee is universal in e. Homogeneity gives $N_w(a,0) \geq H_0\mathsf{M}_w$. The envelope upper-bounds the leakage of every single coherent error, yielding the middle inequality. Finally, scale a margin-optimal allocation by $H_0+\log L$. Each of the L pairs contributes at most $2e^{-H_0-\log L}$, whose sum is e^{-h}. Arbitrarily close feasible allocations handle an unattained infimum. No sup–sum equality is invoked. □

Checkable lower bounds and allocation duality. For $p \in \Delta_L$ put $\mu = G^T p$ and $D_w(p) = (\sum_j w_j \mu_j^2)^{1/2}$. Every such distribution gives the robust lower bound $\mathsf{B}_w \geq H_0/D_w(p)$ by averaging the necessary pair inequalities and applying the weighted Cauchy–Schwarz inequality. Moreover,

$$\widehat{\mathsf{B}}_w = \sup_{p\in\Delta_L} \frac{H_0+H(p)}{D_w(p)}, \qquad \mathsf{M}_w = \frac{1}{\min_{p\in\Delta_L} D_w(p)}. \tag{38}$$

The first identity is the entropy variational argument of Appendix E with C replaced by G; the second uses the same compact convex minimax argument for the minimum gap. Thus any feasible allocation and any exact probability vector form an independently checkable upper/lower certificate. At an exact envelope primal–dual optimum the a posteriori ratio is at most $1+H(p)/(h+\log 2)$, often smaller than the worst-case $\log L$ factor. A numerical solver's stopping status does not certify these inequalities: the actual nonnegative rational witnesses, probability normalization, sums and norm bounds must be checked separately.

Selection among finitely many clock coefficients. Apply the same construction to each member of a specified finite clock set, with identical L, h, σ, w. Choosing the allocation with the smallest certified envelope cost preserves the same approximation factor relative to the best common-query robust design in that finite set. This is a finite-window design rule, not a claim of optimality over all real clocks or all spectra. If verified bounds satisfy $L_c \leq \widehat{\mathsf{B}}_{w,c} \leq U_c$ with $U_c/L_c \leq 1+\epsilon$ for every candidate, the query selected by $\min_c U_c$ costs at most $(1+\epsilon)(1+\log L/(h+\log 2))\min_c \mathsf{B}_{w,c}$. This follows by applying the individual bound to a candidate minimizing the true robust cost. Without such envelope gaps, selecting the smallest upper endpoint still gives a feasible design, but the approximation factor must use the reported certificate gaps rather than assume exact optimization.

D.7 Sharper certificates for the equal-amplitude family

For the uniform key and coefficients $a_j = A/k$, let

$$g_q = \frac{1}{k}\sum_j G_{qj}, \quad g_* = \min_q g_q > 0, \quad U(A) = 2\sum_q e^{-Ag_q}.$$

Let A_{env} be the unique root $U(A) = e^{-h}$ and $A_{\mathrm{coh}}^{\mathrm{eq}}$ the smallest A for which the equal-amplitude query works for every coherent error. One pair, whose own minimizing error is a legal coherent vector, gives

$$\frac{H_0}{g_*} \;\leq\; A_{\mathrm{coh}}^{\mathrm{eq}} \;\leq\; A_{\mathrm{env}}, \qquad \frac{A_{\mathrm{env}}}{H_0/g_*} = 1+\frac{\log N_{\mathrm{eff}}}{H_0}, \quad N_{\mathrm{eff}} = \sum_q e^{-A_{\mathrm{env}}(g_q-g_*)}. \tag{39}$$

Indeed, factoring $e^{-A_{\mathrm{env}}g_*}$ out of the defining sum gives the identity. Here $1 \leq N_{\mathrm{eff}} \leq L$, so this refines the general factor. For an independent lower certificate it is enough to provide *any* one integer q and an upper enclosure of its g_q; verifying that it is a global minimizer is unnecessary. An actual single-error leakage calculation can only strengthen this lower bound. These are certificates for a fixed amplitude ray, not lower bounds against arbitrary reallocated queries.

D.8 WHY THE MAXIMAL-RADIUS CLOCK NEED NOT MINIMIZE AMPLITUDE

Fix $\sigma = \beta/L$ and a fixed fast spectrum, so the physical uncertainty does not change with the clock coefficient. For $\beta < c \le 1/2$, the clock's minimum phase distance at $x = q/L$ is $(c-\beta)x$, attained at the *same* clock error $e_k = -\beta/L$ for every q. To verify this, for $u \in [-\beta, \beta]$ either $(c+u)x \le 1/2$ or

$$1 - (c+u)x \ge 1 - (c+\beta)x \ge (c-\beta)x,$$

where the last inequality is $2cx \le 1$. Thus the lower endpoint minimizes every clock gap simultaneously. For the equal-amplitude query, increasing c on $(\beta, 1/2]$ strictly improves all these gaps, so it strictly decreases both the true coherent robust cost and the envelope cost. With allocation optimized, the cost is nonincreasing; it can be constant if the clock is unused.

Above $1/2$ the trade-off changes. At differentiable points set $x_q = q/L$, $\ell_q = (\|cx_q\| - \beta x_q)_+$, and $\pi_q = e^{-A_{\mathrm{env}} g_q} / \sum_t e^{-A_{\mathrm{env}} g_t}$. The implicit derivative of the finite envelope is

$$\frac{dA_{\mathrm{env}}}{dc} = -A_{\mathrm{env}} \frac{\sum_q \pi_q \, \partial_c g_q}{\sum_q \pi_q g_q}, \tag{40}$$

$$\partial_c g_q = \frac{2\pi}{k} x_q \sin(2\pi \ell_q) \operatorname{sgn}(1/2 - cx_q) \, \mathbf{1}_{\{\ell_q > 0\}}. \tag{41}$$

Use one-sided derivatives at wrap or clipping points. The derivative follows by differentiating $\log U(A, c) = -h$; its denominator is positive. Short displacements $q < L/(2c)$ favor a faster clock, while wrapped longer displacements oppose it. Their *leakage weights*, not their unweighted counts, determine the balance. This diagnoses finite-window costs and does not identify a universally best clock. Maximizing the critical uncertainty radius and minimizing amplitude at a fixed interior uncertainty budget are different objectives.

E CONVEX RESOURCE CERTIFICATES AND PHASE INTERVENTIONS

The convex tools in this section are standard (Boyd & Vandenberghe, 2004). We give the derivations to specify exactly what each numerical certificate verifies.

E.1 SYMMETRIC FIXED-KEY OPTIMIZATION

For S_L, write $C_{qj} = 1 - \cos(2\pi q \theta_j)$ and let w be the fixed key's block-energy fractions. The exact signed-score optimum is

$$B_w^*(L, h) = \min_{a \ge 0} \left\{ N_w(a, 0) : 2 \sum_{q=1}^{L} e^{-(Ca)_q} \le e^{-h} \right\}. \tag{42}$$

It has the exact certificate dual

$$B_w^*(L, h) = \sup_{p \in \Delta_L} \frac{h + \log 2 + H(p)}{\sqrt{\sum_j w_j [(C^T p)_j]^2}}. \tag{43}$$

We derive both statements rather than relying on an optimizer's termination status.

For S_L, evenization replaces (a, b) by $(a, 0)$ without increasing N_w or leakage. Setting each negative a_j to zero also decreases both, because $\cos(2\pi q \theta_j) - 1 \le 0$. Thus (42) is exactly the original signed-coefficient optimum, not a heuristic restricted class.

Write Δ_L for the probability simplex over positive displacements and $H(p) = -\sum p_q \log p_q$, with $0 \log 0 = 0$. For fixed a,

$$\log \left(2 \sum_q e^{-(Ca)_q} \right) = \max_{p \in \Delta_L} \{ \log 2 + H(p) - p^T C a \}.$$

At a fixed budget B, the nonnegative ellipsoid $\{a \ge 0 : N_w(a, 0) \le B\}$ and simplex are compact convex sets. The expression is affine in a and continuous concave in p, so the minimax theorem

applies. Let $\mu = C^T p \geq 0$. The ellipsoid support function is

$$\max_{a \geq 0 : N_w(a,0) \leq B} \mu^T a = B \sqrt{\sum_j w_j \mu_j^2}.$$

Therefore the smallest log-leakage attainable at budget B equals

$$\max_{p \in \Delta_L} \left[\log 2 + H(p) - B \sqrt{\sum_j w_j \mu_j^2} \right].$$

Requiring this to be at most $-h$ is equivalent to (43). If the denominator is zero, the positive numerator gives infinity, corresponding to an unseparable competitor distribution. Infima and suprema are interpreted in the extended real sense.

If the cost were instead $\sum a_j$, the support function would be $B \max_j \mu_j$ and the denominator would be $\max_j (C^T p)_j$. These two duals explain why the coefficient-optimal and uniform-key-optimal numerical budgets cannot be exchanged.

E.2 One shared profile for all addresses

For the full mirror family $\mathcal{C}_{n,r}$, define $\rho_j = \sqrt{a_j^2 + b_j^2}$ and $s_+(t) = \sum_j \rho_j \cos(2\pi\theta_j t)$. For each displacement,

$$\begin{aligned} e^{s_+(t) - s_+(0)} &\leq e^{\sum_j a_j (\cos(2\pi\theta_j t) - 1)} \\ &= \sqrt{e^{s(t) - s(0)} e^{s(-t) - s(0)}} \\ &\leq \tfrac{1}{2}[e^{s(t) - s(0)} + e^{s(-t) - s(0)}]. \end{aligned}$$

The first inequality uses $\rho_j \geq a_j$ and the nonpositive cosine gaps. Summing over any $\mathcal{C}$ yields

$$Z_{\mathcal{C}}(s_+) \leq \tfrac{1}{2}[Z_{\mathcal{C}}(s) + Z_{-\mathcal{C}}(s)].$$

For zero-based addresses, $-\mathcal{C}_{n,r} = \mathcal{C}_{n,n-1-r}$. Taking the maximum gives

$$\max_r Z_{\mathcal{C}_{n,r}}(s_+) \leq \max_r Z_{\mathcal{C}_{n,r}}(s). \tag{44}$$

Phase alignment preserves the original key and every query block norm, hence both A_{coeff} and N_w. It can improve one mirror task at the expense of the other; the guarantee is for their shared worst case. If some active block avoids exact return at every relevant nonzero displacement, all gaps of s_+ are positive. Lemma 4 then establishes finite-amplitude feasibility, not a promise that the original budget suffices.

Consequently, for this all-address task family, optimizing a nonnegative cosine allocation under the constraints

$$\sum_{t \in \mathcal{C}_{n,r}} e^{-C_{|t|} a} \leq e^{-h} \quad \text{for every } r \tag{45}$$

is also exactly the shared signed-score optimum. This is the program used for Table 1; it is not the symmetric problem at $L = n - 1$ and is not a separate optimizer per target.

Here is its dual, useful when the worst address is not known in advance. Let $\pi(r, t)$ be any probability distribution supported on $t \in \mathcal{C}_{n,r}$, with marginal λ_r and conditional $p(t \mid r)$. Define

$$\mathsf{H}(\pi) = H(T \mid R) = \sum_r \lambda_r H(p(\cdot \mid r)), \qquad \mu_j(\pi) = \sum_{r,t} \pi(r, t) C_{|t|,j}.$$

Then the exact all-address fixed-key optimum is

$$B^*_{w,n}(h) = \sup_\pi \frac{h + \mathsf{H}(\pi)}{\sqrt{\sum_j w_j \mu_j(\pi)^2}}. \tag{46}$$

Indeed, maximizing the addresswise log-sum-exp is equivalent to maximizing its mixture over λ. Combining λ with the conditional distributions yields $\mathsf{H}(\pi) - \mu^T a$, a continuous concave function of the joint distribution. Apply the same compact convex minimax argument and ellipsoid support function as above. There is no marginal entropy $H(R)$: mixing addresses does not add an extra softmax over addresses. The $\log 2$ term appears only when positive and negative displacements have already been compressed into symmetric pairs.

E.3 Interpretation of the certificate types

An explicit feasible a supplies a resource upper bound. A valid p or π supplies a lower bound. A float64 primal–dual gap is a numerical diagnostic and is not automatically an interval certificate. The included standard certificates store coefficients and probabilities as exact dyadic rationals, then use outward interval evaluation of frequencies, trigonometric functions, exponentials, logarithms, and norms. The verdict does not rely on the proposing optimizer reporting success. These certificates rely on the interval library and verifier; they are not a proof-assistant development or a certification of GPU multiplication and softmax.

For coherent frequency boxes, a conservative independently checkable certificate can use

$$\ell_{qj} = \left(\|q\theta_j\|_{\mathbb{R}/\mathbb{Z}} - q\sigma \right)_+, \qquad g_q = \frac{8A}{k} \sum_j \ell_{qj}^2.$$

For every installed spectrum in the box, the aligned gap at $\pm q$ is at least g_q. At rational frequency centres and rational σ, each ℓ_{qj} is exactly calculable by modular integer arithmetic. For integer target H, if all $g_q - H \geq 0$, then

$$e^{-(g_q - H)} \leq 2^{-\lfloor g_q - H \rfloor}.$$

An exact upper bound $2 \sum_q 2^{-\lfloor g_q - H \rfloor} \leq 1$ certifies $Z \leq e^{-H}$ for the entire box. Truncating exceptionally small terms upwards keeps the certificate conservative. Failing this sufficient test does not prove infeasibility. The archive includes eighty such inherited certificates and a verifier with an intentionally corrupted negative control.

F Experimental specification, aggregate results, and evidence scope

F.1 Schedules and model permissions

All learning experiments use $k = 8$, $d_{\mathrm{head}} = 16$ and the same supplied target-address interface. Standard RoPE has $\omega_j = 10000^{-j/8}$, $j = 0, \ldots, 7$. The fixed algebraic schedule uses $\theta_j = 2^{j/9}$, $j = 1, \ldots, 8$, interpreted modulo one. The algebraic-plus-clock schedule uses the seven-dimensional algebraic vector $\theta_j = 2^{j/8}$, $j = 1, \ldots, 7$, and $1/[2(n-1)]$ as its last frequency. The standard-plus-clock schedule retains the first seven standard angular frequencies and replaces its final block by this clock (converted to angular units). The latter is an engineering control; the bad-approximability guarantee for algebraic fast blocks is not asserted for standard fast blocks.

There are n payload keys at $0, \ldots, n-1$ and an additional query at position n excluded from the key set. All unrotated keys equal a common K_0; payload enters only V. The query base vector is rotated using the supplied address and current installed spectrum. There is no target value in the query residual, target marker in individual keys, or learned table of address-specific queries. Free-direction training normalizes $\|K_0\| = 1$, $\|Q_0\| = 64\sqrt{d_{\mathrm{head}}}$ at each forward pass. Temperature is one and any scalar gain counts toward the amplitude.

The original protocol samples lengths $32, 64, 128, 256$ in uniform buckets. Within each bucket, its address mixture is one-third boundaries, one-third central-half positions, and one-third uniform addresses. Five model seeds are $11, 23, 37, 53, 71$. AdamW uses weight decay zero and 1024 updates; a separate pilot selected the common learning rate 0.003 from $0.0003, 0.001, 0.003$. Each run processed 2^{22} payload tokens. The main training loss is $-\log \alpha_r$; payloads do not affect that loss and are used for value-readout checks. These token counts are not independent semantic training examples.

The prescribed test lengths are $32, 64, 128, 256, 512, 1024, 2048, 4096$. Reliability means $Z \leq 10^{-2}$ at every address, equivalently $\alpha_r \geq 100/101$, rather than merely correct top-1 selection. Scores depending only on displacement allow complete address enumeration without stochastic payload sampling as a substitute. The main text reports the available final-length aggregates, not invented intermediate curves.

F.2 Interventions and aligned training are different experiments

The phase intervention rotates each original query block to target alignment, preserving its norm and the actual original key. All twenty original profiles have reported strictly negative gap witnesses at length 4096. All twenty become finite-amplitude feasible after phase correction, and thirteen pass with their unchanged budget of 64. The fixed-key optimization intervention then changes query allocation at the evaluation length while retaining each original key and spectrum; all twenty pass at the original budget.

The report describes forty-eight optimized problems over $n = 256, 4096$, including the checkpoint-specific and common-key comparisons, with maximum float primal–dual relative gap 1.24×10^{-9}. These are numerical optimization results. They are not described as interval certificates and are not retrospectively labelled successful training runs.

Aligned training instead starts from the original seeds' block energies and uses separate softmax parameters for query and key energies, with phases aligned from the beginning. It uses the same per-run address manifest, 1024 updates, and unchanged learning rate; it is a different parameterization, not a continuation of a successful unconstrained checkpoint. Its available aggregates are:

Table 3: Reported aligned-training aggregates; all costs use actual QK norm. Ranges describe the five reported calibrated costs, not confidence intervals.

Spectrum	**Pass,** $n = 256$	**Pass,** $n = 4096$	A_{req} **at** 4096
Standard	4/5	4/5	50.532–77.893
Algebraic	5/5	5/5	34.184–51.293
Algebraic + clock	5/5	5/5	39.946–58.269
Standard + clock	5/5	3/5	49.199–81.026

For a fixed profile and fixed spectrum, worst-address leakage is nondecreasing in n: every earlier competitor set is included in the set for the same target at the next length. Hence, in exact reference arithmetic, equal counts $4/5$ at 256 and 4096 for the standard schedule imply equal success sets. The aggregate GPU counts alone do not identify a new out-of-distribution failure. A length-dependent clock changes the profile and is outside this monotonicity statement.

A useful but not reported per-model quantity is the optimization gap $G_s(n) = A_{\mathrm{req},s}(n)/B^*_{w_s,n}(h)$ using the *same learned key* in numerator and denominator. The uniform-key optima in Table 1 do not supply this denominator for a nonuniform learned key. Nor can the original free-direction checkpoint's key optimum be substituted for a separately trained aligned model's optimum. We do not infer relative learning efficiency from unavailable matched gaps.

F.3 Finite standard-spectrum certificates

The original standard-spectrum certificate family uses coefficient cost, with its own per-task key balancing. The six included certified intervals are

Table 4: Standard angular frequencies $10000^{-j/k}$, symmetric task S_L, $h = \log 10^6$. These intervals concern A^*_{coeff}, not a fixed uniform-key optimum.

Head width d_{head}	**Blocks** k	**Per-side** L	**Certified coefficient optimum**
8	4	1024	[267.6141, 267.6144]
8	4	8192	[1068.7062, 1068.7074]
32	16	1024	[58.6052, 58.6055]
32	16	8192	[65.5193, 65.5196]
128	64	1024	[49.0421, 49.0423]
128	64	8192	[53.2618, 53.2621]

At $d_{\mathrm{head}} = 128, L = 8192$, mapping the coefficient-optimal vector to a uniform key gives cost about 212.6786; it is only an upper bound for the new objective. Direct uniform-key optimization yields the included interval $[163.8370, 163.8373]$. The later experimental summary reports $[163.8370, 163.8371]$, but its tightened witness is not in the present archive, so the main numerical certificate claim uses the wider interval. The equal-weight cost is about 487.0934. The approximate ninefold ratio $487.09/53.262$ changes allocation permissions, not frequencies, and is not a common-uniform-key spectral gain.

Additional comparisons for Table 1. At equal allocation, the algebraic schedule's 27.549 budget is 66.54% below the standard schedule's 82.331, and 37.56% below its optimized 44.117. Allocation optimization reduces the algebraic budget by 2.53% and the standard budget by 46.42%. These are fixed-task comparisons under the same uniform key, not a general learned-model ranking.

For the $k = 8$ standard equal-weight profile, the unit-amplitude nearest-neighbour gap is

$$c_1 = \frac{1}{8}\sum_{j=0}^{7}[1 - \cos(10000^{-j/8})] \simeq 0.064354219564.$$

Two adjacent competitors require $A \geq (\log 100 + \log 2)/c_1 \simeq 82.33054$, nearly the full 4096-length requirement 82.33138. One adjacent competitor alone requires $A \geq \log 100/c_1 \simeq 71.55972$. Thus failure at $A = 64$ for this equal-weight baseline is already local and cannot be used as evidence that distant periodic aliases first appear at length 4096.

F.4 Earlier finite construction and frequency-box checks

The initial experimental summary reports one-, two-, and four-block construction tests on lengths $33, 65, \ldots, 65537$. All addresses were checked up to 1025; only specified boundary, near-boundary, and interior subsets were checked at larger lengths. At $n = 65537$, $h = \log 100$, the single-block explicit boundary and centre costs are reported as 544618 and 3.51126×10^8. Their approximately 644.72 ratio is a *construction* ratio, not the exact ratio of oracle optima. We do not reconstruct length curves from these two aggregate values or fit a theorem's exponent to them.

The original frequency-box report covers $k \in \{1, 2, 4, 8\}$ and maximum displacements $64, 256, 1024, 4096, 16384, 65536$. It reports full scans of the finite return radius, intervals for the clock's $L\rho$ containing $1/2$, common-rational failure witnesses, and thirty-six additional integer box certificates. Those additional remote certificate files are not in the current archive. The eighty earlier inherited certificates *are* included and independently checkable. A fixed-clock control is reported, but no missing raw curve is replaced by a theoretical prediction masquerading as measurement.

F.5 Evidence map and replication status

The standalone archive separates three evidence types: mathematical derivations, reproducible finite computations, and reported training or arithmetic aggregates. The inventory below describes files included in this release, not artifacts referenced in an external working directory. The root `README.md` maps each item to its paths and verification command.

The verification driver checks 93 certificate cases, five deliberately reduced-budget negative controls, and fifteen small resource-contract tests. Its full mode also recomputes the four uniform-key budgets and all 450 floating clock configurations. The certificate verifiers check the supplied witnesses independently of the proposing optimizers; they are not proof-assistant formalizations or GPU-pipeline certificates.

Unavailable source artifacts. Original aggregate report documents, remote training executors, original and aligned checkpoints, per-seed records, GPU readout logs, the later tightened uniform-key witness, and the additional remote clock-box files are not included. The reported training and arithmetic values are retained from the manuscript and its transcribed tables, not re-verified against those absent source artifacts. Complete training replication requires those materials; no missing seed values, trajectories, uncertainty estimates, or metric definitions are reconstructed. The main standard uniform-key certificate uses the included wider interval $[163.8370, 163.8373]$.

Result	Included evidence	Verification scope
Table 1	Aggregate CSV, CPU optimizer, and numerical outputs	Four floating budget recomputations; not interval endpoints.
Tables 2–3	Aggregate counts and cost ranges in the manuscript and CSV	Reported results; no checkpoint-level replication.
Standard and uniform-key costs	Exact rational witnesses and independent interval verifiers	Six coefficient-cost cases and one uniform-key case.
Coherent error boxes	Rational witnesses and a modular-integer verifier	Eighty included whole-box certificates.
Clock calibration	450-row reconstructed grid; exact witnesses and interval verifiers	Floating grid diagnostics and six robust-cost certificates.
Table 7	Transcribed staged-arithmetic CSV	Reported discrepancies; reduction formula unavailable.

Interpretation of the training evidence. Five seeds per schedule support a controlled observation, not a universal training comparison. Address enumeration gives deterministic coverage conditional on a checkpoint, not additional independent training replicates. The two algebraic schedules remain separate recipes. The original protocol's stronger empirical-advantage criterion requires matched amplitude improvements at two adjacent extrapolation lengths; the available aggregates do not verify it. Consequently, the training results are mechanism diagnostics, while the resource guarantees rest on the proofs and independently checkable finite certificates.

F.6 Clock coefficients: reported grid and independent finite certificates

The updated experimental manuscript reports a predeclared grid $k \in \{2,4,8\}$, $L \in \{256, 1024, 4096, 16384, 65536\}$, $c \in \{0.25, 0.40, 0.50, 0.60, 0.75\}$. The fast cycles are $2^{j/k} \bmod 1$, $j = 1, \dots, k-1$, and the clock is c/L. For each of these 75 spectra it uses $\sigma = 0.5\min(c, 1-c)/L$, $0.9\min(c, 1-c)/L$, and $0.1/L$, at $h = \log 100$ and $\log 10^6$, totaling 450 records. The candidate set is S_L, containing $2L+1$ positions including the target. The key is uniform and the original query allocation is equal in all blocks.

What the original calibration measures. It solves the sufficient-envelope equation

$$2\sum_{q=1}^{L} \exp\left[-\frac{A}{k}\sum_{j=1}^{k}(1-\cos(2\pi\ell_{qj}))\right] = e^{-h}, \qquad \ell_{qj} = (\|q\theta_j\| - q\sigma)_+.$$

Different summands can require different worst errors. Thus even interval verification of this equation alone does not identify the exact coherent-error optimum. The original report's small bisection interval is precision for this *envelope*, not a certificate of its approximation error. Equation (39) now supplies the missing necessary-cost comparison.

At common physical uncertainty $\sigma = 0.1/L$, the original reported comparisons with $c = 0.5$ are as follows. Every row has 30 matched (k, L, h) settings.

c	Lower envelope cost	Higher envelope cost	Ratio range
0.25	0	30	1.340–5.155
0.40	0	30	1.103–1.695
0.50	reference		1
0.60	30	0	0.688–0.937
0.75	26	4	0.434–1.297

The reported lower-return/theoretical-bound ratio is at least 2.3749328835; the analytic safe amplitude divided by the calibrated envelope amplitude ranges from 17.9085 to 8155.6836. These figures do

not contradict an order theorem, but they preclude treating its explicit universal constants as tight finite-cost predictions. The two normalized-uncertainty comparisons change the physical box with c and are not pooled with this table.

Independent recomputation, not a rerun of remote code. The finite-calibration supplement reconstructs all 450 cases directly from these formulas, using 85-digit algebraic frequencies, extended-precision phase reduction and float64 scalar calibration. It reproduces the displayed ranges. The pair lower bound and envelope upper bound have ratios between 1 and 1.562388 over this grid. Those 450 new ratios are floating diagnostics, not 450 new interval certificates. The right envelope derivative at $c = 0.5$ is negative in all 30 common-physical settings. Appendix D.8 explains the role of weighted unwrapped and wrapped competitors; no universal preference for 0.60 or 0.75 is inferred.

Six new independently checked certificates. At $k = 8, L = 65536, h = \log 10^6, \sigma = 1/(10L)$ we verify three equal-allocation certificates and three allocation-optimized certificates. The reference verifier uses 40-digit outward mpmath interval arithmetic. The phase centers are the symbolic algebraic numbers, not float values silently substituted as exact. Every positive displacement is included, and the factor two covers its negative partner. Both verifier families reject a deliberately halved budget. They do not certify a GPU arithmetic path or constitute proof-assistant formalization.

For the equal-amplitude family, a single integer displacement supplies the lower bound in (39); the full envelope at a rational budget supplies the upper bound. All displayed endpoints are rounded outwards.

Table 5: True coherent-error minimum within the equal-block amplitude ray. All rows use the same physical uncertainty, uniform key and reliable readout criterion. These are not all-query optimal costs.

c	Certified lower	Certified upper	Lower-bound witness q
0.50	283.024	283.279	20031
0.60	208.778	211.605	20031
0.75	178.261	184.823	5946

Thus the original sufficient budgets are within respectively 1.001, 1.014 and 1.037 of the true robust costs in this family, despite the much larger analytic safety factors. The first two rows share a dangerous displacement; increasing the clock improves its short-displacement gap without crossing the clock wrap point. Other returns eventually limit the improvement.

The second certificate family solves (35), converts allocations and dual distributions to exact dyadic rationals, and verifies a feasible envelope upper bound together with the necessary-cost lower bound $H_0/D_w(p)$. Its lower bound permits *every* signed relative query that works for all coherent errors; its upper bound is a particular common query. The same key is fixed at $w_j = 1/8$ in both bounds.

Table 6: True robust common-profile cost B_w with allocation optimized. Unlike Table 5, these lower bounds cover arbitrary signed query coefficients. Each row uses one relative query for the full frequency box.

c	Certified lower	Certified upper	Upper/lower, at most
0.50	168.928	189.526	1.122
0.60	149.036	168.075	1.128
0.75	141.528	161.050	1.139

The upper bounds for $c = 0.60$ and 0.75 are below the lower bound for 0.50, so the improvement over a half-clock is not merely a ranking of conservative upper bounds, even after query allocation is optimized. The intervals for 0.60 and 0.75 overlap; these certificates do *not* order their exact optima. The envelope optimization itself has certified primal–dual ratios below 1.000001; this small numerical gap must not be confused with the larger gap to B_w.

Availability and limitations. The diagnostic CSV, exact witnesses and independent verifiers are included under `finite_calibration/`. The release checks rerun all six interval certificates, all 450 independently reconstructed floating configurations, the six standard coefficient certificates, the uniform-key certificate, and the eighty inherited integer-box certificates. Five intentionally reduced-budget negative controls are rejected. Machine-readable results record exactly these computations. The remote clock-grid interval implementation and training runs are not represented as rerun artifacts; no new GPU or training claim is made by these CPU checks.

G FINITE ARITHMETIC, ABSOLUTE POSITIONS, AND IMPLEMENTATION LIMITS

G.1 PRECISION REQUIRED BY THE CONSTRUCTIONS

The algebraic frequencies have finite symbolic descriptions. Computing the same algebraic number more accurately as length grows does not change its mathematical value, but it does increase computational precision. It is not a permanent fixed-float representation.

If each sine/cosine used in the absolute query and key rotations has error at most $u \le 1/4$, each 2×2 rotation-matrix error has operator norm at most $2u$. For one block,

$$|Q_j^T \widehat{R}_i^T \widehat{R}_\ell K_j - Q_j^T R_i^T R_\ell K_j| \le (4u + 4u^2)\,\|Q_j\|\,\|K_j\| \le 5u\,\|Q_j\|\,\|K_j\|\,.$$

Summing and applying the attention scale bounds the logit error by $5Au$ for the constructed channel. Thus $u \le \nu_{\mathrm{rot}}/(5A)$ is sufficient. Coefficient storage, multiplication, summation, softmax, and value errors need their own budgets; the formula does not certify the whole hardware pipeline.

A sufficient algebraic-root precision for maximum absolute address N_{abs} is

$$B = \log_2 N_{\mathrm{abs}} + \log_2(A+1) + \log_2(1 + 1/\nu_{\mathrm{rot}}) + O(\log(k+1)).$$

For the universal construction on a prefix of length n, this is $(1 + 2/k)\log_2 n + O(\log(k+1) + \log(h+k) + \log(1 + 1/\nu_{\mathrm{rot}}))$. It is an implementation upper bound, not a best possible bit lower bound. Large absolute offsets must not be silently replaced by the relative window size.

Similarly, interpreting a parameter uncertainty box as $\sigma = 2^{-b}$ yields leading precision $(1 + 1/k)\log_2 L$ for a fixed badly approximable spectrum versus $\log_2 L$ for the adaptive construction, at fixed constants and target critical fraction. This comparison requires correctness for *every* frequency in the box. It is not a lower bound for one known deterministic rounding rule.

G.2 THE STAGED ARITHMETIC RESULT

The supplemental experiment keeps the installed FP32 frequency table identical across five arithmetic paths in sixteen spectrum–length–offset configurations. The following entries are transcribed from its report for maximum displacement $L = 4096$ and absolute offset 2^{24}.

Table 7: Reported maximum discrepancy from an FP64 relative-score reference using the same installed frequencies. Columns isolate large-phase storage from phase generation and reduction. These are reported scalar discrepancies, not norm-budget or frequency-box bounds.

	Original	FP64 computation before casting	
Spectrum	**FP32 path**	**Large phase to FP32**	**Reduced phase to FP32**
Standard	26.19690	20.33255	1.80×10^{-5}
Algebraic	120.23590	95.61398	1.60×10^{-5}
Algebraic + clock	126.34136	93.16175	1.24×10^{-5}
Standard + clock	26.19937	20.33389	1.61×10^{-5}

The report describes the metric as a maximum relative-logit discrepancy relative to the FP64 relative-score reference, but does not provide the reduction formula in the supplied summary. The numbers are therefore retained as reported discrepancies, not reinterpreted as dimensionless relative error.

The raw metric definition is an explicit replication item. The original FP32 address-cast path is also reported to create 4096 adjacent collisions in this configuration; that field is repeated across stages for comparison and is not a newly measured collision count for each improved path.

The experiment separates frequency installation from subsequent arithmetic. It does not assert that full FP64 phase computation is the unique or cheapest remedy. Integer range reduction or a different kernel might also preserve the needed information. Nor does it attribute the reduction to a particular spectrum: every schedule benefits. Such distinctions are essential when relating a parameter-box theorem to concrete floating-point RoPE failures (Wang et al., 2024a).

The reported remote environment identifies its device as an RTX 4090 with 49140 MiB frame-buffer memory and no virtualization mode. The drafting archive does not establish the hardware's modification history. The original report lists PyTorch 2.12.1+cu130. Neither device identity nor training-loop timing is used as a performance-comparison claim here.

G.3 A CHANGING CLOCK HAS IMPLEMENTATION COST

For growing prefixes, choose the design horizon H to be the smallest power of two at least $\max\{1, n-1\}$ and use clock $1/(2H)$. Since $n-1 \le H < 2(n-1)$, the asymptotic guarantees change only by fixed factors, and rephasing is needed only when H doubles. In exact arithmetic the fast coordinates remain unchanged; refreshing the clock coordinate of all cached keys at each doubling entails a geometric total of $O(N)$ two-dimensional updates by length N. This is not zero latency or a measured cache optimization.

If fast-coordinate numerical precision changes, consistency may require refreshing those coordinates as well; refreshing all k coordinates by phases gives $O(kN)$ block updates, before accounting for bit complexity and memory movement. Content-dependent keys would also require access to their unrotated block values. The main position-only construction uses a known constant key and does not claim a free drop-in replacement in a pretrained KV cache.

H PROOF DEPENDENCIES AND REMAINING QUESTIONS

Proposition 2 follows from collision counting and the shared finite-grid construction (Appendix B). Theorem 1 combines finite-scale separation, classical lattice covering, and convex phase enclosure for necessity; the classical algebraic vector with a quantified norm bound supplies sufficiency (Appendix C). Theorem 3 combines the exact coherent-box return profile with the balance between fast-frequency and clock separation (Appendix D). Convex duality and mirror-family phase alignment provide the finite diagnostics in Appendix E.

The one-block asymmetric classification and fixed-length sharing examples in the research record are auxiliary, rather than prerequisites for these proofs. The all-length result does not settle the same-dimensional comparison in which the spectrum may be redesigned for each finite horizon. The minimum total dimension under length adaptation and the optimal joint limit near the critical error radius are also distinct from the two minimality statements proved here.

These remaining questions share the resource contract of Section 2. The current results characterize reliable positional routing in that channel; Section 7 records the limits of transferring them to unrestricted attention architectures.