Evidence Holonomy and Entropy Production: From Universal Coding to Irreversibility

Joshua Winters
Independent Researcher
josh@friendmachine.co

August 27, 2025

Abstract

We define an "evidence holonomy" functional on loops of representation transforms applied to sample paths. Using pointwise universality of code lengths for stationary ergodic processes, we prove two reductions: (i) **representation-space holonomy** converges (up to $o(n)$) to the entropy-rate difference $h(Q) - h(P)$; (ii) **KL-holonomy**—implemented by transporting a universal code trained under P to one trained under Q—converges to the relative-entropy rate $\mathsf{d}(P\|Q)$. For finite-state Markov processes, KL-holonomy along the time-reversal loop equals the entropy-production rate σ (bits/step). We show asymptotic code-invariance of KL-holonomy and validate the framework on Arrow-of-Time benchmarks (audio, sensors, finance).

1 Setup and Definitions

Alphabet and path space. Fix a finite alphabet $\mathcal{X}$. Let $\mathcal{X}^n$ denote length-n strings and $\mathcal{X}^{\mathbb{N}}$ the one-sided sequence space with its product σ-algebra. Let P be a stationary ergodic probability measure on $(\mathcal{X}^{\mathbb{N}}, \mathcal{F})$. Write $X_{0:n-1}$ for the length-n prefix of a sample from P and P_n for its law on $\mathcal{X}^n$.

Universal codes. A *universal code* on alphabet $\mathcal{A}$ is a map $\mathcal{E}_{\mathcal{A}} : \bigcup_{n\geq 1} \mathcal{A}^n \to \mathbb{R}_+$ assigning a code length in bits to any finite string, such that for every stationary ergodic law Q on $\mathcal{A}^{\mathbb{N}}$,

$$\frac{1}{n}\Big(\mathcal{E}_{\mathcal{A}}(Y_{0:n-1}) + \log_2 Q_n(Y_{0:n-1})\Big) \xrightarrow[n\to\infty]{Q\text{-a.s.}} 0, \tag{1}$$

with convergence also in $L^1(Q)$. Classical examples include LZ78, Krichevsky–Trofimov mixtures for finite-order Markov models, and CTW [1, 2, 3, 6, 7].

Standing assumptions. Unless stated otherwise, alphabets are finite; processes are stationary and ergodic; and when KL rates are finite we assume absolute continuity $P \ll Q$. All identities are per-symbol, up to $O(1)$ boundary terms (negligible when divided by n). All logs are base 2.

Representation transforms and loops. For each n, let $F_{i,n}$ be a measurable map $F_{i,n} : \mathcal{X}_{i-1}^{n_{i-1}(n)} \to \mathcal{X}_i^{n_i(n)}$, where the alphabets $\mathcal{X}_i$ may differ by step, and $n_0(n) = n$. Define successive images $x^{(0)} = x \in \mathcal{X}^n$ and $x^{(i)} = F_{i,n} \circ \cdots \circ F_{1,n}(x^{(0)}) \in \mathcal{X}_i^{n_i(n)}$. A finite list $\gamma = (F_{1,n}, \ldots, F_{m,n})$ is a *loop at scale* n if $\mathcal{X}_m = \mathcal{X}_0 = \mathcal{X}$ and $n_m(n) = n + O(1)$. We write $L_n = F_{m,n} \circ \cdots \circ F_{1,n}$. When

$n_m(n) \neq n$, evaluations are aligned by truncating/offsetting one argument (e.g., "tail alignment" $X_{1:n-1}$ versus a length $n-1$ loop output); this contributes only $O(1)$ boundary terms, negligible after dividing by n.

Why 'holonomy'? In differential geometry, holonomy measures the failure of a vector to return unchanged after parallel transport around a closed loop; the effect depends on the connection and reflects curvature. Here, the 'vector' is a description length (or log-likelihood) assigned by an observer (a code), the 'connection' is the rule transporting this observer along a loop of representation transforms, and the holonomy is the net change after completing the loop. Zero holonomy corresponds to flatness (e.g., bijective relabelings), whereas coarse-grainings or time-reversal in nonequilibrium systems induce positive 'curvature' detected by nonzero holonomy. This analogy motivates the terminology and clarifies why observer choice acts like a gauge: our KL-holonomy rate is asymptotically gauge-invariant across universal codes.

1.1 Two holonomy functionals

Definition 1.1 (Representation-space holonomy)**.** Given universal codes $\mathcal{E}_{\mathcal{X}_i}$ for intermediate alphabets, define

$$\begin{aligned}\mathrm{Hol}_n^{\mathrm{out},\gamma}(x) &= \sum_{i=1}^{m}\Big(\mathcal{E}_{\mathcal{X}_i}\big(x^{(i)}\big) - \mathcal{E}_{\mathcal{X}_{i-1}}\big(x^{(i-1)}\big)\Big)\\ &= \mathcal{E}_{\mathcal{X}}\big(L_n(x)\big) - \mathcal{E}_{\mathcal{X}}(x).\end{aligned}$$

Definition 1.2 (KL (observer-transported) holonomy)**.** Let P_n be the law of $X_{0:n-1}$ and $Q_n = (L_n)\#P_n$. Define the ideal codelengths $\mathcal{L}_n^{(P)}(x) = -\log_2 P_n(x)$ and $\mathcal{L}_n^{(Q)}(x) = -\log_2 Q_n(x)$. The (ideal) KL holonomy is

$$\mathrm{Hol}_n^{\mathrm{KL},\gamma}(x) \;=\; \mathcal{L}_n^{(Q)}(x) - \mathcal{L}_n^{(P)}(x) \;=\; \log_2 \frac{P_n(x)}{Q_n(x)}.$$

Hence $\frac{1}{n}\mathbb{E}_P[\mathrm{Hol}_n^{\mathrm{KL},\gamma}] = \frac{1}{n}\mathrm{D}(P_n\|Q_n) \to \mathsf{d}(P\|Q)$. *Implementation note.* In experiments we approximate $\mathcal{L}_n^{(P)}$ and $\mathcal{L}_n^{(Q)}$ with universal coders; under standard log-loss consistency for the model class used, the empirical rates converge in L^1 to the ideal ones.

2 Reductions via Universality

Convention. All statements below are per symbol, with $O(1)$ boundary discrepancies (e.g., from alignment) absorbed into the $o(1)$ terms.

Lemma 2.1 (Pointwise reductions)**.** *Assume* (1) *for the relevant laws.*

1. *For* $\mathrm{Hol}^{\mathrm{out}}$*: with* $\mathcal{E}_{\mathcal{X}}$ *universal for both* P *and the pushforward process,*

$$\frac{1}{n}\Big(\mathrm{Hol}_n^{\mathrm{out},\gamma}(X_{0:n-1}) - \log_2 \frac{P_n(X_{0:n-1})}{Q_n(L_n(X_{0:n-1}))}\Big) \xrightarrow[n\to\infty]{P\text{-a.s.}} 0. \tag{2}$$

2. *For* $\mathrm{Hol}^{\mathrm{KL},\gamma}$*:*

$$\frac{1}{n}\Big(\mathrm{Hol}_n^{\mathrm{KL},\gamma}(X_{0:n-1}) - \log_2 \frac{P_n(X_{0:n-1})}{Q_n(X_{0:n-1})}\Big) \xrightarrow[n\to\infty]{P\text{-a.s.}} 0. \tag{3}$$

Both convergences also hold in $L^1(P)$.

Proof. Apply (1) (Barron's strong pointwise coding theorem) to each code/law pair and subtract the limits; see [5, 7, 6]. $\square$

Averaging yields the two central identities.

Theorem 2.2 (Expectation-level reductions)**.** *Under L^1 universality,*

$$\frac{1}{n}\,\mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{out},\gamma}\big] = h(Q) - h(P) + o(1), \tag{4}$$

$$\frac{1}{n}\,\mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{KL},\gamma}\big] = \frac{1}{n}\,\mathrm{D}(P_n\|Q_n) \xrightarrow[n\to\infty]{} \mathsf{d}(P\|Q) \;\geq 0. \tag{5}$$

Proof. Take expectations in (2)–(3) and note $\mathbb{E}_P[-\log_2 P_n(X_{0:n-1})] = H(P_n)$, $\mathbb{E}_P[-\log_2 Q_n(L_n(X_{0:n-1}))] = H(Q_n)$, and $\mathbb{E}_P\big[\log_2 \frac{P_n(X)}{Q_n(X)}\big] = \mathrm{D}(P_n\|Q_n)$. $\square$

Remark 2.3 (Scope)*.* Equation (4) is gauge-invariant and measures net compression or expansion under the loop. Equation (5) is the *irreversibility* functional implemented in our code (KL-rate holonomy): it is observer-transported and non-negative.

3 Canonical Loops and Corollaries

3.1 Gauge invariance for bijective loops

Corollary 3.1 (Gauge invariance)**.** *If each $F_{i,n}$ is a bijection and the loop is the identity on $\mathcal{X}^n$, then*

$$\frac{1}{n}\,\mathrm{Hol}_n^{\mathrm{out},\gamma}(X_{0:n-1}) \to 0 \quad \textit{and} \quad \frac{1}{n}\,\mathrm{Hol}_n^{\mathrm{KL},\gamma}(X_{0:n-1}) \to 0$$

in P-probability and in $L^1(P)$.

Proof. Then $Q_n = P_n$ for all n, so both (4) and (5) vanish. $\square$

3.2 Coarse-graining loops via channels

Let K_n be a (possibly many-to-one) Markov kernel on $\mathcal{X}^n$ and R_n any measurable right-inverse (a "lift") so that $L_n := R_n \circ K_n : \mathcal{X}^n \to \mathcal{X}^n$ is a loop. If $Q_n := L_n\#P_n$ arises from a stationary Q, then (5) gives

$$\frac{1}{n}\,\mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{KL},\gamma}\big] \to \mathsf{d}(P\|Q) \geq 0,$$

i.e. KL holonomy is non-negative by construction (by non-negativity of KL). Moreover, if both P_n and Q_n are mapped through the same observation channel, data processing yields a lower bound on the holonomy of the observed records.

When is the pushforward stationary? For a general sequence of maps L_n, the marginals $Q_n = (L_n)\#P_n$ need not be the n-marginals of any stationary process. A sufficient condition is that the loop arises from a shift-commuting, finite-memory map on the two-sided shift (a sliding-block code): i.e., there exist F and memory m such that $(L_n(x))_t = F(x_{t-m:t+m})$ for all t, and F commutes with the left-shift. Then, if P is stationary, so is the pushforward process Q. The canonical time-reversal loop and coarse-graining channels satisfy this property. In our theorems that invoke entropy-rate limits for Q, we implicitly assume such a stationary extension exists (or restrict to cases where it is direct, e.g. time reversal).

3.3 Time reversal and entropy production for Markov chains

Let P be a stationary Markov chain on $\mathcal{X} = \{1, \dots, k\}$ with transition T and stationary π. Its time-reversal P^{rev} has transitions $T^*_{ji} = \frac{\pi_i T_{ij}}{\pi_j}$.

Canonical time-reversal loop. Let $x_{0:n-1}$ be a path from a stationary finite-state Markov chain P with transition matrix T. Define three maps on paths of length n: (i) the transition encoder E mapping $(x_{t-1}, x_t)_{t=1}^{n-1}$ to the edge sequence; (ii) reversal R mapping an edge sequence $(e_1, \dots, e_{n-1})$ to $(e_{n-1}, \dots, e_1)$; (iii) a state decoder D that reconstructs a path from the reversed edge sequence given the terminal state x_{n-1} as anchor. The loop is $L_n := D \circ R \circ E$. One checks that $(L_n)\#P_n = P_n^{\mathrm{rev}}$, the n-path law of the time-reversed chain.

Algorithm: Time-reversal loop L_n
Input: $x_{0:n-1}$
1. $E \leftarrow ((x_{t-1}, x_t))_{t=1}^{n-1}$
2. $E' \leftarrow \mathrm{reverse}(E)$
3. $\hat{x}_{n-1} \leftarrow x_{n-1}$ (anchor)
4. For $t = n-1$ down to 1:
Set $\hat{x}_{t-1}$ as the unique predecessor such that $(\hat{x}_{t-1}, \hat{x}_t) = E'_{n-t}$
Output: $L_n(x) = \hat{x}_{0:n-1}$

Remark 3.2 (Practical variant). Our implementation uses a length-n–1 variant: encode transitions, reverse, and *decode the second state* of each reversed edge. We then evaluate on $X_{1:n-1}$ to match lengths. This avoids explicit anchoring by x_{n-1} and differs only by $O(1)$ boundary terms, hence the per-symbol limits are unchanged.

Theorem 3.3 (KL holonomy rate equals entropy production)**.** *For the Markov setting above,*

$$\mathsf{d}(P\|P^{\mathrm{rev}}) \;=\; \sum_{i,j} \pi_i T_{ij} \, \log_2 \frac{\pi_i T_{ij}}{\pi_j T_{ji}} \;=\; \sigma \quad (\textit{bits/step}), \tag{6}$$

and the KL-holonomy satisfies

$$\frac{1}{n}\, \mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{KL,time\text{-}rev}}\big] \;\to\; \sigma. \tag{7}$$

Proof. The path log-likelihood ratio between P and the reversed path law under P^{rev} is

$$\log \frac{P_n(X_{0:n-1})}{P_n^{\mathrm{rev}}(R_n(X_{0:n-1}))} = \sum_{t=1}^{n-1} \log \frac{\pi_{X_t}}{\pi_{X_{t-1}}} + \sum_{t=1}^{n-1} \log \frac{T_{X_{t-1}X_t}}{T_{X_t X_{t-1}}}.$$

The stationary term telescopes to $O(1)$; divide by n and take expectations. Identity (6) is standard in stochastic thermodynamics [9, 10]. Equation (7) is (5) with $Q = P^{\mathrm{rev}}$. □

Remark 3.4 (Absolute continuity). The rate in (6) is finite iff $T_{ij} > 0 \Rightarrow T_{ji} > 0$ for all i, j; otherwise $\mathsf{d}(P\|P^{\mathrm{rev}}) = +\infty$.

Remark 3.5 (Why not merely $h(Q) - h(P)$?). For stationary Markov chains, $h(P) = h(P^{\mathrm{rev}})$, so representation-space holonomy would vanish. The KL version (observer-transported) returns the irreversible production σ.

3.4 General ergodic reversal

Let P^* be any stationary time-reversed process absolutely continuous w.r.t. P on cylinders, with finite $\mathsf{d}(P\|P^*)$. Then, by the same argument,

$$\frac{1}{n}\,\mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{KL,time\text{-}rev}}\big] \;\to\; \mathsf{d}(P\|P^*). \tag{8}$$

4 Observer Independence

Theorem 4.1 (Code-robustness of KL holonomy)**.** *Let $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ be universal on $\mathcal{X}$ for the laws appearing in Lemma 2.1. Then, for any fixed loop γ,*

$$\frac{1}{n}\Big|\mathrm{Hol}^{\mathrm{KL},\gamma}_{n,\mathcal{E}^{(1)}}(X_{0:n-1}) - \mathrm{Hol}^{\mathrm{KL},\gamma}_{n,\mathcal{E}^{(2)}}(X_{0:n-1})\Big| \xrightarrow[n\to\infty]{P\text{-}a.s.} 0,$$

and likewise in $L^1(P)$.

Proof. Apply Lemma 2.1 to both codes and subtract. □

5 Code & Data

All code and data are available at: `https://github.com/jbwinters/Evidence-Holonomy`

6 Numerical validation (UEC battery)

We validate the theoretical predictions across window sizes $n \in \{2^9, 2^{11}, 2^{13}, 2^{15}\}$ and multiple random seeds. For synthetic Markov chains, we compare ground-truth entropy production σ with KL-holonomy rate estimates (Table 1, Figure 1). The median relative error converges to under 10% for $n \geq 2^{11}$ across different chain types.

Observer independence was tested by comparing KL-holonomy estimates from different universal coders: KT with varying Markov orders ($R \in \{1, 3\}$) and prior decay parameters. These yield nearly identical rates across windows (Figure 2: $r > 0.99$, mean $|\Delta| < 10^{-6}$ bits/step). We also include an LZ78 *representation-space* baseline that computes $h(Q) - h(P)$ via separate compression rather than cross-entropy. While LZ78 captures similar irreversibility trends, it measures a different functional than KL-holonomy $\mathsf{d}(P\|Q)$.

AoT demos (audio / sensors / finance). For window-level arrow-of-time classification, two choices align AUC with holonomy and our theory: *loop-negatives* (Encode→Reverse→DecodeSecond) instead of literal reversal, and domain preprocessing (`--aot_diff` for audio/sensors, `--aot_logreturn` for finance). These match the time-reversal loop used by the holonomy and avoid negative-KL pathologies. The script logs per-file AUC and bits/step(/s) and writes a scoreboard CSV.

Artifacts and reproducibility. The script writes

- `results/aot_wav.json`, `results/aot_csv.json` (single-file AoT).
- `results/scoreboard.csv`, `results/scoreboard.json` (folder runs).
- `results/summary.json` (aggregated suite summary for the run).

Representative commands and flags for the AoT demos are documented inline in the repository (e.g., `--aot_bins`, `--aot_win`, `--aot_stride`, `--aot_rate`).

Table 1: Entropy production rate σ (bits/step): ground truth vs. estimate.

Chain (states)	n	σ_{true}	σ_{hat}	Rel. error
3-state	2^9	0.175	0.200	15.0%
4-state	2^9	0.673	0.713	3.8%
5-state	2^9	0.944	0.929	3.6%
3-state	2^{11}	0.175	0.186	6.3%[1]
4-state	2^{11}	0.673	0.644	8.8%
5-state	2^{11}	0.944	0.925	5.5%
3-state	2^{13}	0.175	0.174	9.9%
4-state	2^{13}	0.673	0.658	4.8%
5-state	2^{13}	0.944	0.950	4.3%
4-state	2^{14}	0.868	0.859	1.8%
3-state	2^{15}	0.175	0.179	12.6%
4-state	2^{15}	0.673	0.668	0.4%
5-state	2^{15}	0.944	0.947	0.5%

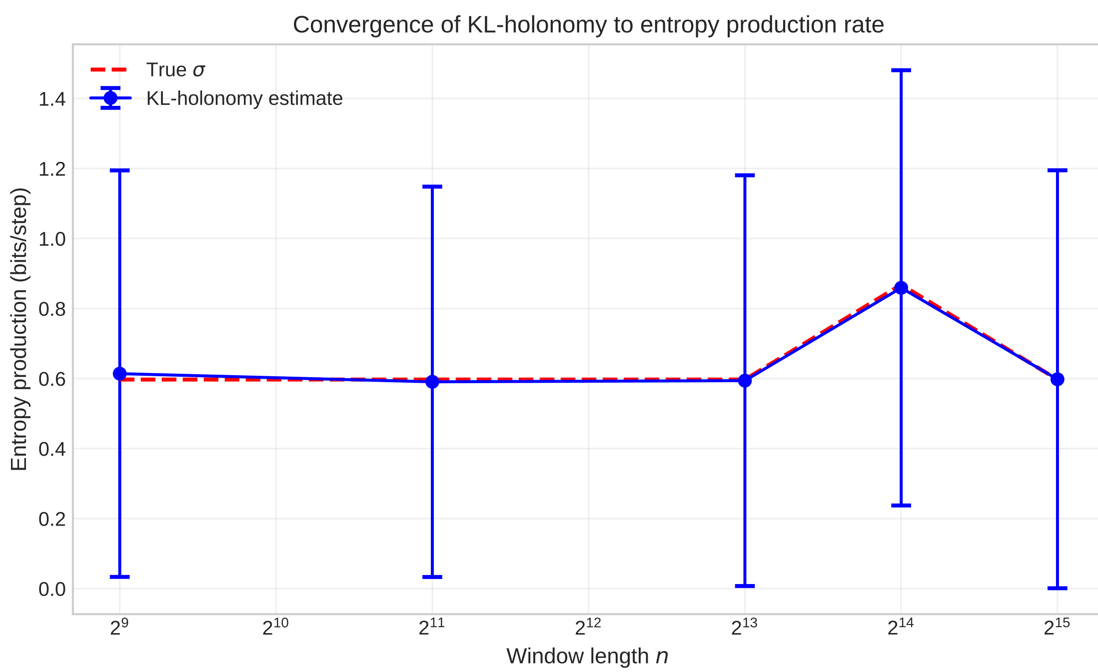

Figure 1: Estimated KL-holonomy rate vs. window length n; horizontal line is σ_{true}.

Discussion

We distinguished two operational regimes. If one evaluates evidence *in the representation reached by the loop*, holonomy reduces to the *entropy-rate difference* $h(Q) - h(P)$ (Theorem 2.2); this yields gauge invariance and detects net compression/expansion by the loop. If instead one *transports the observer* and evaluates evidence against the loop's pushforward law on the *original coordinates*, holonomy equals the *relative entropy rate* $\mathsf{d}(P\|Q)$, recovering irreversibility and, for Markov time reversal, the entropy production rate.

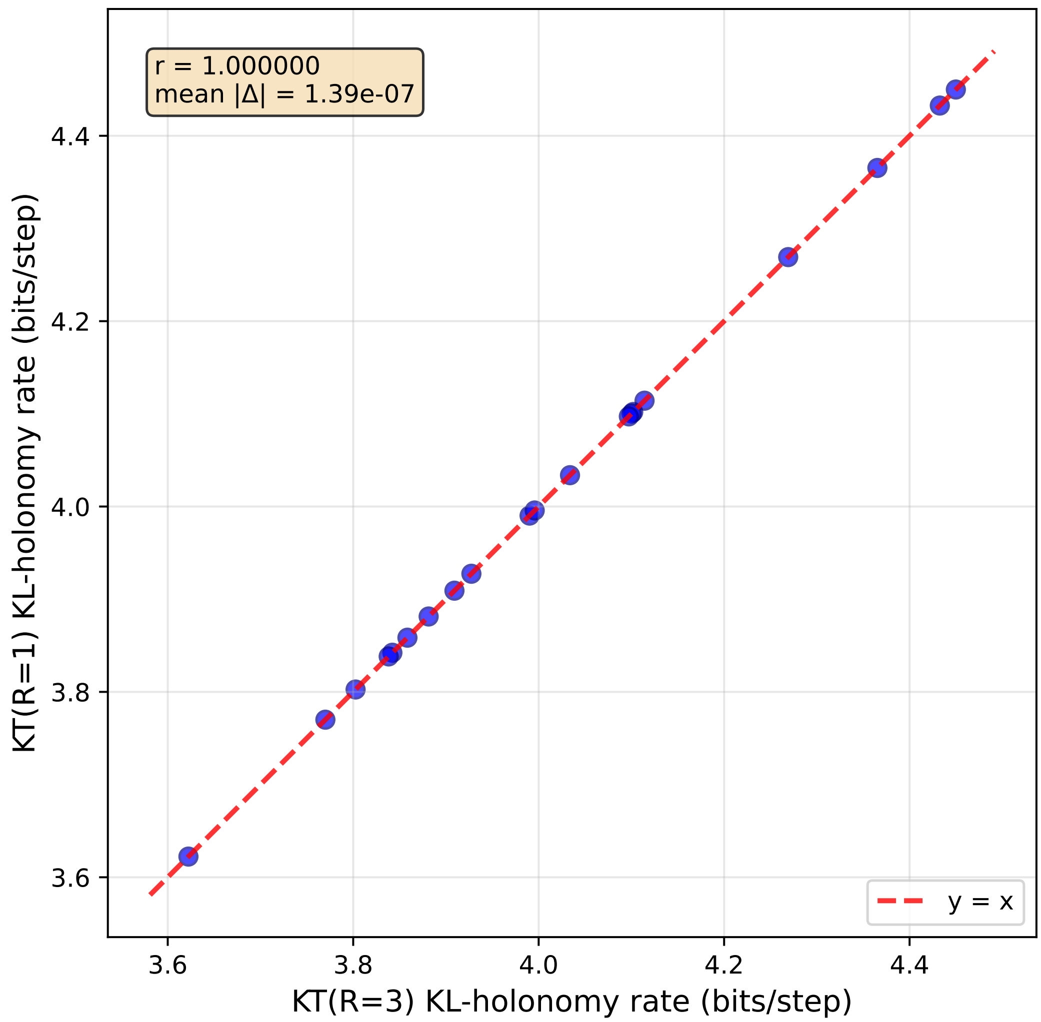

Figure 2: Code invariance: KL-holonomy rate from KT($R = 3$) vs. KT($R = 1$) across windows. Points lie tightly along the diagonal ($r > 0.99$, mean $|\Delta| < 10^{-6}$ bits/step), demonstrating that the KL-holonomy functional is robust to coder hyperparameters.

Technical extensions. The finite-alphabet assumption can be relaxed via quantization and standard approximation. The Markov time-reversal equality extends to hidden Markov models at the level of path measures; holonomy on observed records gives a certified lower bound by data processing (and in the quantum setting by Lindblad/Uhlmann monotonicity [15, 16]). Absolute continuity requirements ensure finite rates (e.g., $\sigma < \infty$ requires $T_{ij} > 0 \Rightarrow T_{ji} > 0$).

Limitations & scope. Our framework requires finite alphabets and stationarity for entropy-rate convergence arguments. Universal code approximations introduce finite-sample error that decreases as $O(\log n/n)$ under standard conditions. The pushforward stationarity condition (sliding-block property) restricts the class of admissible loops but covers the main examples of interest.

Acknowledgements

Portions of the text were drafted or revised with assistance from OpenAI's GPT-5. The author verified all content and takes full responsibility for the paper.

References

[1] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, 24(5):530–536, 1978.

[2] R. E. Krichevsky and V. K. Trofimov, "The performance of universal encoding," *IEEE Trans. Inf. Theory*, 27(2):199–207, 1981.

[3] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens, "The Context-Tree Weighting Method: Basic Properties," *IEEE Trans. Inf. Theory*, 41(3):653–664, 1995.

[4] J. Rissanen, "Modeling by shortest data description," *Automatica*, 14(5):465–471, 1978.

[5] A. R. Barron, "The strong ergodic theorem for densities: generalized Shannon–McMillan–Breiman," *Annals of Probability*, 13(4):1292–1303, 1985.

[6] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths*, Graduate Studies in Mathematics, Vol. 13, American Mathematical Society, 1996.

[7] I. Csiszár and P. C. Shields, "Information theory and statistics: A tutorial," *Foundations and Trends in Communications and Information Theory*, 1(4):417–528, 2004.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006.

[9] J. Schnakenberg, "Network theory of master equation," *Reviews of Modern Physics*, 48(4):571–585, 1976.

[10] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," *Reports on Progress in Physics*, 75:126001, 2012.

[11] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, "Dissipation: The phase-space perspective," *Phys. Rev. Lett.*, 98:080602, 2007.

[12] G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," *Phys. Rev. E*, 60:2721–2726, 1999. (See also *Phys. Rev. E*, 61:2361–2366, 2000.)

[13] T. Hatano and S.-I. Sasa, "Steady-state thermodynamics of Langevin systems," *Phys. Rev. Lett.*, 86:3463–3466, 2001.

[14] É. Roldán and J. M. R. Parrondo, "Estimating Dissipation from Single Stationary Trajectories," *Phys. Rev. Lett.*, 105:150607, 2010.

[15] G. Lindblad, "Completely positive maps and entropy inequalities," *Communications in Mathematical Physics*, 40:147–151, 1975.

[16] A. Uhlmann, "Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory," *Communications in Mathematical Physics*, 54:21–32, 1977.