Streaming Oscillation Traps in Endogenous-Pivot Sequential Extraction

Jack Chaudier Gaffney

Abstract

The companion finite paper establishes a sufficient-condition impossibility result for greedy extraction under endogenous constraints: when development capacity before the forced pivot is smaller than the required prefix length ($j_{\text{dev}} < k$), validity is impossible. This streaming paper adds a new failure mode that does not exist in the finite setting. Under irreversible commitment semantics, phase labels assigned early cannot be revised when the running-max pivot shifts, producing *streaming oscillation traps*. Across 4,200 naturalistic bursty instances, organic traps occur in 2,306 cases (54.9%), and in the tested range $n = 100$ to $n = 1000$ trap rates remain between 52% and 78% across tested cells. Mechanistically, the minimum inter-record gap criterion $\min\_\text{gap} < k$ has perfect recall in our verification regime (100% recall, 0 false negatives; 92.5% accuracy overall). Constructively, deferred commitment yields a smooth quality-latency Pareto frontier: any nonzero patience dominates commit-now, with break-even already at $f = 0.05$. At a practical operating point $f = 0.25$, deferred policy reaches 80.1% effective validity and 14.17 mean score, versus 39.4% and 6.68 for irrevocable commit-now, while approaching finite offline performance (91.5%, 16.58). We recommend deferred running-max commitment with patience $f \approx 0.25$ as the default operating point in this regime.

1 Introduction

The finite paper proved absorbing-state failures for offline greedy extraction with endogenous turning points. Streaming changes the failure geometry. In streaming, labels are assigned as events arrive; if labels are irrevocable and the pivot later shifts, earlier assignments cannot be repaired. This creates a distinct commitment-timing failure mode even when finite extraction would succeed.

This paper contributes:

1. A formal streaming model with irreversible commitment semantics.
2. A sharp adversarial oscillation-trap threshold in terms of effective pre-pivot window size.
3. Organic prevalence measurement: 2,306/4,200 traps (54.9%), with persistently high rates across tested scales.
4. Mechanism verification: $\min\_\text{gap} < k$ gives 100% recall (0 FN) in tested regimes.
5. A deferred-commitment Pareto curve with break-even at $f = 0.05$.
6. TP-consistent scoring to evaluate realizable streaming commitments.

Experiment numbering continues from the finite companion paper: this manuscript reports streaming Experiments 40–49 (plus 48a/48b/48c variants).

Preprint. Under review.

2 Background and Finite Results

We adopt the extraction setup and notation from the finite companion paper (Definitions 1–7). We do not restate the full finite formalism here.

Product automaton view. As in the finite paper, extraction feasibility can be represented on the product automaton between the event graph and grammar DFA state. A search step simultaneously advances in temporal event space and grammar phase state; this makes grammar dead-ends explicit as absorbing regions of the product space.

Endogenous turning point recap (Definition 5 in companion paper). The turning point is not exogenous metadata. It is defined from the extracted set itself as the focal-actor argmax event:

$$\tau(S) = \arg \max_{v \in S,\ a(v)=a^*} w(v).$$

Because τ depends on S, local inclusion decisions can change downstream phase assignments and feasibility.

Finite theorem recap. The Prefix-Constraint Impossibility Theorem states that if the development-eligible prefix count before the forced turning point is below the grammar requirement ($j_{\text{dev}} < k$), greedy extraction has zero valid completions. The finite empirical landscape is feasibility-dominated, with zero-FP theorem behavior in tested boundary regimes and high quality conditional on validity.

What changes in streaming. In finite extraction, one can classify relative to a chosen pivot after the full sequence is known. In streaming, pivot estimates evolve online as a record process. Irreversible commitment means earlier labels may become inconsistent with later pivot updates.

3 Streaming Model

Definition 1 (Streaming Extraction Policy)**.** *A streaming policy π processes events $(e_1, e_2, \ldots)$ in temporal order. At step t, π observes e_t, updates a provisional pivot τ_t (running max-weight focal event seen so far), and assigns a phase label to e_t relative to τ_t. Assigned labels are irrevocable even if τ shifts later.*

Definition 2 (Commit-now Policy)**.** *At each step, assign labels immediately relative to current running-max pivot. If pivot shifts at step t', labels assigned before t' remain locked to the old pivot context.*

Definition 3 (Buffered Policy with Patience f)**.** *Buffer events until step $\lceil f \cdot n \rceil$. At that step, commit to current running-max pivot (if viable: at least k predecessors) and label buffered events from scratch. This is a semi-online policy: it assumes stream length n is known (or estimated) before commitment.*

Definition 4 (Streaming Absorbing Trap)**.** *A streaming run is trapped at step t if, under irreversible commitment semantics, no continuation can produce a grammar-valid sequence. This is the streaming analog of a greatest-fixed-point absorbing set.*

Metrics used throughout. We report four distinct quantities:

- **Streaming-label-lock validity:** grammar validity under true commit-now semantics where labels are irrevocably locked when assigned.
- **TP-consistent extraction validity:** validity after fixing a policy's committed pivot and evaluating extraction under TP-consistent demotion of higher-weight focal competitors.
- **Effective regret:** TP-consistent score shortfall relative to finite baseline, i.e., Regret $=$ Score$_{\text{finite}} -$ Score$_{\pi}$.
- **Committed%:** fraction of instances where the policy makes a pivot commitment. This is a coverage/decision-rate metric, not a validity metric.

ϵ	mean_last_shift_frac	pct_stable_by_half	mean_shifts
0.05	0.481	53%	4.64
0.20	0.445	60%	4.37
0.40	0.340	76%	4.12
0.60	0.233	90%	3.75
0.80	0.121	100%	2.95

Table 1: Pivot stability profile (Exp 40, selected rows). Low-ϵ streams do *not* stabilize trivially early.

$k \setminus$ period	3	5	10	20	50
1	100	100	100	100	100
2	**0**	100	100	100	100
3	**0**	**0**	100	100	100

Table 2: Commit-now validity (%) on adversarial oscillation traces. Commit-delayed and finite validity are 100% in all tested cells; k=2, period = 3 gives 0% vs 100% delayed/finite. In this construction, `osc_period` is a raw cycle measure and maps to effective pre-pivot capacity via $p_{\text{eff}} = \max(1, \lfloor \texttt{osc\_period}/2 \rfloor)$ (matching the Exp 41 generator).

4 Pivot Dynamics as a Record Process

Experiment 40 measures running-max pivot dynamics across $\epsilon \in [0.05, 0.95]$. Mean shifts are near logarithmic-record expectations (observed 4.64 at low-ϵ regimes vs $\ln(N_{\text{focal}})$ scale), and stabilization is delayed when front-loading is weak.

5 Oscillation Traps

5.1 Adversarial construction (Exp 41)

Hand-constructed traces with alternating pivot spikes reveal a sharp period boundary. Commit-now fails exactly when oscillation period is too short for required development accumulation.

Proposition 1 (Oscillation Trap Threshold)**.** *Let p_{eff} denote the maximum number of non-pivot events that can be assigned DEVELOPMENT between consecutive pivot shifts under commit-now. Trapping occurs iff $p_{\text{eff}} < k$.*

For the alternating-spike traces in Exp 41, the generator sets the first spike index as $\texttt{first\_spike} = \max(1, \lfloor \texttt{osc\_period}/2 \rfloor)$, so the effective pre-pivot capacity is $p_{\text{eff}} = \max(1, \lfloor \texttt{osc\_period}/2 \rfloor)$ in this construction, which matches the observed boundary in Table 2.

Proof sketch. Between consecutive pivot shifts, commit-now has a bounded pre-pivot window before labels lock relative to the next pivot. By definition this window contributes at most p_{eff} DEVELOPMENT slots. If $p_{\text{eff}} < k$, the grammar prefix requirement is unreachable and the run is trapped. If $p_{\text{eff}} \geq k$, the adversarial construction supplies enough local pre-pivot capacity to satisfy the prefix requirement.

5.2 Organic prevalence (Exp 43)

We evaluate on 4,200 naturalistic bursty traces (7 epsilon values $\times$ 3 k values $\times$ 200 seeds). Organic traps are common:

$$2306/4200 = 54.9\%.$$

By k: 38.9% ($k = 1$), 58.0% ($k = 2$), 67.9% ($k = 3$). Peak observed cell: $\epsilon = 0.40, k = 3$ with 77% trap rate. Finite validity remains high (83.5–97.5%), while commit-now streaming validity drops to 15.5–67.0%.

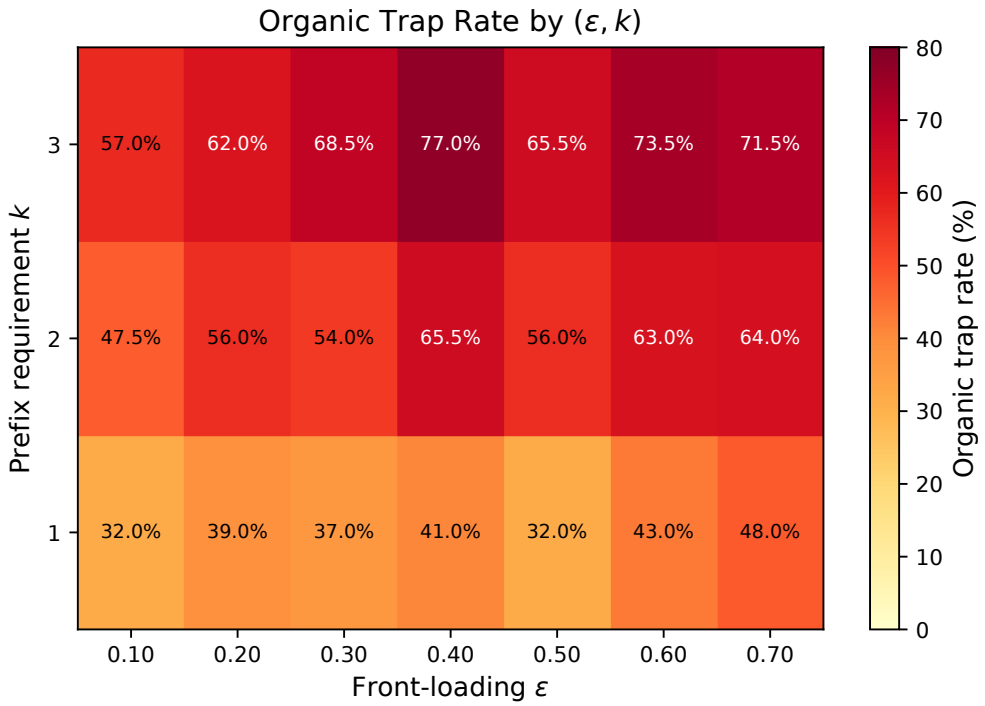

Figure 1: Organic trap rate heatmap by (ϵ, k) (Exp 43).

	Pred trap	Pred no trap
Actual trap	1040 (TP)	0 (FN)
Actual no trap	124 (FP)	490 (TN)

Table 3: Mechanism confusion matrix (Exp 44 aggregate).

5.3 Mechanism verification (Exp 44)

Using finite-valid runs in the verification subset ($n = 1654$), criterion $\text{min\_gap} < k$ yields:

$$\text{TP} = 1040,\ \text{FP} = 124,\ \text{FN} = 0,\ \text{TN} = 490.$$

Accuracy is 92.5%; recall is 100.0%; false negatives are zero. The 124 false positives are tight-gap cases where $\text{min\_gap} < k$ occurs but the realized sequence still avoids a trap. The criterion is therefore necessary for trapping in this tested regime (100% recall, zero false negatives) but not sufficient (124 false positives).

5.4 Scale behavior in the tested range (Exp 45)

Across $n \in \{100, 200, 500, 1000\}$, trap rates stay in the same broad band (roughly 52–78% by cell), and mean minimum gap remains near 1.0.

Proposition 2 (Record-gap scaling in an i.i.d. reference model)**.** *Under i.i.d. continuous weights, the minimum gap between consecutive records is $O(1)$ as sequence length grows. In particular, the first inter-record gap has $\Pr(gap = 1) = 1/2$.*

Proof sketch. For continuous i.i.d. draws, the event that the next sample is a new record has probability $1/2$ at the first opportunity, so one-step record gaps are common. Over length-n streams, the number of record opportunities grows on the order of $\log n$ (Rényi; Nevzorov). As more record opportunities accumulate, the probability that at least one inter-record gap equals 1 increases, so the minimum observed gap remains bounded near a constant rather than growing with n. Thus the minimum gap is $O(1)$.

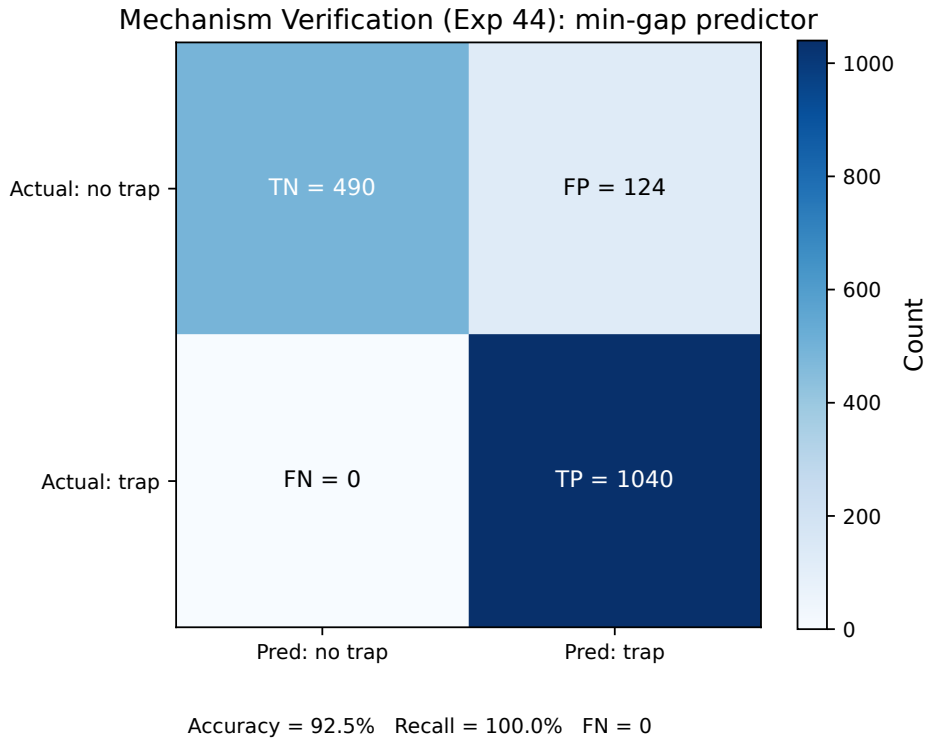

Figure 2: Mechanism verification visualization (Exp 44). FN remains zero.

Clarification for our synthetic regime. Our Exp45 generators are bursty/front-loaded and therefore not literally i.i.d. The proposition above is a reference-model result that explains why constant-scale minimum gaps are plausible. In this paper's generator family, the scale claim is empirical: trap rates stay in the same broad band across tested n, and mean minimum gap remains near 1.0.

6 Constructive Policies

6.1 Policy definitions

Commit-now is the baseline: it labels each event immediately relative to the current running-max pivot and never revises prior labels. Naive buffered waits, then commits to the first viable focal event; first-viable-record (FVR) similarly waits for viability but requires the pivot candidate to be a record-setting focal event.

Deferred running-max with patience f waits until $\lceil f \cdot n \rceil$, then commits to the current running-max pivot and labels from that fixed commitment. Varying f traces the quality-latency frontier studied below.

6.2 Buffered commitment latency (Exp 46)

Buffered commitment is low-latency in relative stream fraction:

- Median commit fraction: 0.01–0.03 across all k.
- 90th percentile: 0.07–0.14.
- Trap-only subset: median 0.02–0.04; 95th percentile 0.21–0.27.

Latency is therefore empirically small (but not zero-cost) compared with the validity gain for moderate f.

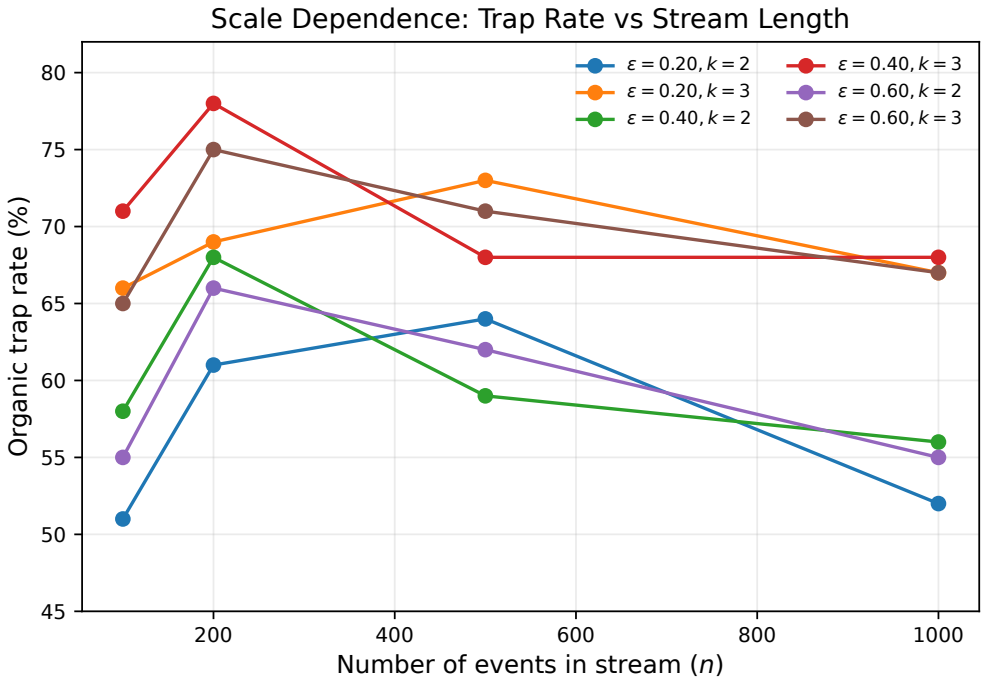

Figure 3: Trap rate vs stream length n (Exp 45). Curves remain largely flat/non-monotone rather than decaying with n.

6.3 Pivot arrival distribution (Exp 49)

From 4,200 raw pivot diagnostics:

- Global max pivot arrival: $p_{50} = 0.28$, $p_{75} = 0.48$, $p_{90} = 0.69$.
- By ϵ, median arrival shifts from 0.43 ($\epsilon = 0.10$) to 0.14 ($\epsilon = 0.70$).
- Running-max convergence ratios: record 1 = 72.9%, record 2 = 84.9%, record 3 = 91.8%.
- Mean records per instance: 4.41.

6.4 TP-consistent scoring methodology

Raw override scoring can evaluate non-realizable pivots because higher-weight focal events can still claim TP after policy commitment. This creates an argmax inconsistency where the evaluated pivot differs from the policy's committed pivot, so the policy is scored on trajectories that violate its own commitment semantics.

In the streaming diagnostics (Exp 48b diagnostics), 84.0% of raw extraction failures (14,567/17,342) are attributable to this inconsistency mode. The issue is most severe for non-global-max policies (FVR and low-patience deferred variants), because their committed pivot is often not the highest-weight focal event in the final selected set.

TP-consistent scoring fixes this by demoting higher-weight focal competitors to non-focal status for evaluation. Demoted events remain available as ordinary resolution events, but they are prevented from claiming TURNING_POINT identity.

This matters because we want to answer: "What performance does policy π achieve *if it actually commits to the pivot it selected online*?" Under this protocol, the largest gains appear in early-commit policies (e.g., FVR and deferred $f = 0.10$, roughly +8.5pp and +8.6pp validity uplift versus raw scoring), while finite and commit-now baselines remain largely stable.

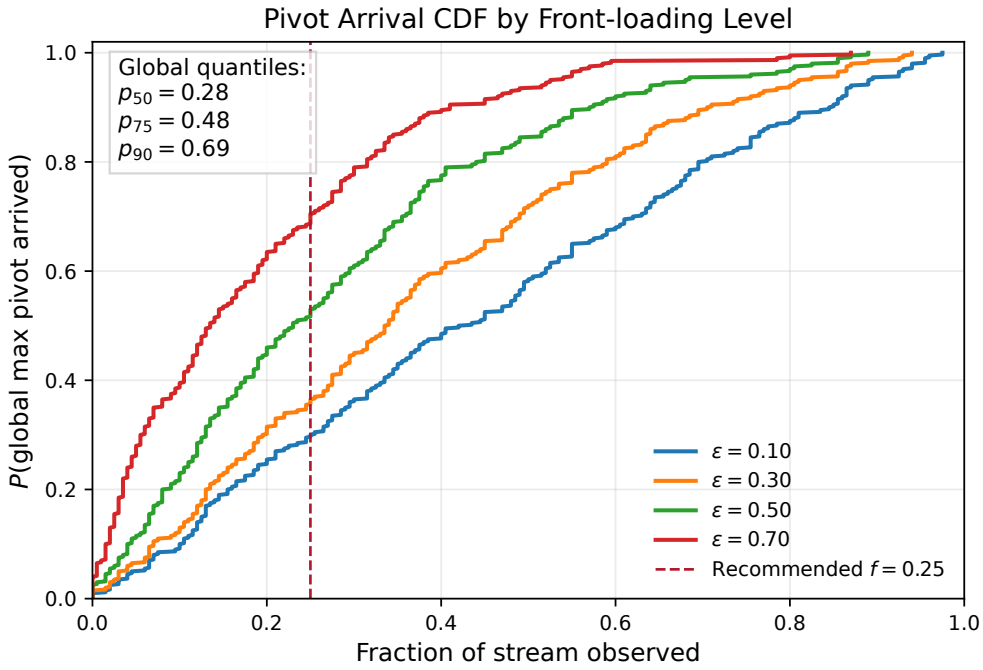

Figure 4: CDF of global-max pivot arrival fraction by ϵ (Exp 49). Dashed line marks $f = 0.25$.

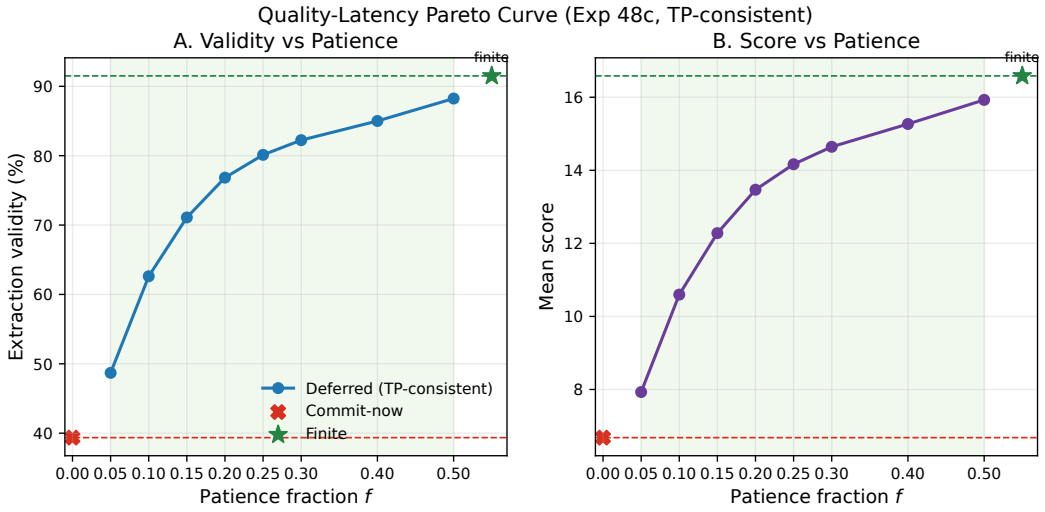

Figure 5: TP-consistent deferred-commitment Pareto curve (Exp 48c aggregate). Any $f > 0$ dominates commit-now.

6.5 Quality-latency Pareto curve (Exp 48c)

Aggregating all (ϵ, k) cells under TP-consistent metrics gives a smooth monotone frontier.

Key finding: deferred dominates commit-now from $f = 0.05$ onward. Practical operating point: $f \approx 0.25$ gives 80% validity and 85% of finite quality.

Higher k requires more patience because later-arriving viable pivots need more predecessor events before commitment can satisfy prefix constraints. The dependence is clear in the selected slices: at $\epsilon = 0.10, k = 3$, even $f = 0.25$ reaches only 53.5% validity, while $\epsilon = 0.70, k = 1$ reaches 91.5%

Policy	Effective Valid%	Effective Score	Effective Regret
Commit-now (streaming)	39.4	6.68	9.90
Deferred ($f = 0.10$)	62.6	10.60	5.99
Deferred ($f = 0.25$)	80.1	14.17	2.42
Deferred ($f = 0.50$)	88.2	15.93	0.66
Finite (offline)	91.5	16.58	0.00

Table 4: Streaming-aware policy comparison (headline table).

ϵ	k	$f = 0.05$	$f = 0.10$	$f = 0.25$	$f = 0.50$
0.10	1	53.5	70.5	88.0	92.5
0.10	2	22.0	43.5	68.0	83.5
0.10	3	16.0	26.5	53.5	76.0
0.70	1	80.0	88.0	91.5	93.5
0.70	2	65.0	75.5	87.0	89.5
0.70	3	50.5	64.0	81.0	85.0

Table 5: Deferred TP-consistent extraction validity (%) by (f, k) for two ϵ slices. Higher k generally requires more patience.

at the same patience. This confirms that the operating-point recommendation is condition-dependent: $f \approx 0.25$ is a strong default, but harder cells (high k, slower pivot arrival) may require larger f.

7 Discussion

7.1 Front-loading duality

Front-loading simultaneously raises trap risk and accelerates pivot convergence. The same concentration mechanism that produces finite absorbing conditions also increases streaming pivot volatility early.

7.2 Relation to finite hierarchy

Finite hierarchy (greedy $\rightarrow$ VAG $\rightarrow$ BVAG $\rightarrow$ TP-solver) addresses within-extraction failure classes. Streaming policies address commitment timing across time-revealed data. The hierarchies are orthogonal: *when* to commit vs *how* to solve after commitment.

7.3 Connection to optimal stopping

Deferred commitment is closely related to optimal-stopping and secretary-style tradeoffs: the policy waits through an observation phase and then commits based on prefix evidence. Our setting differs because the objective is not only to maximize pivot quality, but also to preserve downstream grammar feasibility under irreversible label lock. This adds a structured feasibility layer on top of the classical stop-now versus wait tradeoff.

7.4 Limitations

TP-consistent demotion introduces second-order coupling through pool construction. Demoting high-weight focal competitors removes their injection privilege, which can shrink candidate pools and create new timespan or coverage failures. Empirically, this is visible in naive buffered policy: its validity decreases from 17.2% to 16.7% under demotion, showing that argmax-consistency fixes can partially trade off against candidate-pool breadth.

All results are on synthetic bursty graphs, so cross-domain external validity is still open. Buffered policies here are semi-online because they assume known (or forecasted) stream length for $\lceil f \cdot n \rceil$ commitment timing.

7.5 Applications

In constrained LLM decoding, the streaming "pivot" analog is an early schema/type commitment made from a partial token prefix. If this commitment is made too early and cannot be revised, later tokens can be forced into invalid structures even when a valid completion exists under delayed commitment.

In incident triage, the pivot analog is a locked root-cause hypothesis. The prefix-constraint analog is evidence of precursor events that should appear before escalation to a final cause label. Early lock-in can therefore create streaming traps where accumulating evidence cannot repair an incorrect commitment.

In process mining, the pivot analog is an early fixed reference activity used to align subsequent events into phases. When reference selection is committed before enough prefix context is observed, downstream phase assignment can become path-dependent and irrecoverably inconsistent.

8 Conclusion

Streaming endogenous-pivot extraction exhibits a failure mode not captured by finite offline analysis: oscillation traps under irreversible commitment. Organic traps are common (54.9%), and in the tested range they remain high across scales, with mechanism linked to record-gap structure and zero-FN detection in our verification regime. Deferred commitment yields a smooth tradeoff and dominates commit-now even at minimal patience ($f = 0.05$), with $f \approx 0.25$ providing a strong practical balance and our recommended default operating point.

References

1. Anonymous. *Absorbing States in Greedy Search: When Endogenous Constraints Break Sequential Extraction*. Companion finite paper, 2026.
2. B. Korte and L. Lovász. Greedoids and linear objective functions. *SIAM J. Algebraic and Discrete Methods*, 1981.
3. G. Irnich and G. Desaulniers. Shortest path problems with resource constraints. In *Column Generation*, 2005.
4. W. M. P. van der Aalst. *Process Mining: Data Science in Action*. Springer, 2016.
5. T. Scholak, N. Schucher, and D. Bahdanau. PICARD: Parsing incrementally for constrained auto-regressive decoding. *EMNLP*, 2021.
6. B. T. Willard and R. Louf. Efficient guided generation for constrained language models. *arXiv*, 2023.
7. G. Gange and P. J. Stuckey. The argmax constraint. *CP*, 2020.
8. A. Rényi. The theory of order statistics. *Acta Mathematica Academiae Scientiarum Hungaricae*, 1962.
9. V. B. Nevzorov. *Records: Mathematical Theory*. AMS, 2001.
10. J. P. Gilbert and F. Mosteller. Recognizing the maximum of a sequence. *Journal of the American Statistical Association*, 1966.
11. T. S. Ferguson. Who solved the secretary problem? *Statistical Science*, 1989.

A Full Experiment List (Appendix A)

B Per-cell Data Tables (Appendix B)

Selected per-cell tables are included in the main text (Table 2, Table 5) and figures. Full per-cell tables for all Exps 40–49 and 48a/b/c will be included in the camera-ready appendix.

C Hypothesis Status Updates (Appendix C)

- **H1 (superseded):** "Deferred at $f = 0.10$ is near-optimal." This was an artifact of a bug in an earlier evaluation path; corrected TP-consistent analyses show a broader tradeoff frontier.
- **H2 (refined):** "Buffered commitment is free." Updated conclusion: latency cost is often low, but quality and realizability tradeoffs are real and policy-dependent.

Exp	Name	One-line key finding
40	Pivot stability	Running-max shifts are nontrivial at low ϵ; stabilization is delayed.
41	Adversarial oscillation trap	Sharp threshold in effective pre-pivot capacity ($p_{\text{eff}} < k$) for commit-now failure.
42	Truncation shadow	Finite quality rises with truncation horizon and then plateaus in higher-T regimes (roughly $T \geq 300$).
43	Organic oscillation prevalence	2,306/4,200 organic traps (54.9%), strongest at high k.
44	Mechanism verification	$\min\_\text{gap} < k$ achieves 100% recall and 0 FN in tested subset.
45	Scale dependence	Trap rates remain high from $n = 100$ to $n = 1000$; min gap stays near 1.
46	Commit latency	Buffered commit latency is low (median 0.01–0.03 overall).
47	Policy regret	Early commitment has large regret relative to finite baseline.
48a	Smart policies (raw)	Initial policy comparison before TP-consistent correction.
48b	Smart policies (fixed)	Bug-fixed policy sweep improves consistency of reporting.
48c	Smart policies (consistent)	TP-consistent Pareto frontier; deferred dominates commit-now.
49	Pivot diagnostics	Global max arrival quantiles: 0.28/0.48/0.69 (p50/p75/p90).

Table 6: Streaming experiment catalog used in this draft.