

From Dependency Geometry to Linear Traces

Serialization as an Information-Losing Realization of Computation

J. R. Landers

Abstract. A computation is naturally represented by a partial order of operations, while an execution assigns those operations to moments in time. We formalize executions as *realizations* of an intrinsic dependency poset. Optimal parallel execution and complete serialization become opposite endpoints of a single discrete spectrum: every execution length between the critical-path depth and the total work is attainable. We then study an inverse problem. If only an order-only serial trace is observed, how much of the dependency geometry remains identifiable? For a trace of n labeled operations, we show that the number of compatible dependency posets lies between $2^{\lfloor n^2/4 \rfloor}$ and $2^{\binom{n}{2}}$, so the trace can conceal $\Theta(n^2)$ bits of structural information. As a consequence, the ideal average parallelism cannot be estimated from the ordering alone within a worst-case multiplicative factor smaller than $\sqrt{n}$. Finally, an exact-uniform sampling experiment on 13,800 random height-two posets quantifies recovery from repeated traces. Recovery time grows with both system size and dependency density, and an independently estimated rare-orientation statistic strongly predicts instance difficulty (Spearman correlation 0.859). The results organize work-span analysis, partial-order semantics, information loss, and repeated-trace recovery into a compact inverse theory of serialization.

1 Introduction

A serial trace records that one operation occurred before another. It does not, by itself, say that the first operation *had to* occur before the second. This distinction separates temporal observation from computational necessity.

Consider a computation with two independent additions followed by a multiplication. Its dependency structure is a fork-join graph, yet a serial execution records a line. Both representations compute the same value, but the line suppresses the independence of the additions. The missing structure is precisely the structure that determines the computation’s theoretical parallelism.

Classical parallel-computation analysis begins with a known computation graph and studies its work, span, and scheduling cost [1, 2]. Partial-order and trace semantics likewise distinguish causal order from interleaved observations [6, 7, 4]. We reverse the usual direction. Starting from an observed serial ordering, we ask what dependency structures remain possible and what parallelism can still be inferred.

The paper makes five connected contributions:

- (i) it treats a dependency poset as the intrinsic geometry of a unit-cost computation and an execution as a temporal realization of that geometry;
- (ii) it proves that realizations occupy every integer execution length between optimal parallel depth and complete serialization;
- (iii) it defines the information loss of an order-only trace through the size of its inverse fiber and proves quadratic ambiguity bounds; and
- (iv) it gives a minimax impossibility bound for estimating ideal parallelism from a single linearization; and
- (v) it derives finite-sample recovery bounds and measures recovery and fiber decay under exact-uniform repeated traces.

The conceptual claim is concise: *computation is partially ordered; execution adds temporal order*. An optimally parallel realization adds only the order forced by dependency. A serial realization adds a total order. All less efficient parallelizations lie between these endpoints.

2 Dependency geometry and realizations

Let $P = (V, \prec)$ be a finite poset of operations. The relation $u \prec v$ means that v cannot be executed before u . We write

$$n = |V|, \quad h = \text{ht}(P), \quad w = \text{wd}(P),$$

where h is the size of a longest chain and w is the size of a largest antichain. Under a unit-cost model, n is total work, h is unlimited-processor execution time, and

$$\text{par}(P) = \frac{n}{h}$$

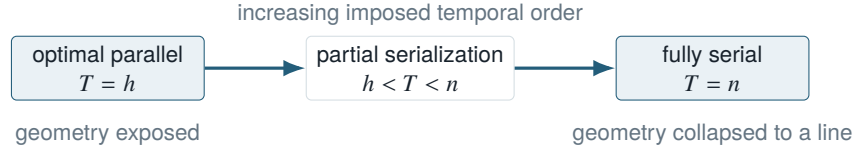


Figure 1: Executions of one dependency poset form a spectrum from minimum-depth realization to complete serialization.

is the ideal average parallelism. This is the familiar work-span interpretation of a computation DAG, expressed at the level of its transitive closure.

Definition 1 (Realization). A T -step realization of P is a surjective map

$$\tau : V \longrightarrow [T] = \{1, \dots, T\}$$

such that

$$u \prec v \implies \tau(u) < \tau(v).$$

The set $\tau^{-1}(t)$ is the collection of operations executed simultaneously at time t .

Surjectivity simply removes empty time steps. Every fiber $\tau^{-1}(t)$ is an antichain. A realization may impose additional temporal precedence between incomparable operations, but it may not reverse an intrinsic dependency.

For $h < n$, define the normalized *serialization coordinate*

$$\sigma_P(\tau) = \frac{T - h}{n - h} \in [0, 1].$$

Thus $\sigma_P(\tau) = 0$ at optimal parallel depth and $\sigma_P(\tau) = 1$ for a serial execution. If $h = n$, then P is already a chain and the two endpoints coincide.

3 The realization spectrum

The normalization above would be less meaningful if only a few intermediate execution lengths were possible. In fact, the entire discrete interval occurs.

Theorem 2 (Realization Spectrum). *Let P be an n -element poset of height h . A T -step realization of P exists if and only if*

$$h \leq T \leq n.$$

Consequently every integer execution length between optimal parallel execution and complete serialization is realizable.

Proof. Any chain of length h must occupy h strictly increasing time steps, so $T \geq h$. Surjectivity gives $T \leq n$.

For existence, assign each operation v its depth

$$d(v) = 1 + \max\{d(u) : u \prec v\},$$

with the maximum of the empty set equal to 0. The depth classes

$$L_j = \{v : d(v) = j\}, \quad j = 1, \dots, h,$$

are antichains and define an h -step realization.

If a current time layer contains at least two operations, split it into two nonempty consecutive layers. No two operations in the original layer are comparable, so the split introduces only permissible extra temporal order. All later layers may be shifted forward by one step. Each split increases the execution length by exactly one. Repeating until all layers are singletons produces realizations of every length $h, h + 1, \dots, n$. $\square$

The theorem turns the serial/parallel dichotomy into a single ordered family. One may equivalently view a realization as an extension of the intrinsic partial order by a weak temporal order. Moving rightward in Figure 1 adds comparabilities that are operational rather than logically necessary.

This spectrum is discrete for a fixed finite computation. The word *continuum* is therefore best understood asymptotically or conceptually: as n grows, the possible normalized coordinates can become arbitrarily fine.

4 Serialization as an inverse problem

Let

$$\pi = (v_1, \dots, v_n)$$

be a serial ordering of the labeled operations. We intentionally assume an *order-only observation*: the trace records operation identities and positions but omits operand provenance, read/write sets, messages, and synchronization metadata. The ordering is a linear extension of the true poset P .

Define the inverse fiber

$$\mathfrak{P}(\pi) = \{Q : \pi \text{ is a linear extension of the poset } Q\}.$$

Every $Q \in \mathfrak{P}(\pi)$ is a dependency geometry consistent with the same observed serial trace.

Definition 3 (Combinatorial dependency loss). The dependency loss of an order-only trace is

$$\mathcal{L}(\pi) = \log_2 |\mathfrak{P}(\pi)|.$$

It is the number of bits needed, in the worst case, to distinguish the hidden poset from the other structures compatible with the trace.

A probabilistic version replaces the logarithm of the fiber size by conditional entropy $H(P \mid \Pi = \pi)$ under a chosen prior. The combinatorial definition is prior-free and sufficient for the present theory.

Theorem 4 (Quadratic Serialization Ambiguity). *For every linear ordering π of n labeled operations,*

$$2^{\lfloor n^2/4 \rfloor} \leq |\mathfrak{P}(\pi)| \leq 2^{\binom{n}{2}}.$$

Equivalently,

$$\lfloor n^2/4 \rfloor \leq \mathcal{L}(\pi) \leq \binom{n}{2}.$$

In particular, an order-only serial trace can conceal $\Theta(n^2)$ bits of dependency structure.

Proof. Every comparability of a poset compatible with π must be one of the $\binom{n}{2}$ forward pairs (v_i, v_j) with $i < j$. Hence there are at most $2^{\binom{n}{2}}$ possible relations, proving the upper bound.

For the lower bound, let $m = \lfloor n/2 \rfloor$ and divide the trace into consecutive blocks

$$A = \{v_1, \dots, v_m\}, \quad B = \{v_{m+1}, \dots, v_n\}.$$

For each subset $R \subseteq A \times B$, define a poset whose only nontrivial comparabilities are the pairs in R . Such a relation is transitive because all comparabilities go from A to B , with no nontrivial comparabilities within either block. Every resulting poset admits π as a linear extension, and distinct choices of R produce distinct posets. Therefore

$$|\mathfrak{P}(\pi)| \geq 2^{|A||B|} = 2^{\lfloor n^2/4 \rfloor}.$$

Taking base-two logarithms yields the second display. □

Remark 5 (Edges versus dependencies). If the hidden object is an arbitrary set of forward DAG edges, there are exactly $2^{\binom{n}{2}}$ possibilities. For dependency partial orders, however, transitive closure identifies many edge descriptions and excludes non-transitive relations. The theorem therefore states bounds rather than the tempting but incorrect exact poset count $2^{\binom{n}{2}}$.

The theorem formalizes a simple observational gap. A serial trace records *what happened first*; dependency geometry records *what had to happen first*. The former generally does not determine the latter.

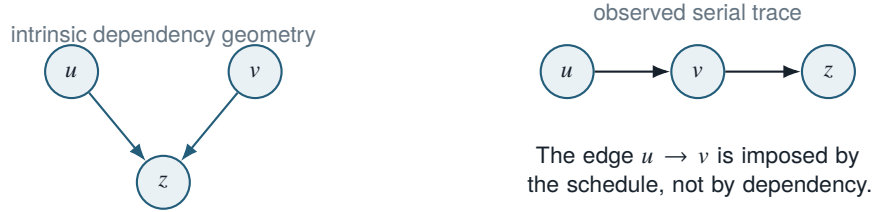


Figure 2: The fork-join computation and one order-only serial observation.

5 Hidden parallelism is not identifiable

Since ideal average parallelism is $n/\text{ht}(P)$, ambiguity in the hidden poset induces ambiguity in the possible speedup.

Corollary 6 (Non-identifiability). *For any fixed serial ordering π of n operations, there exist compatible posets $P_1, P_2 \in \mathfrak{P}(\pi)$ such that*

$$\text{par}(P_1) = n, \quad \text{par}(P_2) = 1.$$

Thus ideal parallelism cannot be determined from an order-only linearization.

Proof. The n -element antichain and the chain

$$v_1 \prec v_2 \prec \cdots \prec v_n$$

are both compatible with π . Their heights are 1 and n , respectively. $\square$

This gives a quantitative estimation limit.

Proposition 7 (Minimax estimation barrier). *Let an estimator observe only π and output $\hat{p}(\pi) > 0$ for the ideal average parallelism. Its worst-case multiplicative error over $\mathfrak{P}(\pi)$ is at least*

$$\boxed{\sqrt{n}}.$$

The bound is attained, for the two extreme cases, by the uninformative estimate $\hat{p} = \sqrt{n}$.

Proof. The same trace may have true parallelism 1 or n . If the estimate is $a > 0$, its multiplicative error on these cases is at least

$$\max \left\{ a, \frac{n}{a} \right\} \geq \sqrt{n},$$

where the inequality follows because the product of the two terms is n . $\square$

The barrier is not computational. Even an all-powerful inference procedure cannot overcome information absent from the observation. To predict useful parallelization, traversal must reveal additional structure: operands, read/write provenance, synchronization, communication, or repeated executions whose differing linearizations expose incomparability.

6 A minimal fork-join example

Consider

$$f(a, b, c, d) = (a + b)(c + d),$$

with operations

$$u = a + b, \quad v = c + d, \quad z = uv.$$

The true dependency poset has

$$u \prec z, \quad v \prec z, \quad u \parallel v.$$

It has $n = 3$, height $h = 2$, width $w = 2$, and ideal average parallelism $3/2$.

The optimally parallel realization is

$$\{u, v\} \longrightarrow \{z\},$$

with $T = 2$ and $\sigma_P = 0$. The serial realization

$$\{u\} \longrightarrow \{v\} \longrightarrow \{z\}$$

has $T = 3$ and $\sigma_P = 1$.

Suppose the order-only trace is $\pi = (u, v, z)$. Exactly seven posets are compatible with this ordering: the antichain; the three one-comparability posets; the two two-comparability posets allowed by transitivity; and the chain. Hence

$$\mathcal{L}(\pi) = \log_2 7 \approx 2.81 \text{ bits.}$$

Among the seven candidates are the antichain, with ideal parallelism 3; the true fork-join poset, with ideal parallelism $3/2$; and the chain, with ideal parallelism 1. The observed line therefore preserves executability but not the dependency geometry needed to identify the best realization.

7 Relation to prior viewpoints

The proposed framework is a synthesis and inversion of established ingredients.

Work and span. Brent’s simulation principle and later work-stealing analyses characterize execution time from known work and critical-path length [1, 2]. Here the critical path itself is latent: the question is what can be inferred when only one serial ordering is retained.

Partial-order semantics. Lamport’s happened-before relation and Mazurkiewicz trace theory distinguish causality from interleaving and represent concurrent behavior through partial orders or dependence graphs [6, 7, 4]. Our observation model fixes one interleaving and measures the size of the inverse class of causal structures consistent with it.

Linear extensions and dimension. The combinatorics and computational complexity of linear extensions are classical [3, 8]. Poset dimension asks for a family of linear extensions whose intersection recovers a known poset. Section 8 studies the complementary random question: how many independent uniform linear extensions are needed before their intersection recovers the poset?

Entropy and partial information. Entropy-based methods have quantified the uncertainty of linear extensions of a given poset in sorting under partial information [5]. Our direction is reversed: the linear extension is given and the uncertainty lies in the compatible poset.

The novelty claim should therefore be modest. The individual tools are standard, and the counting bounds are elementary. The contribution is the compact *inverse theory of serialization*: dependency geometry, a spectrum of realizations, an information-loss fiber, an inference limit for hidden parallelism, and a reproducible repeated-trace experiment.

8 Recovery from repeated order-only traces

Let $\pi_1, \pi_2, \dots$ be independent, uniformly sampled linear extensions of a fixed poset $P = (V, \prec_P)$. Write $<_{\pi_i}$ for the strict total order in trace i and define the *agreement order*

$$R_k = \bigcap_{i=1}^k <_{\pi_i}.$$

Each R_k is a poset, $\prec_P \subseteq R_k$, and the sequence decreases monotonically. The compatible fiber after k traces is

$$\mathfrak{P}_k = \{Q : \pi_1, \dots, \pi_k \text{ are linear extensions of } Q\} = \{Q : \prec_Q \subseteq R_k\}.$$

Thus $|\mathfrak{P}_k|$ is the number of transitive subsets of R_k . Define the random recovery time

$$T(P) = \min\{k \geq 1 : R_k = \prec_P\}.$$

Recovery identifies the true order as the intersection of the traces, but it does not reduce $\mathfrak{P}_k$ to $\{P\}$. At recovery the fiber consists of all subposets of P , the ambiguity that orderings alone cannot remove.

Proposition 8 (Dimension floor). *For every sample path,*

$$T(P) \geq \dim(P).$$

Proof. The first $T(P)$ traces have intersection exactly P , so they form a realizer of P . Poset dimension is the minimum size of such a family. $\square$

There is also a direct finite-sample description of the recovery tail. Let $\mathcal{I}(P)$ be the unordered incomparable pairs of P . For $\{x, y\} \in \mathcal{I}(P)$, put

$$a_{xy} = \Pr(x <_{\pi} y),$$

where π is one uniform random linear extension.

Proposition 9 (Rare-orientation recovery bounds). *If P is not a chain, then for every $k \geq 1$,*

$$\max_{\{x,y\} \in \mathcal{I}(P)} (a_{xy}^k + (1 - a_{xy})^k) \leq \Pr(T(P) > k) \leq \sum_{\{x,y\} \in \mathcal{I}(P)} (a_{xy}^k + (1 - a_{xy})^k).$$

Proof. Recovery fails by time k exactly when some incomparable pair has appeared in only one orientation across all k traces. For a fixed pair this event has probability $a_{xy}^k + (1 - a_{xy})^k$. The lower bound keeps any one pair; the upper bound is the union bound over all incomparable pairs. $\square$

Writing $q_{xy} = \min(a_{xy}, 1 - a_{xy})$ makes the mechanism transparent: recovery is slowed by pairs whose rarer orientation has small probability. The median waiting scale for one such pair is approximately $\log(2)/q_{xy}$ when q_{xy} is small.

8.1 Experimental design

The lead model is the height-two random family used in the lower-bound construction of the Quadratic Serialization Ambiguity theorem. Split $V = A \sqcup B$ with $|A| = \lfloor n/2 \rfloor$ and include each relation $a \prec b$, $(a, b) \in A \times B$, independently with probability

$$p \in \{0.1, 0.25, 0.5, 0.75\}.$$

We used $n \in \{8, 10, 12, 14, 16, 18, 20\}$. The number of independently generated posets per (n, p) was respectively

$$1000, 500, 500, 400, 250, 300, 500.$$

The $n = 20$ total combines two independently seeded runs of 200 and 300 instances.

Linear extensions were sampled *exactly uniformly*. For each P , a downset dynamic program counts the extensions from every placed-set state. If S is the set already placed, an available minimal element x is selected with probability proportional to the number of extensions from $S \cup \{x\}$. Integer weights are used throughout; no floating-point approximation enters the sampler. Each instance was sampled until recovery or $k_{\max} = 10,000$. All 13,800 instances recovered.

The implementation was tested against literal enumeration through six elements, the identity-trace count $|\mathfrak{P}(\pi)| = 7$ at $n = 3$, the identity $|\mathfrak{P}_k| = 2^{|R_k|}$ for height-at-most-two agreement orders, and brute-force linear-extension counts and probabilities. Exact fiber counting was used for all $n = 8$ agreement orders. For larger n , $|R_k|$ is reported as an upper bound on $\log_2 |\mathfrak{P}_k|$ until R_k becomes height at most two, at which point equality is exact. Exact dimensions were computed for all 4000 instances at $n = 8$. Confidence intervals for means are nonparametric 95% bootstrap intervals with 5000 resamples.

For an independent difficulty diagnostic, every instance in the second run received 2000 additional uniform traces after recovery. These traces estimate

$$\widehat{q}_{\min} = \min_{\{x,y\} \in \mathcal{I}(P)} \min(\widehat{a}_{xy}, 1 - \widehat{a}_{xy})$$

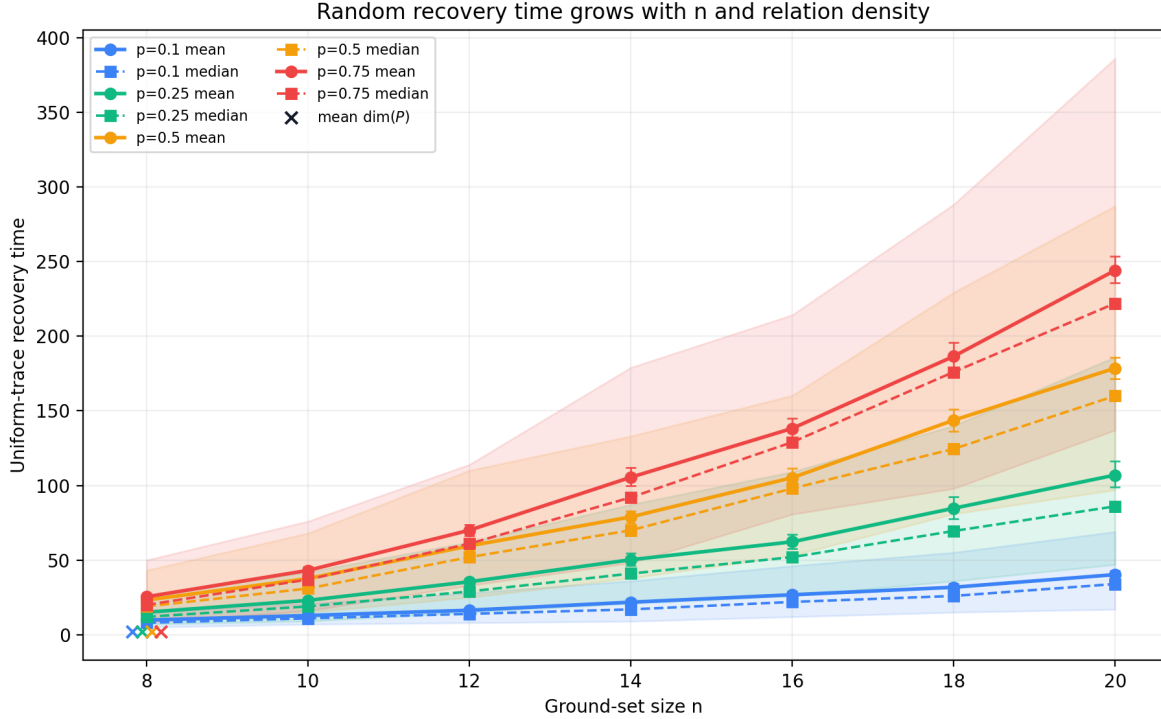
without the stopping bias of the recovery sequence.

8.2 Recovery, fiber decay, and rare pairs

Figure 3 and Table 1 show a smooth increase with both n and p . At $n = 20$, for example, mean recovery rises from 40.3 traces at $p = 0.1$ to 244.1 at $p = 0.75$. The distributions are right-skewed, so medians are consistently smaller. The two independent $n = 20$ runs replicate closely: their means at $p = 0.75$ are 246.0 and 242.8, and at $p = 0.25$ they are 110.4 and 104.6.

Table 1: Mean recovery time, with median in parentheses.

n	$p = .10$	$p = .25$	$p = .50$	$p = .75$
8	9.8 (8)	15.1 (12)	23.3 (19)	25.5 (20)
10	12.9 (11)	23.0 (19)	37.6 (31)	43.1 (37)
12	16.4 (14)	35.5 (29)	59.6 (52)	70.0 (61)
14	21.8 (17)	50.2 (41)	78.9 (70)	105.5 (92)
16	26.8 (22)	62.3 (52)	105.2 (98)	138.2 (129)
18	31.8 (26)	84.7 (69.5)	143.7 (124.5)	186.5 (176)
20	40.3 (34)	107.0 (86)	178.6 (160)	244.1 (222)

**Figure 3:** Uniform-trace recovery time. Solid and dashed curves are seed-level means and medians; shaded regions are the 10th–90th percentiles, and error bars are 95% bootstrap confidence intervals for the means. Crosses show mean exact dimension where computed.

At $n = 8$, mean dimension ranges only from 2.000 to 2.305 across the four densities, while mean recovery ranges from 9.830 to 25.457. Every instance satisfies the dimension floor, with mean random-sampling gaps from 7.830 to 23.152 traces. Dimension is therefore a hard optimum, not a typical uniform-sampling cost.

Figure 4 shows that most imposed comparabilities disappear quickly, followed by a density-dependent tail as the agreement order approaches P . The exact fiber curve and the pair-count proxy coincide once the agreement order has height at most two and end at the subposet floor $|P|$ in logarithmic units.

The rare-orientation diagnostic explains much of the remaining variability. Across the 6000 independently diagnosed instances, the rank correlation between observed recovery and $\log(2)/\hat{q}_{\min}$ is 0.859 (Figure 5). The longest run, an $n = 20$, $p = 0.25$ instance recovered at 1679 traces. Exact counting gives $q_{\min} = 0.00409$ for its hardest orientation, compared with 0.02850 for the median instance in the same cell. The hard pair alone has median waiting scale 170; surviving to 1679 traces has probability about 0.001 for that pair, a rare but plausible tail event across hundreds of instances and many incomparable pairs.

These curves are empirical finite-size laws for one explicit prior on posets and one explicit trace distribution. They should not be read as universal asymptotics. Uniform linear extensions are a clean reference measure, but real schedulers need not sample them; the height-two family is theoretically motivated but structurally special; and

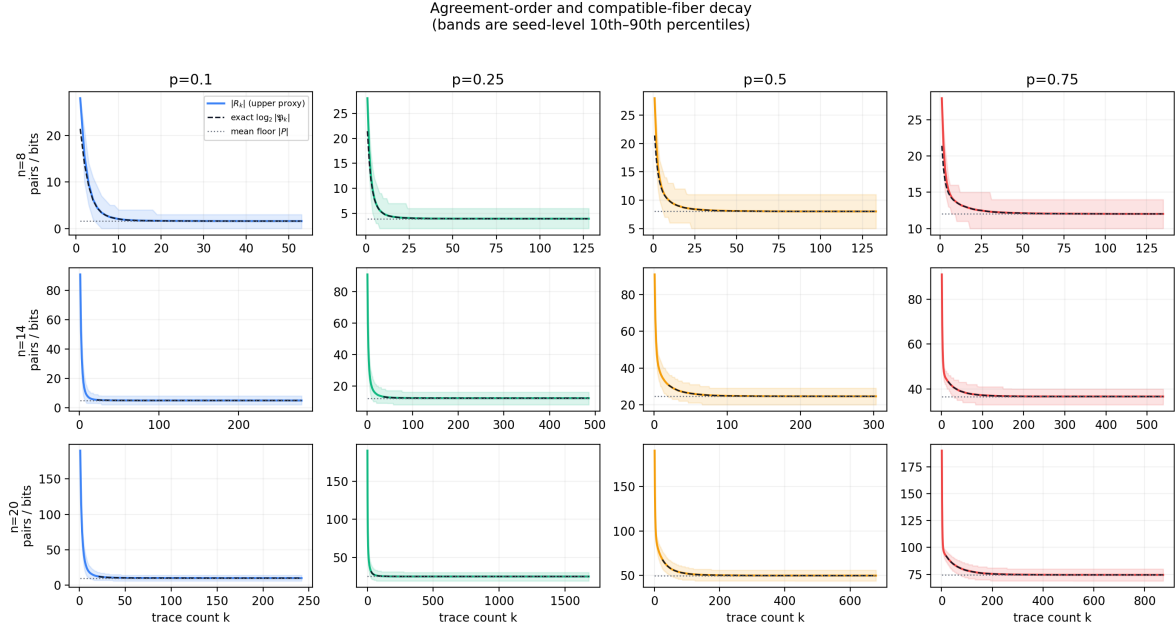


Figure 4: Agreement-order and fiber decay for representative sizes. Colored curves show $|R_k|$ with seed-level 10th–90th percentile bands; dashed black curves show exact $\log_2 |\mathfrak{P}_k|$ where every instance in the plotted aggregate is exact; dotted curves show the mean irreducible floor $|P|$.

arbitrary-order exact fiber counting did not scale beyond the smallest experimental size with the auditable counter used here.

9 Extensions and limitations

The order-only model is intentionally austere. Real traces often expose enough metadata to reconstruct substantial portions of the dependency graph. Let Z denote such side information. A probabilistic extension would study

$$H(P \mid \Pi, Z),$$

while a prior-free analogue would count the compatible fiber after conditioning on Z . This gives a direct formalization of the minimum provenance that must be added back to reveal useful parallelism.

The repeated-trace experiment uses the clean baseline of independent uniform linear extensions. A natural next step is to replace that distribution by a scheduler model, allowing trace probabilities to depend on processor count, queueing policy, or operation duration. The rare-orientation bounds remain valid after replacing a_{xy} by the corresponding scheduler-induced orientation probabilities, provided traces remain independent.

The unit-cost, unlimited-processor model also omits communication, memory contention, and heterogeneous operation costs. Weighted operations replace n by total work and h by maximum weighted-chain length. With a fixed processor budget p , the left endpoint should be the optimal p -processor makespan rather than h . These refinements change the scheduling metric but not the observational distinction between necessary and imposed order.

Finally, the serialization coordinate measures elapsed layers, not every notion of inefficiency. Two schedules of equal length can differ in processor utilization, communication, or energy. Its purpose is narrower: to locate a realization between intrinsic depth and a line.

Reproducibility. The accompanying repository contains the exact sampler, counter, tests, deterministic seed configuration, event-level outputs, plotting code, and a validation manifest with content hashes. Raw and aggregated artifacts are retained separately so that every table and figure can be regenerated from the recorded traces.

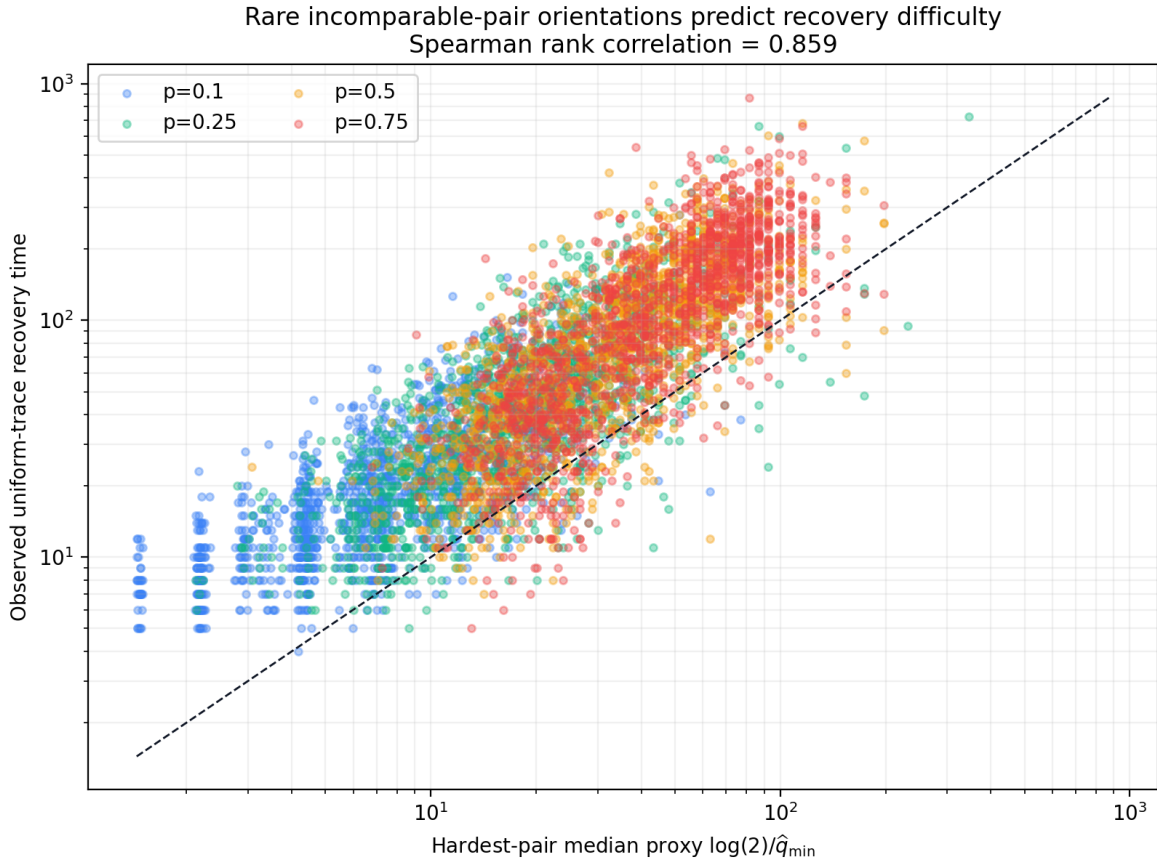


Figure 5: Observed recovery against the independently estimated hardest-pair median scale. Both axes are logarithmic; the dashed line is equality. The estimate uses traces drawn after recovery and therefore does not inherit the recovery stopping rule.

10 Conclusion

A computation does not begin as a sequence. It begins as a set of operations constrained by dependency. Execution embeds that partial order into time. Optimal parallel execution adds only the order required by the computation; complete serialization adds a total order; every intermediate execution length is realizable.

When the result of this embedding is observed only as a line, the underlying geometry is not merely visually hidden. The inverse class contains exponentially many candidate posets and carries quadratically many bits of ambiguity. At the extremes, the same trace is compatible with ideal parallelism 1 or n , forcing a $\sqrt{n}$ minimax estimation barrier.

Repeated traces can recover the intersection order, but the number required is not controlled by dimension alone. In the uniform reference model it is governed by rare orientations of incomparable pairs. The experiment makes this gap quantitative: random recovery lies well above the dimension floor, grows with size and density, and exhibits predictable right tails.

The central principle is therefore:

Serialization preserves execution order but not dependency geometry.

Parallelism is not created by parallel hardware. It is revealed when the dependencies suppressed by linearization are retained or recovered.

References

- [1] R. P. Brent. The parallel evaluation of general arithmetic expressions. *Journal of the ACM*, 21(2):201–206, 1974.

- [2] R. D. Blumofe and C. E. Leiserson. Scheduling multithreaded computations by work stealing. *Journal of the ACM*, 46(5):720–748, 1999.
- [3] G. Brightwell and P. Winkler. Counting linear extensions. *Order*, 8:225–242, 1991.
- [4] V. Diekert and G. Rozenberg, editors. *The Book of Traces*. World Scientific, 1995.
- [5] J. Kahn and J. H. Kim. Entropy and sorting. *Journal of Computer and System Sciences*, 51(3):390–399, 1995.
- [6] L. Lamport. Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM*, 21(7):558–565, 1978.
- [7] A. Mazurkiewicz. Introduction to trace theory. In V. Diekert and G. Rozenberg, editors, *The Book of Traces*, pages 3–41. World Scientific, 1995.
- [8] W. T. Trotter. *Combinatorics and Partially Ordered Sets: Dimension Theory*. Johns Hopkins University Press, 1992.