Ordered Patch Theory

Appendix T-1: Stability Filter — Full Rate-Distortion Specification

Anders Jarevåg

April 3, 2026 | DOI: 10.5281/zenodo.19300777

Original Task T-1: Stability Filter — Full Rate-Distortion Specification Problem: Shannon's Rate-Distortion theory requires: a source X, a reproduction alphabet, and a distortion function $d(x, \hat{x})$. The preprint invokes $R_{pred}(D)$ without specifying these three elements for OPT's substrate. **Deliverable:** A complete $(\mathcal{X}, \hat{\mathcal{X}}, P_X, d)$ specification for OPT's rate-distortion problem.

This revision distinguishes **excess entropy** from **statistical complexity**, proves the predictive-KL identity at finite horizon, proves the general lower bound $R_{T,h}(D) \geq E_{T,h} - D$, and states an exact equality criterion for when that lower bound is attained. $C_{\max}$ remains an empirical parameter rather than a quantity derived from the rate-distortion formalism.
Closure status: PARTIALLY RESOLVED. The four-tuple specification, the predictive-KL identity, and the general lower bound $R_{T,h}(D) \geq E_{T,h}(\nu) - D$ are established with an exact equality criterion. The earlier generic closed-form claim $R(D) = C_\mu - D$ has been retracted; the correct result is the lower bound. $C_{\max}$ remains an empirical parameter rather than a quantity derived from the rate-distortion formalism.

---

§0. Formulation Level

Working formulation. Fix $T, h < \infty$. Let $X := X_{1:T}$ denote the past block and $Y := X_{T+1:T+h}$ the future look-ahead block under a fixed computable stationary ergodic measure $\nu \in \mathcal{M}$. Define the finite-horizon predictive information

$$E_{T,h}(\nu) := I(X; Y).$$

When the infinite-horizon limit exists, define the excess entropy

$$E_\nu := I(\overleftarrow{X}; \overrightarrow{X}).$$

If S denotes the full ϵ-machine causal state, define the statistical complexity

$$C_{\mu,\nu} := H(S).$$

These are distinct quantities. The finite-horizon rate-distortion problem in this appendix is stated in terms of $E_{T,h}$, not $C_{\mu,\nu}$. The Solomonoff measure ξ enters only as the meta-prior weighting (preprint Eq. 1): individual $R(D)$ curves are computed per-measure ν. Results that require the full mixture ξ are stated separately.

---

§1. The Complete Four-Tuple Specification

1.1 Source X and Distribution P_X

Fix a computable stationary ergodic measure $\nu \in \mathcal{M}$ on $\{0,1\}^\infty$. The source is the process $(X_t)_{t\geq 1}$ distributed according to ν. For the meta-prior role, ξ from preprint Eq. (1) weights each such ν by $w_\nu \approx 2^{-K(\nu)}$. We write $P_X = \nu$ for a fixed member of $\mathcal{M}$. All results below apply per-measure ν; the Solomonoff connection enters through the dominance bound in §4.

1.2 Reproduction Alphabet $\hat{X}$

For fixed T, h, define a finite-horizon predictive equivalence relation on past blocks:

$$x \sim_h x' \iff \nu(Y \in A \mid X = x) = \nu(Y \in A \mid X = x') \quad \text{for all measurable } A \subseteq \{0,1\}^h.$$

Let S_h be the equivalence class of X under $\sim_h$. Then S_h is the minimal sufficient statistic for predicting Y from X at horizon h.

The full ϵ-machine causal state S is the infinite-horizon object obtained when one passes to semi-infinite pasts and the full future. This appendix uses S_h for finite-horizon derivations and reserves S for the full causal-state limit.

Computability status. For general computable ν, this appendix does not claim exact computability of the predictive-state partition. It is treated as an idealized measurable object. Exact computability is asserted only for explicitly identified subclasses such as finite-memory processes.

1.3 Distortion Function $d_h(x, z)$

The distortion function is the KL predictive divergence:

$$d_h(x,z) := D_{\mathrm{KL}}\big(P_\nu(Y \mid X = x) \,\|\, P_\nu(Y \mid Z = z)\big).$$

Here Z is a representation variable produced by an encoder $p(z \mid x)$. When $Z = S_h$, this is the exact predictive-state distortion; when Z is a coarsening or stochastic code, $P_\nu(Y \mid Z = z)$ is the induced predictive law.

Complete Four-Tuple

Element	Definition
X	$(X_t)_{t\geq 1}$ — stationary ergodic process under $\nu \in \mathcal{M}$
$\hat{X}$	S_h — finite-horizon predictive states
P_X	ν — fixed computable member of $\mathcal{M}$; Solomonoff ξ is the meta-prior
$d_h(x, z)$	$D_{\mathrm{KL}}(P_\nu(\cdot\|x)\|P_\nu(\cdot\|z))$ — KL predictive divergence over horizon h

---

§2. Derivation of $R_{T,h}(D)$ under the Four-Tuple

The rate-distortion function for the four-tuple of §1 is:

$$R_{T,h}(D) = \min_{p(z|x):\mathbb{E}[d_h(X,Z)]\leq D} I(X;Z)$$

2.1 The KL Distortion Identity

Let $X := X_{1:T}$, $Y := X_{T+1:T+h}$, and let Z be any representation produced by an encoder $p(z \mid x)$. Since $Z - X - Y$ is a Markov chain,

$$\mathbb{E}[d_h(X,Z)] = \mathbb{E}[D_{\mathrm{KL}}(P(Y \mid X)\|P(Y \mid Z))] = H(Y \mid Z) - H(Y \mid X) = I(X;Y \mid Z).$$

Equivalently,

$$\mathbb{E}[d_h(X,Z)] = I(X;Y) - I(Z;Y) = E_{T,h}(\nu) - I(Z;Y).$$

Therefore the distortion constraint $\mathbb{E}[d_h(X,Z)] \leq D$ is equivalent to

$$I(Z;Y) \geq E_{T,h}(\nu) - D.$$

2.2 The Information Bottleneck Reformulation

The distortion constraint restricts the space of allowable encoders to those satisfying $\mathbb{E}[d_h(X,Z)] \leq D$. This corresponds precisely to bounding $I(Z;Y)$ from below, giving the constrained Information Bottleneck problem. Because the achievable region $\{(I(Z;Y), I(X;Z))\}$ is convex under standard time-sharing arguments, strong duality holds. This permits an exact reformulation using the Information Bottleneck Lagrangian (Tishby, Pereira & Bialek 1999 [28]):

$$\mathcal{L}[p(z|x)] = I(X;Z) - \beta \cdot I(Z;Y)$$

with the Lagrange multiplier β determined by D. The IB Lagrangian traces the Pareto frontier of compression rate vs. predictive fidelity.

2.3 Main Theorem: General Lower Bound and Equality Criterion

We establish the bound for the rate-distortion function:

Proposition (general lower bound and equality criterion).
For any encoder $p(z \mid x)$, let

$$D := \mathbb{E}[d_h(X, Z)].$$

Then

$$I(X; Z) = E_{T,h}(\nu) - D + I(X; Z \mid Y).$$

Consequently,

$$R_{T,h}(D) \geq E_{T,h}(\nu) - D.$$

For compact finite reproduction alphabets where continuity guarantees the infimum over encoders is attained, equality at a given distortion D holds if and only if there exists an encoder achieving that distortion with

$$I(X; Z \mid Y) = 0.$$

For deterministic encoders $Z = g(X)$, this is equivalent to

$$H(Z \mid Y) = 0.$$

At zero distortion, the minimal sufficient statistic S_h achieves

$$R_{T,h}(0) = I(X; S_h) = H(S_h).$$

Note that this $H(S_h)$ zero-distortion rate sits strictly above the lower bound $E_{T,h}$ in general. The difference is the non-negative gap $H(S_h) - E_{T,h} = H(S_h|Y)$. This gap physically represents structural 'stored information' in the past that the future window alone fails to recover. Equality holding at zero distortion ($H(S_h|Y) = 0$) is a highly degenerate case generically false for complex processes.

In the full causal-state limit,

$$R(0) = C_{\mu,\nu} = H(S).$$

This equals E_ν only in special cases; in general $E_\nu < C_{\mu,\nu}$.

2.4 Behaviour for Coarser Reproduction Alphabets

For any deterministic coarsening $Z = g(S_h)$,

$$I(X; Z) = I(Z; Y) + I(X; Z \mid Y) = E_{T,h}(\nu) - D + I(X; Z \mid Y) \geq E_{T,h}(\nu) - D.$$

The nonnegative slack term $I(X; Z \mid Y)$ vanishes only when the coarsened representation is recoverable from the future window Y. Hence coarser alphabets generally produce rate-distortion curves strictly above the line $E_{T,h} - D$. The line is a universal lower bound, not a generic achieved envelope. Any practically computable codec uses a finite-memory approximation to the causal states and therefore has a curve above this bound.

2.5 Boundary Evaluations

Limit	Value	Interpretation
$D = 0$	$R_{T,h}(0) = I(X; S_h)$	Exact predictive-state compression; maximum information preserved
$D = E_{T,h}$	$R_{T,h}(E_{T,h}) = 0$	Trivial representation; all predictive information discarded
$D = D_{\min}$	$R_{T,h}(D_{\min}) \geq E_{T,h}(\nu) - D_{\min}$	Minimum lower bound for viable observer; Stability Filter threshold

(Note: In the infinite-horizon limit, the zero-rate point is at distortion E_ν, not at $C_{\mu,\nu}$)

---

§3. $C_{\max}$ — Characterisation and Barriers

3.1 Infinite-Horizon Convergence Lemma

The main theorem (§2.3) establishes the lower bound $R_{T,h}(D) \geq E_{T,h}(\nu) - D$ for finite (T, h). We now show this extends to the infinite-horizon setting.

Lemma (Infinite-horizon extension). Let ν be a stationary ergodic measure on $\{0,1\}^\infty$. Then:

1. $E_{T,h}(\nu) = I(X_{1:T}\,;\, X_{T+1:T+h})$ is non-decreasing in both T and h (by the data-processing inequality: conditioning on longer blocks cannot decrease mutual information between past and future under stationarity).
2. The limit $E_\nu := \lim_{T,h\to\infty} E_{T,h}(\nu)$ exists (possibly $+\infty$) by monotone convergence.
3. For each fixed $D \geq 0$, the sequence $R_{T,h}(D)$ is non-decreasing in T (longer pasts cannot reduce the optimal compression rate) and non-decreasing in h. *Proof sketch for monotonicity in h:* The distortion function decomposes as $d_{h+1}(x,z) = D_{\mathrm{KL}}(P_\nu(\cdot \mid x) \,\|\, P_z(\cdot \mid z))$ over $h+1$ future steps, which can be written via the chain rule as $d_h(x,z) + D_{\mathrm{KL}}(P_\nu(X_{T+h+1} \mid x, X_{T+1:T+h}) \,\|\, P_z(X_{T+h+1} \mid z, X_{T+1:T+h}))$. Since the second term is non-negative, $d_{h+1} \geq d_h$ pointwise. Therefore the constraint set $\{P(z|x) : E[d_{h+1}] \leq D\} \subseteq \{P(z|x) : E[d_h] \leq D\}$, and minimising over a smaller feasible set cannot decrease the rate: $R_{T,h+1}(D) \geq R_{T,h}(D)$.
4. Therefore $R_\nu(D) := \lim_{T,h\to\infty} R_{T,h}(D)$ exists.

Since $R_{T,h}(D) \geq E_{T,h}(\nu) - D$ holds at every finite stage, and both sides converge monotonically, the bound passes to the limit:

$$R_\nu(D) \geq E_\nu - D$$

This is the infinite-horizon lower bound invoked in Propositions T-1a and T-1c below. **Note:** For processes with $E_\nu = +\infty$ (e.g., high-order de Bruijn cycles as $k \to \infty$), the bound is trivially satisfied; such processes are excluded from the observer-compatible set $O_{C_{\max},D_{\min}}$ for any finite $C_{\max}$.

3.2 Partition of M by the Stability Filter — Proposition T-1a

Proposition T-1a (non-trivial partition).
Fix empirical $C_{\max} > 0$, $\Delta t > 0$, and $D_{\min} \geq 0$. Define

$$O_{C_{\max},D_{\min}} := \{\nu \in \mathcal{M} : R_\nu(D_{\min}) \leq C_{\max}\Delta t\}.$$

Then both $O_{C_{\max},D_{\min}}$ and its complement are non-empty.

Proof. The constant process lies in $O_{C_{\max},D_{\min}}$ because it has $E_\nu = 0$ and $R_\nu(D) = 0$.
For the complement, choose a binary de Bruijn-cycle process of order k: a stationary ergodic binary process of period 2^k with uniform phase, in which every length-k word appears exactly once per cycle. For this process,

$$E_\nu = C_{\mu,\nu} = k.$$

Hence

$$R_\nu(D_{\min}) \geq k - D_{\min}.$$

Choosing $k > C_{\max}\Delta t + D_{\min}$ gives $R_\nu(D_{\min}) > C_{\max}\Delta t$, so $\nu \notin O_{C_{\max},D_{\min}}$. □

3.3 Definition/Characterisation of $C_{\max}$ — T-1b

Definition T-1b (empirical bandwidth parameter).
$C_{\max}$ is taken as an empirical conscious-access bandwidth parameter external to the rate-distortion formalism. Given $C_{\max}$, define the observer-compatible class

$$O_{C_{\max},D_{\min}} := \{\nu \in \mathcal{M} : R_\nu(D_{\min}) \leq C_{\max}\Delta t\}.$$

If one wishes to summarize a separately specified reference class $\mathcal{O}_{ref}$, define

$$C_{max}^{ref} := \frac{1}{\Delta t} \sup_{\nu \in \mathcal{O}_{ref}} R_\nu(D_{\min}).$$

This is a summary statistic of a chosen class, not the definition of the class itself.

3.4 The Non-Emergence Barrier — Proof Sketch T-1c

Proof sketch T-1c (no finite universal bound from ξ alone).
The Solomonoff semimeasure ξ assigns positive prior weight to every computable measure $\nu \in \mathcal{M}$. The class $\mathcal{M}$ contains stationary ergodic binary processes with arbitrarily large excess entropy E_ν (for example, the de Bruijn family above). Since

$$R_\nu(D_{\min}) \geq E_\nu - D_{\min},$$

there is no finite support-wide upper bound on $R_\nu(D_{\min})$ derivable from ξ alone. Any finite $C_{\max}$ therefore requires additional empirical or class-restricting input beyond the bare Solomonoff prior. □

---

§4. Connection to the Solomonoff Meta-Prior

The four-tuple of §1 and the $R(D)$ derivation of §2 are stated per-measure ν. The Solomonoff connection — how the meta-prior ξ weights observer-compatible streams — is a structural correspondence rather than a derivation.

For any observer-compatible $\nu \in O_{C_{\max},D_{\min}}$, the rate-distortion equilibrium ensures the compressed stream $z_{0:T}$ is the Stability Filter's selected representation. The Solomonoff prior ξ assigns this ν weight $w_\nu \approx 2^{-K(\nu)}$: simpler (lower K) observer-compatible processes are exponentially more likely under ξ. This is the formal expression of the parsimony argument (Appendix T-4): the Stability Filter, operating on ξ, selects the simplest codec that fits within the bandwidth.

The dominance bound from T-4b applies directly: for any computable physics measure ν with $K(\nu) < \infty$:

$$-\log \xi(y_{1:T}) \leq -\log \nu(y_{1:T}) + K(\nu)$$

This ensures the OPT meta-prior ξ never assigns lower probability to observer-compatible streams than any fixed computable physics model, up to the model's own description length $K(\nu)$.

---

§5. The Experiential Bit Quantum h^* (Preview of E-1)

Given an empirical choice of $C_{\max}$ and an empirical conscious update window Δt, define

$$h^* := C_{\max}\Delta t.$$

For $C_{\max} \approx 10$ bits/s and $\Delta t \in [50, 80]$ ms,

$$h^* \approx 0.5\text{–}0.8$$

bits per conscious moment.

Any stationary ergodic process $\nu \in \mathcal{M}$ satisfying $E_{T,h}(\nu) - D_{\min} > h^*$ will legally trigger Narrative Decay. This is because $R_{T,h}(D_{\min}) \geq E_{T,h} - D_{\min} > h^* = C_{\max}\Delta t$, explicitly violating the compatibility criterion. However, this is a *sufficient* condition for collapse, not a strictly necessary one: because the lower bound is rarely tight ($R_{T,h} > E_{T,h} - D_{\min}$ generically per §2.4), processes can undergo Narrative Decay even when $E_{T,h} - D_{\min} \leq h^*$. This provides the quantitative prediction for E-1; the sensitivity to the choice of $\Delta t \in [40, 300]$ ms is discussed in the E-1 appendix.

---

§6. Closure Summary

T-1 Deliverables — Revised Status

1. The four-tuple is specified in a finite-horizon predictive setting.
2. The predictive-KL identity is derived correctly.
3. The generic theorem $R(D) = C_\mu - D$ is replaced by the correct lower bound

$$R_{T,h}(D) \geq E_{T,h} - D$$

   together with an exact equality criterion $I(X; Z \mid Y) = 0$.
4. Zero-distortion coding is characterized by the minimal sufficient statistic S_h, and in the full causal-state limit $R(0) = C_{\mu,\nu}$.
5. $C_{\max}$ is treated as empirical, not internally derived.
6. $h^* = C_{\max}\Delta t$ is an empirical parametrization, not a theorem from §2.

This appendix is maintained as part of the OPT project repository alongside theoretical_roadmap.pdf.