TEMPORAL QUOTIENTS FOR CERTIFIED COUNT-PRESERVING SPIKE-RETIMING SMOOTHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Moving spike times can change a spiking neural network's prediction even when event counts remain fixed. We study how to certify robustness to such retiming while preserving event counts and distinct timestamps. We introduce temporal quotient smoothing for discretized event streams under a global budget on total spike displacement. A shared random partition maps each block's events on each line to a common configuration with the same count and distinct timestamps. If retiming crosses no sampled boundary, the transformed inputs coincide, yielding a finite-vote certificate for the smoothed classifier. To control the displacement introduced by smoothing, we minimize expected transport for a fixed event source and placement rule under boundary-hit constraints. A compact dual hierarchy and feasible kernels bound the remaining transport improvement; matching bounds establish periodic-kernel optimality for selected uniform fixed-count families. At global displacement budget $B = 2$ on the evaluated fixed checkpoints, Period-5 raises certified accuracy over all-or-none smoothing from 53.8% to 62.8% on DVS-Gesture with VGGSNN and from 25.6% to 76.8% on N-MNIST. A compact DVS classifier gives a small reversal, while source and decoder fitting can lower both transport and certified accuracy. Temporal quotients thus make count-preserving kernel design verifiable, with classification benefits that depend on the source, classifier, and budget.

1 INTRODUCTION

Spiking neural networks (SNNs) process event streams over time, so their predictions can depend on when spikes arrive. Recent retiming attacks exploit this dependence: moving existing spikes can change a prediction while preserving the count on every event line (Yu et al., 2026). These attacks motivate certification for discretized streams with distinct timestamps on each line. We study this problem under a global budget on the sum of spike displacements across the entire input.

Existing approaches to SNN robustness include certified bounds and training-based defenses (Liang et al., 2022; Ding et al., 2024). Randomized smoothing offers certificates for a predictor formed by aggregating a fixed classifier's outputs over random transformations. However, independent changes to time-bin occupancies can alter event counts, while directly quantizing timestamps can merge distinct spikes. Splitting-noise smoothing aligns nearby inputs through shared quantization cells and supports efficient derandomization (Levine & Feizi, 2021). Adapting this idea to retiming requires transformations that preserve both per-line counts and distinct timestamps.

In this paper, we introduce **temporal quotient smoothing** (Figure 1). We partition the timeline into blocks and, on each event line, replace each block's events by a shared placement with the same count and distinct timestamps. The key observation is that moving spikes within a block leaves its count unchanged. Under a shared partition and placement rule, retimed inputs therefore have identical projections whenever no spike crosses a sampled boundary. Classifier votes and a bound on this boundary-hit probability then certify the smoothed prediction.

We then optimize the distribution over partitions to control the displacement introduced by smoothing. For a fixed event source and placement rule (decoder), we minimize expected total displacement, or *transport*, under a prescribed boundary-hit bound. The number of partitions grows exponentially with the time horizon. We address this obstacle with compact linear programs (LPs) for dual lower bounds

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.

and globally feasible kernels for upper bounds. Their gap bounds how much a candidate's transport can still improve.

Experiments evaluate the classification consequences under matched boundary-hit bounds. At global displacement budget $B = 2$ on the evaluated fixed checkpoints, Period-5 raises certified accuracy over all-or-none smoothing from 53.8% to 62.8% on DVS-Gesture with VGGSNN and from 25.6% to 76.8% on N-MNIST. A compact DVS classifier gives a small reversal, while source and decoder fitting can lower both transport and certified accuracy. These findings show why geometric kernel design must be accompanied by evaluating the resulting classifier.

Contributions. 1) We construct a count-preserving temporal quotient and derive a finite-vote certificate for global spike displacement. 2) We develop an exact compact formulation of the bounded-span dual for rational additive interval costs and prove periodic-kernel optimality for the uniform fixed-count families in Table 1. 3) We evaluate fixed checkpoints on MNIST, DVS-Gesture, and N-MNIST, together with training on a synthetic source matched to the theory, identifying both classification benefits and their dependence on the source, classifier, and budget.

2 Related Work

Spike timing robustness and certified smoothing. Certified-bound propagation, training-based defenses, and count-preserving timing controls address SNN robustness (Liang et al., 2022; Ding et al., 2024; Liu et al., 2024; Wang et al., 2026; Wu et al., 2024; Hindman & Cantley, 2026; Trong Luu et al., 2026). The closest work studies capacity-one retiming attacks (Yu et al., 2026); we adopt its event model while designing an event-valid smoothing kernel around a fixed classifier. Randomized smoothing for norm, sparse, edit, transport, deformation, and temporal perturbations is well established (Lee et al., 2019; Zhang et al., 2020; Bojchevski et al., 2020; Huang et al., 2023; 2024; Levine & Feizi, 2020; Alfarra et al., 2022; Delattre et al., 2026; Goli & Farnia, 2024; Liu et al., 2026). DSSN is the closest smoothing construction: a shared scalar shift permits exact quantized vote evaluation (Levine & Feizi, 2021). Our quotient instead uses block-count configurations to preserve event counts and distinct timestamps. Optimal smoothing laws have also been studied for norm-bounded perturbations (Yang et al., 2020). Unlike prior smoothing distribution optimization over generic perturbation spaces, our design variables are probability laws over count-preserving temporal partitions with event-valid constraints and boundary-hit certification constraints. The resulting objective couples source-dependent transport minimization with the exact collision conditions required for spike-retiming certificates. The DSSN-JCR experiment is one specified joint-repair control, not a claim about all count-preserving extensions.

Distances, verification, and compact optimization. Rank-matched transport complements van Rossum and related spike distances (van Rossum, 2001; Sihn & Kim, 2019; Sotomayor-Gómez et al., 2023). Formal SNN verification, quotient abstractions, Wasserstein medoids, and rainbow packing provide related tools (Seong et al., 2024; Gupta et al., 2024; Jazi et al., 2026; Cuturi & Doucet, 2014; Sukegawa et al., 2024; Cheng et al., 2023). Finite-state methods support our capacity proofs (De Schutter, 2000; Nowak & Charron-Bost, 2014; Merlet et al., 2014; Klimann et al., 2004). Dynamic-programming and bounded-treewidth extended formulations provide the general optimization tools (Martin et al., 1990; Kolman & Koutecký, 2015; Tiwary et al., 2020). Our partition dual records both recent cut decisions and the unfinished block's start: the latter is needed when interval costs extend beyond the local span. This gives an explicit exact formulation for arbitrary rational interval costs.

3 Problem Setup

3.1 Capacity-one event lines

Let $[T] = \{0, \dots, T-1\}$ denote a discrete time horizon. An input has L pixel–polarity lines. Line ℓ is a subset $x_\ell \subseteq [T]$ containing at most one event at each time, and $K_\ell = |x_\ell|$ is its event count. For two inputs with the same count on every line, write their ordered event times as

$$x_\ell = (a_{\ell,1} < \cdots < a_{\ell,K_\ell}), \qquad x'_\ell = (a'_{\ell,1} < \cdots < a'_{\ell,K_\ell}). \tag{1}$$

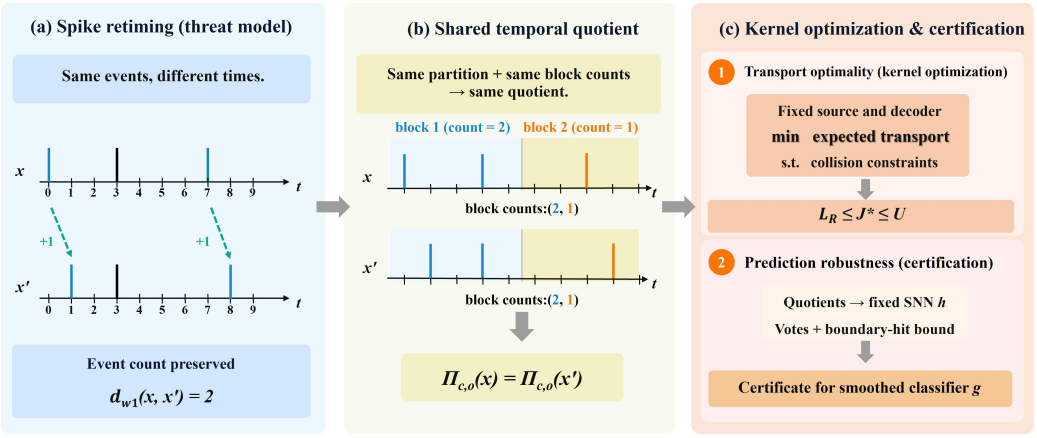

Figure 1: Overview of temporal quotient smoothing. Spike retiming preserves event counts but changes event times (a). A shared temporal partition produces identical quotient representations for inputs with the same block counts (b). The optimized kernel provides transport guarantees, while classifier votes and boundary-hit bounds yield certification of the smoothed classifier (c).

The integer transport distance is

$$d_{W_1}(x, x') = \sum_{\ell=1}^{L} \sum_{k=1}^{K_\ell} |a_{\ell,k} - a'_{\ell,k}|. \tag{2}$$

In one dimension, this rank-matched expression equals the minimum total displacement over all matchings. A unit displacement crosses one temporal boundary. Hence, an optimal adjacent-move path with $d_{W_1}(x, x') \le B$ crosses at most B distinct boundaries, although repeated crossings still consume the total-displacement budget. For example, $B = 2$ allows two unit moves, including moves on different event lines. Certificates on natural event streams apply after binning and binarization; the resulting event-number bins do not correspond to a fixed duration in milliseconds.

Relation to retiming budgets. If at most M spikes move and each moved spike shifts by at most J bins, then $d_{W_1}(x, x') \le MJ$, and an optimal path crosses at most $\min\{MJ, T-1\}$ distinct boundaries. Thus Theorem 3 at total-displacement budget MJ covers the intersection of the tamper-count-M and per-spike-jitter-J constraints. Because the proof depends only on the set of crossed boundaries, the certificate check need consider subsets only up to size $\min\{MJ, T-1\}$. Either constraint alone can still allow paths crossing all $T-1$ boundaries. Therefore, the empirical $B = 2$ results below do not by themselves certify a tamper-count ball or a per-spike-jitter ball.

3.2 Smoothed Classifier and Target

Let $\mathcal{X}$ be the capacity-one input space, and let h be a deterministic SNN classifier on $\mathcal{X}$. A random choice Z specifies a time partition and a canonical placement rule; $\Pi_Z(x)$ denotes the resulting count-preserving input. For class c, define the vote probability and the smoothed classifier by

$$p_c(x) = \Pr_Z[h(\Pi_Z(x)) = c], \qquad g(x) = \arg\max_c p_c(x), \tag{3}$$

with a fixed tie rule. Our goal is an exactly enumerable sufficient condition for $g(x') = g(x)$ for every capacity-one, equal-count input x' satisfying $d_{W_1}(x, x') \le B$. The guarantee applies to the smoothed classifier g, rather than to the unsmoothed classifier h, and the threat model excludes event insertion, deletion, polarity changes, and spatial relocation.

4 Count-Preserving Temporal Quotients

4.1 Partition Projection

Let $E = \{0, \dots, T-2\}$ be the temporal boundaries. A cut set $C \subseteq E$ partitions $[T]$ into contiguous blocks. Given a line and a block, the quotient first counts the input events in that block and then

places the same number k of events at canonical, distinct positions within the block. Because this operation is performed separately on every line and block, every output preserves each line count K_ℓ and the capacity-one constraint (Figure 1).

For a block of length m with k events, let $\mathcal{A}_{m,k}$ contain all increasing k-tuples in $\{0, \dots, m-1\}$. The total displacement from these states to a candidate placement $q \in \mathcal{A}_{m,k}$ is

$$\mathcal{R}(q) = \sum_{a \in \mathcal{A}_{m,k}} \sum_{r=1}^{k} |a_r - q_r|. \tag{4}$$

Proposition 1 (Fixed-count block W_1 medoid)**.** *Choose the lower median at every rank, or choose the upper median at every rank, of the order statistics over $\mathcal{A}_{m,k}$. Either common orientation gives strictly increasing coordinates and jointly minimizes $\mathcal{R}(q)$.*

The objective separates over ranks. Stochastic ordering of the order statistics makes the coordinate-wise medians a valid capacity-one state, while the lower and upper choices form a reflection-balanced pair of decoders. This proposition supplies the canonical placement rule used by the quotient; its proof appears in Appendix A.2.

4.2 Collision certificate

Let $S \subseteq E$ be the set of distinct boundaries crossed by a valid retiming path. Write $Z = (C, o)$, where C is the sampled cut set and o selects the lower or upper median in every block. We couple the clean and retimed inputs using the same Z; when the orientation is fixed, we write Π_C for the corresponding projection.

Lemma 2 (Quotient collision)**.** *If $C \cap S = \varnothing$, then $\Pi_C(x) = \Pi_C(x')$.*

If $C \cap S = \varnothing$, every adjacent move stays within one partition block, so the block-line counts are identical for x and x'. Let μ be the cut-set distribution and let η be the input-independent orientation law. The probability that the sampled cuts hit S is

$$\delta_S(\mu) = \Pr_{C \sim \mu}[C \cap S \neq \varnothing]. \tag{5}$$

Let $A = g(x)$ be the clean input's top class and let $c \neq A$ be a competitor. We retain the observed votes on no-hit partitions and assign the entire hit probability to the competitor, which gives the conservative bounds:

$$\underline{p}_S(A) = \Pr_{C \sim \mu, o \sim \eta}[h(\Pi_{C,o}(x)) = A,\ C \cap S = \varnothing], \tag{6}$$

$$\overline{p}_S(c) = \Pr_{C \sim \mu, o \sim \eta}[h(\Pi_{C,o}(x)) = c,\ C \cap S = \varnothing] + \delta_S(\mu). \tag{7}$$

Theorem 3 (Boundary-subset collision certificate)**.** *If $\underline{p}_S(A) > \max_{c \neq A} \overline{p}_S(c)$ for every nonempty $S \subseteq E$ with $|S| \leq B$, then $g(x') = A$ for every valid equal-count retiming x' with $d_{W_1}(x, x') \leq B$.*

The certificate uses only predictions of transformed versions of the clean input. For a period w, a uniformly sampled phase selects one residue class of boundaries to cut, giving $\delta_S \leq \min(B, w, T-1)/w$. For $1 \leq B \leq T-1$, the smallest period with a nonvacuous worst-case probability gap is $w = 2B+1$. With both median orientations, inference uses at most $2(2B+1)$ nominal deterministic SNN evaluations, independent of L and the number of events; phases with identical empty cuts may be deduplicated when $T < 2B+1$. The remaining certificate check depends only on the residue set represented by S; at most $\sum_{j=1}^{B} \binom{2B+1}{j}$ patterns are checked, independent of T, L, and the event count. At $B = 2$, this construction gives Period-5, an analytically tractable member of the partition family optimized below.

Projection and medoid-computation costs are given in Appendix B.

5 Kernel Design and Optimality Certificates

5.1 Complete partition LP

Fix an input source distribution $\mathcal{D}$ and decoder law η. For a cut set C, define $f(C) = \mathbb{E}_{x \sim \mathcal{D}, o \sim \eta}[d_{W_1}(x, \Pi_{C,o}(x))]$ as the expected event displacement induced by that partition. Let

μ_C be the probability assigned to C, let δ be the allowed probability of hitting any set of at most B crossed boundaries, and let $\mathbf{1}[\cdot]$ denote the indicator. Kernel design then solves

$$\min_{\mu} \quad \sum_{C \subseteq E} f(C)\mu_C \tag{8}$$

$$\text{s.t.} \quad \sum_{C} \mathbf{1}[C \cap S \neq \varnothing]\mu_C \leq \delta, \qquad S \subseteq E,\ 1 \leq |S| \leq B, \tag{9}$$

$$\sum_{C} \mu_C = 1, \qquad \mu_C \geq 0. \tag{10}$$

Denote the optimum transport cost by J^*. For uniform fixed-count masks, let $c_{\mathrm{blk}}(m)$ be a length-m block's expected contribution to transport. Then f is block additive:

$$f(C) = \sum_{I \in \mathrm{blocks}(C)} c_{\mathrm{blk}}(|I|). \tag{11}$$

The sum runs over the blocks induced by C. For position-dependent sources, write $c(a, b)$ for the expected contribution of interval $[a, b)$.

Proposition 4 (Finite-horizon certificate saturation)**.** *If $B \geq T - 1$, $f(C) \geq 0$ for every $C \subseteq E$, and $f(E) = 0$, then every feasible kernel has $\mu_\varnothing \geq 1 - \delta$. The all-or-none kernel with masses $1 - \delta$ on $\varnothing$ and δ on E is globally optimal, with value $(1 - \delta)f(\varnothing)$.*

The constraint indexed by $S = E$ forces the stated mass on the empty cut set (Appendix A.3). This saturation result concerns only the collision constraints: total event displacement can still exceed $T - 1$.

The closed-form results use the calibrated operating point $\delta = B/(2B + 1)$ induced by period $w = 2B + 1$. This choice fixes a common geometric comparison across kernels; it does not optimize classification accuracy over δ.

5.2 A COMPACT DUAL HIERARCHY

The complete LP has exponentially many partition variables. Its dual has one capacity constraint per partition; restricting the dual to nearby boundary sets permits a compact representation. Define the span of a boundary set S as $\max S - \min S$. For an integer span limit $0 \leq R \leq T - 2$, retain only the constraints indexed by $\mathcal{H}_R = \{S \subseteq E : 1 \leq |S| \leq B,\ \max S - \min S \leq R\}$. Removing the other constraints gives a relaxed optimum $L_R \leq J^*$. A globally feasible candidate of cost U gives an upper bound, so $U - L_R$ bounds the candidate's remaining possible improvement.

Theorem 5 (Compact local-dual hierarchy)**.** *Let $f(C) = \sum_{[a,b) \in \mathrm{blocks}(C)} c(a, b)$ for arbitrary rational interval costs $c(a, b)$, $1 \leq B \leq T - 1$, and rational $0 \leq \delta \leq 1$. The strongest dual supported on $\mathcal{H}_R$ has an exact LP formulation with $O(T2^R + T^2 + |\mathcal{H}_R|)$ variables and constraints. Its optimum is L_R, and $L_0 \leq \cdots \leq L_{T-2} = J^*$. For any globally feasible kernel of cost U, $L_R \leq J^* \leq U$.*

The compact state (i, m, r) records the next boundary index i, a bit mask m of the last R decisions (one means no cut), and the trailing no-cut run length r. The recent bits determine which local dual terms a decision completes, while $i - r$ identifies the unfinished block's start. The run length is necessary because, for example at $R = 2$, prefixes ending in two and five no-cut decisions have identical recent bits but different interval costs at the next cut. When the recent bits contain a cut, they already determine r; only an all-no-cut suffix needs additional states. This yields at most $2^R + i + 1$ states per layer.

Shortest-path potentials encode the partition capacities. The associated dual path flow recovers a locally feasible kernel, after which the omitted collision rows determine whether that kernel is globally feasible. This yields a useful diagnostic: a cheaper local kernel rules out a matching span-R dual, whereas a cheaper globally feasible kernel disproves the candidate's optimality. Appendix A.1 proves the equivalence and gives the exact rational repair procedure.

5.3 FEASIBLE KERNELS FROM BOUNDARY COLORING

To complement the dual lower bounds, we construct and optimize a family of globally feasible kernels. Assume $B < p \leq T - 1$. Assign every boundary to one of p nonempty colors and let C_r denote

color class r. Uniformly sampling a color gives worst-case collision probability $\rho = B/p$ for a set of at most B crossed boundaries, whereas the target is $\delta = B/(2B+1)$. If $B < p \le 2B+1$, mix the coloring kernel with $C = \varnothing$ using coloring weight $p/(2B+1)$. If $p \ge 2B+1$, mix it with $C = E$ using coloring weight $p(B+1)/[(2B+1)(p-B)]$. The weight is one at $p = 2B+1$. Because the other component is fixed in each case, distortion minimization at fixed T, B, p reduces to minimizing $\sum_r f(C_r)$.

For boundary i, write $s_i = f(\{i\})$, and for $i < j$ define the pair saving $a_{ij} = s_i + s_j - f(\{i,j\})$. Block additivity then gives

$$\sum_{i \in C} s_i - f(C) = \sum_{r=1}^{|C|-1} a_{c_r, c_{r+1}}, \qquad C = \{c_1 < \cdots < c_{|C|}\}. \tag{12}$$

Theorem 6 (Coloring to path cover)**.** *Optimizing a p-color boundary partition is equivalent to choosing $T - 1 - p$ forward arcs with maximum total saving, subject to indegree and outdegree at most one. These arcs form a vertex-disjoint directed path cover with p components, and every such cover defines a coloring.*

This fixed-cardinality matching problem has $O(T)$ vertices and $O(T^2)$ arcs, and admits an integral flow formulation. It optimizes the coloring family, which supplies a feasible upper bound for the complete LP. Equality with the complete LP requires a matching dual and is established only for the families covered below. Appendix A.8 develops these duals for the following families.

5.4 When periodic kernels are optimal

We write $J_{T,K}$ for the optimum for one line with K events, uniformly distributed over its $\binom{T}{K}$ valid configurations. This is the exchangeable fixed-count source used for the closed-form results. For a single spike, a block of length m contributes $\lfloor m^2/4 \rfloor / T$ to the expected transport. A packing of nearby boundary pairs supplies the matching dual (Appendix A.8).

Theorem 7 (All-length one-spike optimum)**.** *For $B = 2$, $\delta = 2/5$, and every $T \ge 6$, the period-five kernel globally solves the complete partition LP for the exchangeable one-spike source, with*

$$J_{T,1} = \frac{6}{5} - \frac{2}{T}. \tag{13}$$

The local packing proves every $T \ge 7$; an exact rational full-LP dual closes $T = 6$.

With two spikes, block counts are coupled by the fixed total. At $B = 2$, a reflection-symmetric radius-two packing uses slack within boundary runs to cover endpoint corrections. Its validity for all horizons reduces to finitely many scalar affine inequalities (Appendix A.12).

Theorem 8 (All-length two-spike optimum)**.** *For $B = 2$, $\delta = 2/5$, and every $T \ge 8$, the period-five kernel globally solves the complete partition LP for the exchangeable two-spike source, with*

$$J_{T,2} = \frac{6(T-2)^2 + 4}{5\binom{T}{2}}. \tag{14}$$

The onset is sharp: at $T = 7$, period five has value $22/15$, whereas the complete-LP optimum $51/35$ is attained by a calibrated period-six mixture.

For three spikes, we obtain a closed dual for the residue class $T = 5q + 3$ by assigning weights to six reflection classes of boundary pairs at distances one and two. The proof combines a finite-state certificate for runs of length at most four with a deletion argument: removing the third boundary of a longer run cannot decrease its capacity defect. Repeated deletion therefore reduces every mask to the finite-state case.

Theorem 9 (Residue-class three-spike optimum)**.** *For $B = 2$, $\delta = 2/5$, and every integer $q \ge 2$, the period-five kernel globally solves the complete partition LP for the exchangeable three-spike source at $T = 5q + 3$, with*

$$J_{5q+3,3} = \frac{375q^3 + 125q^2 + 40q - 2}{5\binom{5q+3}{3}}. \tag{15}$$

The same statement holds for $K = T - 3$ by spike-complement symmetry, but the result does not cover the other horizon residues or arbitrary source laws.

Table 1: Kernel-design guarantees. Periodic optima use uniform-conditional block medoids; the hierarchy accepts arbitrary given rational interval costs and rational $\delta \in [0, 1]$. Other rows use $\delta = B/(2B + 1)$. Uniform means uniform conditional on K; numerical sources are uniform $K = 1, 3$ and nonuniform $K = 1$.

Source or cost	B	Horizon T	Guarantee
Arbitrary interval costs	$1, \ldots, T - 1$	Any $T \geq 2$	Lower bound (Thm. 5)
Uniform $K = 1$	2	$T \geq 6$	Exact optimum (Thm. 7)
Uniform $K = 2$	2	$T \geq 8$	Exact optimum (Thm. 8)
Uniform $K = 3$	2	$5q + 3,\ q \geq 2$	Exact optimum (Thm. 9)
Uniform $K = 1$	3 / 4 / 5	$T \geq 9$ / 19 / 23	Exact optimum (Thms. 15,17)
Exchangeable, all counts	2	$8, \ldots, 14$	Finite exact (App. B.5)
Three prescribed laws	2	$64, 128, 256$	Numerical bounds (App. B.1)

Appendix A.13 proves the result and examines other residues, including numerical near-optimality at $T = 25$; these numerical observations are not claimed as exact optimality results.

Table 1 summarizes the scope of the guarantees. Complement symmetry covers $K = T - 1, T - 2, T - 3$ in the respective families, and mixtures of counts sharing an optimizer inherit the same optimizer (Corollary 18, Appendix A.14). For other counts, horizon residues, and source laws, the compact hierarchy provides instance-specific bounds rather than an automatic exact-optimality claim.

5.5 From transport to classifier votes

The LP measures geometric event displacement, whereas a class vote measures how much projected-input probability falls in the classifier's decision region. These objectives need not agree: by the classical total-variation characterization, distinct projected-input laws can admit classifiers with either sign of vote difference (Proposition 12, Appendix A.5). Thus transport optimality alone cannot impose a classifier-uniform ordering of pointwise vote utility, and this result does not by itself imply an ordering of multi-class certified accuracy. The fixed-checkpoint experiments below test when transport optimization happens to improve certification and when it does not.

6 Experiments

We ask how tightly kernel transport can be certified and when lower transport improves certified classification. The first study varies sources, horizons, and event counts. The second compares collision-matched kernels on fixed MNIST, DVS-Gesture, and N-MNIST classifiers, with controls for source fitting, classifier training, and displacement budget. *Certified accuracy* is the fraction of inputs classified correctly and certified at total-displacement budget B; certificate rate includes certified errors. Transport sums event displacements, while accuracies weight inputs equally. Appendix B gives training protocols, selection rules, and seed results.

6.1 Choosing kernels across sources and horizons

We first test whether the compact formulation can certify kernels beyond the closed-form families, including high-count and position-dependent sources (Table 2). At $B = 2$, a span-two dual certifies Period-5 within 6.4×10^{-9} of the complete-LP optimum for $T \in \{13, 25, 41\}$ and $K \in \{1, 2, 3, \lfloor T/2 \rfloor\}$. At $T = 41$, the compact formulation uses 901 states rather than 2^{40} partition columns. On small instances, its bounds agree within 10^{-7} with an independently constructed all-partition LP (Appendix B.1).

The hierarchy also separates two reasons for a nonzero gap. At $B = 3, T = 25, K = 3$, span three is provably insufficient, whereas span six certifies Period-7 within 4×10^{-9}. At $T = 41, K = 20$, in contrast, a feasible calibrated Period-9 mixture improves on Period-7 by 0.591%, showing that the candidate itself can be suboptimal. The appendix reports the full grid, including unresolved cases. Long-horizon CPU tests reduce the global gap below 6.6×10^{-8} at $T = 256, R = 2$ for three source laws; span-four solves reach the prescribed time limit (Appendix B.1).

DVS-Gesture. On all 1,176 training clips, a 64-state span-two dual bounds Period-5's full-LP suboptimality below 8.2×10^{-10} per line across the observed count distribution. Training-only

Table 2: Transport costs and certified optimality gaps. P is the Period-$(2B+1)$ cost; U is the best globally verified candidate cost. Costs are expected W_1 per line. Gaps $U - \underline{L}$ use the unrounded exact repaired lower bound $\underline{L}$ (Appendix B.1). The nonuniform source has $p_t \propto 1 + (t \bmod 4)$.

Source (T, K)	B	R	Periodic P	Candidate U	Gap bound
High event counts					
Uniform $(25, 12)$	2	2	6.710261	6.710261	$< 1.3 \times 10^{-9}$
Uniform $(41, 20)$	2	2	11.115947	11.115947	$< 3.7 \times 10^{-9}$
Increasing the dual span					
Uniform $(25, 3)$	3	3	4.055901	4.055901	< 0.019876
Uniform $(25, 3)$	3	6	4.055901	4.055901	$< 3.7 \times 10^{-9}$
Improving on the periodic kernel					
Uniform $(41, 20)$	3	6	14.227777	14.143743	$< 1.6 \times 10^{-8}$
Nonuniform $(41, 1)$	2	4	1.150495	1.126733	$< 7.1 \times 10^{-9}$

Table 3: Native event classification at global budget $B = 2$. Transport is mean total displacement per input; accuracies (%) weight inputs equally. Conv–LIF entries are means $\pm$ sample standard deviations over three training seeds; VGGSNN and N-MNIST each use one checkpoint. All kernels have ten nominal atoms and collision allowance $2/5$. AON has at most three distinct projections and Period-5 at most ten; nominal atom counts do not equalize computation.

DVS-Gesture: 288 shared test inputs

		VGGSNN Base: 95.1		Conv–LIF Base: 83.6 ± 0.7	
Kernel	Transport ↓	Smoothed acc.	Certified acc.	Smoothed acc.	Certified acc.
All-or-none	36 741.0	62.2	53.8	83.1 ± 0.2	81.7 ± 0.7
Calibrated p4	29 758.1	84.7	50.7	83.2 ± 0.5	79.5 ± 1.6
Period-5	25 723.3	85.4	62.8	83.3 ± 0.7	80.4 ± 1.1

N-MNIST: 1,000 test inputs; base accuracy: 99.6

Kernel	Transport ↓	Smoothed acc.	Certified acc.
All-or-none	2249.0	38.7	25.6
Calibrated p4	1835.1	97.0	25.6
Period-5	1600.2	95.3	76.8

selection retains Period-5, whose test transport is 30.0% below AON (Appendix B.2). This result establishes geometric near-optimality for the empirical source; it does not by itself establish a corresponding gain in certified accuracy.

6.2 Kernel comparisons on fixed SNNs

Native event classification. On the single N-MNIST checkpoint, Period-5's paired gain over AON and p4 is 51.2 certified-accuracy points (95% interval $[48.5, 53.9]$). All 512 gains over p4 occur on inputs that p4 classifies correctly but cannot certify. For both kernels at $B = 2$, the boundary-subset certificate holds exactly when all ten atom predictions agree (Appendix B.9); Period-5's projected inputs produce more unanimous correct votes under this criterion.

Changing the DVS classifier. The same DVS projections yield different utility orderings across classifiers. AON has higher certified accuracy than Period-5 at all three Conv–LIF seeds (Table 3), where its count-only templates retain high accuracy. These reversals are small, and all three paired 95% intervals include zero (Appendix B.9). Because both architecture and training differ from VGGSNN, this comparison concerns the trained classifier and does not isolate an architectural effect.

Fitting the kernel and decoder. We keep each classifier fixed and fit transport costs to the nonexchangeable $T = 13$ adjacent-triplet source. The fitted kernel lowers test transport by 1.34% but reduces certified accuracy by 0.05–0.15 points; two paired 95% intervals include zero (Appendix B.7). In a separate $T = 10$ control, decoder fitting lowers transport by 27.82% and certified accuracy at every evaluated seed (Appendix B.7). These controls show that fitting the transport objective need not improve certification.

Training on the theorem source. We additionally train the same MNIST architecture on uniformly sampled three-spike lines at $T = 13$, matching the residue-class theorem's source. This changes

classifier training; each subsequent kernel comparison holds the trained classifier fixed. Across three training seeds on one fixed test source draw, Period-5 reaches $96.51 \pm 0.06\%$ certified accuracy versus $92.37 \pm 2.46\%$ for AON, with mean displacement per event of 0.8341 versus 1.1638. A second temporal draw preserves the ordering at every checkpoint (Appendix B.6). This establishes a utility gain for the evaluated source and architecture.

Comparison with joint-count repair. In the matched MNIST comparison, Period-5 lowers transport and raises certified accuracy from 69.1% to 93.4% relative to the specified ten-atom DSSN-JCR adaptation with $m = 10$ nonempty reveal atoms (Table 7). The additional $m = 2, \ldots, 10$ frontier tests sensitivity within this repair family (Table 6). Members with $m = 5, \ldots, 10$ share Period-5's maximum B2 exposure of $4/10$, while $m = 2, 3, 4$ have smaller maximum exposure. The full frontier therefore spans different exposure levels. These comparisons do not represent all possible joint-count repairs.

6.3 Attack budgets and timing dependence

Larger displacement budgets. Under the matched construction, the mean certified-accuracy advantage over AON falls from 22.98 points at $B = 1$ to 4.58 at $B = 4$, with one seed reversing. At $T = 10, B = 9$, the constraints cover all boundary sets and AON becomes transport-optimal (Proposition 4); the observed advantage is therefore budget-dependent. Figure 3 in Appendix B.8 gives the budget-specific curves.

Timing dependence and attack budgets. Unrestricted count-preserving retiming lowers base accuracy by 16.5–18.5 points on latency-MNIST and from 95.1% to 19.1–19.4% on DVS-Gesture. In the official $B = 2000$ attack on eleven prespecified DVS inputs, nine base predictions flip; AON certifies three flipped inputs, whereas Period-5 certifies none (Appendix B.11). Thus low-budget benefits do not automatically extend to large natural-data budgets. This probe is diagnostic; Appendix B.10 gives a constructed LIF control.

7 Limitations

Closed-form optima assume uniform-conditional sources, and fixed-span dual tightness is instance-dependent. Our discretized, fixed-checkpoint experiments leave cross-architecture and raw-event generality open. Conditional on an active pixel, uniform-K3 timing is sampled independently of intensity and label; that study does not test label-bearing temporal information. The DSSN comparison covers one repair family, and nominal atom counts do not equalize inference cost. Timing-retaining certified protection at damaging natural-data budgets remains unresolved.

8 Conclusion

We introduced temporal quotient smoothing to certify count-preserving spike retiming in discretized event streams. Shared block counts connect valid event transformations to a certificate for the smoothed classifier under a global total-displacement budget. For a fixed source and decoder, compact dual bounds and feasible kernels make transport quality verifiable, with exact periodic optima for the proved families. Experiments show gains in certified accuracy on evaluated checkpoints, alongside reversals that demonstrate the need to assess classifier votes as well as transport. Obtaining useful certificates while retaining timing information at damaging budgets on natural event data remains an open challenge. Our results provide a basis for designing and evaluating event-valid smoothing with explicit geometric guarantees and measured classification benefits.

AI USE STATEMENT

We used generative AI tools (e.g., [ChatGPT, Gemini]) to assist with code implementation, experimental design and analysis, and manuscript editing. All code, data analysis pipelines, and textual outputs were thoroughly verified and validated by the authors. The validated codebase is made publicly available for verification, and the authors retain full responsibility for the integrity of the code, data, and scientific claims.

REPRODUCIBILITY STATEMENT

Appendix A gives symbolic proofs and finite-check reductions. Appendix B specifies experimental protocols, complete seed results, limitations of the comparisons, and reproduction entry points. The accompanying workspace contains exact certificates, result records, and a separate provenance index.

REFERENCES

Motasem Alfarra, Adel Bibi, Naeemullah Khan, Philip HS Torr, and Bernard Ghanem. Deformrs: Certifying input deformations with randomized smoothing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 6001–6009, 2022.

Aleksandar Bojchevski, Johannes Gasteiger, and Stephan Günnemann. Efficient robustness certificates for discrete data: Sparsity-aware randomized smoothing for graphs, images and more. In *International Conference on Machine Learning*, pp. 1003–1013. PMLR, 2020.

Yangyang Cheng, Jie Han, Bin Wang, and Guanghui Wang. Rainbow spanning structures in graph and hypergraph systems. In *Forum of Mathematics, Sigma*, volume 11, pp. e95. Cambridge University Press, 2023.

Marco Cuturi and Arnaud Doucet. Fast computation of wasserstein barycenters. In *International conference on machine learning*, pp. 685–693. PMLR, 2014.

Bart De Schutter. On the ultimate behavior of the sequence of consecutive powers of a matrix in the max-plus algebra. *Linear Algebra and its Applications*, 307(1-3):103–117, 2000.

Blaise Delattre, Hengyu Wu, Paul Caillon, Wei Yang Bryan Lim, and Yang Cao. Certified robustness under heterogeneous perturbations via hybrid randomized smoothing. *arXiv preprint arXiv:2605.12876*, 2026.

Jianhao Ding, Zhiyu Pan, Yujia Liu, Zhaofei Yu, and Tiejun Huang. Robust stable spiking neural networks. *arXiv preprint arXiv:2405.20694*, 2024.

Hossein Goli and Farzan Farnia. Certified adversarial robustness via partition-based randomized smoothing. *arXiv preprint arXiv:2409.13546*, 2024.

Arkaprava Gupta, Sumana Ghosh, Ansuman Banerjee, and Swarup Kumar Mohalik. Configuring safe spiking neural controllers for cyber-physical systems through formal verification. In *2024 22nd ACM-IEEE International Symposium on Formal Methods and Models for System Design (MEMOCODE)*, pp. 103–107. IEEE, 2024.

Lucas S Hindman and Kurtis D Cantley. Tempo: a stochastic benchmarking protocol for evaluating temporal robustness in spiking neural networks. *Neuromorphic Computing and Engineering*, 6(3): 034002, 2026.

Zhuoqun Huang, Neil G Marchant, Keane Lucas, Lujo Bauer, Olga Ohrimenko, and Benjamin Rubinstein. Rs-del: Edit distance robustness certificates for sequence classifiers via randomized deletion. *Advances in Neural Information Processing Systems*, 36:18676–18711, 2023.

Zhuoqun Huang, Neil G Marchant, Olga Ohrimenko, and Benjamin IP Rubinstein. Cert-ed: Certifiably robust text classification for edit distance. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 10813–10835, 2024.

Nikan Zandian Jazi, Elisabetta De Maria, and Christopher Leturc. A formal tool for verification of probabilistic spiking neural networks based on quotient abstractions. *arXiv preprint arXiv:2606.20674*, 2026.

Ines Klimann, Sylvain Lombardy, Jean Mairesse, and Christophe Prieur. Deciding unambiguity and sequentiality from a finitely ambiguous max-plus automaton. *Theoretical Computer Science*, 327 (3):349–373, 2004.

Petr Kolman and Martin Kouteckỳ. Extended formulation for csp that is compact for instances of bounded treewidth. *arXiv preprint arXiv:1502.05361*, 2015.

Guang-He Lee, Yang Yuan, Shiyu Chang, and Tommi Jaakkola. Tight certificates of adversarial robustness for randomly smoothed classifiers. *Advances in Neural Information Processing Systems*, 32, 2019.

Alexander Levine and Soheil Feizi. Wasserstein smoothing: Certified robustness against wasserstein adversarial attacks. In *International conference on artificial intelligence and statistics*, pp. 3938–3947. PMLR, 2020.

Alexander J Levine and Soheil Feizi. Improved, deterministic smoothing for l_1 certified robustness. In *International Conference on Machine Learning*, pp. 6254–6264. PMLR, 2021.

Ling Liang, Kaidi Xu, Xing Hu, Lei Deng, and Yuan Xie. Toward robust spiking neural network against adversarial perturbation. *Advances in Neural Information Processing Systems*, 35:10244–10256, 2022.

Shijie Liu, Tansu Alpcan, Christopher Leckie, and Sarah Erfani. Fortifying time series: Dtw-certified robust anomaly detection. *Advances in Neural Information Processing Systems*, 38: 165773–165797, 2026.

Yujia Liu, Tong Bu, Jianhao Ding, Zecheng Hao, Tiejun Huang, and Zhaofei Yu. Enhancing adversarial robustness in snns with sparse gradients. *arXiv preprint arXiv:2405.20355*, 2024.

R Kipp Martin, Ronald L Rardin, and Brian A Campbell. Polyhedral characterization of discrete dynamic programming. *Operations research*, 38(1):127–138, 1990.

Glenn Merlet, Thomas Nowak, and Sergeĭ Sergeev. Weak csr expansions and transience bounds in max-plus algebra. *Linear Algebra and its Applications*, 461:163–199, 2014.

Thomas Nowak and Bernadette Charron-Bost. An overview of transience bounds in max-plus algebra. *arXiv preprint arXiv:1405.3458*, 2014.

Baekryun Seong, Jieung Kim, and Sang-Ki Ko. Towards efficient formal verification of spiking neural network. *arXiv preprint arXiv:2408.10900*, 2024.

Duho Sihn and Sung-Phil Kim. A spike train distance robust to firing rate changes based on the earth mover's distance. *Frontiers in Computational Neuroscience*, 13:82, 2019.

Boris Sotomayor-Gómez, Francesco P Battaglia, and Martin Vinck. Spikeship: A method for fast, unsupervised discovery of high-dimensional neural spiking patterns. *PLOS Computational Biology*, 19(7):e1011335, 2023.

Noriyoshi Sukegawa, Shohei Suzuki, Yoshiko Ikebe, and Yoshito Hirata. On computing medians of marked point process data under edit distance. *Journal of Optimization Theory and Applications*, 200(1):178–193, 2024.

Hans Raj Tiwary, Martin Kouteck'y, and Petr Kolman. Extension complexity, mso logic, and treewidth. *Discrete Mathematics & Theoretical Computer Science*, 22, 2020.

Nhan Trong Luu, Duong Trung Luu, Nam Ngoc Pham, and Thang Cong Truong. Why temporal spike order reversal drops spiking network accuracy and how to partially mitigate it. *Neuromorphic Computing and Engineering*, 6(3):034016, 2026.

Mark CW van Rossum. A novel spike distance. *Neural computation*, 13(4):751–763, 2001.

Jihang Wang, Dongcheng Zhao, Ruolin Chen, Qian Zhang, and Yi Zeng. Boosting the robustness-accuracy trade-off of snns by robust temporal self-ensemble. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 26285–26293, 2026.

Keming Wu, Man Yao, Yuhong Chou, Xuerui Qiu, Rui Yang, Bo Xu, and Guoqi Li. Rsc-snn: Exploring the trade-off between adversarial robustness and accuracy in spiking neural networks via randomized smoothing coding. In *Proceedings of the 32nd ACM international conference on multimedia*, pp. 2748–2756, 2024.

Greg Yang, Tony Duan, J Edward Hu, Hadi Salman, Ilya Razenshteyn, and Jerry Li. Randomized smoothing of all shapes and sizes. In *International conference on machine learning*, pp. 10693–10705. PMLR, 2020.

Yi Yu, Qixin Zhang, Shuhan Ye, Xun Lin, Qianshan Wei, Kun Wang, Wenhan Yang, Dacheng Tao, and Jiang Xudong. Time is all it takes: Spike-retiming attacks on event-driven spiking neural networks. In *International Conference on Learning Representations*, volume 2026, pp. 37613–37646, 2026.

Dinghuai Zhang, Mao Ye, Chengyue Gong, Zhanxing Zhu, and Qiang Liu. Black-box certification with randomized smoothing: A functional optimization based framework. *Advances in Neural Information Processing Systems*, 33:2316–2326, 2020.

A Proof Details

Section A.1 proves the general compact formulation; Sections A.2–A.6 give the projection, certification, and utility arguments. The periodic optima use the common dual construction in Section A.8, followed by the K1–K3 proofs and the mixture closure in Section A.14.

A.1 Constructive bounded-span partition duals

Write $n = T - 1$ and $D = E \setminus C$ for the uncut boundaries. Assign a nonnegative weight λ_S to each retained boundary set and a free offset θ. The dual of the locally constrained primal in Theorem 5 is

$$\max_{\lambda \geq 0, \theta \in \mathbb{R}} \quad (1-\delta) \sum_{S \in \mathcal{H}_R} \lambda_S - \theta, \tag{16}$$

$$\text{subject to} \quad \sum_{S \in \mathcal{H}_R : S \subseteq D} \lambda_S \leq f(E \setminus D) + \theta, \qquad D \subseteq E. \tag{17}$$

Extending λ by zero outside $\mathcal{H}_R$ gives a full-LP dual. The local primal itself need not be feasible for the full LP.

Sufficient state and size. After processing i boundary bits $d_j = \mathbf{1}[j \in D]$, keep (i, m, r): m holds the most recent R bits, newest at the least significant position, and r is the number of trailing ones. The unfinished block starts at $a = i - r$. Choosing $d_i = 0$ closes $[a, i+1)$ at cost $c(a, i+1)$ and resets r to zero. Choosing $d_i = 1$ extends the run and subtracts every λ_S with $\max S = i$ whose earlier members are ones in m. In both cases update m to the most recent R bits. At $i = n$, charge $c(n - r, T)$ and enter a common sink.

Every path from the empty prefix to the sink represents a cut set and has weight $f(C) - \sum_{S \subseteq E \setminus C} \lambda_S$. Each interval and multiplier is charged exactly once. Prefixes sharing (i, m, r) have identical future contributions: the unfinished block start is known, and any support completed later can only inspect the preceding R bits. Thus merging prefixes preserves the minimum over all 2^n cut sets. At layer i, when m is not all ones, r is its trailing-one count; otherwise r has at most $i+1$ possibilities. There are at most $2^R + i + 1$ reachable states per layer, including $R = 0$, and two outgoing decision edges per nonterminal state. The DAG has $O(T2^R + T^2)$ vertices and edges. Also $|\mathcal{H}_R| \leq n \sum_{j=0}^{\min(B-1,R)} \binom{R}{j}$.

Potential formulation and equivalence. For edge $e : u \to v$, let a_e be its interval charge and $\mathcal{A}_e$ the supports it completes. Give all vertices free potentials p_v, fix $p_{\text{source}} = 0$, and impose

$$p_v - p_u + \sum_{S \in \mathcal{A}_e} \lambda_S \le a_e, \qquad p_{\text{sink}} + \theta \ge 0.$$

Keep the objective of (16) and $\lambda \ge 0$. Summing edge inequalities along each path gives $p_{\text{sink}} \le f(C) - \sum_{S \subseteq E \setminus C} \lambda_S$; the sink inequality implies every capacity. Conversely, when all capacities hold, shortest-path distances from the source supply feasible potentials, including $p_{\text{sink}} \ge -\theta$. The graph is acyclic and all states are reachable. This proves exact equivalence and the claimed formulation size. Each edge contains at most $\sum_{j=0}^{\min(B-1,R)} \binom{R}{j}$ multiplier coefficients; matrix sparsity must also be included in an implementation's resource accounting.

Hierarchy and witnesses. The row families are nested and $\mathcal{H}_{T-2}$ contains every row, so finite-LP strong duality proves $L_0 \le \cdots \le L_{T-2} = J^*$. The dual of the potential LP has nonnegative unit source–sink flow and expected support-completion mass at least $1 - \delta$ for each S. Decomposing the flow into paths gives a probability distribution with $\Pr[C \cap S \ne \varnothing] \le \delta$ on $\mathcal{H}_R$. If such a locally feasible law costs $V < U$, weak duality rules out every span-R certificate attaining U, not just the fitted multiplier vector. Only a separate global collision check can turn the law into a counterexample to the candidate kernel's global optimality.

Exact repair. For any nonnegative rational proposal λ, recompute

$$m(\lambda) = \min_{C \subseteq E} \left[f(C) - \sum_{S \subseteq E \setminus C} \lambda_S \right]$$

on the DAG with exact arithmetic. Then $\theta = -m(\lambda)$ is feasible, and $\underline{L} = (1 - \delta) \sum_S \lambda_S + m(\lambda)$ is an exact global lower bound. This assertion does not depend on floating-point LP accuracy. The repaired objective is 1-Lipschitz in the ℓ_1 multiplier norm: for each cut set, each multiplier coefficient is either $1 - \delta$ or $-\delta$, and taking a minimum preserves this bound. For a rationalized, normalized path mixture with maximum local collision $q > \delta$, multiply its masses by δ/q and place the remainder on $C = \varnothing$. This restores all local rows exactly. Applying the same operation using the maximum *global* collision instead gives a globally feasible upper bound, which can be loose. An exact gap below ϵ establishes ϵ-optimality, not exact equality of optima.

A.2 Block medoids and collision certificates

Proof of Proposition 1. For $a, q \in \mathcal{A}_{m,k}$, one-dimensional optimal transport matches equal ranks, hence

$$W_1(a, q) = \sum_{r=1}^{k} |a_r - q_r|.$$

Summing over a separates $\mathcal{R}(q)$ into k scalar absolute-deviation objectives. Each is minimized by any median of the corresponding order statistic. Enumerate the $N = \binom{m}{k}$ states and let x_i and y_i be rank r and rank $r + 1$ in state i. Pointwise, $y_i \ge x_i + 1$. If $x_{(j)}$ and $y_{(j)}$ are the two sorted multisets, then $y_{(j)} \ge x_{(j)} + 1$ for every j: otherwise at least j values of y would lie below $x_{(j)} + 1$, and their paired x values would all lie below $x_{(j)}$, a contradiction. Applying this fact at the common lower-median index, or at the common upper-median index, gives $q_r < q_{r+1}$. Thus both common orientations are coordinatewise minimizers in $\mathcal{A}_{m,k}$. □

Proposition 10 (Coordinatewise event-validity barrier)**.** *Fix $1 \le K \le T - 1$. Let a deterministic binary coordinatewise atom have $Q(x)_t = \phi_t(x_t)$, where $\phi_t \colon \{0, 1\} \to \{0, 1\}$. If $|Q(x)| = K$ for every binary x with $|x| = K$, then Q is one of: the identity; an input-independent binary template of count K; or, only when $T = 2K$, the coordinatewise complement. The classification applies atomwise to any randomized coordinatewise smoother whose every output is event-valid.*

Proof. Put $b_t = \phi_t(0)$ and $a_t = \phi_t(1) - \phi_t(0)$. For every K-subset $S \subset [T]$, count preservation is

$$\sum_{t=0}^{T-1} b_t + \sum_{t \in S} a_t = K.$$

Any two coordinates occur in two K-subsets that differ only by swapping those coordinates, so all a_t equal one $a \in \{-1, 0, 1\}$. If $a = 1$, every ϕ_t is the identity. If $a = -1$, every ϕ_t is complementation and the count equation requires $T - K = K$. If $a = 0$, each ϕ_t is constant and the constants form a fixed count-K template. These cases are also sufficient. Applying the deterministic statement after conditioning on a randomized atom proves the final sentence. □

Proof of the quotient-collision lemma. An equal-count pair of increasing event-time tuples admits an optimal adjacent-move realization that preserves capacity one: first move every left-going event to its target in increasing rank order, then move every right-going event in decreasing rank order. The ordered target tuple keeps each move strictly between its already placed neighbor and its unmoved neighbor. Its number of unit moves is the rank-matched distance in Equation (2). If no crossed boundary lies in C, each move stays within one block, so every block-line event count is unchanged. The decoder is deterministic given those counts and the shared median orientation. Thus clean and retimed inputs map to the same projected event set. □

Proof of Theorem 3. If $x' = x$, the conclusion is immediate. Otherwise fix the adjacent-move realization above and let the nonempty set S contain its distinct crossed boundaries. Equation (2) implies $|S| \le B$. Under the shared-cut coupling, the collision lemma makes the two projected inputs identical on $C \cap S = \varnothing$. Therefore the probability that x' receives class A is at least $\underline{p}_S(A)$, while the probability of any $c \neq A$ is at most $\overline{p}_S(c)$: all cut-hitting mass is assigned to c. The assumed strict inequalities make A the unique top class for x'. The argument uses only the set S, so it covers every valid retiming within budget. □

A.3 Finite-horizon certificate saturation

Proof of Proposition 4. When $B \ge T - 1$, the complete LP contains the constraint indexed by $S = E$. Because $C \cap E \neq \varnothing$ exactly when $C \neq \varnothing$, this constraint gives

$$\sum_{C \neq \varnothing} \mu_C \le \delta, \qquad \mu_\varnothing \ge 1 - \delta.$$

Nonnegative costs therefore imply $\sum_C f(C)\mu_C \ge f(\varnothing)\mu_\varnothing \ge (1-\delta)f(\varnothing)$. The all-or-none kernel attains this bound because its only other atom is E and $f(E) = 0$. □

A.4 Margin-conditioned utility bridge

Let $s_c(z)$ be the logits of h, let $y = h(x)$ have margin $m_x = s_y(x) - \max_{c \neq y} s_c(x) > 0$, and define $\Delta_x(z) = [m_x - s_y(z) + \max_{c \neq y} s_c(z)]_+$.

Proposition 11 (Margin-conditioned utility bridge)**.** *If $\Delta_x(z) \le L_x d_{W_1}(x, z)$ for every quotient atom $z = \Pi_Z(x)$, then*

$$\Pr_Z[h(\Pi_Z(x)) \neq y] \le \min\left\{1, \frac{L_x \mathbb{E}_Z[d_{W_1}(x, \Pi_Z(x))]}{m_x}\right\}. \tag{18}$$

Proof of Proposition 11. Write $m_y(z) = s_y(z) - \max_{c \neq y} s_c(z)$. If $h(z) \neq y$, then $m_y(z) \le 0$ under any fixed tie rule, and hence $\Delta_x(z) = [m_x - m_y(z)]_+ \ge m_x$. Therefore

$$\begin{aligned}
\Pr_Z[h(\Pi_Z(x)) \neq y] &\le \Pr_Z[\Delta_x(\Pi_Z(x)) \ge m_x] \\
&\le \frac{\mathbb{E}_Z[\Delta_x(\Pi_Z(x))]}{m_x} \\
&\le \frac{L_x \mathbb{E}_Z[d_{W_1}(x, \Pi_Z(x))]}{m_x},
\end{aligned}$$

where the second line is Markov's inequality. Intersecting the right-hand side with the trivial probability bound one proves Equation (18). □

A.5 SHARP UTILITY AND STABILITY–CORRECTNESS SEPARATION

Let $Q_{\mu,x}$ be the law of the projected input $\Pi_Z(x)$ when the partition is sampled from μ and the decoder law is fixed.

Proposition 12 (Total-variation characterization applied to quotient laws)**.** *For any kernels μ, ν and fixed x,*

$$\sup_{h:\mathcal{X}\to\{0,1\}} \left| \Pr_{Q_{\mu,x}}[h(z)=1] - \Pr_{Q_{\nu,x}}[h(z)=1] \right| = \mathrm{TV}(Q_{\mu,x}, Q_{\nu,x}). \tag{19}$$

If the laws differ, deterministic binary classifiers attain both signs. Hence a label-free distortion ordering alone cannot impose one pointwise vote-utility ordering uniformly over classifiers; this statement is about vote probabilities and does not by itself order multi-class certified accuracy.

Proof of Proposition 12. Write $q_\mu(z)$ and $q_\nu(z)$ for the two finite projected-input laws and let $A_+ = \{z : q_\mu(z) > q_\nu(z)\}$. The binary classifier $h_+(z) = \mathbf{1}[z \in A_+]$ gives

$$\Pr_{Q_{\mu,x}}[h_+(z)=1] - \Pr_{Q_{\nu,x}}[h_+(z)=1] = \sum_{z\in A_+} (q_\mu(z) - q_\nu(z)) = \mathrm{TV}(Q_{\mu,x}, Q_{\nu,x}).$$

Every indicator expectation differs by at most total variation, so this is the supremum. The complementary classifier $1 - h_+$ attains the negative of the same value. When the laws differ, the value is positive and both pointwise utility orders are therefore realized. □

Corollary 13 (Stability does not imply correctness)**.** *If a collision kernel has $\delta_S < 1/2$ for every nonempty $|S| \le B$, a constant classifier passes Theorem 3 on every input. On examples whose ground-truth label differs from that constant, its predictions are certified but wrong.*

Proof. For the constant class A, the lower probability is $1-\delta_S$ and every competitor upper probability is δ_S. Their strict ordering follows from $\delta_S < 1/2$. The certificate concerns stability of g, not agreement with the ground truth. □

A.6 SOURCE-LAW CLOSURE AND FEASIBLE REGRET

Fix a projection decoder independently of source law P, let c_P be its expected partition-cost vector, and write $F_P(\mu) = \langle c_P, \mu\rangle$ on the feasible kernel polytope $\mathcal{M}$.

Proposition 14 (Fixed-decoder source-law closure and feasible regret)**.** *If one $\mu_\star$ minimizes every endpoint objective F_{P_i}, it minimizes F_P for every P in their convex hull. Moreover, for endpoint optimizers μ_P, μ_Q and $\Delta = c_Q - c_P$,*

$$0 \le F_Q(\mu_P) - F_Q(\mu_Q) \le \max_{\mu\in\mathcal{M}} \langle \Delta, \mu\rangle - \min_{\mu\in\mathcal{M}} \langle \Delta, \mu\rangle. \tag{20}$$

Proof of Proposition 14. The projection decoder is fixed independently of the source law, so $c_{\sum_i \theta_i P_i} = \sum_i \theta_i c_{P_i}$ for convex weights θ_i. For any feasible μ and convex weights θ_i, endpoint optimality gives $F_{P_i}(\mu_\star) \le F_{P_i}(\mu)$. Multiplying by θ_i and summing, using linearity of the cost in the source law, proves $F_{\sum_i \theta_i P_i}(\mu_\star) \le F_{\sum_i \theta_i P_i}(\mu)$.

For the regret statement, insert and subtract the two P objectives:

$$\begin{aligned} F_Q(\mu_P) - F_Q(\mu_Q) &= F_P(\mu_P) - F_P(\mu_Q) + \langle \Delta, \mu_P - \mu_Q\rangle \\ &\le \langle \Delta, \mu_P\rangle - \langle \Delta, \mu_Q\rangle \\ &\le \max_{\mu\in\mathcal{M}} \langle \Delta, \mu\rangle - \min_{\mu\in\mathcal{M}} \langle \Delta, \mu\rangle. \end{aligned}$$

The left inequality is optimality of μ_Q under Q.

Finally, along an affine source path, optimality of a fixed $\mu_\star$ is equivalent to $\langle c_{P_\alpha}, \mu - \mu_\star\rangle \ge 0$ for every $\mu \in \mathcal{M}$. Each condition is affine in scalar α, so their intersection is a closed interval. Adding α to a complete-LP dual and enforcing primal–dual equality recovers its two endpoints by linear optimization. □

A.7 PATH-COVER REDUCTION

For each color class $C = \{c_1 < \cdots < c_k\}$, expanding the interval costs in Equation (11) shows

$$f(C) = \sum_{i \in C} s_i - \sum_{r=1}^{k-1} a_{c_r, c_{r+1}}.$$

Across all colors, $\sum_C \sum_{i \in C} s_i = \sum_{i \in E} s_i$ is constant. The chosen successor arcs have indegree and outdegree at most one and increase in boundary index, hence form vertex-disjoint directed paths. A partition of $T-1$ vertices into p nonempty paths has exactly $T-1-p$ arcs. Conversely, the vertices of every such path define one color class, and sorting each path recovers its consecutive saving arcs. Maximizing total saving is therefore equivalent to minimizing coloring cost. Splitting each boundary into a left and right copy turns the selection into a cardinality-constrained bipartite matching; the matching polytope is integral, which gives the min-cost-flow algorithm.

A.8 CRITICAL RAINBOW DUALITY

At $p = 2B+1$, let $F(D) = f(E \setminus D)$. A period-coloring kernel is globally optimal if nonnegative weights λ_S on rainbow B-subsets satisfy

$$\sum_{S \subseteq D} \lambda_S \le F(D) + \theta, \qquad D \subseteq E, \tag{21}$$

$$\sum_{S \subseteq E \setminus C_r} \lambda_S = f(C_r) + \theta, \qquad r = 0, \ldots, p-1. \tag{22}$$

These relations reconstruct a feasible complementary dual, while every complementary full-LP dual has this form. The K1 and K2 constructions at $B = 2$ use $p = 5$ and $\theta = 0$; the K1 construction at $B = 3$ uses $p = 7$ and $\theta = 0$.

Let $y_S \le 0$ be the dual variable for collision constraint S, ν the free normalization multiplier, and $\lambda_S = -y_S$. Under the critical $p = 2B+1$ coloring, a constraint with $|S| < B$ is strict because its hit probability is at most $(B-1)/p < B/p$. For $|S| = B$, equality holds exactly when the boundaries have B distinct colors. Complementary slackness therefore restricts nonzero λ_S to rainbow B-sets. Put $L = \sum_S \lambda_S$ and $\theta = L - \nu$. For a cut set C and $D = E \setminus C$, the dual constraint is

$$\nu - \sum_{S : C \cap S \neq \varnothing} \lambda_S \le f(C).$$

Substitution gives Equation (21). Tightness on every positive phase column gives Equation (22). Conversely, these two families reconstruct a feasible complementary dual. Summing the p phase equalities counts each rainbow B-set in exactly $p - B$ complements, so

$$(p-B)L = \sum_{r=0}^{p-1} f(C_r) + p\theta.$$

At $\delta = B/p$, the dual objective is $\nu - \delta L = p^{-1} \sum_r f(C_r)$, exactly the coloring primal objective. Weak duality proves global optimality.

A.9 ALL-LENGTH K1 PACKING

For one uniform spike, a length-m block is occupied with probability m/T; conditional position is uniform, and its unnormalized median deviation is $\lfloor m^2/4 \rfloor$. This proves the block cost used in Theorem 7.

Scale all packing weights by T. First place unit weight on every pair at distance one or two. If $D \subseteq E$ has maximal selected runs of lengths $s_1, \ldots, s_q$, define

$$H(s) = \left\lfloor \frac{(s+1)^2}{4} \right\rfloor, \quad u(s) = \max(s-1, 0) + \max(s-2, 0), \quad \sigma(s) = H(s) - u(s).$$

A distance-two pair additionally joins consecutive runs separated by one unselected boundary. Group runs connected by such gaps. A component of q' runs has unit-packing slack

$$1 + \sum_{\text{runs } s \text{ in component}} (\sigma(s) - 1),$$

because $\sigma(s) = 1$ for $s = 1, 2, 3, 4$ and $\sigma(s) \geq 2$ for $s \geq 5$. Thus each component has at least one unit of scaled slack, and at least two if it contains a run of length five or more.

The remaining five $1/3$ corrections lie on adjacent edges in the six-boundary window $\{0, \ldots, 5\}$. If a component selects at most three, its unit slack pays their mass. Selecting four or five forces a run of length at least five: runs of length at most four can contain at most three adjacent edges in this window after accounting for a separating gap. The resulting two units of slack cover at most $5/3$. This proves every subset capacity.

For a phase complement $D_r = E \setminus C_r$, all runs have length at most four and form one local-pair component, so the unit packing has slack exactly one. Removing color r deletes two of the five correction-cycle edges and retains three; their total mass is one and fills that slack. All phase equalities hold. There are $(T-2) + (T-3) = 2T - 5$ unit pairs and correction mass $5/3$. Critical rainbow duality therefore gives $(3/5)(2 - 10/(3T)) = 6/5 - 2/T$. This packing applies for $T \geq 7$. At $T = 6$, the period-five primal has value $13/15$, and an exact rational full-LP dual has the same value; this establishes Theorem 7 at its first horizon.

A.10 All-length $B = 3$ K1 packing

Theorem 15 (All-length $B = 3$ one-spike optimum)**.** *For $B = 3$, $\delta = 3/7$, and every $T \geq 9$, the period-seven kernel globally solves the complete partition LP for the exchangeable one-spike source, with*

$$J^{(3)}_{T,1} = \frac{12}{7} - \frac{4}{T}. \tag{23}$$

The onset is sharp: at $T = 8$, period seven has value $17/14$, whereas the all-or-none feasible kernel has value $8/7$.

For $B = 3$, put $p = 7$ and scale the dual weights by T, writing $w_S = T\lambda_S$. For $T \geq 11$, define the local rainbow triples

$$A_i = \{i, i+1, i+2\}, \qquad B_i = \{i, i+1, i+3\}, \qquad C_i = \{i, i+2, i+3\}.$$

Begin with $w(A_i) = 3$ at every legal anchor. Replace the first four A weights by $(2, 1, 2, 2)$ and the last four, in increasing-anchor order, by $(2, 2, 1, 2)$. Give unit weight to B_i at its first two and last four legal anchors, and to C_i at its first four and last two legal anchors. All other weights vanish. The endpoint lists are disjoint for $T \geq 11$. Reflection exchanges the endpoint corrections, and every supported triple has span at most three, so the construction is rainbow modulo seven. The A, B, and C masses are $3T - 19$, 6, and 6, respectively. Hence

$$L = \sum_S \lambda_S = \frac{3T - 7}{T}, \qquad \nu - \frac{3}{7}L = \frac{4}{7}L = \frac{12}{7} - \frac{4}{T}. \tag{24}$$

Phase equalities. For each period-seven cut class C_r, the scaled equality is

$$\sum_{S \subseteq E \setminus C_r} w_S = T f(C_r). \tag{25}$$

Exact integer evaluation verifies Equation (25) for the seven residue bases $T = 11, \ldots, 17$. Under $T \mapsto T + 7$, the left endpoint correction is unchanged and the reflected correction shifts by seven, so both retain their residues. The consecutive-triple layer gains seven anchors; exactly four of their triples avoid a fixed residue, adding $4 \cdot 3 = 12$. The critical partition gains one length-seven block, whose scaled K1 cost is $\lfloor 7^2/4 \rfloor = 12$. Thus all seven phase equalities propagate to every $T \geq 11$.

Subset capacities. For $D \subseteq E$, let ℓ range over the lengths of its maximal selected runs, and put $h(m) = \lfloor m^2/4 \rfloor$. The scaled capacity is

$$TF(D) = \sum_{\text{runs}} h(\ell + 1).$$

Giving weight three to every consecutive triple A_i places $3\max(\ell - 2, 0)$ inside a run of length ℓ. Its runwise slack is

$$s(1) = 1, \qquad s(2) = 2, \qquad s(\ell) = \left\lfloor \frac{(\ell - 5)^2}{4} \right\rfloor \quad (\ell \geq 3), \tag{26}$$

which is nonnegative.

Relative to this base, the left endpoint correction is

$$\begin{aligned} Q_L(D) = & -\mathbf{1}[A_0 \subseteq D] - 2\mathbf{1}[A_1 \subseteq D] - \mathbf{1}[A_2 \subseteq D] - \mathbf{1}[A_3 \subseteq D] \\ & + \mathbf{1}[B_0 \subseteq D] + \mathbf{1}[B_1 \subseteq D] + \sum_{i=0}^{3} \mathbf{1}[C_i \subseteq D]. \end{aligned} \tag{27}$$

Membership in Equation (27) depends only on the seven boundary bits $\{0, \ldots, 6\}$. An exhaustive 128-pattern integer lemma shows that $Q_L(D)$ is at most the sum of $s(\ell)$ over runs completed in this window, plus $\min_{q \geq \ell_{\text{tail}}} s(q)$ for a run leaving the window. These tail minima are zero for lengths one through six and one for length seven. The right endpoint obeys the reflected lemma.

For $T \geq 17$, the endpoint windows are disjoint. If their run families are disjoint, the allocated slack is disjoint. If one run connects the windows, a nonzero allocation from one endpoint requires a seven-boundary tail and at least one unit of actual slack; nonzero allocations from both endpoints require run length at least 14 and at least two units. The two corrections therefore never exceed the base slack. Complete exact enumeration of all 64,512 subsets closes the overlapping lengths $T = 11, \ldots, 16$. This proves every capacity inequality, not only interval or solver-active rows.

Equations (25) and the subset capacities give the explicit zero-θ dual for every $T \geq 11$. Independently reconstructed exact rational duals cover $T = 9, 10$. At $T = 8$, the all-or-none kernel gives the strict feasible comparison $8/7 < 17/14$. Together with Equation (24), critical rainbow duality proves Theorem 15 and its sharp onset.

To extend a capacity proof to longer horizons, we track the largest excess of dual reward over block cost for each possible boundary history. A max-plus update adds a transition reward and maximizes over predecessor states. The quantity κ below is the allowed increase in this excess when one bulk period is appended, not an event-displacement budget.

Theorem 16 (Conditional finite-state max-plus reduction)**.** *Fix a period p, one horizon residue, and an additive target increment κ. Suppose a finite state set $\mathcal{S}$ exactly represents every unrestricted capacity candidate: histories assigned the same state have the same legal next transitions and incremental rewards. Let b be the exact post-prefix value vector, let A be a fixed max-plus period operator, and let Φ be a fixed right-end operator followed by componentwise maximization. Let $v_q(s)$ be the maximum accumulated reward at state s after q bulk periods, and M_q the maximum after the right-end completion. Thus*

$$v_q = A^q b, \qquad M_q = \Phi(v_q). \tag{28}$$

Assume that A and Φ are monotone and additively homogeneous, and that for one finite q_0,

$$v_{q_0+1} \leq v_{q_0} + \kappa\mathbf{1} \tag{29}$$

componentwise, with absent states represented by $-\infty$. Suppose also that the exact finite bases through $q_0 + 1$ equal a proposed target sequence τ_q satisfying $\tau_{q+1} = \tau_q + \kappa$, and that a compatible feasible saturated witness supplies $M_{q+1} \geq M_q + \kappa$ and the required phase and dual-mass equalities. Then $M_q = \tau_q$ for every $q \geq 0$. Checking these conditions for every residue proves all-length complete-dual feasibility; with the matching primal, weak duality proves global optimality.

Proof. Apply A to Equation (29). Monotonicity and additive homogeneity give

$$v_{q_0+2} = Av_{q_0+1} \leq A(v_{q_0} + \kappa\mathbf{1}) = v_{q_0+1} + \kappa\mathbf{1}.$$

Induction gives the same componentwise inequality for every later pair. Applying Φ preserves it, hence $M_{q+1} \leq M_q + \kappa$. The saturated witness gives the reverse inequality. Equality on the finite bases therefore propagates to every q, and the phase and mass identities reconstruct the complete dual objective. □

A sufficient event-partition state. Applying the theorem requires an explicitly constructed dual and a sufficient finite state for the chosen B and K. For the event-partition residual, such a state follows when the mask alphabet is finite, reward supports have bounded span, at least one maximizing path has a uniform finite penalty-memory bound, the bulk translates with period p, endpoint corrections have fixed width, and the state records every recent symbol and memory value needed for the next reward and transition. Histories represented by the same state must have identical legal continuations and incremental rewards. Finite-state max-plus representations and transience are classical (De Schutter, 2000; Nowak & Charron-Bost, 2014; Merlet et al., 2014; Klimann et al., 2004). The theorem specializes these tools to the event-partition dual; constructing the dual and proving state sufficiency remain specific to each application.

A.11 ALL-LENGTH $B = 4, 5$ K1 STATE-VECTOR PROOF

Theorem 17 (State-vector one-spike optima)**.** *For $B \in \{4, 5\}$, $\delta = B/(2B + 1)$, and respectively $T \geq 19, 23$, the period-$(2B+1)$ kernel globally solves the complete partition LP for the exchangeable one-spike source, with*

$$J_{T,1}^{(B)} = \frac{B(B+1)}{2B+1} - \frac{B(B+1)}{3T}. \tag{30}$$

Fix $B \in \{4, 5\}$, put $p = 2B + 1$, and scale the candidate dual offset as $\tau_T = T\theta_T$. Let T_0 be the first certified horizon and R_{run} the maximum run length needed by a maximizing mask. The per-period target increment is $\kappa = \tau_{T+p} - \tau_T$. The exact candidates have

$$\begin{aligned}(B, p, T_0, R_{\mathrm{run}}, \kappa) &= (4, 9, 19, 11, 45/16),\\ (B, p, T_0, R_{\mathrm{run}}, \kappa) &= (5, 11, 23, 14, 154/25),\end{aligned} \tag{31}$$

Their nonzero weights are nonnegative rainbow B-sets of span at most p, with periodic bulk and fixed reflected endpoint corrections. The complete coefficient tables are in `research/paper/proof_certificate/b45_coefficients.json`; the standalone `verify.py` in that directory reconstructs every residue's transitions and base vectors. The JSON tables and verifier are part of the submitted artifact and discharge only finite arithmetic obligations; state sufficiency and max-plus induction remain the analytic steps below. The reference command is `pythonverify.py--partb45--outputverification.json` with Python 3.10 and SymPy 1.13.3. Equivalently,

$$\tau_T = \frac{5}{16}(T - 19) \quad (B = 4), \qquad \tau_T = \frac{14}{25}(T - 23) \quad (B = 5).$$

Each dual is specified by bulk, left-end, and right-end coefficient tables. A bulk entry (r, s, w) assigns weight w to $a + s$ for every legal anchor $a \equiv r \pmod p$. Left corrections are fixed supports; a right correction is a left template reflected by $i \mapsto T - 2 - i$. Equal supports coalesce. Endpoint corrections may be signed, but every expanded horizon weight must be nonnegative. The B4 table has 65 bulk, five left, and six right entries; B5 has 127, four, and seven. For example, the B4 entry $(0, (0, 1, 2, 3), 38/27)$ assigns $38/27$ to $\{a, a + 1, a + 2, a + 3\}$ for every legal $a \equiv 0 \pmod 9$. Translating the bulk entries and adding the endpoint corrections defines the dual at each horizon. The expanded weights are nonnegative, are supported on rainbow sets of span at most p, and retain the bulk periodicity. It remains to establish the complete capacity inequalities for every $T \geq T_0$.

Uniform run bound. Let D have selected runs of lengths ℓ. Its scaled K1 capacity is $F_T(D) = \sum_\ell \lfloor (\ell + 1)^2/4 \rfloor$. Removing the central boundary of a run of length ℓ, with $a = \lfloor (\ell - 1)/2 \rfloor$ and $b = \ell - 1 - a$, reduces capacity by

$$\Delta_F(\ell) = \left\lfloor \frac{(\ell+1)^2}{4} \right\rfloor - \left\lfloor \frac{(a+1)^2}{4} \right\rfloor - \left\lfloor \frac{(b+1)^2}{4} \right\rfloor.$$

The lost support reward is at most $79/4$ for B4 and $156/5$ for B5, the exact all-length incident-weight bounds including positive endpoint corrections. The four parity formulas for Δ_F show strict domination from $\ell = 12$ and 15, respectively. Hence no maximizing mask has a run longer than $R_{\mathrm{run}} = 11$ or 14.

Why scalar prefix deletion is insufficient. Deleting one leading period and comparing each large mask only with its shifted suffix has exact maximum defect $57/16$ at B4 and $502/75$ at B5. These exceed κ by $3/4$ and $8/15$. Thus a maskwise inequality that discards the smaller prefix state's slack is false. The componentwise state comparison below retains the information needed for induction.

Finite-state decomposition. Process boundaries from left to right. A state is $s = (u, r)$, where u is the last p selection bits and $0 \le r \le R_{\mathrm{run}}$ is the current run length. A transition omits or selects the next boundary, subtracts the incremental run capacity, and adds every support weight whose final required bit is reached. The fields are constrained: if $u \neq 1^p$, then r is the trailing-one count of u; if $u = 1^p$, then $r \in \{p, \dots, R_{\mathrm{run}}\}$. Thus the fixed domain has

$$|\mathcal{S}| = 2^p - 1 + (R_{\mathrm{run}} - p + 1),$$

which is 514 states at B4 and 2,051 at B5. The span and run bounds make this state sufficient: equal states have identical legal next symbols, incremental run penalties, and newly completed support rewards.

For a concrete B4 transition, let the recent-bit field be $u = \texttt{001101111}$ and let $r = 4$. Appending zero produces $(\texttt{011011110}, 0)$; appending one produces $(\texttt{011011111}, 5)$. In either branch, the transition subtracts the run- capacity increment and adds exactly those support weights whose rightmost bit has just arrived. Every support has span at most $p = 9$, so no older bit can affect this reward. This example makes the state invariant explicit: u determines newly completed supports, while r determines the run penalty.

For every horizon residue $j \in \{0, \dots, p-1\}$, exact translation of the support tables decomposes the transition sequence into a fixed p-step left operator, j pure-bulk steps, q copies of a fixed p-step bulk operator A_j, and a fixed p-step right operator. Let $v_q(s)$ be the max-plus value vector immediately before the right operator. Direct exact bases give, for every residue, an index $q_0 \le 1$ satisfying

$$v_{q_0+1}(s) \le v_{q_0}(s) + \kappa \quad \text{for every reachable state } s. \tag{32}$$

The finite comparison covers all nine B4 residues over 514 states and all eleven B5 residues over 2,051 states. Every reachable successor has a corresponding predecessor value, and the maximum component slack in Equation (32) is exactly zero for each residue. These inequalities follow from the exact post-prefix vectors, sparse period transitions, fixed right-end transitions, and both finite scalar bases at the first qualifying $q_0 \le 1$. A separate implementation reconstructs each object directly from the support templates.

The exact transitions establish the state-sufficiency and fixed-operator hypotheses of Theorem 16. Equation (32) is its cycle inequality. Exact finite bases supply the initial equalities, while every canonical phase-complement mask is saturated and its objective and τ_T both increase by κ under $T \mapsto T + p$. Writing M_T for the maximum reward minus capacity at horizon T, the theorem gives $M_T = \tau_T$ for all $T \ge T_0$.

All complete capacities are therefore feasible. Every phase equality also holds, and the dual-mass identity in the critical-rainbow argument makes the dual objective equal to the critical-period primal. Direct phase averaging gives

$$J^{(B)}_{T,1} = \frac{B(B+1)}{2B+1} - \frac{B(B+1)}{3T}.$$

Critical rainbow duality proves Theorem 17.

A.12 ALL-LENGTH K2 PACKING

Closed block cost. For the uniform two-spike source, define

$$A_1(m) = \lfloor m^2/4 \rfloor, \tag{33}$$

$$r_m = \min\{r \ge 1 : r(r+1) \ge \tbinom{m}{2}\}, \tag{34}$$

$$A_2(m) = 2 \sum_{j=1}^{m-1} j|r_m - j|. \tag{35}$$

The block cost scaled by $\binom{T}{2}$ is

$$h_T(m) = (T-m)A_1(m) + A_2(m). \tag{36}$$

For a block with one of the two global spikes, summing the conditional single-spike medoid cost gives $(T-m)A_1(m)$. If both spikes lie inside, rank separation gives reflected weighted-median problems. One rank has weights $1, \dots, m-1$; its median condition is

$$r(r-1) \le \binom{m}{2} \le r(r+1),$$

which yields r_m and $A_2(m)$. This proves Equation (36). The equivalent polynomial form is

$$A_2(m) = \frac{m(m-1)(2m-1) - 3r_m m(m-1) + 2r_m(r_m+1)(r_m-1)}{3}.$$

Explicit weights. Let $n = T-1$ and $a = T-2$. For $j = 0, 1, 2$, define reflection-paired endpoint classes

$$U_j = \{(j, j+1), (n-2-j, n-1-j)\},$$
$$X_j = \{(j, j+2), (n-3-j, n-1-j)\}.$$

Every interior adjacent pair $(i, i+1)$, $3 \le i \le n-5$, receives scaled weight $W_T = 2a - 2/3$. The distance-two endpoint weights are

$$x_0 = a, \qquad x_1 = 2a/3, \qquad x_2 = (a+2)/3.$$

The adjacent endpoint weights u_0, u_1, u_2 use the seeds

T	u_0	u_1	u_2
13	$44/3$	14	17
14	16	$46/3$	$56/3$
15	$52/3$	$50/3$	$61/3$
16	$56/3$	18	22
17	20	$58/3$	$71/3$

and recurrence

$$(u_0, u_1, u_2)(T+5) = (u_0, u_1, u_2)(T) + (20/3, 20/3, 25/3).$$

Assign u_j to U_j, x_j to X_j, and zero elsewhere. All supported pairs have distance one or two and are rainbow modulo five.

Run capacity. For a selected run of s uncut boundaries, put

$$P(s) = A_1(s+1), \quad Q(s) = (s-1)P(s) - A_2(s+1), \quad H_T(s) = aP(s) - Q(s).$$

The base slack after assigning W_T to adjacent pairs is $S_T(s) = H_T(s) - (s-1)W_T$. Direct substitution gives

$$S_T(1) = a, \quad S_T(2) = 2/3, \quad S_T(3) = 4/3, \quad S_T(4) = 0.$$

For $s \ge 5$, $a \ge s-1$ and $P(s) - 2(s-1) \ge 1$. To make the needed pairwise-distance bound explicit, let $M = \sum_{j=1}^{m-1} j = \binom{m}{2}$. For any integer r, the triangle inequality gives

$$\sum_{j,k=1}^{m-1} jk|j-k| \le \sum_{j,k=1}^{m-1} jk\big(|j-r| + |k-r|\big) = 2M \sum_{j=1}^{m-1} j|j-r|.$$

Taking $r = r_m$, using $A_2(m) = 2\sum_j j|j-r_m|$, dividing by M, and evaluating the finite polynomial sum yields

$$A_2(m) \ge \frac{(m-2)(m+1)(2m-1)}{15}$$

For the substitution $m = s + 1$, put $d = a - (s - 1) \geq 0$. The definitions give

$$S_T(s) = d\big[P(s) - 2(s-1)\big] + A_2(s+1) - 2(s-1)^2 + \tfrac{2}{3}(s-1)$$
$$\geq d\big[P(s) - 2(s-1)\big] + \frac{(s-1)(2s^2 - 25s + 42)}{15}.$$

Both terms are nonnegative for $s \geq 11$, proving $S_T(s) \geq 0$ on the unbounded range. For $s = 5, \ldots, 10$, the exact lower margins are respectively

$$2/3,\ 16/3,\ 12,\ 68/3,\ 124/3,\ 68.$$

The same bound yields $S_T(s) \geq S_T(5) = a - 10/3$. Its remaining exact margins for $s = 6, \ldots, 11$ are $11/3, 28/3, 19, 110/3, 187/3, 92$. For the full run $n \geq 12$, it also gives $S_T(n) \geq 2S_T(5)$.

View the final packing as the uniform base plus a correction in each five-boundary endpoint window. For a left-window selection pattern $z = (z_0, \ldots, z_4) \in \{0,1\}^5$, the selected left correction is

$$\Delta_T(z) = \sum_{j=0}^{2} (u_j - W_T) z_j z_{j+1} + x_0 z_0 z_2 + x_1 z_1 z_3 + x_2 z_2 z_4.$$

Let $\mathcal{R}^-(z)$ contain the maximal one-runs in z that end before position four, and define the conservative allocation

$$R_T(z) = \sum_{I \in \mathcal{R}^-(z)} S_T(|I|) + \mathbf{1}[z = (1,1,1,1,1)] S_T(5).$$

Thus no slack is assigned to a suffix run that may continue into the interior. For each seed $T_0 \in \{13, \ldots, 17\}$ and $T = T_0 + 5q$, both allocated slack and $\Delta_T(z)$ are affine in q. Exact expansion of all $5 \cdot 32$ cases gives nonnegative intercept and slope. The minimum intercept and five-step increment margins are

T_0	$\min_z\{R_{T_0}(z) - \Delta_{T_0}(z)\}$	$\min_z$ five-step increment
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0

Inactive endpoint patterns attain zero. Exact expansion of the 160 affine rows establishes all 320 intercept and slope inequalities. The right endpoint follows by reflection. A completed local run cannot connect the two windows. Hence the allocations are disjoint unless both endpoint patterns are all ones and their suffix/prefix runs connect; this forces the full run $D = E$, in which case $S_T(n) \geq 2S_T(5)$ pays both allocations. Hence all subset capacities hold.

Phase equality and onset. For each residue, both sides of Equation (22) are quadratic in $q = (T - T_0)/5$: the count of interior pairs or length-five blocks is affine, and their weight or cost is affine. Exact equality at $q = 0, 1, 2$ for each seed proves the identity for all $q \geq 0$. Rainbow duality therefore proves Theorem 8 for $T \geq 13$. Complete rational LP certificates cover $T = 8, \ldots, 12$. At $T = 7$, exact primal–dual solutions give $51/35 < 22/15$, establishing the sharp onset. The optimal mixture assigns mass $9/10$ to the uniformly phased period-six kernel and $1/10$ to the full-cut identity kernel. The pure period-six cost is $34/21$, so the mixture costs $(9/10)(34/21) = 51/35$; its maximum boundary-hit probability is $(9/10)(1/3) + 1/10 = 2/5$.

A.13 Residue-class K3 packing

Put $T = 5q + 3$ and scale the exchangeable K3 objective by $M_T = \binom{T}{3}$. Let $A_k(m)$ be the total rankwise lower-medoid cost over all k-subsets of a block of length m. The scaled cost contributed by a length-m partition block is

$$H_T(m) = \sum_{k=0}^{3} \binom{T-m}{3-k} A_k(m). \tag{37}$$

This is the K3 analogue of Equation (36): choose the k spikes inside the block, choose the other $3 - k$ spikes outside it, and sum the exact block-medoid transport.

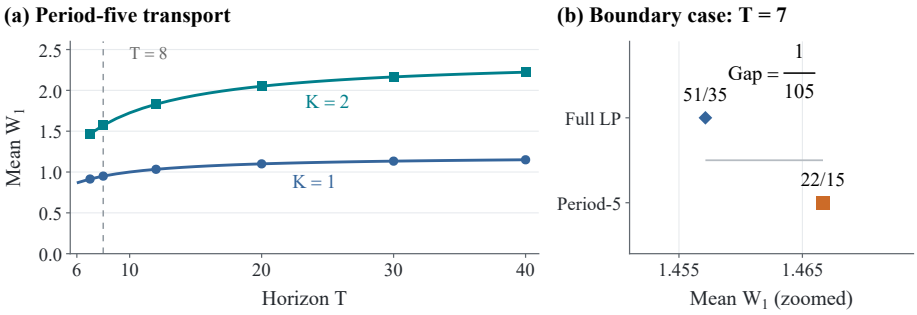

Figure 2: Exact $B = 2$ objectives and the sharp K2 onset. The K1 formula is globally optimal from $T = 6$; the K2 period-five formula is globally optimal from $T = 8$. At $T = 7$ the unrestricted LP improves period five by exactly $1/105$; panel (b) uses a zoomed horizontal scale.

Six-class rainbow dual. Reflection maps the left endpoint i of a distance-d pair to $T - 2 - i - d$. For $d \in \{1, 2\}$, define $c_d = 3(1 - d) \bmod 5$ and the three reflection orbits

$$O_{d,0} = \{c_d\}, \qquad O_{d,j} = \{c_d - j, c_d + j\} \bmod 5, \quad j \in \{1, 2\}.$$

A pair $(i, i + d)$ has type (d, j) when $\{i, T - 2 - i - d\} \bmod 5 = O_{d,j}$. Every such pair is rainbow modulo five. Assign every type-(d, j) pair weight $\lambda_{d,j} = u_{d,j}/M_T$, where

d	j	$u_{d,j}$
1	0	$(75q^3 + 160q^2 + 41q - 4)/(6(q + 1))$
1	1	$(75q^3 + 100q^2 - 7q + 8)/(6(q + 1))$
1	2	$(75q^4 + 165q^3 + 35q^2 - 11q + 4)/(6q(q + 1))$
2	0	$(75q^4 + 15q^3 + 65q^2 + 37q - 8)/(6q(q + 1))$
2	1	$(75q^3 + 20q^2 - 11q + 4)/(6(q + 1))$
2	2	$5q(5q + 1)/2$

After $q = z + 1$, every numerator and denominator has nonnegative coefficients, so all weights are nonnegative. Direct residue counting gives zero slack in all five phase equalities. Summing the phase objectives gives

$$\frac{375q^3 + 125q^2 + 40q - 2}{5M_T},$$

which is Equation (15).

Capped-run capacities. For $D \subseteq E$, define the scaled capacity defect (all quantities below are in the $M_T = \binom{T}{3}$ scaled units)

$$\Delta_q(D) = M_T \left(\sum_{\{i,j\} \subseteq D} \lambda_{ij} - F(D) \right).$$

Only distance-one and distance-two pairs occur. Gaps of at least two unselected boundaries split D into components with additive defects. Within a component, selected runs are separated by exactly one unselected boundary. Write $w_d(s) = M_T \lambda_{s,s+d}$ and define the initial-run reward

$$a_s(\ell) = \sum_{d=1}^{2} \sum_{i=0}^{\ell-d-1} w_d(s + i) - H_T(\ell + 1), \qquad \ell \in \{1, 2, 3, 4\},$$

where an empty sum is zero. The state $e \in \{0, \ldots, 4\}$ is the residue of the last selected boundary. Appending a run of length ℓ sends it to $e + \ell + 1 \bmod 5$, adds reward $w_2(e) + a_{e+2}(\ell)$, and consumes $\ell + 1$ physical positions. This five-state multigraph represents every capped-run component. Its elementary cycles have at most five states and consume positive multiples of five positions.

Keeping the physical span yields four successive cycle tiers. After scaling rewards by $6q(q+1)$ and shifting $q = z + 16$, all normalized cycle and access/exit polynomials have nonpositive coefficients. The only zero cycles are respectively the self-loops indexed by (e, ℓ) as $(0, 4), (4, 4)$; $(2, 4)$; $(1, 4), (3, 4)$; and the five-state all-three cycle. Their exact per-period means are

$$\mu_1 = 75q^3 + 90q^2 + 15q, \qquad \mu_2 = 75q^3 + 80q^2 + 37q - 8,$$
$$\mu_3 = 75q^3 + 20q^2 - 11q + 4, \qquad \mu_4 = (50q^3 + 40q^2 + 70q)/4.$$

For coverage, classify a walk by the first tier whose critical self-loop it uses, and split at that loop's state. After deleting critical self-loops, erase repeated-state cycles from the prefix and suffix separately. Each erased cycle has nonpositive reward after subtracting its number of periods times μ_i, so erasure preserves an upper bound. The remaining prefix and suffix are simple paths through the critical state; all such access/exit pairs are enumerated. For start residue s and end residue e, their total physical span is at most $5q + e - 5\mathbf{1}[e > 1] - s + 1$. The positive μ_i bounds unused periods. Walks avoiding every removed self-loop use the final tier's simple-path endpoint bound. These cases cover all component walks and hence all masks. The standalone proof package gives the exact row-generation rules and reconstructs every polynomial. The largest shifted cycle and endpoint coefficients are zero (the final-tier endpoint maximum is $-125/2$). Exact dynamic programs close $q = 3, \ldots, 15$; the $q = 2$ base enumerates all 2^{12} boundary masks. Hence every capped-run mask has $\Delta_q(D) \le 0$ for every $q \ge 2$. Period-five phase complements attain zero, so this is a capacity certificate rather than only failure to find a violator.

Reduction of arbitrary runs. Consider a selected run of length $n - 1$ and delete its third boundary. Its partition block of length n becomes blocks of lengths three and $n - 3$, so the scaled capacity loss is

$$L_T(n) = H_T(n) - H_T(n-3) - H_T(3).$$

The largest removed incident dual mass, in the same scaled units, is

$$I_2 = \frac{150q^3 + 35q^2 + 19q - 2}{3q}.$$

Thus deletion cannot decrease Δ_q once $L_T(n) \ge I_2$. To prove this inequality without a horizon cutoff, set $u = T - n$ and $P_T(n) = L_T(n) - L_T(6)$. Exact expansion gives

$$P_T(n) = \alpha(n)\binom{u}{2} + \beta(n)u + \gamma(n),$$

where $\alpha(n) = 3(n-6)/2$ for even n and $\alpha(n) = (3n-19)/2$ for odd n. The special case $n = 7$ is $P_T(7) = \binom{u}{2} + 3u - 2 > 0$ for $u \ge 1$.

The coefficient of u is explicitly

$$\beta(n) = A_2(n) - A_2(n-3) - 3A_1(n-3) - 7n + 20.$$

Let $r(s)$ be the least integer satisfying $r(s)(r(s)+1) \ge \binom{s}{2}$. Then $A_1(s) = \lfloor s^2/4 \rfloor$ and

$$A_2(s) = \big(2s^3 - 3s^2 r - 3s^2 + 3sr + s + 2r^3 - 2r\big)/3, \qquad r = r(s).$$

Substitution in the defining quadratic gives $2(s-1)/3 \le r(s) \le \lfloor (3s-1)/4 \rfloor$ for $s \ge 4$; $s = 3$ is direct. Comparing the quadratic at $r(s)+1$ and $r(s)+3$ gives $r(s+3) - r(s) \in \{2, 3\}$. With $s = n - 3$, substituting the upper bound for increment two and the lower bound for increment three in $\beta(n)$ yields

	$r(n) - r(n-3) = 2$	$r(n) - r(n-3) = 3$
n even	$(n-6)(2n-7)/2$	$(n-4)(11n-62)/12$
n odd	$(4n^2 - 38n + 81)/4$	$(11n^2 - 106n + 239)/12$

as lower bounds. The quadratic in r is minimized at the upper endpoint for increment two, and is increasing above the lower endpoint for increment three. The displayed bounds are nonnegative for even $n \ge 6$ and odd $n \ge 7$; $n = 6$ is direct. Eliminating the cubic outer-median term gives four lower bounds on $\gamma(n)$, indexed by the parity of n and by whether the outer median changes under $n - 3 \mapsto n$:

Parity	Change	Lower bound on $\gamma(n)$	Onset
even	0	$(87n^3 - 1739n^2 + 10873n - 23729)/288$	11
even	1	$(203n^3 - 3888n^2 + 23832n - 51912)/648$	10
odd	0	$(87n^3 - 1775n^2 + 11053n - 24269)/288$	12
odd	1	$(203n^3 - 3969n^2 + 24237n - 53127)/648$	11

Each polynomial has nonnegative coefficients after shifting by its displayed onset. Direct medoid recomputation gives $\gamma(6), \ldots, \gamma(11) = 0, -2, 15, 28, 70, 116$, closing the finite gap. Finally,

$$L_T(6) - I_2 = \frac{75q^3 - 175q^2 + 34q + 4}{6q}.$$

For $q = z+3$, its numerator is $75z^3+500z^2+1009z+556 > 0$. At $q = 2$, the required $n = 7, \ldots, 13$ cases are exact and have minimum margin $86/3$.

Deleting the third selected boundary maps a run of length ℓ to runs of lengths two and $\ell - 3$. For $q \geq 3$, repeated deletion of every $\ell \geq 5$ strictly decreases total excess above four and terminates in the capped-run family. For $q = 2$, deleting every $\ell \geq 6$ terminates in the exact $T = 13$ base. Because no deletion decreases the defect, a positive unrestricted defect would imply a positive capped-run or base defect, which the preceding certificates exclude. All complete capacities and phase equalities therefore hold. Critical rainbow duality and weak duality prove Theorem 9. Independent constructions give the same block medoids, six pair classes, polynomial coefficients, deletion margins, and finite base values.

Failed extensions of the fitted dual. At $T = 8$, Period-5 is optimal by a separate full-LP dual, but the symbolic residue-three dual above has four capacity violations. A separately fitted residue-zero quadratic selector feasible at $T = 10, 15, 20$ has 12 violated capacities at its held-out horizon $T = 25$, with minimum normalized slack $-53767/6555000$. This rejects that selector, not the candidate kernel. Automatic span-two synthesis at $T = 25, K = 3$ instead yields an exactly repaired global lower bound within $11/4600000000$ of Period-5, illustrating why a failed formula need not indicate an inferior kernel.

A.14 Complement symmetry and mixture closure

Corollary 18. *At $B = 2$ and $\delta = 2/5$, one common period-five kernel is optimal for every exchangeable source supported on $\{0, 1, 2, T-2, T-1, T\}$, analytically for $T \geq 13$ and with exact finite verification for $T = 8, \ldots, 12$. For $T = 5q + 3$ and $q \geq 2$, the support may be enlarged to $\{0, 1, 2, 3, T-3, T-2, T-1, T\}$.*

Proof of Corollary 18. Represent a capacity-one state in a block of length m by its binary indicator and let A_t and Q_t be the two prefix counts through position t. The one-dimensional transport identity gives

$$W_1(a, q) = \sum_{t=0}^{m-2} |A_t - Q_t|.$$

For the binary complements a^{c} and q^{c}, the prefix difference is $(t+1-A_t)-(t+1-Q_t) = -(A_t-Q_t)$. Hence $W_1(a^{\mathsf{c}}, q^{\mathsf{c}}) = W_1(a, q)$. Complementation bijects the k-event block states with the $(m-k)$-event block states, so their optimal medoid costs are equal.

Applying this identity independently in every partition block shows that the fixed-count partition objectives obey $f_{T,K}(C) = f_{T,T-K}(C)$ for every cut set C. Thus the K1 and K2 optima also cover $K = T-1$ and $K = T-2$; on the residue family in Theorem 9, K3 also covers $K = T-3$. The $K = 0$ and $K = T$ objectives vanish for every cut set.

Finally, let an exchangeable source have count probabilities π_K on the six stated counts. Its partition objective is $f_\pi(C) = \sum_K \pi_K f_{T,K}(C)$. If $\mu_\star$ is the common period-five optimizer and μ is any feasible partition distribution, componentwise optimality gives

$$\sum_C \mu_C f_\pi(C) = \sum_K \pi_K \sum_C \mu_C f_{T,K}(C) \geq \sum_K \pi_K \sum_C (\mu_\star)_C f_{T,K}(C).$$

Therefore $\mu_\star$ minimizes every such mixture. The K1 and K2 analytic proofs establish the six-count statement for $T \geq 13$; the exact K2 certificates close $T = 8, \ldots, 12$. When $T = 5q + 3$ and $q \geq 2$, Theorem 9 adds the K3 and $T - 3$ components and proves the eight-count statement. □

B Experimental Methods and Additional Results

Sections B.1–B.2 evaluate kernel optimization; Sections B.3–B.7 give the classifier protocols, baselines, and fitting controls. Native-event comparisons follow in Section B.9, with timing and attack controls in Sections B.10 and B.11. Section B.12 indexes the data, code, and exact certificates.

B.1 Compact synthesis across sources and spans

We compare the compact formulation with independently constructed all-partition LPs using the same collision rows. The comparison includes 54 controls at $T = 5, \ldots, 10$, $B = 1, 2, 3$, and $R = \min(B, T - 2)$, using uniform $K = 1$, uniform $K = \lfloor T/2 \rfloor$, and a position-dependent $K = 1$ source. Another 108 use $T = 5, \ldots, 8$, the same budgets and source families, full span $R = T - 2$, and $\delta \in \{1/3, 2/5, 1/2\}$. All 162 objective differences are below 10^{-7}.

The larger uniform grid fixes $T \in \{13, 25, 41\}$, $K \in \{1, 2, 3, \lfloor T/2 \rfloor\}$, $B = 2$ at $R = 2, 4$, and $B = 3$ at $R = 3, 6$, giving 48 configurations. Six further configurations use the same horizons with $K = 1$, $B = 2$, $R = 2, 4$, and $p_t \propto 1 + (t \bmod 4)$ under a fixed lower uniform-median decoder. The collision level is $\delta = B/(2B + 1)$ throughout this larger grid. The primary candidate is the uniformly phased period-$(2B + 1)$ kernel. We report every configuration, including positive and unresolved gaps.

Numerical solutions and rational certificates. SciPy 1.11.4/HiGHS proposes multipliers under a 30-second per-solve limit. Rounding nonnegative weights to multiples of 10^{-9} and recomputing the intercept with exact integer-scaled arithmetic gives feasible lower bounds. Path-flow masses are rationalized, normalized, and repaired as in Appendix A.1. An independent whole-block verifier reconstructs small interval costs by enumerating source states; it agrees with the one-boundary DAG on every dual minimum and local primal bracket, and confirms every exhaustively checked capacity. The largest local bracket width is 1.59×10^{-8}. In Table 2, repaired lower bounds round to the displayed candidate costs at six decimals except at $(T, K, B, R) = (25, 3, 3, 3)$, where the lower bound is 4.036025. Gap bounds use the unrounded rational certificates.

Global optimality from local constraints. All 24 uniform B2 configurations certify Period-5 within 6.4×10^{-9}. In particular, span two succeeds at $T = 25, 41$ for $K = 3$ and $K = \lfloor T/2 \rfloor$, outside the current closed-form families. At $T = 41, R = 2$, the DAG has 901 states and 1,758 edges, versus 2^{40} implicit partition columns. Increasing to $B = 3, R = 6$ uses 2,899 states. The finite grid leaves the sufficiency of span two at arbitrary B2 horizons open.

Insufficient locality versus an inferior candidate. At $T = 25, K = 3, B = 3$, the span-three local primal costs about 4.036025, strictly below Period-7's 4.055901, but violates a global collision row. Consequently every span-three dual falls short of Period-7's cost, even though the cheaper local mixture is globally infeasible. Span six certifies Period-7 within 4×10^{-9}. More strongly, the exact span-three local upper bound $3249/805$ and span-six global lower bound $40812499963/10062500000$ imply $J^* - L_3 \geq 199999963/10062500000 > 0.01987577$, a strictly positive gap between the local and global optima. Conversely, at $B = 3$ and $(T, K) \in \{(13, 3), (13, 6), (25, 12), (41, 20)\}$, the span-six flow yields a globally feasible mixture that genuinely improves on Period-7 by 1.603%, 3.168%, 1.259%, and 0.591%, respectively. The recovered mixture is the calibrated alternative $(6/7)\,\mathrm{Period9} + (1/7)\,\mathrm{Identity}$ up to rational repair, with matching full-LP lower bounds to numerical precision.

For the nonuniform source, span-four flows recover nonperiodic five-color kernels improving on Period-5 by 3.106%, 2.346%, and 2.065% at $T = 13, 25, 41$. All have verified global gaps below 10^{-8}. Global verification checks every size-B collision row, which dominates smaller rows by nonnegative masses. Choosing the better globally feasible candidate closes 43 of 54 gaps within 10^{-6}; the remaining 11 stay inconclusive at their prescribed spans. The full grid is retained.

Longer horizons. We extend synthesis to $T = 64, 128, 256$ at $B = 2$, $\delta = 2/5$, and $R = 2, 4$. Each horizon uses three fixed sources: uniform $K = 1$, uniform $K = 3$, and the position-dependent $K = 1$ law above. All 18 cases run in fresh sequential processes on the Ryzen 7 5700X, with one BLAS/HiGHS thread, a 15-second solver limit and a 90-second process limit. No cases are replaced. A separate $T = 8$ control agrees with the full partition LP.

Table 4: Long-horizon synthesis at $B = 2$. Each time (seconds) and peak working set (MiB) is the maximum over three prescribed source laws; column maxima can come from different cases. Repair combines dual repair and primal recovery. Verification includes the independent DP and global collision checks. Closed counts global gaps at most 10^{-6} out of three. A dash denotes no certificate after the 15-second solver limit.

T	R	Graph	Solve	Repair	Verify	MiB	Closed
64	2	0.02	0.08	0.02	0.17	66.2	2/3
64	4	0.04	0.32	0.03	0.18	69.4	3/3
128	2	0.05	0.52	0.07	1.35	93.8	3/3
128	4	0.10	2.96	0.09	1.69	103.4	3/3
256	2	0.42	4.26	0.31	12.42	201.3	3/3
256	4	0.29	≥ 15	—	—	221.4	0/3

At $T = 256, R = 2$, the LP has 33,151 states and 229,750 matrix nonzeros. All three sources obtain global gaps below 6.6×10^{-8}. The corresponding span-four solves reach the time limit before producing certificates. At $T = 64$, the nonuniform source leaves a span-two global gap of 0.0225, which span four closes exactly. Overall, 15 cases produce verified bounds and 14 close the global gap to 10^{-6}.

Table 4 separates synthesis from verification. The nonuniform $T = 256, R = 2$ verification takes 12.42 seconds, compared with 4.26 seconds for solving: global checks can dominate even when the local formulation is small. Every size-two collision row is checked; recovered flows are globally repaired before contributing an upper bound. The CSV records cost construction, LP assembly, matrix dimensions, local and global gaps, and both worker and process-tree memory for every case. Memory is the worker's operating-system peak working set, including Python and libraries. These single-run timings characterize this implementation, not a speedup over another solver.

B.2 Compact design for a natural event source

Empirical interval costs incorporate the full count and timing distribution of the official DVS128Gesture split: 1,176 training clips from 23 subjects and 288 test clips from six other subjects. Ten-bin integration by event number followed by thresholding retains all $2 \times 128 \times 128$ pixel-polarity lines per clip. Event conservation concerns these binary inputs; preprocessing can merge raw sensor events.

The decoder is fixed before fitting: equal probabilities on the lower and upper uniform-conditional block medoids, with one shared orientation for each transform. If $H(m)$ counts training lines with binary mask m, the exact empirical interval cost is

$$c(a,b) = \frac{1}{2\sum_m H(m)} \sum_m H(m) \sum_{o \in \{\text{lower},\text{upper}\}} W_1(m \cap [a,b), q_o(a,b,|m \cap [a,b)|)) \,.$$

Here q_o is the fixed canonical configuration in the interval. Linearity of expectation makes these costs additive across blocks even when lines are dependent. We fit only the partition law, at $B = 2$ and $\delta = 2/5$, with prescribed spans $R = 2, 4, 8$. The deployment rule compares the globally repaired full-span law with Period-5 using exact training cost, breaking ties in favor of Period-5. The selected law is frozen before loading the test clips.

All-count finite-source certificate. At span two, 64 states and 89 LP variables bound Period-5's suboptimality by less than 8.2×10^{-10} per line. Spans four and eight give bounds below 1.1×10^{-9} after rational repair and agree with the all-512-partition LPs within 10^{-7}. Rounding explains the variation across repaired bounds; the exact hierarchy is monotone. Training selection retains Period-5 (Table 5).

Table 5: DVS-Gesture geometry under the fixed decoder at $B = 2$, $\delta = 2/5$. Values are mean total event displacement per clip. The fitted selection equals Period-5; no classifier is evaluated in this experiment.

Source	Period-5	Frozen selection	All-or-none
Training, 1,176 clips	22,880.60	22,880.60	33,389.01
Test, 288 clips	25,723.31	25,723.31	36,740.99

Only 13.03% of the test events lie on one- or two-event lines, so most event mass is outside the closed-form low-count families. The optimum concerns the training histogram; held-out costs are reported for each of the six test subjects. An independent implementation rebuilds masks and medoids from all 1,464 clips, reproduces the per-subject costs, and checks all 512 partition capacities and 36 size-two collision rows at each span. Including loading, this CPU check takes 21.4 seconds; each solve and rational/full-column check takes less than 0.05 seconds in this single run.

B.3 Models, inputs, and evaluation

Latency encoding and training. We use the official MNIST split of 60,000 training and 10,000 test images. A pixel of normalized intensity $v > 0$ emits one spike at $t(v) = \lfloor 9(1 - v) + 1/2 \rfloor$ in a ten-bin horizon; zero-valued pixels remain silent. The classifier consists of two Conv–BN–LIF–MaxPool stages with 32 and 64 channels, 3×3 convolutions, unit padding, and pooling factor two. A 128-unit LIF layer follows the flattened $64 \times 7 \times 7$ representation, then a ten-class linear readout. Convolutional and linear layers have no bias. LIF nodes have $\tau = 2$, threshold one, reset zero, `decay_input=False`, and `PiecewiseLeakyReLU(c=0)`. Logits are averaged over time, and network state is reset between evaluations.

Training uses 12 epochs, batch size 256, cross-entropy on averaged logits, AdamW with learning rate 10^{-3} and weight decay 10^{-4}, cosine decay to 10^{-5}, and CUDA automatic mixed precision. We retain only the final checkpoint for each seed: 20260731, 20260732, and 20260733. Tables abbreviate these seeds as 731, 732, and 733. There is no early stopping, model sweep, or test-based replacement. Fashion-MNIST uses the same architecture and training schedule in a single-seed pilot; its limited absolute accuracy is reported without extrapolating to multiple seeds.

Three-spike transfer. The $T = 13$ experiment reuses these ten-bin-trained checkpoints. For an active pixel with original latency t, define

$$a(t) = 1 + \left\lfloor \frac{10t}{9} + \frac{1}{2} \right\rfloor, \qquad x = \{a(t) - 1, a(t), a(t) + 1\}.$$

Silent pixels remain silent, and each active line contains three distinct, adjacent spikes. This adjacent-triplet source differs from the theorem's uniform law over three-element subsets while matching its horizon and count. The experiment measures transfer to the new representation with the ten-bin-trained model and decoder held fixed: the recurrent dynamics run for 13 steps, and the prediction uses the average of those logits.

Metrics and randomness. All finite kernel atoms are evaluated exactly, so the reported votes have no Monte Carlo sampling error. Smoothed accuracy measures correctness of the top vote. Certified accuracy measures the fraction of the entire population that is both correct and satisfies the boundary-subset certificate. Certificate rate includes certified errors. Class ties use the smallest index. Transport is the sum of rank-matched event displacements over all lines, averaged over inputs and kernel atoms; it is not normalized by event count unless explicitly stated. Thus transport weights high-event-count inputs more heavily than the input-weighted accuracy metrics. Kernel randomness is shared across lines, as required by the certificate. Three-seed entries in Tables 3 and 7 report means and sample standard deviations over training seeds evaluated on the same test set.

B.4 Kernel comparisons and fairness

Periodic and all-or-none kernels. Period-5 samples each of the five boundary residues with probability $1/5$. Each cut set has lower and upper medoid orientations, giving ten equally weighted

atoms. At $B = 2$, its maximum crossed-boundary collision mass is $2/5$. The all-or-none kernel (AON) assigns mass $3/5$ to the empty cut set and $2/5$ to the full cut set, with balanced orientations. Its ten-atom implementation duplicates some transforms. Equal nominal atom counts control the reported vote budget, but distinct projections and minimum computation differ: AON has at most three distinct transformed inputs, while Period-5 uses ten. We therefore do not interpret these comparisons as equal wall-clock or energy cost. AON is a collision-matched geometric control.

Train-fitted four-color kernel. At $T = 10$, the train-source p4 kernel uses the empty cut and the four color classes $\{0, 4\}$, $\{1, 6\}$, $\{2, 5, 8\}$, and $\{3, 7\}$, each with mass $1/5$. Balanced medoid orientations again give ten atoms. The coloring uses the training time histogram without labels and is not refit on test. The separate raw-event and Fashion control uses $\{4\}$, $\{1, 6\}$, $\{0, 3, 7\}$, and $\{2, 5, 8\}$ instead. These are different prescribed comparators and are not interchanged after evaluation.

DSSN with joint count repair. Our defined adaptation of DSSN (Levine & Feizi, 2021) uses $q = 1$ splitting, singleton-reveal atoms, and a joint repair that minimizes conditional transport while preserving the line count. Repair costs are fit to the official training histogram without labels. The primary comparison has ten nonempty atoms, each with mass $1/10$, and maximum B2 exposure $4/10$. Joint repair couples temporal coordinates and thus lies outside the coordinatewise barrier in Proposition 10. The baseline tests this particular adaptation, not all count-preserving extensions of DSSN.

For a line $x \subseteq [T]$, let $V_j \subseteq [T]$ be the positions revealed by atom j, sampled uniformly from ten atoms. The observation is $O_j(x) = (V_j, x \cap V_j, |x|)$. In the $q = 1$ splitting representation, revealed zeros and ones have values $1/4$ and $3/4$; hidden coordinates have value $1/2$. For an observed subset v and count k, define

$$\mathcal{C}_j(v, k) = \{z \subseteq [T] : |z| = k,\ z \cap V_j = v\}.$$

The training histogram gives a conditional law $P_j(\cdot \mid v, k)$ on this set, using a uniform law if its training mass is zero. Joint repair is

$$r_j(v, k) \in \operatorname*{arg\,min}_{y \subseteq [T],\, |y| = k} \mathbb{E}_{z \sim P_j(\cdot \mid v, k)} W_1(z, y).$$

Ties choose the smallest integer mask $\sum_{t \in y} 2^t$. The output preserves count but need not preserve the revealed bits. The same atom is shared across lines, and identical observations produce identical repaired outputs.

For the primary $m = 10$ kernel, $V_j = \{j\}$ for $j = 0, \ldots, 9$. Crossing boundary e can change only observations revealing e or $e + 1$. Thus a boundary set S exposes at most $|\bigcup_{e \in S} \{e, e + 1\}|/10$ of the atom mass. At $B = 2$, two nonadjacent boundaries attain the maximum $4/10$. For $m < 10$, the nonempty V_j form a partition of $[T]$ chosen to minimize training repair cost; the other atoms reveal nothing. The finite search breaks ties lexicographically in the canonical tuple of block masks, ordering blocks by their least member. Exposure is the fraction of atoms whose reveal sets intersect an endpoint of a crossed boundary.

To test sensitivity within this adaptation, we evaluate every prespecified member with exactly $m = 2, \ldots, 10$ nonempty reveal atoms and $10 - m$ atoms that reveal nothing. Every member uses ten nominal atoms. The members with $m = 5, \ldots, 10$ share Period-5's maximum B2 exposure of $4/10$, whereas $m = 2$ has maximum exposure $2/10$ and $m = 3, 4$ have maximum exposure $3/10$. Matching the maximum exposure does not imply identical exposure for every crossed-boundary subset. We therefore report the full frontier as a sensitivity analysis over the number of nonempty reveal atoms and the resulting exposure, transport, and classification utility. No test outcome selects m. These comparisons concern the specified repair family.

B.5 Exact geometry and source assumptions

Proposition 19 (Finite-horizon exchangeable-source optimality)**.** *For every $T = 8, \ldots, 14$ at $B = 2$ and $\delta = 2/5$, the same period-five primal solves the complete partition LP for every temporally exchangeable binary source.*

At $T = 13$, $K = 3$, $B = 2$, the complete LP over 4,096 partitions and 78 collision constraints agrees with the exact Period-5 cost $1789/715$; AON costs $192/55$. Appendix A.13 proves the Period-5

Table 6: Sensitivity frontier for the prescribed DSSN-JCR adaptation at $T = 10$, $B = 2$. Accuracy entries are three-checkpoint means (%). Every member uses ten nominal atoms; m is the number of nonempty reveal atoms. Members with $m = 5, \ldots, 10$ match Period-5's maximum exposure of $4/10$; $m = 2$ and $m = 3, 4$ have smaller maxima of $2/10$ and $3/10$, respectively. Thus the full frontier includes different maximum exposure levels. Period-5 has transport 149.394, smoothed accuracy 97.710%, and certified accuracy 93.387%.

m	Transport ↓	Smoothed acc.	Certified acc.	Maximum exposure
2	333.266	75.217	75.217	$2/10$
3	334.053	75.217	75.217	$3/10$
4	334.863	75.217	75.217	$3/10$
5	335.703	75.313	72.183	$4/10$
6	336.616	79.240	70.490	$4/10$
7	336.616	80.257	68.887	$4/10$
8	337.880	83.263	68.660	$4/10$
9	340.849	85.303	68.457	$4/10$
10	345.235	86.737	69.130	$4/10$

optimum under the uniform law on the $\binom{13}{3} = 286$ masks; these per-line objectives differ from the image-level transports in Table 7. Exact component duals for every count at $T = 8, \ldots, 14$ share the same Period-5 primal, so componentwise equality extends optimality to arbitrary nonnegative count mixtures; complement symmetry supplies reflected counts, while empty and full lines have zero distortion. For the two official latency-MNIST one-spike histograms, the fixed uniform decoder also recovers Period-5, and their train–test total variation is 0.0075; Proposition 14 covers their convex mixtures. The failed quadratic extension at $T = 25$ is documented with the three-spike proof in Appendix A.13.

B.6 Kernel utility across training seeds

Table 7 separates the baseline comparison at $T = 10$ from source fitting at $T = 13$. The accompanying seed tables retain smoothed accuracy, certified accuracy, and, where recorded, certificate rate (Section B.12).

Table 7: MNIST kernel comparisons on all 10,000 test images at $B = 2$. Accuracies (%) are means $\pm$ sample standard deviations over seeds 731–733. Transport is mean total displacement per image. Each comparison holds the classifier fixed and uses ten nominal kernel atoms with collision mass $2/5$. The $T = 13$ rows use the adjacent-triplet transfer code, not the uniform-K3 source of the residue-class theorem; the classifier is transferred from the $T = 10$ training code.

Kernel	Transport ↓	Smoothed acc.	Certified acc.
Baseline comparison: latency code, $T = 10$; base: 98.55 ± 0.18			
All-or-none	297.97	82.82 ± 18.16	73.86 ± 21.07
DSSN-JCR specified adaptation	345.23	86.74 ± 4.91	69.13 ± 6.90
Period-5	149.39	97.71 ± 0.05	93.39 ± 0.42
Source fitting: three-spike code, $T = 13$; base: 97.80 ± 0.42			
All-or-none	1091.68	95.35 ± 0.57	95.35 ± 0.57
Period-5	386.14	98.18 ± 0.48	96.98 ± 0.83
Train-source LP	380.97	98.19 ± 0.48	96.87 ± 0.85

The theorem-aligned closed-loop audit separately verifies the uniform-K3 objective $1789/715$, the complete-LP agreement, and the $2/5$ collision allowance before evaluating the transferred adjacent-triplet source below. At $T = 13$, Period-5 has higher certified accuracy than AON at all three seeds by 1.42, 2.35, and 1.11 points (mean 1.63, sample standard deviation 0.64 points). This is a descriptive three-seed effect summary, not an inferential population estimate. AON certifies all its predictions, including errors. Only aggregate counts and prediction digests are available for this transfer experiment, so paired-bootstrap intervals cannot be reconstructed. At $T = 10$, the p4

kernel gains 0.18–0.23 smoothed-accuracy points over Period-5 at every checkpoint, but its certified accuracy ranges from 49.43% to 87.60%, versus 92.93–93.75% for Period-5.

Source-matched uniform-K3 training. To align the classifier source with the residue-class theorem, we trained the same two-convolution architecture directly on $T = 13$ inputs. For every active pixel line, each training update samples one of the $\binom{13}{3} = 286$ three-spike masks uniformly and independently; silent lines remain silent. The three seeds are 20260910–20260912, with the same 12-epoch AdamW schedule as the latency-MNIST replication. The primary evaluation uses a fixed uniform K3 draw generated with seed 20260913; its full mask hash is recorded in the run JSON for reproducibility. We then replay the same checkpoints on a second fixed draw (seed 20260914). The replay of seed 20260913 matches the training record exactly; the second draw is a source-sampling sensitivity check. On the primary draw, mean image-level transport is 378.167 for Period-5 and 527.646 for AON; the means of per-image displacement per event are 0.8341 and 1.1638. Across both draws, transport ranges from 378.166 to 378.167 per image for Period-5 and from 527.478 to 527.646 for AON. The complete per-input atom predictions, certificates, masks, hashes, and environment record are stored in `results/T13_uniform_k3_source_seeds_arrays.npz` and its companion JSON. Period-5 and AON use ten nominal atoms and collision allowance $2/5$.

Table 8: Source-matched uniform-K3 $T = 13$ training on 10,000 MNIST test images. Accuracy entries are percentages; mean $\pm$ sample standard deviation over three training seeds. All entries use source seed 20260913. Transport/event averages each image's displacement per event.

Kernel	Base acc.	Smoothed acc.	Certified acc.	Transport/event
All-or-none	98.31 ± 0.07	92.37 ± 2.46	92.37 ± 2.46	1.1638
Period-5	98.31 ± 0.07	98.17 ± 0.06	96.51 ± 0.06	0.8341

On source seed 20260914, Period-5 certifies 96.66%, 96.59%, and 96.59% at the three training seeds, compared with 96.57%, 96.45%, and 96.52% on the primary draw; AON remains at 94.94%, 92.14%, and 90.03%. The draws share images, labels, and active spatial support, resampling only the temporal K3 masks. Thus the second draw measures source-sampling sensitivity, not an independent test population or training run. Since these masks are independent of intensity and label conditional on an active pixel, the result tests utility under the theorem's source law without testing label-bearing temporal information; spatial support and event counts can still carry class information, and only one architecture is evaluated.

B.7 Transport fitting and classification utility

Source fitting. The adjacent-triplet source occupies only ten of 286 three-spike masks, giving total variation at least $1 - 10/286 = 138/143$ from the uniform law. To optimize its actual transport, we recompute costs from the official training histogram with the fixed, equally weighted low/high block medoids. The complete LP with 4,096 columns and 78 B2 constraints selects five cut sets, each with probability $1/5$:

$$\{0, 5\}, \quad \{3, 8\}, \quad \{4, 9\}, \quad \{1, 6, 10\}, \quad \{2, 7, 11\}.$$

Boundaries are zero-indexed. The cut sets partition the twelve boundaries, so any two-boundary set hits at most two fifths of the mass. Rational reconstruction verifies primal feasibility and all dual inequalities exactly, with coincident primal and dual objectives $3554381/9375$ per training image. Period-5 instead costs $115327831/300000$: its training excess is 1.38%. The LP kernel is transport-optimal for the empirical training law and fixed decoder.

The frozen kernel lowers test transport from 386.14147 to 380.97443 (1.34%). Both kernels use ten equally weighted atoms and the same three unchanged checkpoints. Certified accuracy falls at each checkpoint (Table 9), although two paired intervals include zero.

Decoder fitting. Replacing uniform block medoids with medoids fit to the training time histogram lowers transport by 27.82%, yet Period-5 certified accuracy falls from 93.75, 92.93, and 93.48% to 91.29, 90.34, and 92.89%. The partition kernel and classifier are unchanged. This is a direct counterexample to treating reduced projection transport as a sufficient condition for improved certified accuracy.

Table 9: Actual-source LP versus Period-5 at $T = 13$, $B = 2$. Differences are LP minus Period-5 in percentage points. Intervals are per-checkpoint paired 95% percentile bootstrap intervals using 20,000 resamples of the 10,000 test inputs (seed 20260906); they are not simultaneous intervals across seeds.

	Train-source LP accuracy (%)		LP minus Period-5 (points), with 95% interval	
Seed	Smoothed	Certified	Smoothed acc. difference	Certified acc. difference
731	98.46	97.26	$+0.04\ [-0.05, 0.13]$	$-0.15\ [-0.26, -0.04]$
732	98.47	97.45	$-0.02\ [-0.10, 0.06]$	$-0.05\ [-0.15, 0.05]$
733	97.63	95.89	$+0.01\ [-0.09, 0.11]$	$-0.13\ [-0.28, 0.02]$

Classifier-aware kernel selection. We select one kernel from 14,721 canonical four- and five-color candidates on 2,000 calibration images, then evaluate it on a disjoint 2,000-image confirmation set. Relative to Period-5, certified accuracy changes by -0.25 points, with paired 95% bootstrap interval $[-0.90, +0.40]$; smoothed accuracy changes by -0.20 points. Thus the calibration advantage vanishes on confirmation for this selection rule.

Conditional margin bound. For 200 fixed N-MNIST records, a per-input constant fit to the same ten evaluated atoms makes a conditional bound nonvacuous on 86 of 100 confirmation positions. That is an oracle diagnostic, not independent prediction. Estimating one constant on the separate 100 calibration positions instead covers 90 confirmation inputs at the prescribed $\alpha = 0.10$, but only three are both covered and nonvacuous. The practical geometry-to-margin guarantee remains weak.

B.8 Larger budgets and saturation

The matched-budget comparison compares critical-period kernels with AON at $B = 1, \ldots, 4$ on the same latency-MNIST checkpoints, without retraining. Period-$(2B + 1)$ uses $2(2B + 1)$ nominal atoms. Its certified accuracy falls from $96.84 \pm 0.08\%$ at $B = 1$ to $78.44 \pm 14.69\%$ at $B = 4$ (mean $\pm$ sample standard deviation over three seeds). The budget-four range, 61.51–87.80%, exposes checkpoint sensitivity. The figure data include all per-seed metrics and summary statistics.

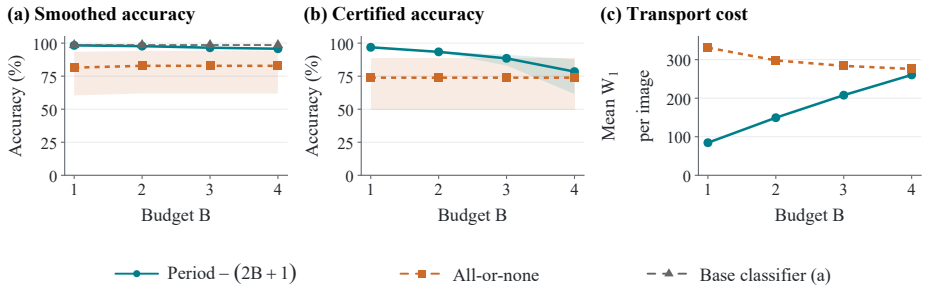

Figure 3: Matched-budget latency-MNIST comparison. Curves show means and ranges over three seeds; each B uses a separately calibrated kernel with collision allowance $B/(2B + 1)$. The figure compares budget-specific constructions rather than one fixed kernel across budgets.

Matched all-or-none comparison. The AON comparison holds the collision allowance fixed at each budget. The B-matched kernel assigns mass $B/(2B + 1)$ to identity and $(B + 1)/(4B + 2)$ to each median, matching the critical-period collision allowance. Its predictions follow by reweighting the identity and two empty-cut median votes from the saturated-budget evaluation, using the same checkpoints and test labels.

The AON certificate holds exactly when the two median predictions agree. If both vote for a class, its invariant mass is $(B + 1)/(2B + 1)$, exceeding all mutable identity mass. If they disagree, a competitor has at least as much invariant mass as the predicted class. Assigning all identity mass to that competitor makes the certificate bound nonpositive. Thus AON's certified set is independent of $B \geq 1$ for this calibration. Its certified accuracies are 88.52%, 49.72%, and 83.34% at the three seeds. At $B = 4$, Period-9 gains 11.79 and 2.68 points at seeds 732 and 733 but loses 0.72 points at seed 731.

For the DVS-Gesture VGGSNN comparison in Table 3, the paired certified-correct counts are 144 for both kernels, 37 for Period-5 alone, 11 for AON alone, and 96 for neither. AON's equal-median consensus certifies the same 155 correct inputs at saturated budgets.

At $T = 10$, at most nine distinct temporal boundaries can be crossed. For effective budget nine and $\delta = 9/19$, Proposition 4 makes AON exactly optimal with objective $11665/4864$, compared with $19113/4864$ for Period-19 under the source used in that calculation. At seed 732, AON nevertheless has certified accuracy 49.72% and certificate rate 87.76%, including 3,804 stable but incorrect predictions.

We compute the projected-input laws by merging identical transformed inputs before summing atom weights. AON and Period-19 differ on all 10,000 inputs, with mean total variation 0.473676. Across 30,000 seed–input pairs, the lower-transport AON kernel assigns less mass to the true class on 228 pairs and more on 6,058. Both signs occur at every checkpoint, and 215 pairs attain the bound in Proposition 12. These comparisons measure true-class vote mass at each input; population accuracy depends on the winning class across the input distribution.

B.9 Native-event comparisons across classifiers

Other data and checkpoints. The single-seed Fashion-MNIST pilot gives 68.88% smoothed and 40.41% certified accuracy with Period-5. An initial N-MNIST cohort has 200 binary ten-bin test inputs, 20 per class, excluding smoothing-development indices. Period-5 trails the prescribed p4 by 2.50 smoothed-accuracy points, reversing the latency-MNIST ordering. The fixed N-MNIST SNN has three 128-channel Conv–BN–LIF stages, two max pools, a 256-unit LIF layer, and a ten-class readout. DVS-Gesture uses the official VGGSNN (channels 64, 128, 256, 256, 512, 512, 512, 512; average pooling; 11-class readout) on all 288 test inputs. Both models average logits over ten steps.

Kernel ordering across DVS-Gesture classifiers. To test classifier dependence, we train a compact Conv–LIF network on the same binary ten-bin DVS inputs: three 64-channel Conv–BN–LIF stages, two max pools, a 256-unit spiking layer, and an 11-class readout (1,126,656 parameters). An internal 4×4 average pool produces 32×32 frames. Projection, event-count checks, and transport all operate on the original 128×128 grid *before* pooling. LIF settings match Section B.3: $\tau = 2$, threshold one, zero hard reset, `decay_input=False`, and the piecewise-leaky-ReLU surrogate with $c = 0$. States reset per batch; logits average over ten steps.

Subjects 01–19 supply 984 optimization examples, subjects 20–23 supply 192 validation examples, and the official test subjects 24–29 supply 288 examples. Each of three prescribed seeds trains for 48 epochs with batch size 32, AdamW learning rate 10^{-3}, zero weight decay, and a 48-epoch cosine schedule. Training uses float32, disabled TF32, deterministic cuDNN, and four CPU threads, without augmentation or parameter search. For seeds 20260907–20260909, checkpoint selection maximizes validation correct count, then minimizes validation loss, then chooses the earliest epoch. All three exceed the prespecified 80% validation threshold; test evaluation follows selection.

The frozen Period-5, AON, and p4 laws use $B = 2$, collision allowance $2/5$, and ten nominal atoms. Exact integer votes cover all 46 boundary subsets of size at most two; cumulative counts independently verify transport and event preservation. AON exceeds Period-5 in correct certificates at each compact-model seed, while Period-5 exceeds p4 (Table 10).

AON's templates agree on 283, 283, and 282 compact-model inputs, and with the base on 271, 269, and 275. Each template correctly classifies 235–239 inputs, versus 157 and 172 for VGGSNN. Count-only projection therefore preserves much more of the compact model's accuracy. Both architecture and training differ in this comparison.

Paired percentile intervals resample the six test subjects with all their clips over all $6^6 = 46{,}656$ ordered bootstrap draws (Table 11). They measure subject variation for each fixed checkpoint; seed standard deviations measure training variation. All three Period-5–AON intervals include zero, leaving the population-level ordering unresolved.

Larger N-MNIST comparison. The same frozen N-MNIST checkpoint and kernels are evaluated on 1,000 additional inputs, 100 per class. Selection precedes inference and excludes 1,408 smoothing-development indices and the initial 200 comparison indices. Earlier clean evaluation may have

Table 10: DVS-Gesture predictions on the same 288 test inputs at $B = 2$. All entries are counts. Compact-model seed labels 907–909 abbreviate 20260907–20260909. The VGGSNN row group uses one existing checkpoint. Each kernel acts on the same binary event grid before model-specific processing. Training and clean accuracy differ between the two models.

Kernel	Smoothed correct	Correct certified	Wrong certified
VGGSNN; base correct: 274/288			
All-or-none	179	155	107
Calibrated p4	244	146	11
Period-5	246	181	17
Conv–LIF, seed 907; base correct: 243/288			
All-or-none	240	237	46
Calibrated p4	240	233	36
Period-5	242	235	34
Conv–LIF, seed 908; base correct: 239/288			
All-or-none	239	236	47
Calibrated p4	241	230	39
Period-5	240	231	41
Conv–LIF, seed 909; base correct: 240/288			
All-or-none	239	233	49
Calibrated p4	238	224	41
Period-5	238	229	41

Table 11: Paired correct-certificate differences for the compact DVS SNN. Gain and loss count inputs certified correctly by exactly one of the two kernels, with Period-5 as the first kernel. Differences and 95% paired subject-bootstrap intervals are in percentage points.

Seed	Comparator	Gain	Loss	Period-5 difference [95% interval]
907	All-or-none	2	4	$-0.69\ [-2.78, 0.69]$
908	All-or-none	3	8	$-1.74\ [-3.57, 0.44]$
909	All-or-none	3	7	$-1.39\ [-3.33, 0.31]$
907	Calibrated p4	2	0	$+0.69\ [0.00, 1.67]$
908	Calibrated p4	3	2	$+0.35\ [0.00, 1.19]$
909	Calibrated p4	6	1	$+1.74\ [0.60, 3.33]$

included these inputs. All kernels share the decoder, ten nominal atoms, collision allowance $2/5$ at $B = 2$, input frames, and certificate implementation.

Table 3 reports the complete comparison. Period-5 and each comparator correctly certify 256 inputs in common; Period-5 alone certifies another 512, and 232 are correctly certified by neither. Thus each paired gain is 51.2 points. A class-stratified paired bootstrap with 20,000 resamples gives a 95% interval of $[48.5, 53.9]$ points. Period-5's smoothed-accuracy differences are $+56.6$ points versus AON ($[53.9, 59.3]$) and -1.7 versus p4 ($[-2.7, -0.8]$). These intervals measure input variation for the fixed checkpoint.

The vote decomposition (Table 12) attributes all 512 additional correct certificates over p4 to inputs that p4 classifies correctly but cannot certify. The gain over AON comprises 113 correct uncertified and 399 incorrect certified predictions.

Table 12: N-MNIST vote outcomes at $B = 2$: counts among the same 1,000 inputs. Correctness concerns the smoothed prediction. Certified errors are stable incorrect labels, not violations of the certificate.

Kernel	Correct certified	Correct uncertified	Wrong certified	Wrong uncertified
All-or-none	256	131	613	0
Calibrated p4	256	714	1	29
Period-5	768	185	4	43

For Period-5 and calibrated p4, the implemented boundary-subset test at $B = 2$ passes exactly when all ten atom predictions agree. To see this, group the atoms into their five lower/upper decoder pairs. The nonempty cut sets are disjoint, so two boundaries can expose any two nonempty pairs. If at most eight atoms vote for the winner, two such pairs can leave at most four fixed winner votes; their margin against four mutable atoms is nonpositive. With nine winner votes, exposing two all-winner pairs leaves five fixed winner votes, one fixed rival vote, and four mutable atoms, giving margin zero. Ten equal votes give margin $1/5$. This argument also covers p4's one empty-cut pair, which is always fixed. An integer check over all 1,023 nonempty winner masks per kernel independently verifies the criterion, assigning each other atom a distinct rival label to obtain the most favorable competing-vote counts.

Among p4's correct predictions, 561 inputs have eight or nine correct votes and thus fail certification. Its two full-bin median atoms are both correct on only 256 inputs, imposing a count-only bottleneck that Period-5 avoids. Both kernels' certified inputs also pass the global collision-gap bound: the gain comes from transformed-input agreement, rather than a tighter vote-bound calculation. This analysis uses the frozen votes without changing the cohort or certification threshold.

Selection uses NumPy seed 20260907 and bootstrapping seed 20260908. Independent integer-vote and cumulative-count calculations reproduce the certificates, transport, and event preservation. Per-input predictions support the paired intervals and vote decomposition.

In the original 200-input N-MNIST cohort, one- or two-event lines account for 32.32% of event mass (10.64% for one-event lines alone). These count-support measurements do not establish exchangeability. The distribution-free collision certificate applies across count strata; Appendix B.2 reports the corresponding all-count DVS source analysis.

B.10 Timing-dependence controls

Timing dependence. We permute the association between active pixels and their event times while preserving active support, line counts, capacity one, and the global time histogram. This removes pixel–time correspondence without changing occupancy statistics. Table 13 reports the mean permuted accuracy and paired losses. A distinct occupancy control averages logits over the ten constant-time encodings of the same active support. The deterioration under permutation supports timing dependence, although the remaining high accuracy shows that the task also contains strong spatial information.

Table 13: Timing-destruction control on the full MNIST test set. Losses are original minus mean-permuted accuracy in percentage points; intervals are paired 95% bootstrap intervals for base-accuracy loss. The last column is constant-time mean-logit accuracy (%).

Seed	Base acc. loss	95% interval	P5 smoothed acc. loss	P5 certified acc. loss	Occupancy acc. (%)
731	17.99	[17.352,18.628]	2.30	9.68	89.07
732	18.52	[17.890,19.164]	2.50	11.44	55.93
733	16.48	[15.856,17.116]	3.21	9.08	89.16

Exhaustive timing-dependent LIF protection. This deterministic mechanism test isolates timing: each input has two labeled lines, exactly one spike per line, and $T = 13$. Positive clean inputs are all 13 simultaneous pairs (t, t); negative inputs are all 56 ordered pairs with separation at least six. Opposite-class clean centers have minimum transport six, exceeding $2B = 4$, so their threat balls do not overlap. No count- or support-only classifier can exceed 50% balanced accuracy.

A single LIF neuron with $\tau = 2$, threshold one, reset zero, no bias, `decay_input=False`, and two weights of 0.75 predicts the positive class if it spikes at any time. Parameters and the full clean population are fixed before execution, with no training or parameter search. Its membrane recurrence makes gaps zero and one trigger a spike; gaps of two or more do not. Both base clean classes are correct, but budget two can separate any simultaneous pair enough to suppress the spike.

The finite state space permits exact robust accuracy, rather than an estimate from attack search. It contains $13^2 = 169$ states and 745 clean–attack pairs within total displacement two, including clean states. SpikingJelly and an independent NumPy LIF simulator agree on every state. Exact atom votes give the smoothed predictions, and enumeration of each threat ball determines whether its certificate is valid.

Table 14: Constructed timing-only test. Accuracy averages the two class accuracies equally; certification concerns the corresponding smoothed classifier. A dash denotes that no quotient certificate is applied. The gap rule is a transparent task-specific reference, not a learned SNN.

	Balanced accuracy (%)		
Classifier	Clean	Robust	Certified
Base LIF	100	50	—
LIF + Period-5	100	100	100
LIF + AON	50	50	50
Gap ≤ 3 rule	100	100	—

Period-5 correctly certifies all 69 inputs, including all 13 base-vulnerable positives, with zero false certificates. AON's majority empty-cut mass maps both classes to coincidence. Moving both events to time six likewise erases the label information while preserving line counts and support, yielding 50% balanced accuracy. The gap rule also attains 100% robust accuracy: this finite test establishes timing-dependent protection, not superiority over task-specific rules. The fully enumerated population has no sampling uncertainty.

B.11 Natural-event timing dependence and attack budgets

We reuse the official DVS-Gesture VGGSNN and all 288 test inputs, without training, checkpoint selection, or changing the ten-bin preprocessing. Every diagnostic preserves per-line counts, binary capacity, and spatial and polarity support. A shared random time permutation preserves whole frames, whereas independent permutations within each line disrupt cross-line synchrony as well as order and do not preserve the framewise histogram. Three fixed randomizations give the linewise results in Table 15. These are distribution-shift diagnostics: large reordering may change gesture semantics and is not asserted to be a label-preserving bounded attack.

Linewise permutations demonstrate this checkpoint's sensitivity to temporal structure at fixed counts. The count templates provide exactly invariant inputs but incur a substantial clean-accuracy cost in the same model. Separately training a count-only classifier would answer the different question of how much predictive information the counts alone contain.

Table 15: Fixed-model timing dependence on the full DVS-Gesture test set. Macro accuracy equally weights the eleven classes. Lower and upper count templates use the empty-cut uniform medoids, so their predictions depend only on per-line counts. No template is selected from test performance.

Input to fixed SNN	Accuracy (%)	Macro accuracy (%)
Clean	95.14	94.70
Shared time permutation	87.50	86.36
Linewise permutation 1	19.44	19.70
Linewise permutation 2	19.10	19.70
Linewise permutation 3	19.44	20.27
Lower count template	54.51	50.95
Upper count template	59.72	56.63

Official retiming attack. We use the official L_1 attack of Yu et al. (2026) and independently check that output candidates are binary, capacity-one, count-preserving, and within the rank-matched transport budget. At $B = 2$ on 274 clean-correct DVS-Gesture VGGSNN examples, the 40-step attack with seed zero and no random start changes no prediction. This small-budget execution does not reproduce the original repository's 2,000–16,000 budget regime. At $B = 3$ on 512 latency-MNIST inputs per seed, it finds one base-model flip among 1,536 evaluations but no certified example exhibiting a smoothed-label violation. A finite unsuccessful attack is not a proof of robustness, nor do these runs demonstrate a broad attack-strength advantage.

A destructive-budget probe. We select the smallest-index clean-correct input per class before attacks and evaluate all eleven at $B = 2000$, retaining the B2 attack's 40 steps, projection, capacity, penalty settings, seed zero, and no random start. Two previously obtained examples are independently revalidated. This prespecified probe diagnoses protection under damaging attacks, rather than estimating population attack success.

Table 16 reports every input. Nine base predictions are changed. At effective boundary budget nine, AON assigns identity mass $9/19$ and mass $5/19$ to each empty-cut median. It correctly certifies inputs 120, 240, and 264 despite their base-model failures. Both median templates depend only on line counts and give the correct label on these inputs. AON's protection therefore equals count-only consensus, as predicted by the AON criterion in Appendix B.8. Five other inputs have certified but incorrect AON labels. Period-5 gives the correct attacked label on the same three base-vulnerable inputs but has no B2000 certificate.

Table 16: The complete prespecified DVS-Gesture probe at $B = 2000$. All base clean predictions equal the label. The next three columns are predictions on the same legal attack. Count-template predictions are unchanged by the attack; "correct cert." requires both correctness and certification. AON uses collision allowance $9/19$; Period-5 is unchanged.

		Prediction after attack			Count template low/high	AON correct and certified
Index	Label	Base	Period-5	AON		
0	0	10	10	10	2/0	No
24	1	2	2	2	2/2	No
48	2	5	5	5	5/5	No
72	3	1	5	5	5/5	No
96	4	4	4	4	4/4	Yes
120	5	4	5	5	5/5	Yes
144	6	4	4	5	5/5	No
168	7	4	4	5	5/5	No
192	8	8	8	8	8/8	Yes
240	9	10	9	9	9/9	Yes
264	10	9	10	10	10/10	Yes

Every saved witness is binary, capacity-one, and count-preserving, with rank-matched transport at most 2000. An independent cumulative-mass calculation checks transport, and both count templates are checked for bitwise invariance. No certified AON label changes. These attacks target the base

model and are evaluated on the smoothers as transfer attacks. The earlier B2 candidates at indices zero and 24 preserve the clean-correct base and Period-5 predictions, whereas their B2000 candidates change both. The additional B8000 witness at index zero changes both predictions with rank-matched transport 7996 and displacement sum 8000.

Both clean inputs already fail the B2 certification test. More generally, Period-5's full-boundary certificate is vacuous for $B \geq 5$, because a five-boundary set can hit every phase. At $T = 10$, complete partition collision constraints saturate at nine distinct boundaries. Its worst-case collision rows may include boundary sets unreachable from a particular input at the same displacement budget. Exploiting this input-dependent reachability could help close the gap between small-budget certification and damaging natural-data attacks.

B.12 Reproduction and computational requirements

SNN training and inference use an RTX 3090; LPs and exact checks use CPU. The compact-LP grid takes approximately 20 seconds and the 1,000-input N-MNIST evaluation 182 seconds in single runs with costs and projections cached. N-MNIST's fifteen distinct projections form two forward batches of at most ten per input. Section B.1 reports the separate long-horizon timings.

Projection complexity. For $N = \sum_{\ell=1}^{L} K_\ell$ total events, projection with cached block medoids costs $O(N + L(|C| + 1))$ time. A length-m, count-k medoid takes $O(k(m - k + 1))$ binomial-count evaluations and $O(k)$ memory. Given block medoids, direct projection does not require a table over all temporal masks. The ten-bin experiments optionally precompute the 2^{10} mask outputs; this accelerator costs $O(T2^T)$ time and memory and is not part of the general-horizon complexity claim. Exact rational verification runs on CPU; enumerating the unrestricted partition LP is reserved for small horizons.

Reproduction materials. `README.md` gives commands, dependencies, batch sizes, resource measurements, and the appendix-to-file mapping. Additional result records are grouped by experiment name in `results/appendix_results.json`. Seed tables are in `research/paper/data/`, and budget-sweep data in `research/paper/latex/figures/data/`. The standalone `research/paper/proof_certificate/` contains the B4/B5 and K3 checks. From that directory, `pythonverify.py--outputverification.json` runs all finite obligations with Python 3.10 and SymPy 1.13.3; Paired analyses use per-input predictions; the original three-spike transfer retains aggregates only. The compact archive provides Uniform-K3 training and inference with the supplied checkpoints and specified environment; other dataset-dependent comparisons are supplied as saved results. The source-matched uniform-K3 run is reproduced with `python-mexperiments.run_t13_uniform_k3_training`; its per-seed summaries are in `research/paper/data/t13_uniform_k3_training.csv` and the full run record is `results/T13_uniform_k3_training.json`. The independent source-draw evaluation is reproduced with `python-mexperiments.evaluate_t13_uniform_k3_source_seeds`; it refuses to overwrite completed outputs unless `--overwrite` is supplied. The result JSON and per-input NPZ are `results/T13_uniform_k3_source_seeds.json` and `results/T13_uniform_k3_source_seeds_arrays.npz`, respectively.